UNIVERSITY OF SOUTHAMPTON

FACULTY OF MEDICINE

CLINICAL AND EXPERIMENTAL SCIENCES

VOLUME I OF I

PHENOTYPIC AND GENOTYPIC ANALYSIS OF *STREPTOCOCCUS PNEUMONIAE* DIVERSITY DURING THE INTRODUCTION OF PNEUMOCOCCAL CONJUGATE VACCINES IN THE UK

BY

REBECCA A GLADSTONE

Thesis for the degree of Doctor of Philosophy

March 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF MEDICINE

MOLECULAR MICROBIOLOGY

Thesis for the degree of Doctor of Philosophy

PHENOTYPIC AND GENOTYPIC ANALYSIS OF *STREPTOCOCCUS PNEUMONIAE* DIVERSITY DURING THE INTRODUCTION OF PNEUMOCOCCAL CONJUGATE VACCINES IN THE UK

REBECCA ASHLEY GLADSTONE

Streptococcus pneumoniae is a major global pathogen capable of causing pneumonia and meningitis, particularly at the extremes of age. Pneumococcal conjugate vaccines (PCV) to protect against invasive pneumococcal disease were implemented from 2006 in the UK. Asymptomatic carriage occurs in approximately one third of children aged 4 years and under in the UK, whilst colonisation is known as a precursor to disease.

Vaccination against a limited number of the >90 serotypes was hypothesised to result in serotype replacement in carriage which, would in turn affect disease epidemiology. This thesis set out to characterise carried pneumococci from children aged 4 years and under, collected each winter from the introduction of 7-valent PCV in 2006 through to winter 2010/11, which extended one year after the introduction of 13-valent PCV. This work sought to detect changes in serotypes and genotypes, using traditional phenotypic and genotypic techniques, as well as to further validate and utilise next generation sequencing for determining typing data, and characterisation of the gene diversity targeted by current and investigational vaccines.

Whole genome sequence analysis allowed robust derivation of typing that contributed to epidemiological analysis. Carriage remained stable, as did the number of serotypes and sequence types (STs) observed each year. Vaccine serotypes (VT) significantly decreased with a concomitant increase in non-vaccine serotypes (NVT). Significant decreases for VT 6A, 6B, 19F and 23F and increases for NVT 21, 23B, 33F and 35F were observed. Significant changes in prevalence were also observed for genotypes associated with these serotypes. Novel sequence variation was observed in capsular defining genes and pneumolysin, which has potential to affect the efficacy of vaccines against these targets.

Near complete serotype replacement occurred during the introduction of PCV7 and is likely to continue for PCV13. This resulted in a greater number of NVT pneumococci colonising individuals, which could result in disease cases. Others have consequently observed serotype replacement in disease in the UK. The future of control of pneumococcal disease through vaccination will therefore require a response to this replacement, through targeting the most prevalent replacing serotypes or broader pneumococcal vaccines taking into account the variation within potential targets.

Contents

List of Tables

List of Figures

DECLARATION OF AUTHORSHIP

I, Rebecca A Gladstone declare that the thesis entitled:

'Phenotypic and genotypic analysis of *Streptococcus pneumoniae* diversity during the introduction of pneumococcal conjugate vaccines in the UK'

and work presented in the thesis are both my own, and have been generated by me as the result of my own original research.

I confirm that this work was done wholly or mainly while in candidature for a research degree at this University;

- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- Where I have consulted the published work of others, this is always clearly attributed;
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself summarised on the following pages;
- Parts of this thesis introduction have been published as: (Gladstone *et al*. 2011; Gladstone *et al.* 2012b), the full text and my additional publications related to this subject but not included in this thesis can be found in the Bibliography.

Signed:

Date: 27/09/2013.....

Clarification of Contribution

The pneumococcal carriage study is an on-going collection of isolates. Analysis of the first three years of the study has been reported elsewhere by past members of the research team (Tocheva *et al.* 2010; Tocheva *et al.* 2011; Tocheva *et al.* 2013). This section declares and defines the analysis performed by Rebecca Gladstone or others (Table 1).

Determination of serotype and MLST of pneumococci collected during winters (October to March) of 2006/07 through to 2008/09 was carried out using traditional methods by Dr Anna Tocheva. Rebecca Gladstone performed all analyses of S. pneumoniae collected during winters 2009/10 and 2010/11. DNA extraction for whole genome sequencing and consequent genome analysis for all years 1-5 was carried out by Rebecca Gladstone for this thesis. All data from years 2006/07 through to 2010/11 was then analysed together by Rebecca Gladstone to produce a full analysis for five years of carriage study data.

Table 1. Details of carriage study isolate analyses

Year of Study	Winter	Comments	Analysis performed by
1	2006/7	*S. pneumoniae* typing	Dr Anna Tocheva
2	2007/8	*S. pneumoniae* typing	Dr Anna Tocheva
3	2008/9	*S. pneumoniae* typing	Dr Anna Tocheva
4	2009/10	*S. pneumoniae* typing	Rebecca Gladstone
5	2010/11	*S. pneumoniae* typing	Rebecca Gladstone
1-3	2006/7-2008/9	Time series analysis of *S. pneumoniae*	Dr Anna Tocheva (Tocheva *et al.* 2011; Tocheva *et al.* 2013)
1-5	2006/7-2010/11		Rebecca Gladstone
1-5	2006/7-2010/11	Preparation of DNA for genomic sequencing and whole genome analysis	Rebecca Gladstone

ACKNOWLEDGEMENTS

I would like to take this opportunity to offer sincere thanks to my supervisors Dr Stuart Clarke, Dr Jo Jefferies and Dr Saul Faust for giving me the opportunity to study towards this doctoral degree, helping me progress thus far and fot their support and encouragement. I would like to acknowledge staff at Public Health England microbiology services Southampton and the research nurses, BM/BMedSci students for their hard work in the collection and processing of isolates. I would like to give special thanks to the children and their parents for participation in the carriage study. I would like to thank the staff at the University of Birmingham's centre for genomics for allowing me to gain valuable expertise under their supervision particularly Dr Nick Loman and Professor Mark Pallen for allowing me to visit and learn. Additionally I would like to thank Dr Stephen Bentley and the Sanger Institute for making the sequencing of the carriage study isolates feasible and thank them for inviting me to attend the institution to utilise their expertise and resources for my analysis of the sequences. Grants received from the Society of General microbiology and Public Health England (PHE) enabled my attendance at highly relevant conferences and training placements, which were invaluable to me. I would like to gratefully acknowledge the financial support of Pfizer for this research and my stipend. Finally I cannot neglect to acknowledge my gratitude for all of my treasured friends and family who all believed in me and were responsible for my continued sanity.

1. INTRODUCTION

1.1 *Streptococcus pneumoniae*

Streptococcus pneumoniae, also known as the pneumococcus, was first isolated in 1880 by both Pasteur and Stenberg independently (Pasteur *et al.* 1881; Sternburg 1881). It is one of the oldest recognised pathogens, and was clinically demonstrated in the 1880's to be a major cause of pneumonia. Its association with pneumonia earned the gram-positive bacterium the title *"Captain of the men of death"*, quoted by Sir William Osler in the early 1900's. These encapsulated diplococci have nonetheless contributed a great deal to the advancement of science in microbiology, immunology (Tillett & Francis 1930) and genetics. Work by Frederick Griffith in the 1920's revealed that bacteria were capable of transferring genetic information by studying the conversion of an avirulent pneumococcal strain to a pathogenic strain (Griffith 1928). This went on to influence the work of Avery and colleagues and subsequently led to the confirmation of DNA being the hereditary material, at the centre of all biology (Avery 1944). The volume of knowledge generated through research into *S. pneumoniae* is a consequence of its well-established pathogenic nature.

1.2 Pneumococcal disease

The global yearly burden of serious pneumococcal disease was estimated at 14.5 million episodes and an excess of 800,000 child deaths each year based on estimates for the year 2000 (O'Brien *et al.* 2009). While the majority of this burden lies in the developing world (Johnson *et al.* 2010), the burden within the UK is still substantial (Melegaro *et al.* 2006; Miller *et al.* 2011b; Ladhani *et al.* 2013). The pneumococcus has a spectrum of manifestations in the human host ranging from asymptomatic colonisation and acute otitis media to fatal cases of sepsis and meningitis.

1.2.1 Meningitis

Meningitis is a medical emergency due to its potential to cause permanent disability and be rapidly fatal, due to cerebral and systemic complications (Grimwood *et al.* 2000; Kastenbauer & Pfister 2003). It involves the inflammation of the meninges, the protective layers that surround the brain and spinal cord, when the infectious agent has managed to invade the cerebral spinal fluid. Meningitis has a number of bacterial, viral and fungal aetiologies. *S. pneumoniae* is reported as the second most common cause

of bacterial meningitis within the UK in 2009-2010 (Meningitis Research Foundation 2010) and has been reported to have a mortality rate in the UK of approximately 7.5% in children under five years old, whilst a European study suggested 30% mortality in those over 64 years old (Kastenbauer & Pfister 2003; Johnson *et al.* 2007).

1.2.2 Pneumonia

An infection of the lung, pneumonia causes breathing difficulties though inflammation of the infected tissue. The etiological link between *S. pneumoniae* and this disease had contributed to its taxonomy by the early 1900's (Chester 1901). Untreated cases of pneumonia can progress to empyema (the build-up of pus in the pleural cavity) or pneumococcal sepsis. Cases of pneumonia are hard to define clinically and the World Health Organisation (WHO) has been working towards a standardised definition (Cherian *et al.* 2005). Diagnosis is still heavily dependent upon x-ray interpretation. Even microbiological confirmation currently lacks sensitivity, but potentially this could be increased with the use of non-culture detection (Lahti *et al.* 2006). It is plausible that the incidence of pneumonia is under reported, yet community associated pneumonia, most commonly caused by *S. pneumoniae*, is still the leading infectious cause of death worldwide (Anevlavis & Bouros 2010).

1.2.3 Invasive pneumococcal disease

Invasive pneumococcal disease (IPD) is typically defined by the isolation of the bacteria from blood, cerebral spinal fluid or a normally sterile site. This therefore includes pneumonia with sepsis, sepsis alone or meningitis in the presence or absence of bacteraemia. These clinical presentations can be amalgamated under the term IPD for the purpose of simplifying the evaluation of disease burden in populations but the exact definition can differ between studies dependant on the method used to positively identify pneumococci such as culture or PCR positivity (Avni *et al.* 2010).

1.2.3.1 Risk groups

Certain populations are at a higher risk of IPD and other pneumococcal diseases. Infants under two years of age, for whom a UK study of an unvaccinated population estimated 15 pneumococcal meningitis cases per 100,000 (Melegaro *et al.* 2006) are a primary risk group. The elderly are also at high risk of invasive disease, approximately 45 cases of IPD per 100,000 occurred in persons over 65 years in Scotland for 1999-2001 and an adjusted estimate of 28 cases per 100,000 for 2009-10 in those ≥65 in England and Wales (Kyaw *et al.* 2003; Miller *et al.* 2011b). Additional risk groups for

serious pneumococcal infection include children between two and five years old and those that are immunosuppressed or immunocompromised due to genetic or infectious conditions, treatment or substance misuse (Burman *et al.* 1985; Kyaw *et al.* 2003). Smoke exposure and overcrowded conditions are thought to contribute to high carriage rates and increased risk of disease in these circumstances (Hoge *et al.* 1994; Lee *et al.* 2010).

1.2.4 Acute otitis media

S. pneumoniae is also major causative agent of acute otitis media (AOM), (Bluestone *et al.* 1992; Rovers *et al.* 2004), an infection of the middle ear with a 61% global incidence of all cause AOM in children aged 1-4 years reported for the year 2005 (Monasta *et al.* 2012). AOM is a major reason for childhood attendance at clinics, however the causative agent is challenging to unequivocally identify due to the polymicrobial nature of the environment and the difficulty of correctly sampling the middle ear (Casey *et al.* 2009).

This common cause of childhood illness is associated with high morbidity and economic burden, particularly as reoccurring or prolonged infection can result in impaired hearing and learning ability and even mortality, estimated as 90/100,000 children aged 1-4 years in 2005 from complications (Teele *et al.* 1990; World Health Organisation 2004; Melegaro *et al.* 2006; Monasta *et al.* 2012). Treatment options are a topic of debate with two particular clinical trials published in 2011 in The New England Journal of Medicine (NEJM) reporting evidence for antimicrobial treatment. This sparked a stream of discussion and replies in NEJM on the validity of the findings, suggesting they overstated the case for antibiotics versus the ‘wait and see prescription’ (Hoberman *et al.* 2011; Tahtinen *et al.* 2011).

1.3 Pneumococcal carriage

Despite the potential of pneumococci to cause infections, the bacterium is predominantly commensal in nature and can colonise the upper respiratory tract without clinical presentation. Colonisation is thought to be an essential step in the progression to disease, although such progression only occurs in a minority of carriers (Bogaert *et al.* 2004a). Humans have long been considered the only natural host for *S. pneumoniae*, yet carriage has been observed in animals that have close contact with people such as pets, zoo and laboratory animals (van der Linden *et al.* 2009).

In developed countries such as the UK, asymptomatic carriage occurs in around 30-40% of children less than five years old as is reported in this thesis and other work (Roche *et al.* 2007; Tocheva *et al.* 2011). Developing countries and population groups including native populations in developed countries have been reported to have higher carriage and consequent disease rates (Lloyd-Evans *et al.* 1996; Woolfson *et al.* 1997; Watson *et al.* 2006; Millar *et al.* 2009; O'Brien *et al.* 2009)

The respiratory tract is also occupied by additional commensals, some of which are potential pathogens. It is becoming apparent that pneumococci exist in a polymicrobial environment in the nasopharynx (Gladstone *et al.* 2010; Bogaert *et al.* 2011). This includes other alpha haemolytic streptococci, *Haemophilus influenzae*, *Moraxella catarrhalis, Staphylococcus aureus* and potentially *Neisseria meningitidis* (Kononen *et al.* 2002; Emlickov *et al.* 2006; Christensen *et al.* 2010). This multi species environment is complex and there are reports suggesting that pneumococci have different specific abilities to co-colonise with other respiratory bacteria (Bogaert *et al.* 2004b; Margolis *et al.* 2010; Lebon *et al.* 2011; Lijek *et al.* 2012; Spijkerman *et al.* 2012). This has implications for the impact of pneumococcal vaccines on the commensal niche.

1.3.1 Detection of carriage

The epidemiology of pneumococcal carriage can be evaluated through the collection and characterisation of pneumococci that are colonising individuals in a population, in both a culture dependent and independent manner. This can be achieved by a variety of methods; each differing in their sensitivity dependent upon the variables chosen.

1.3.1.1 Sampling site

S. pneumoniae can be isolated from the nasal passage, nasopharynx and oropharynx (Capeding *et al.* 1995; Rapola *et al.* 1997; Watt *et al.* 2004). Nasal washings (Abdullahi *et al.* 2007) or nasopharyngeal aspirates also permit collection of pneumococci, although aspirates are not usually available from healthy individuals (Rapola *et al.* 1997; Leach *et al.* 2008). Additionally oropharyngeal swabbing has been demonstrated to be significantly poorer than nasopharyngeal swabbing for isolation of pneumococci in children (Capeding *et al.* 1995; Watt *et al.* 2004) whilst the combined sampling of the oropharynx and nasopharynx is suggested for adults (Watt *et al.* 2004). Nasal swabs are reported to give comparable results to nasopharyngeal swabs in Finland, although significant differences in isolation rates may be masked by small study size (Rapola *et al.* 1997). Abdullahi *et al* reported that nasal washing is of increased sensitivity than nasopharyngeal swabbing but noted that nasal washings were less well tolerated by

participants (Abdullahi *et al.* 2007), demonstrating that ethical limitations may prevent use of methods with superior sensitivity.

No single sample can detect all pneumococcal colonisation, therefore sampling a combination of sites or repeated sampling of a single site could increase detection rates (Capeding *et al.* 1995; Watt *et al.* 2004; Abdullahi *et al.* 2007). The added benefit of detecting pneumococci or additional serotypes using multiple sites may not outweigh the discomfort/inconvenience to study subjects or change the clinical course of care for those with disease who may be sampled for carriage.

The most sensitive sample type differs with age group, with implications for sampling carriage in the elderly to evaluate PCV use in this risk group (Hendley *et al.* 1975; Boersma *et al.* 1993; Watt *et al.* 2004). No single study has compared all relevant sampling sites to determine the most sensitive and pragmatic sample source for any age group. Despite this, the current consensus for pragmatic and feasible epidemiological research is that the nasopharyngeal swab alone is adequate for sampling pneumococcal carriage. Consequently carriage rates may be underestimated.

1.3.1.2 Swab type

The swab material has not been examined extensively for its influence on detection sensitivity. Two studies from the 1930's and 1940's suggested that fatty acids contained within cotton are inhibitory to pneumococci (Wright 1934; Pollock 1948). No further studies have confirmed or disputed the impact of cotton swabs. A study by Rubin *et al* has however demonstrated differences in both culture detection rate and culture-independent PCR sensitivity that vary with swab composition for pneumococci (Rubin *et al.* 2008). Calcium alginate swabs were observed to be superior to rayon or Dacron® for culture and PCR detection, whilst Dacron®, was inferior to both calcium alginate and rayon for culture detection (Rubin *et al.* 2008). However, during the 1990's, studies investigating the effect of calcium alginate on PCR gave conflicting results (Wadowsky *et al.* 1994; Poddar *et al.* 1998). Nylon flocked swabs have recently come to market, the multi-length fibres of these swabs are proposed to improve sample absorbency and release of specimens, however there is little literature on their use for nasopharyngeal bacterial sample collection, with one study on *S. aureus* reporting no difference in sensitivity to rayon swabs (De Silva *et al.* 2010). Any unknown impact of swab composition could be greater in studies with longer transportation periods and those that store the original swab for retrospective analysis. As there appear to be differences between swabs it would be advantageous for the scientific community to

determine which swab type is least detrimental to identification of pneumococci in samples.

1.3.1.3 Swab transport medium

Swab transport medium is responsible for maintaining viability of bacteria before analysis whether it is transferred to a plate for culture or to storage medium. O'Brien *et al* reported that, when compared to direct plating on solid media, use of the WHO-recommended STGG liquid medium resulted in subsequent equivalent rates of recovery (O'Brien *et al.* 2001). Additionally greater numbers of pneumococci can be recovered from STGG than SGG over four days at 20-30ºC (Charalambous *et al.* 2003). Gray described the superiority of Egg-thioglycolate-antibiotic (ETA) medium to Amies or Stuarts media in situations where rapid processing or freezing (a requirement of numerous media types for maintaining pneumococci) is not possible (Gray 2002). Relatively few studies have been performed to determine the optimal media for transportation and storage of pneumococci, comparing a limited number of media types (O'Brien *et al.* 2001; Gray 2002; Charalambous *et al.* 2003). Comprehensive investigations of the differing abilities of transport media to preserve pneumococci and other respiratory microorganisms would provide evidence to allow appropriate selection of media for carriage studies.

1.3.1.4 Storage and recovery

Archived isolates are an excellent resource for research, but require that pneumococci recovered from samples that have been stored for long durations are representative of the original sample. Direct storage of the swab or swab contents lends itself to high-sensitivity molecular techniques independent of culture that do not require abundant presence of viable bacteria. This genotypic approach has its limitations in extrapolation to true phenotype as it detects only the presence of genes not their expression. Additionally in a species known to readily exchange genetic material the detection of genes considered to be species specific is not equivalent to phenotypic species designation.

Alternatively, storage of single colonies allows researchers to return to a specific cultured isolate. The WHO recommends the selection of two morphologically distinct presumptive pneumococcal colonies. Long term storage of isolates at -70 to -80ºC is considered the gold standard (Gibson & Khoury 1986; O'Brien *et al.* 2003). It is important to note that the act of long term storage itself or frequent freeze thaw cycles has been shown to alter the phenotype of pneumococci, for example susceptibility to

optochin (Robson *et al.* 2007). The WHO recommends STGG medium for long term storage, both for the original swab and for primary cultures. This medium has been shown to give enhanced ability for recovery of pneumococci on solid media after nine weeks storage when compared to direct plating, particularly for samples with low colony counts (O'Brien *et al.* 2001). Tryptone-deficient SGG has however been shown to be as good as STGG for maintaining viability in long term storage at -70ºC (Siberry *et al.* 2001; Charalambous *et al.* 2003) which could reduce costs associated with carriage studies. This latter point is important for developing countries where PCV evaluation may increase as GAVI (formerly The Global Alliance for Vaccines and Immunisation) supports PCV immunisation programs.

Rubin *et al* go on to describe enhanced recovery of plated pneumococci when the plate is inoculated with a swab that has been stored in STGG, when compared to inoculation of a plate with a volume of STGG from a vial that contains the swab (Rubin *et al.* 2008). Any study using the supernatant from a stored swab for further analysis could benefit from sample vortexing, ensuring dispersal of microbial matter from the swab, endorsed by the WHO (O'Brien *et al.* 2003). An alternative method for enhancement of pneumococcal recovery is use of a broth enrichment step, either prior to plating or prior to direct molecular assessment of serotype presence. Broth enrichment for culture and PCR has been reported to have an increased sensitivity for detection of multiple serotype colonisation. Results from these genotypic and phenotypic techniques are complementary in nature as no one technique detected all serotypes detected (Kaltoft *et al.* 2008; Carvalho *et al.* 2010). Broth enrichment detects more multiple serotypes and low density groups however it may still introduce some unknown culture bias with a potential for differential enrichment of some low-density groups due to the competitive growth environment. Specimens are often shared and transported between laboratories for research, in this situation the WHO recommend Dorset egg media for transportation of pneumococcal strains, Inverarity *et al* point out that this product cannot be imported into certain countries due to avian influenzae related concerns about egg products and suggest Amies charcoal media to be a suitable alternative (Inverarity D *et al.* 2006). Again there are not extensive evidence-based comparisons of long-term storage options of recovery for pneumococci. Other media or recovery amendments such as enrichment may have an equal or superior ability to STGG for maintaining pneumococci, not yet investigated.

In an era where the importance of polymicrobial interactions within the respiratory niche is now being realised, identification of colonisation events with multiple potential pathogens is becoming more common and desirable but is not standard. Optimal

detection of each organism within a panel is still necessary and the description of the best method for simultaneous detection would be most valuable to a broad range of carriage studies. Ideally detection of pneumococci would involve the use of a standardised method that had been shown experimentally to be proficient at all stages for preserving and reporting the original specimen. Widespread use of such a method would augment the value of collected data. The benefits and disadvantages of high sensitivity detection over the increased validity of standardised study comparison need to be carefully considered. When assessing a polymicrobial niche it also needs to be recognised that viral and fungal agents may also co-colonise, and a description of this broad range of microbes would add further to the level of complexity regarding a detection protocol.

1.4 Pneumococcal typing

The pneumococcal population is diverse and dynamic worldwide (Hausdorff *et al.* 2005). Identification of pneumococci and classification below the species level is valuable for defining outbreaks, global spread and additional epidemiological analysis including vaccine evaluation. A variety of techniques based on both phenotypic and genotypic properties are available to identify and distinguish pneumococci. Presumptive identification of pneumococci is based on colonial morphology, alpha haemolysis on blood agar, bile (deoxycholate) solubility and susceptibility to optochin (Bowen *et al.* 1957; Pikis *et al.* 2001). As the pneumococcus is such a diverse population, further sub species classification is needed.

1.4.1 Serotyping

Sub-classification of pneumococci based on the polysaccharide capsule is a gold standard approach for distinguishing between pneumococci due to the associations of the capsule with invasive potential, disease manifestation and role as a primary virulence determinant (Hausdorff *et al.* 2000; Brueggemann *et al.* 2004; Hausdorff *et al.* 2005). Capsular type can be deduced by a variety of methods both phenotypic and genotypic each with their own benefits and disadvantages.

1.4.1.1 Phenotypic serotyping

The pneumococcus is typed based on the immunological properties of the polysaccharide that encapsulates it. The capsule is considered a primary determinant of pneumococcal immunogenicity (Melin *et al.* 2010a; Melin *et al.* 2010b; Lammers *et*

al. 2011). The Quellung reaction was first described in 1902 by Neufeld and is used for phenotypic classification of the capsule based on the swelling effect of antibodies specific to the polysaccharide motifs (Neufeld 1902). This swelling reaction allows pneumococci to be classified into serogroups and serotypes. Serotypes belong to a serogroup when they share cross-reactive antibodies due to similar capsular structure and shared antigens. Currently 46 serogroups and 96 serotypes have been documented, the latest additions being serotypes 6C, 6D, 6F, 6G (Park *et al.* 2007b; Jin *et al.* 2009; Oliver *et al.* 2013b), 11E (Calix & Nahm 2010) and 20B (Calix *et al.* 2012). Specific antibodies against polysaccharide capsules can also be utilised in latex and traditional slide agglutination assays (Kronvall 1973; Lafong & Crothers 1988) and in multiplex cytometric bead assays (Pickering *et al.* 2002; Yu *et al.* 2011).

1.4.1.2 Genotypic serotype inference

In the molecular era capsule type can also be deduced genotypically, for example by using the PCR method developed at the Centers for Disease Control and Prevention (CDC). The CDC protocol includes primers for the 40 most commonly observed serogroup/types in IPD, with schemes specific to the geographical regions of the US, Latin America and Africa (Pai *et al.* 2006; Centers for Disease Control and prevention 2012). However, Carvalho and colleagues have reported that non-pneumococcal isolates tested PCR positive for some pneumococcal capsules due to their close relation (Carvalho *et al.* 2012). This phenomenon could potentially confound pneumococcal typing results and is an argument for confirmation of capsule expression through phenotypic methods.

Multiple serotype colonisation occurs in some individuals (Brugger *et al.* 2009) but detection of the lesser serotypes presents its own sensitivity issues, broth enrichment combined with PCR (Carvalho *et al.* 2010) and microarray have been shown as capable methods for detection of pneumococcal capsular biosynthesis genes allowing identification of multiple serotype colonisation with enhanced sensitivity when compared to traditional culture based phenotypic techniques (Brugger *et al.* 2010; Newton *et al.* 2011; Turner *et al.* 2011; Glennie *et al.* 2013).

1.4.1.3 Importance of serotyping

Pneumococci genomes contain mobile genetic elements including bacteriophage and are subject to intra- and inter- species horizontal gene transfer and the homologous recombination of genetic material, all of which add to the plasticity of the pneumococcal genome (Romero *et al.* 2009; Croucher *et al.* 2011). For this reason serotypes can be

associated with, but not exclusively linked to genetic lineages as recombination events in the capsular locus can allow pneumococci to switch their serotype (Coffey *et al.* 1991; Coffey *et al.* 1998; Silva *et al.* 2006). Pneumococcal serotypes have been observed to have differing associations with carriage, disease and clinical outcome (Brueggemann *et al.* 2003; Brueggemann *et al.* 2004; Greenberg *et al.* 2011), with a minority of the >90 known serotypes responsible for the vast majority of disease cases. Serotype 3 and serogroup 15 are examples of this as they are commonly found in carriage but cause comparatively few disease cases. Conversely serotypes such as 1 and 5 are often observed in invasive disease cases and are particularly associated with outbreaks (Smith *et al.* 1993; Brueggemann *et al.* 2003; Brueggemann *et al.* 2004). Serotypes can be associated with specific disease manifestations such as the serotype 1 association with complicated pneumonia and empyema (Eastham *et al.* 2004; De Schutter *et al.* 2006; Greenberg *et al.* 2011). Several studies have therefore calculated odds ratios to determine the invasive potential of an individual serotype to cause a case of invasive disease, taking into account the prevalence in carriage (Brueggemann *et al.* 2004; Hanage *et al.* 2005b; Bättig *et al.* 2006; Greenberg *et al.* 2011). As serotypes can be associated with but not limited to a specific genetic background, it is also important to gather pertinent genotypic information on circulating strains.

1.4.2 Multi-locus sequencing typing

Multi locus sequence typing (MLST) was first proposed as a sequence based epidemiological tool for the analysis of *S. pneumoniae* isolates in 1998 (Enright & Spratt 1998). MLST is widely used for the surveillance of *S. pneumoniae* (Jefferies *et al.* 2004; McChlery *et al.* 2005; Miller *et al.* 2011b) and a variety of other pathogens. This typing method is an adaptation of the earlier multi locus enzyme electrophoresis (Selander *et al.* 1986) and uses the sequences of seven conserved house-keeping genes to classify isolates and elucidate the relationship between strains. The determined sequence type (ST) can be associated to varying degrees with a particular serotype or serotypes. The occurrence of isolates within a serotype from a ‘new’ genetic background previously associated with another specific serotype can be indicative of capsular switch events. MLST has proved very useful for collecting epidemiological data for routine comparison, tracking the spread of novel variants or antibiotic resistant clones (Silva *et al.* 2006; Lambertsen *et al.* 2010; Tocheva *et al.* 2013). MLST can aid outbreak investigation (Vainio *et al.* 2009) and adequately identify and track genetic lineages, though its resolution is limited to the seven sequenced loci representing the genetic background of the strain. Isolates involved in the outbreak

often belong to the same ST and therefore look identical. MLST therefore does not allow easy identification of the direction of spread and ultimately the source of the outbreak. Clonal complexes (CC) can be defined as a group of STs that are not identical but share a set number of the seven alleles, for example five or six. Comparison between members of a CC allows the expansion or emergence of clones to be put in the context of their genetic background with the associated metadata of the member STs. The algorithm eBURST (electronic-based on related sequence types) (Feil *et al.* 2004), is used to examine the relationships within clonal complexes. This web-based BURST utilises allelic profiles but does not take into account the sequence data contained, and it cannot draw conclusions on the relationships between different alleles of the same loci.

The standard seven gene MLST is designed to include genes which are not subject to external evolutionary pressure and therefore regrettably tells us nothing of genes of clinical importance including virulence determinants, antibiotic resistance cassettes and vaccine targets that can be readily exchanged and highly variable. However, additional advances in the field of genomics enable larger quantities of epidemiologically valuable data to be generated by alternative means.

1.4.3 Whole genome sequencing

The genome of an organism contains complete information for its survival and reproduction. The whole genetic repertoire therefore includes genes of clinical importance and those genes that could be utilised for epidemiological purposes such as virulence profiling and outbreak tracing due to their importance in disease or their level of conservation respectively.

The bacteriophage phi X174 was the first genome to be fully sequenced, in 1977 (Sanger *et al.* 1977a), this made use of the ‘plus and minus’ sequencing method (Sanger & Coulson 1975). Alternative sequencing methods were fast becoming available including methods by Maxam and Gilbert and further advances by Sanger (Maxam & Gilbert 1977; Sanger *et al.* 1977b). Developed in the 1970’s, Sanger sequencing fast became the gold standard for determining the nucleotide sequence of a specific region of the genome, and is still utilised for a multitude of applications including MLST (Sanger *et al.* 1977b; Enright & Spratt 1998). Sanger sequencing employs analogues of dNTPs (deoxynucleoside triphosphates) which cause specific termination of DNA extension by the DNA polymerase, the sequencing is divided into four reactions each containing all four standard dNTPs and one modified ddNTP that

specifically terminates the chain at A, T, C or G. This produces an array of DNA fragments, which can be separated by gel electrophoresis, and a combination of the four reactions allows the sequence to read from gel images. Adaptations of the Sanger method vastly increased the automation, rapidity and value of this technique, notably the conjugation of the ddNTPs with radiolabels and later fluorophores which resulted in the dye termination method (Smith *et al.* 1985) and the use of capillaries for electrophoresis. These advances encouraged by the drive for completed genomes ultimately led to the publication of the first whole bacterial genome sequence, *Haemophilus influenzae* Rd KW20, in 1995. This was also the first use of a random fragment library, also known as shotgun sequencing, a method that is designed to avoid cloning biases (Fleischmann *et al.* 1995).

The Human Genome Project greatly accelerated developments in sequencing. The quest to determine the whole sequence of the human genome first started with debates in 1985 and formally became a worldwide effort in 1990 (Watson & Cook-Deegan 1991). In that year the head of the National Centre for Human Genome Research, James Watson was quoted by the New York Times **"***It is essentially immoral not to get it [the human genome sequence] done as fast as possible***"**. The size and complexity of the human genome put considerable strain on the current sequencing technologies and forced the development of high throughput, massively parallel technologies to allow its completion in 2003 (International Human Genome Sequencing Consortium 2004).

1.4.3.1 Next generation sequencing technologies

Conventional capillary based sequencing can perform up to 96 parallel reactions, the limitation of further extending this method being mainly one of cost (Shi *et al.* 1999). The ability of high throughput technologies to process millions of sequence reads in parallel is what defines “next generation” methodologies. While several platforms exist they are all high throughput and use massively parallel sequencing chemistry (Metzker 2009); working from fragment libraries to produce sequence reads. Read length, however, does differ between sequencers and few are comparable yet to the 650-800 base pairs (bp) of capillary sequencers without loss of base calling accuracy (Metzker 2009). The read length produced by next generation technologies varies considerably dependant on the platform (Metzker 2009). The massively parallel approach goes some way to overcome this limitation while commercial researchers continue work towards increasing the read length produced by their platforms. A number of competitive platforms currently exist on the market, each with their own chemistry,

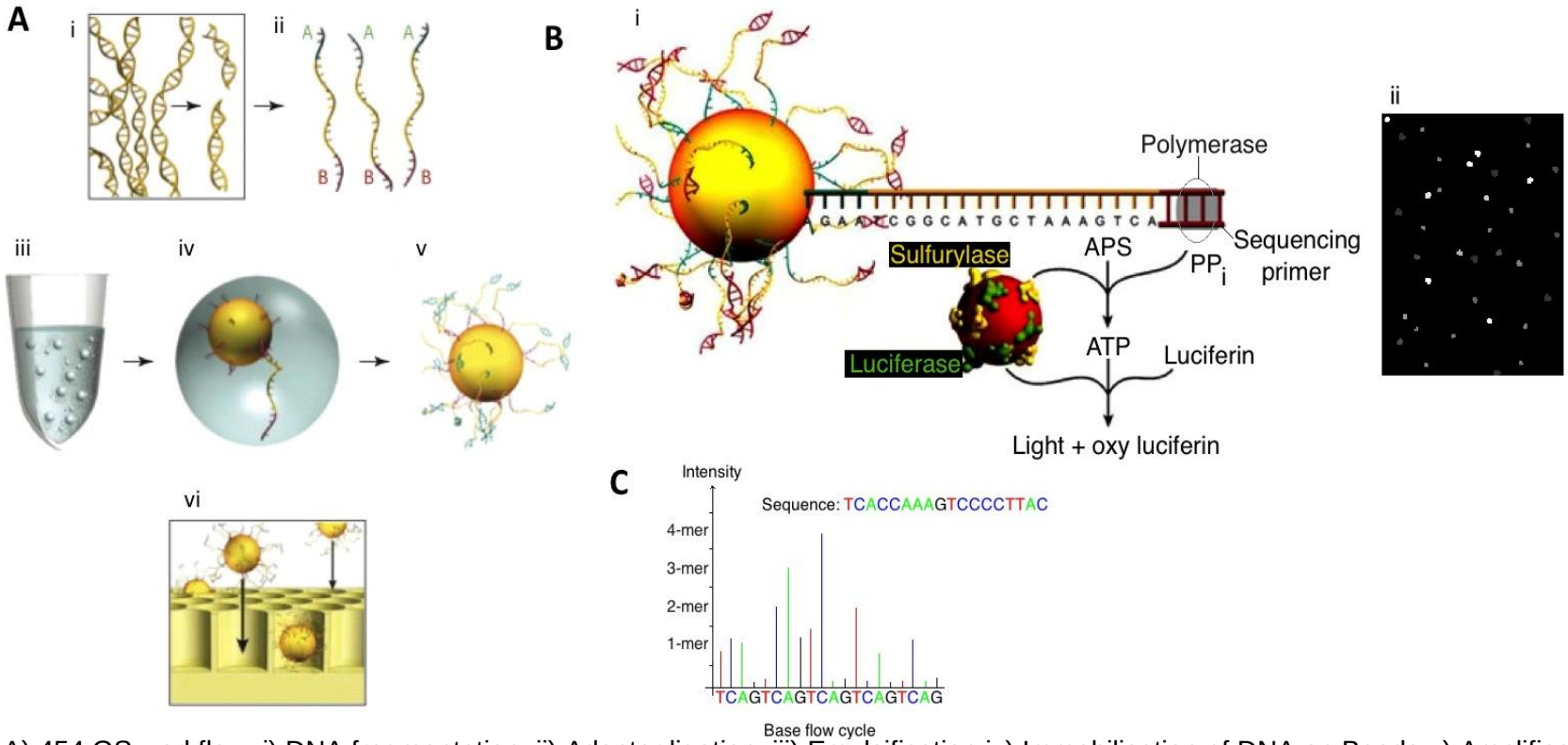
areas of strength and potential biases. The key platforms used within this thesis and related work will be described in detail.

1.4.3.1.1 454 Genome Sequencers

In 2007 and 2010 the GS FLX and the bench top GS Junior systems respectively were released by 454 Life Sciences (Roche). Both adhere to the workflow 'One Fragment = One Bead = One Read' coined by 454. Essentially this enormous simplification refers to the wet chemistry of the GS system (Figure 1) that comprises of four main steps. Step one - the generation of a single stranded DNA template library, containing the 'fragment', which is individually immobilized onto a single 'bead'. Step two - the clonal amplification of this library using emulsion based PCR, which results in several millions copies of the original 'one fragment' on each bead. This is then followed by the sequencing-by-synthesis reaction, where an individual 'bead' occupies a single well of the picotitre reaction plate. The picotitre plate (PTP) is for all intents and purposes a flow cell, which has the four nucleotide bases sequentially incubated within its wells and is washed between cycles. 454 sequencers use Pyrosequencing® technology where incorporation of bases results in a release of light, in this case, via the chemiluminescent enzyme luciferase (Nyren & Lundin 1985; Hyman 1988). Recorded by the instrument camera, the signal strength is equal to that of the number of bases incorporated; this limits the GS system in accurately calling the number of bases in a homopolymeric tract. The imaging of the sequential reactions in this single well produces the third step 'one read', which in the 454 system is an average of 450bp, the longest consistent read length of next generation platforms. All of these steps occur in parallel with one bead in each of the 1,000,000 wells of the PTP, allowing the formation of contiguous and overlapping DNA reads (contigs) and the mapping of series of contigs to form scaffolds to deduce the original sequence. 454 additionally offers a paired end protocol that helps to overcome gaps between reads and contigs forming a more complete scaffold. This involves the circularisation of DNA fragments of a specific length (3, 8 or 20kb) using specialised adaptors, further fragmentation and selection of fragments including the now internalised adaptors. These paired end fragments are now a library of paired stretches of DNA (for example A and B) that were originally, dependant on the library prepared, 3, 8 or 20kb apart, linked by the adaptors. This allows two contigs that do not overlap but that contain DNA 'A' and DNA 'B' to be placed in the scaffold, based on the scaffold location of one of pair. Importantly, paired end sequencing allows inclusion of reads that do not overlap resulting in a more complete scaffold but does not necessarily mean a more complete sequence as the 3, 8 or 20kb sequence between 'A' and 'B' is not known. The superior read length

combined with paired end data, makes 454 GS systems a good choice for *de novo* assembly of reads, which does not require a template to map against. The last step in sequencing, but arguably the most multifaceted, follows with the acts of assembly, annotation and further bioinformatic analysis of the data.

Figure 1. 454 pyrosequencing chemistry



A) 454 GS workflow, i) DNA fragmentation, ii) Adapter ligation, iii) Emulsification iv) Immobilisation of DNA on Beads v) Amplification vi) Picotitre plate loading. B) Pyrosequencing reaction i) Sequencing by synthesis ii) Signal image. C) Base calling intensity flowgram. Figure 1,A and B,i modified in presentation and text from (Droege & Hill 2008; Mardis 2008), B ii and C representations produced for this thesis.

1.4.3.1.2 Illumina sequencing technology

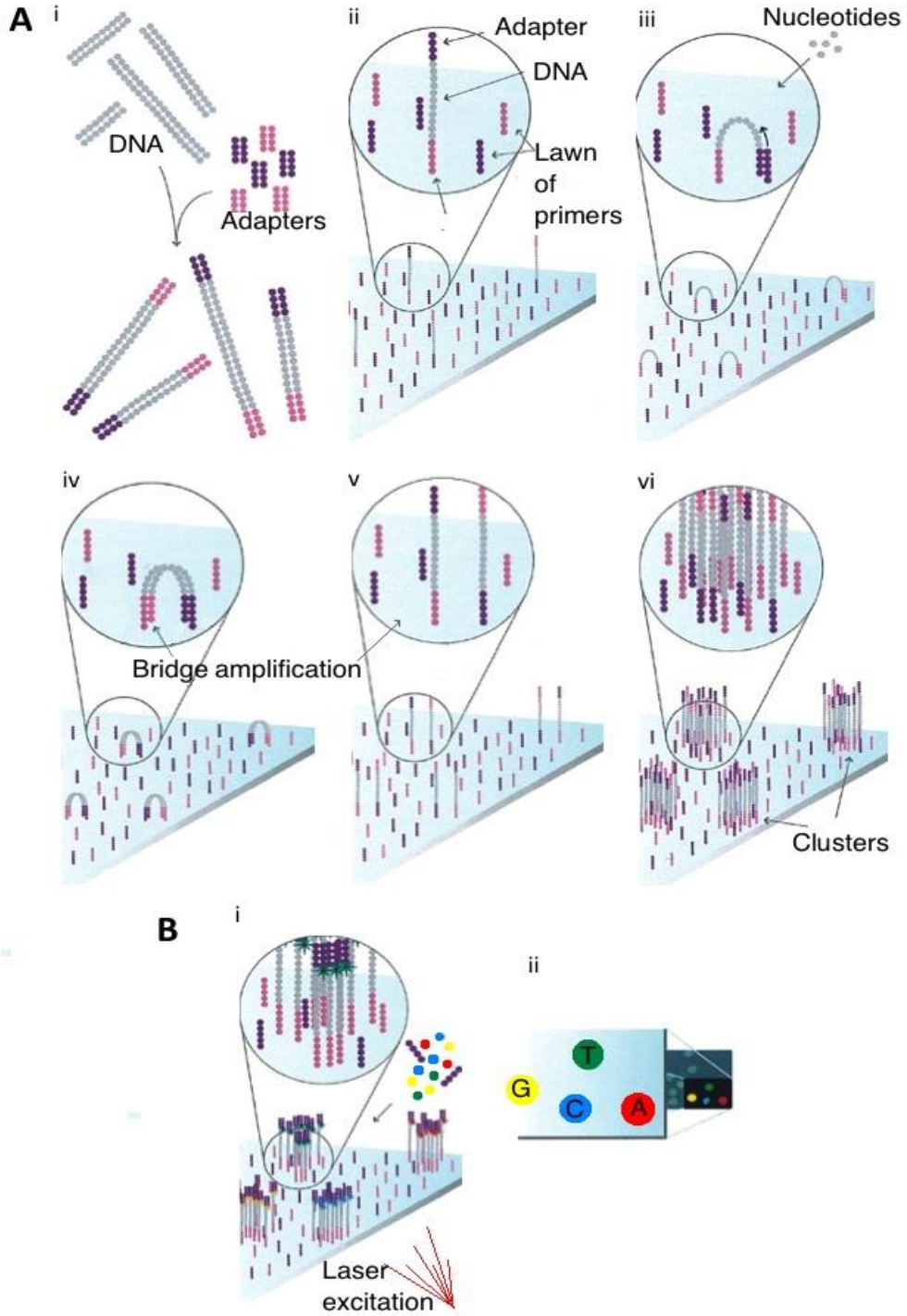
All Illumina platforms share the same chemistry as each other and include the Genome Analyser IIx, the high capacity HiSeq 2500, 2000 and 1000, the 'personal sequencer' MiSeq and finally the HiScan SQ, all of which have slightly different applications. With similarities to the 454 technology, libraries are fragmented and adaptors ligated. In contrast to 454 technology, fragment amplification occurs via the immobilisation of the fragments on a solid surface and cycles of bridge amplification and denaturation, which form clusters of fragments attached to the surface before addition of the sequencing primer (Figure 2). Illumina sequencing chemistry is another example of sequencing by synthesis, but in this case the technology utilises reversible terminator nucleotides labelled with individual fluorescent dyes for each of the four bases. Each cluster is provided with the modified components to allow extension and sequencing in a fluid environment. A single base is incorporated into each cycle due to the inactivation of the 3' hydroxyl group, and this incorporation is detected by an imaging step where the fluorescence of a particular cluster is used to call the base. Finally this is followed by a chemical step to reverse the termination by removal of the fluorescent group. Due to the base by base approach of this technology homopolymeric tracts do not have such a dramatic effect on the data quality as with 454, though in this situation the reads that are produced are at the bottom end of the Illumina read length spectrum of 35-250bp. Illumina technology offers both short insert (200-500bp) paired end and long insert (2-5kb) mate pair libraries which can be combined to greatly improve the overall genome coverage. These protocols, particularly mate pair, differ from 454 paired ends but the principle of linking regions of the DNA, to allow better assembly of these short reads including de novo assembly, remains the same. The shorter read lengths of Illumina sequencing chemistry are also balanced by the increased throughput; in excess of 50 million reads are produced in parallel per flow cell.

1.4.3.1.3 Additional platforms

454 and Illumina sequencing platforms are the most established sequencing technologies at present and microbial whole genome projects and sequencing centres predominantly use these technologies (Lewis *et al.* 2010; Mutreja *et al.* 2011; Everett *et al.* 2012; Croucher *et al.* 2013; Holden *et al.* 2013) as did this thesis and a pilot study of work related to this thesis (Loman NJ *et al.* 2013). However alternative next generation sequencing technologies exists including the Ion Torrent platform, which differs from other technologies primarily in that it utilises a specialised silicon chip to detect the release of hydrogen ions upon base incorporation which is independent of imaging

(Rothberg *et al.* 2011) and the SOLiD platform which is an example of sequencing by ligation (Valouev *et al.* 2008).

Figure 2. Illumina sequencing chemistry



A) Illumina work flow i) Fragmentation and adapter ligation ii) Immobilisation of strands on flow cell iii) Bridge formation iv) Production of double strands v) Denaturing of double strands vi) Clusters generatation complete. B) Sequencing by synthesis i) Laser excitation, ii) Signal image. Select images modified in presentation and text to illustrate key stages from (Mardis 2008)

1.4.3.2 Third generation technologies

Further innovation mean that technologies described above may well be superseded by a third generation of sequencers. Characteristics of the 3rd generation include the sequencing of bases, as they are incorporated real time and the use of single molecules that circumvent the need for PCR amplification reducing pre-sequencing preparation. A number of third generation technologies exist; HeliScope, Helicos Biosciences, GridION and MinION from Oxford Nanopore, and PacBio, Pacific Biosciences.

The PacBio platform promises to produce longer read lengths of up to 1.5kb in considerably shorter run times of around 20 minutes. It does this by using a real time single DNA molecule sequencing method where the polymerase is continuously monitored by a zero-mode wave detector to detect the continuous incorporation of bases (Levene *et al.* 2003). The PacBio long reads, which lend itself well to de novo assembly, combined with the intrinsic sequencing errors of all platforms, has led to combinations of technologies such as PacBio and Illumina being used for whole genome sequencing (WGS) characterisation (Okoniewski *et al.* 2013). In addition the PacBio real time single molecule sequencing had an unexpected significant scientific advantage, the detection of the methylation states of bases that can be applied to epigenetics (Clark *et al.* 2012). The Nanopore MinIon has the potential to be a truly portable example of a sequencer (Oxford Nanopore Technologies 2012). The Heliscope was the first single molecule sequencer to reach the market but Loman *et al* suggest there may not be a future for this platform due to a failing business plan (Helicos Biosciences 2012; Loman *et al.* 2012).

Looking to the future, a joint venture between the Chemistry department and the Optoelectronics Research Centre (OPC) at the University of Southampton secured funding from the Biotechnology and Biological Sciences Research Council (BBSRC) in 2012 to develop an optical DNA sequencer that reads DNA directly (University of Southampton 2012). Further and current information on sequencing platforms, the system workflow advances and comparison between technologies which are constantly improving and evolving in the fast paced technological race can be found on company websites and in a number of reviews (Mardis 2008; Ansorge 2009; Metzker 2009; Loman *et al.* 2012). This race is exemplified in the pursuit of the $1000 genome which represents progress towards personalised medicine and routine human genome sequencing and for which even prizes have been offered (Wade 2006; Kedes & Campany 2011).

1.4.3.3 Bioinformatics

Once the sequencing has been performed bioinformatic analysis is required to handle the resultant data, perhaps the largest constraint of the entire process of genome sequencing. The task of deriving biologically meaningful genetic data from raw sequences is still developing. Mardis has pointed out satirically that for the generation of the $1000 genome the bioinformatics may cost $100,000, to illustrate that there are still great costs within the complete process of sequencing (Mardis 2010). User-friendly bioinformatic advances have not matched the surge of development in the wet chemistry of sequencing, and those that do exist are perhaps premature over simplifications of a complex analytical process. Short read lengths prove challenging for assembly software designed for the longer read length Sanger sequence data, and to date only limited software is available that can easily handle these short sequences as well as the longer reads. High-throughput sequencing has become accessible and consequently the demand for bioinformaticians, computing resources and high performance software is greater than ever before. New pipelines, programs and algorithms specific tailored to short read sequence data and the characteristic errors of platforms are required. Software that is open source has the potential to hasten development in this field and benefit the wider scientific community more than protected technology (Pop & Salzberg 2008). Whole genome data itself is widely published and available, however, whole genome sequences are not routinely 'finished' or even 'whole' which involves sequencing through the gaps, confirming sequence data and permanently annotating the genome, therefore many published genomes contain regions of unknown sequence and fragmented assemblies (NCBI 2011).

The increasing availability and substantial power of next generation technologies is revolutionising the field of genomics, the power of such technologies is increasing rapidly and the possible applications are almost limitless. Utilising such technologies in the name of clinical microbiology, epidemiology and public health is now gathering momentum with translational results becoming reality, with great future potential for tracing and tracking outbreaks among the likely public health applications (Didelot *et al.* ; Rohde *et al.* 2011). Discerning the complete sequence of any microorganism provides genotypic detail which could potentially be used to determine the serotype, MLST and countless other specifics which are of epidemiologic importance. The presence or absence of virulence determinants and heterogeneity of vaccine candidates could also be described from the data. Determining serotype and pneumococcal typing via methods such as MLST, is essential for epidemiological surveillance but a whole genome approach can not only provide this information but also provide a wealth of

information for evolutionary studies, phylogenetics and the description of pneumococcal pan genome. The comparative use of multiple sequences of an organism, serotype or sequence type, combined with the high resolution this data offers allows detailed analysis of the relationship and relatedness of strains which can be used to track outbreaks, origins and mechanisms of antimicrobial resistance, and evaluate vaccine pressure (Lewis *et al.* 2010; Croucher *et al.* 2011; Mullick *et al.* 2011; Everett *et al.* 2012; Croucher *et al.* 2013; Holden *et al.* 2013).

1.5 Control of pneumococcal disease

The high economic burden and potentially fatal nature of pneumococcal disease requires medical intervention. Antibacterial agents can be used to treat pneumococcal infection, the route of administration dependent on the clinical manifestation. Empirical prescription of antibiotics based on clinical condition is however a relatively non-specific intervention. In the presence of microbiological confirmation or high suspicion of pneumococcal aetiology, narrower spectrum antibacterials are usually administered. However, antibiotic resistance emerged, in particular to penicillin, soon after its first usage in the 1960's (Hansman & Bullen 1967). There are a rising number of pneumococcal lineages which exhibit antibiotic resistance, particularly associated with serotype 19A and 23F in the USA and Spain respectively, making antibiotic treatment more difficult (McGee *et al.* 2001; Centers for Disease Control and prevention 2007; Croucher *et al.* 2009; Pillai *et al.* 2009; Jones *et al.* 2010; Croucher *et al.* 2011). Song *et al* comprehensively reviewed the relationship between pneumococcal serotype and antibiotic resistance, particularly the global observation that after PCV7 introduction there was an increase in non-vaccine type antibiotic non-susceptibility relating to the phenomenon of serotype replacement in countries with a high level of vaccine uptake (Song *et al.* 2012). This increase could be predominantly explained by the expansion of previously documented antibiotic resistant serotype specific lineages in the US rather than increased acquisition of resistance (Gertz *et al.* 2010) with the exception of a 19A clonal complex 271 resulting from a suspected capsular switch event which was not detected before PCV7 (Pai *et al.* 2005). However this CC271 multi drug resistant 19A was observed in the first year of PCV7 use in the US and suggests that the capsule switch event may well have occurred pre-PCV7 and been selected for as has been concluded for other lineages (Croucher *et al.* 2013). Serotype independent vaccines were suggested as a method to further control pneumococcal resistance by Song *et al* by preventing replacement and therefore associated rises in NVT antibiotic resistance (Song *et al.* 2012). However antibiotic resistance in pneumococci has been shown to

have a statistical association with country specific antibiotic usage in Europe, therefore tighter regulation of antibiotic usage that could reduce non-susceptibility for more than just pneumococcal disease would also seem an important policy (van de Sande-Bruinsma *et al.* 2008).

Protection from disease is a better means of preventing morbidity, mortality and the associated financial, personal and social costs of treatment. Therefore a number of pneumococcal vaccines have been developed by academics and the pharmaceutical industry.

1.5.1 Pneumococcal vaccines

The first pneumococcal vaccine trials occurred as early as 1911, using whole-cell heat-killed pneumococci in miners in the absence of understanding serotype specific disease potential (Wright *et al.* 1914). Grabenstein and Klugman give an account of the progression of pneumococcal vaccine as a centenary review describing the initial development of both multi-valent purified free polysaccharide and protein-conjugated polysaccharide vaccines (Avery & Goebel 1929) in the 1930's (Grabenstein & Klugman 2012). In the advent of antibiotics Grabenstein and Klugman also reviewed the clinicians preference for antimicrobial therapy rather than mass vaccination and consequent decreased vaccine development and use (Grabenstein & Klugman 2012), until the early 1980's with the development of a high valency polysaccharide vaccine.

The 23-valent pneumococcal polysaccharide vaccine (PPV, Pneumovax^{TM}) manufactured by Merck was first licensed over a quarter of a century ago. This vaccine includes the polysaccharide capsule from 23 of the most common disease causing serotypes (Table 2). Pneumovax^{TM} however, relies on T-cell independent immune stimulation by polysaccharides (Stein 1992). This mechanism is under developed in children of less than two years old (Stein 1992), a major at risk group for pneumococcal disease (O'Brien *et al.* 2009). PPV efficacy in those over 65 years of ages has been recently called into question due to a deficit of supportive evidence (Huss *et al.* 2009; Department of Health 2011b; Department of Health 2011a; Department of Health 2011c). However current recommendations within the UK is that PPV is recommended for use for the elderly and is administered to other younger adult risk groups including the immunocompromised (Didelot *et al.* ; Melegaro & Edmunds 2004a; Huss *et al.* 2009). To reduce pneumococcal disease in young children of <2 years of age, vaccines were developed to trigger T-cell immunity against polysaccharides via conjugation to an adjuvant protein.

Introduction

The early development of protein-pneumococcal polysaccharide conjugation in the 1930's (Avery & Goebel 1929) did not readily advance to clinical value until the end of the 20th century. Prevenar™, a seven valent pneumococcal conjugate vaccine (PCV7, Pfizer) was developed and subsequently licensed in 2000 but unlike the US it was not added to the UK childhood immunisation schedule until 2006 (Department of Health 2006). This vaccine contains the capsular polysaccharide of seven serotypes, conjugated to CRM197, a non-toxic diphtheria variant carrier protein (Eskola et al., 2001). This immunogenic protein increases the vaccine efficacy in young children by inducing a T-cell-dependent response (Black et al., 2000; Rennels et al., 1998). The seven serotypes included in the vaccine – 4, 6B, 9V, 14, 18C, 19F and 23F – were selected as prior to use of the vaccine they caused the majority of invasive disease in the US (Hausdorff et al., 2000) (Table 2). These serotypes were also associated with antibiotic resistance prior to vaccine implementation (Hicks et al., 2007; Tyrrell et al., 2009).

In the conjugate vaccine era there is debate over the continued use of PPV regarding its use alongside conjugate vaccines and its complete replacement by conjugate vaccines. There is evidence to suggest blunting of the PCV response in adults previously vaccinated with PPV when compared to those who had only received PCV, this observation did not result in PCV inferiority to PPV for the shared serotypes and therefore the clinical extrapolation of this phenomenon is limited (de Roux *et al.* 2008; Musher *et al.* 2008; Jackson *et al.* 2013). Additionally one dose of PPV after the PCV course is recommended for children >2 <5 years old to broaden serotype immunity (Department of Health 2012). The clinical benefit of this additional vaccination in children between 2 and 5 has also been called into question with evidence suggesting that there was no significant increase in the number of children that had protective antibody concentrations for the serotypes in PPV but not in PCV7, with a IgG1 response after PCV7 with PPV rather than IgG2 with PCV7 or PPV alone (Uddin *et al.* 2006; Balmer *et al.* 2007). A recent review of the subject after seven years of discussion highlights the lack of evidence and agreement on PPV use in children (Borrow 2013).

Second generation pneumococcal conjugate vaccines followed, GlaxoSmithKline's (GSK) 10-valent PCV (Pneumococcal *H. influenzae* protein D conjugate vaccine (PHiD-CV), Synflorix™) and Prevenar 13™ (PCV13, Pfizer) were licensed in 2010. PHiD-CV and PCV13 are now both licensed in the UK, with the latter superseding Prevenar in the UK childhood immunisation schedule in April 2010 (Department of

Health 2010). These vaccines target additional serotypes that are important to current disease incidence and are not well targeted by PCV7 (Table 2).

The PHiD-CV developed by GSK includes two capsular polysaccharide types conjugated to either diphtheria (serotype 19F) or tetanus toxoid (serotype 18C), and eight others (1, 4, 5, 6B, 7F, 9V, 14, 23F) conjugated to non-typeable *Haemophilius influenzae* (NTHi) protein D (Wysocki *et al.* 2009) (Table 2). NTHi protein D was included in PHiD-CV to provide potential protection against AOM caused by NTHi. Carriage of NTHi and episodes of AOM were shown to reduce in a clinical trial of an 11-valent PHiD-CV experimental vaccine but the aetiology of AOM is inherently difficult to determine. (Prymula *et al.* 2006; Prymula *et al.* 2009; Wysocki *et al.* 2009). An effect on carriage of NTHi and any real effect on NTHi AOM could positively influence the impact and cost effectiveness of the GSK pneumococcal vaccine (Kim *et al.* 2010; Giglio *et al.* 2011; Robberstad *et al.* 2011). Originally the GSK experimental vaccine included 11 serotypes but the inclusion of serotype 3 was later rejected due to a lack of inducible immunogenicity in clinical trials (Prymula *et al.* 2006). PCV13 developed by Wyeth (now Pfizer) targets the same pneumococcal serotypes as the PHiD-CV plus three additional serotypes (3, 6A and 19A), all conjugated to the immunogenic diphtheria toxoid (Scott *et al.* 2007). There is still some debate surrounding the immunogenicity of serotype 3 in PCV13, however studies have shown using opsonophagocytic assays that functional antibodies are induced (Centers for Disease Control and prevention 2010; Yeh *et al.* 2010).

1.5.1.1.1 Use of conjugate vaccines

Whilst conjugate vaccines are freely available on the market it is predominantly countries of the developed world that have introduced them into routine immunisation schedules, notably the UK and the US (Centers for Disease Control and prevention 2011; Department of Health 2012; Gladstone *et al.* 2012a). However as a large burden lies in the developing world introducing vaccination schedules in these countries is arguably more important (Johnson *et al.* 2010). GAVI have been going some way to rectify this disparity by supporting vaccination schedules in ~100 GAVI eligible countries. This has been through advance market commitments to purchase vaccines and agreements to fixed sustainable prices with pharmaceutical companies (GAVI 2011; Gladstone *et al.* 2012a).

1.5.1.2 Vaccines in development

An investigational 15-valent PCV is being developed by Merck (Skinner *et al.* 2011). This investigational vaccine is formulated in a similar manner to PCV13, except for the inclusion of two additional serotypes 22F and 33F (Skinner *et al.* 2011). Expanding valency of vaccines is a response to serotype replacement observed globally for carriage (Hanage *et al.* 2011; Tocheva *et al.* 2011) and disease caused by PCV7 types and by extrapolation, expected for PCV13 types (Hicks *et al.* 2007; Miller *et al.* 2011b). Serotypes 22F and 33F were selected for inclusion due to increases in their prevalence in IPD, resulting in these serotypes being primary causes of IPD in the US after PCV13 vaccine types (Pilishvili *et al.* 2010; Skinner *et al.* 2011). A phase I trial of the 15-valent experimental vaccine (NCT01215175) has been completed for healthy adults and toddlers. A phase II, randomised, multicentre, double blind clinical trial of the vaccine is now being carried out by Merck (NCT01215188) to assess its non-inferiority to Prevenar 13™ in healthy infants with its primary outcome completed in July 2012 (U.S National institutes of Health 2012). The uptake of the next generation conjugate vaccines will in part depend on the extent to which disease replacement is observed after widespread routine use of PCV13 –specifically upon changes in the epidemiology of the additional serotypes. The ability to further expand the valency of conjugates in response to replacement is limited both by manufacturing restraints and hypothetical antigenic overload. Antigenic overload is a theoretical issue as vaccines are routinely given concomitantly with immunogens from a wide range of pathogens, which could blunt the immune response of each component, however there are those that argue that there is no evidence for this hypothesis (Gregson & Edelman 2003). Finally and perhaps most importantly there are prohibitive costs per dose incurred through expanding the valency.

1.5.1.2.1 Serotype independent vaccines

Serotype dependant vaccines do not prime broad-spectrum pneumococcal immunity. They are composed of purified, free or conjugated, polysaccharides from a limited set of serotypes. Serotype-independent vaccines would circumvent this issue, and a number are in development including killed whole cell vaccines (WCV), with formulations being tested in mouse models to determine the optimal aluminium adjuvant stimulation of the desired cytokine profiles (Hogenesch *et al.* 2011; Lu *et al.* 2011). The administration of killed whole cell vaccine has already been shown to illicit both T-cell dependant IL-17A and antibody responses (Lu *et al.* 2011).

Another alternative is the use of highly conserved, immunogenic pneumococcal surface proteins that are involved in bacterial virulence. Single and multi-component protein-based vaccines are being investigated by analysing the protective capacities of a number of different proteins, protein combinations, and protein fusions to maximise synergistic effects. Protein candidates include pneumococcal surface protein A (PspA), pneumococcal surface protein C (PspC), pneumolysin (Ply) and caseinolytic protease (ClpP) (Hamel *et al.* 2004). At present a number of candidate proteins are being evaluated in murine models, including the demonstration of passive immunity from polyclonal antibodies against specific pneumococcal proteins (Cao *et al.* 2007; Ogunniyi *et al.* 2007; Morsczeck *et al.* 2008; Cao *et al.* 2009; Olafsdottir *et al.* 2011; Zhang *et al.* 2011). Protein based vaccines could potentially give broad protection from pneumococcal infection independent of serotype and invasiveness.

1.5.1.2.2 Mechanisms of protection

Establishing protection against otitis media and carriage would ideally require stimulation of a mucosal antibody response (Zhang *et al.* 2002). If proven to be safe in administration, efficacious and capable of conferring indirect effects to a vaccinated population serotype independent vaccines could offer greater potential for controlling pneumococcal disease. Additional advantages of serotype independent vaccines could include the induction of T-cell responses and ease of vaccine formulation resulting in reduced production costs when compared to conjugate vaccines.

Additional approaches are aimed at stimulating the immune system in a similar way to natural colonisation. The WHO only supports vaccine delivery via the intramuscular or subcutaneous routes due to the current lack of evidence of the immunogenicity and safety of alternative delivery (World Health Organisation 2008). Despite this, research is being carried out into mucosal delivery of vaccines (Malley *et al.* 2001; Douce *et al.* 2010). Malley and colleagues have shown protection against colonisation and sepsis in mouse models with both mucosal and peritoneal immunisation with WCV (Malley *et al.* 2001; Lu *et al.* 2011). Additional research is investigating the potential of pneumolysin as an adjuvant for improving the mucosal stimulation by proteins which do not normally induce strong immune responses (Douce *et al.* 2010). However, broad-spectrum protein or whole cell vaccines could potentially have major effects on overall pneumococcal carriage with as yet unknown clinical significance of non-pneumococcal bacterial replacement.

Table 2. Details of licensed pneumococcal vaccines, serotype inclusion and current use within the UK

Vaccine	Manufacturer	Serotypes	UK Status
PPV (Pneumovax™)	Merck	1, 2, 3, 4, 5, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19F, 19A, 20, 22F, 23F, 33F	Licensed for use in adult high risk groups
PCV7 (Prevenar™)	Pfizer	4, 6B, 9V, 14, 18C, 19F, 23F	Licensed, superseded by PCV13
PHiD-CV (Synflorix™)	GSK	**1**, 4, **5**, 6B, **7F**, 9V, 14, 18C, 19F, 23F	Licensed
PCV13 (Prevenar 13™)	Pfizer	**1, 3,** 4, **5**, **6A**, 6B, **7F**, 9V, 14, 18C, **19A**, 19F, 23F	Licensed, included in childhood immunisation schedule

Bold underlined text indicates vaccine serotypes within conjugate vaccines that are additional to those contained within PCV7.

1.5.2 Serotype replacement

Replacement is an epidemiological phenomenon that occurs when targeting a sub population where the decrease in vaccine types is offset to some extent by a rise in prevalence of the subtypes not targeted that replace the vaccine types. Serotype replacement was predicted in advance of PCV use (Lipsitch 1997). Following PCV7 introduction in the US, a shift in the prevalent serotypes circulating in the population and causing disease was observed, termed 'serotype replacement' (McEllistrem *et al.* 2003). This was predicted to be mirrored in the UK (Spratt & Greenwood 2000).

Although there has been a dramatic reduction in vaccine type (VT) IPD, the phenomenon of "replacement" has occurred in both disease and carriage in the UK (Miller *et al.* 2011b; Tocheva *et al.* 2011). Replacement disease is due to the increase in non-vaccine serotype (NVT) IPD cases offsetting the decrease in VT IPD. The incidence of IPD per 100,000 in the population of England and Wales between 2000 and 2006 was 16.1, in 2008/09 it had reduced to 10.6 per 100,000 (Miller *et al.* 2011b). Importantly, there was an increased incidence of IPD caused by PCV7 NVT detected in all age groups, particularly involving serotypes 7F, 19A and 22F as early as 2009 (Public Health England 2009). An increase of 100% was observed for serotype 7F for all age groups, 372 cases to 742 cases between 2000-2006 and 2008-2010 after adjustments were made for reporting bias (Miller *et al.* 2011b). Serotype 7F is however, included in PCV13, 22F though is not targeted by the PHiD-CV or PCV13. 22F has also dramatically increased in IPD prevalence in children less than two years of age in England and Wales and was ranked 6th in children aged less than five years for Scotland in 2009 (Shakir 2009; Miller *et al.* 2011b).

The inclusion of a serotype in a vaccine was previously assumed to infer some level of cross protection against other members of the same serogroup the so-called vaccine related types (VRT), this assumption has not borne out (Hausdorff *et al.* 2000; Jacobs *et al.* 2008). Since PCV7 implementation VRT 19A IPD incidence has increased in the US, UK and elsewhere despite the fact that the related serotype 19F was included in the vaccine (Jacobs *et al.* 2008; Pillai *et al.* 2009; Miller *et al.* 2011b). The 19F polysaccharide is known to be the least immunogenic of the PCV7 VTs (Pletz *et al.* 2008). In addition cross reaction of antibodies stimulated by PCV7 against 19F on 19A has also been shown to be weak *in vitro* (Lee *et al.* 2009).

Since PCV7 introduction an increase in the prevalence of serotype 6C have been observed for both carriage and disease worldwide (Nahm *et al.* 2009; Green *et al.*

2011; Miller *et al.* 2011b; Tocheva *et al.* 2011). The rapid and clonal expansion of 6C was observed to be significant within the first three years of the on-going pneumococcal carriage study in Southampton (Tocheva *et al.* 2010; Tocheva *et al.* 2011). This is despite the presence of the 6B polysaccharide in PCV7 that does provide protection against 6A (Vakevainen *et al.* 2001). Serotype 6B cross reactivity does not extend to 6C, therefore the PCV7 elicits negligible or no immune protection against this serotype (Park *et al.* 2007a). The inclusion of 6A in PCV13 (6A also included in PHiD-CV) has been shown to offer cross-functional antibodies against 6C and should go some way to protecting against 6C disease (Cooper *et al.* 2010; Cooper *et al.* 2011).

Serotypes related to the vaccine types have contributed to serotype replacement more than was perhaps first expected (Hanage 2008; Jacobs *et al.* 2008; Ladhani *et al.* 2008). It could be supposed that vaccine related types were in a prime position, sharing immunological epitopes and to a lesser extent genotype due to genome plasticity, to fill the specific niche vacated by their counterpart in an environment under vaccine pressure.

It must however be noted that serotype distribution can fluctuate substantially in the absence of vaccination. A highly significant increase in serotype 1 was observed within the UK prior to routine PCV7 immunisation highlighting that other factors are also involved in pneumococcal serotype dynamics (Kirkham *et al.* 2006; Jefferies *et al.* 2010a). The most common serotypes causing IPD vary with both location and time (Harboe *et al.* 2010; Jefferies *et al.* 2010a).

Although not the only influence, PCV7 is likely to have been playing an important part in serotype fluctuations in the UK by reducing VTs and creating a niche, which is being filled by NVT or potentially other species of bacteria. On-going surveillance, typing and research will help uncover the reasons why certain species and strains appear better at filling this niche than others, and why some cause invasive disease while others cause little or none at all.

Traditionally the serotypes chosen for vaccine inclusion have been based on the rank order incidence of disease. These serotypes are often the most prevalent in carriage but they are not necessarily those with the highest potential for invasiveness. By targeting the serotypes in rank order of disease incidence, any serotype replacement that occurs may result in increased prevalence of a particularly invasive serotype. One study highlighted that moderately prevalent NVT serotypes 3, 8, 33 and 38 all had a similar potential to cause invasive disease when compared to the VT 6B, 19F and 23F

(Brueggemann *et al.* 2003), previously responsible for a considerable proportion of IPD in the pre vaccine era (Miller *et al.* 2011b). If a shift in carriage prevalence occurred, for example, reduced VT 23F with increased NVT 8, this could then potentially result in serotype 8 having a similar disease incidence as VT 23F previously. NVT 19A and 7F that have recently increased as a cause of IPD in the UK (Miller *et al.* 2011b) have also previously been reported to be associated with invasive disease (Brueggemann *et al.* 2003; Sjostrom *et al.* 2006). This may indicate that greater virulence of certain serotypes has been central to the rise in case numbers rather than an expansion of clones with these serotypes. Examination of the genetic background of serotypes causing invasive disease suggests the acquisition of an alternative capsular locus has occurred in some cases which has allowed vaccine escape (Brueggemann *et al.* 2007; Temime *et al.* 2008; Hanage *et al.* 2011; Golubchik *et al.* 2012). Capsular variants are likely to have existed in the in pneumococcal population before vaccine implementation and the selective pressure resulted in the expansion not the occurrence (Croucher *et al.* 2011).

The presence of additional serotypes in PHiD-CV and PCV13, as compared to PCV7 should, to some degree, help to protect against the emergence of some previously under-represented serotypes with significant invasive potential.

1.5.3 Vaccine coverage

Historically, PCV7 covered 90% of the serotypes causing IPD in the US. Crucially, even before PCV7 introduction the disease serotype coverage in the UK was much lower. In Scotland only 76.5% of all IPD cases in those aged less than five years old were calculated to be covered by PCV7 (Clarke *et al.* 2006), whilst in the developing world coverage was suggested to drop as low as 45% for IPD in Asia (Hausdorff *et al.* 2000).

Based on national (England and Wales) surveillance data, the percentage of serotypes causing cases of IPD covered by PCV7, PHiD-CV and PCV13 were calculated (Public Health England 2009). In 2007-2008 only 24% of IPD cases in those under 5 years old were caused by serotypes covered by PCV7, in stark contrast to the 76.5% UK estimate based on IPD coverage in this age group prior to vaccine implementation (Clarke *et al.* 2006). The serotype coverage of IPD in children under five years of age for PHiD-CV and PCV13 was 53% and 74% respectively for 2007/8, a dramatic decrease from the 81% and 92% 2005/6 coverage. A fundamental observation is that the potential coverage of PHiD-CV and PCV13 had already decreased prior to

implementation due to the routine use of PCV7 and the associated serotype replacement, as well as shifts in pneumococcal epidemiology caused by other non-vaccine factors.

A decrease in VT IPD incidence has been seen in children over the age of five and adults, who are largely unvaccinated (Public Health England 2012). Herd protection against VT pneumococci has been induced in the UK population as an indirect effect of infant PCV7 immunization (Miller *et al.* 2011b). Herd protection is primarily due to reduced exposure through decreased carriage and transmission from the vaccinated population. These indirect effects contribute extensively to the overall impact and cost-effectiveness of vaccination (Melegaro & Edmunds 2004b; Centers for Disease Control and Prevention 2005). Without such effects, the PCV introduction may not have been considered economically viable in the UK (Melegaro & Edmunds 2004b). The level of herd protection has been suggested to increase with the number of doses given (Haber *et al.* 2007). Prior to this observation, surveillance carried out in the US demonstrated a 42% fall in the incidence of IPD in infants, 90 days old (Carter 2006), indicating that herd protection can also extend to those not yet old enough to be vaccinated or to have completed the vaccination course. A US model also predicted that even incomplete coverage and/or limited dose schedules would still confer indirect protection (Haber *et al.* 2007).

1.5.4 Vaccine evaluation

Clinical trials and mathematical models offer a basis for prediction of vaccine impact. Studies using such algorithms suggested that the PHiD-CV will be at least as effective as PCV7 in protecting against pneumococcal invasive disease worldwide (Hausdorff *et al.* 2009). The design of such an algorithm is complex due to the fluidity and dynamics of the pneumococcal population and has to be based on assumptions that may affect the model output. This is not untrue of the Hausdorff model, which did not take account of the positive impact of herd immunity, the impact on NTHi AOM or the negative influence of serotype replacement on vaccine success. Despite being a valuable tool, the conclusions of any model should be treated with caution.

In clinical trials the study population will naturally be exposed to multifaceted epidemiological factors that will affect vaccine impact. Therefore clinical trials should reflect more accurately, the ‘true’ impact of a vaccine even under their tightly controlled conditions, but this does not include herd protection, which requires high coverage in the population. A German clinical study, powered to show immunological non-inferiority

showed that PCV13 should be just as efficacious as PCV7 at protecting against the seven serotypes included within the PCV7, as well as inducing sufficient immunity for the further VT (Kieninger D. M *et al.* 2008). Importantly PCV13 was shown to induce an opsonophagocytic response to serotype 19A that indicates it will be efficient in preventing cases of serotype 19A invasive disease (Kieninger D. M *et al.* 2008). Surveillance of carriage has an important place in vaccine evaluation for a number of reasons, carriage is a precursor to disease so changes in carriage will affect disease epidemiology even when serotypes has differing potential for disease (Brueggemann *et al.* 2003), additionally the carriage state is vastly more common than disease and statistically significant changes can be detected rapidly due to the larger sample size available (Tocheva *et al.* 2010; Miller *et al.* 2011b). Clinical trials offer valuable but incomplete information on the potential outcome of a vaccination campaign, therefore vaccines require efficacy trials and both disease incidence and carriage surveillance to evaluate effectiveness after introduction.

Both PHiD-CV and PCV13 are likely to be effective in further reducing IPD and non-invasive pneumococcal disease, however the additional serotypes unique to PCV13 were of clinical importance in the UK and contributed to the decision to replace PCV7 with PCV13 (Public Health England 2009; Department of Health 2010). Nonetheless, the relative effect of PHiD-CV compared to PCV13 immunisation on the combined prevention of pneumococcal and NTHi invasive or all cause disease has not yet been tested in a head to head trial. As well as the effect of routine vaccination on invasive disease it also affects carriage. Carried serotype prevalence for the most part can be related to invasive disease cases despite individual serotype invasive propensities as high incidence of colonisation offers increased opportunities to progress to disease (Miller *et al.* 2011b; Tocheva *et al.* 2011). Public Health England (PHE) has reported the enhanced surveillance of IPD serotypes in England and wales (Miller *et al.* 2011b; Miller *et al.* 2011a)

1.5.4.1 Disease specific impact

The impact of conjugate vaccination on specific disease manifestations such as pneumonia is harder to quantify but studies such as the Clinical Otitis Media and Pneumonia Study (COMPAS NCT00466947) have been designed to try to answer these questions (U.S National institutes of Health 2012). Pneumococcal pneumonia cases are difficult to define as they can result in pneumococcal invasion of bacteria into the blood (Jacups & Cheng 2011). Despite this variation, studies have indicated that PCV7 efficacy for clinical and x-ray defined pneumonia is between 5% and 25% (Black

et al. 2002; Lucero *et al.* 2009). In addition, the incidence of pneumonia-related hospitalisation in the US of children under two years old was observed to decrease by up to 39% after PCV7 introduction (Grijalva *et al.* 2007; Centers for Disease Control and Prevention 2009). All-cause pneumonia cases remained stable in the US during PCV7 introduction, a potential example of replacement disease contributed to by other species not targeted by PCVs, but it is however also possible that it is due to an increase in reporting after vaccine deployment (Centers for Disease Control and Prevention 2009).

Enhanced disease surveillance during 2004-2009 revealed changes in the molecular epidemiology of meningitis with serotype 7F with ST191 being the most prevalent cause of meningitis by 2008-2009 (Pichon *et al.* 2013). Bacterial pneumonia and empyema admissions in England were on the increase before the introduction of PCV7 and began decreasing between 2006 and 2008 (Koshy *et al.* 2010). Conversely another recently published study of pneumonia aetiology in children suggested that pneumococcal pneumonia in children of less than 16 years of age was comparable between the pre and post PCV with PCV7 non-vaccine serotypes continuing to be a significant cause of childhood pneumonia in the UK (Elemraid *et al.* 2013). However the age of the subjects is important, 25% of the subjects were not eligible for vaccination due to age, whilst the majority of participants were under the age of 5 and all serotypes recovered are included in PCV13.

Pneumococcal conjugate vaccines have been shown to offer some protection against AOM (Eskola *et al.* 2001; Prymula *et al.* 2006). As the clinical diagnosis of AOM is to some extent subjective and the definitive identification of the causative agent difficult, it is important to consider that data from these studies is useful but not necessarily representative of the true situation. AOM incidence has been shown to decrease in the US post PCV7 (Eskola *et al.* 2001; Black *et al.* 2004; Block *et al.* 2004; Prymula *et al.* 2006). However US cases of otitis media due to NVT were also seen to increase in incidence in the post-PCV7 era, a 10% rise in NVT was reported by one study (Block *et al.* 2004), along with observations of capsular switch events (McEllistrem *et al.* 2003). Serotype data for AOM in the UK and Europe is scarce (Rodgers *et al.* 2009). Data from the US suggests that serotype replacement, capsular switch and species replacement may limit the effectiveness of pneumococcal vaccination against otitis media in the UK. *H. influenzae,* specifically NTHi, is also an important cause of AOM (Casey *et al.* 2009). NTHi AOM was seen to increase by 15%, following wide-spread PCV7 vaccination in the US (Casey *et al.* 2009), a situation which the PHiD-CV may be

better able to control. It could be supposed that increases in NTHi carriage and AOM may fill part of the niche left by pneumococcal VT in the UK.

Vaccination is essential for the prevention of pneumococcal disease. As currently licenced vaccines target only a subpopulation of dynamic and diverse *S. pneumoniae* serotypes that are not fixed to genetic lineages, vaccine pressure is likely to cause epidemiological shifts that need to be monitored and reacted to.

1.6 Study rationale & design

As *S. pneumoniae* is responsible for a substantial burden of disease worldwide, it is a humane responsibility to act to control and prevent pneumococcal disease. Interventions to this end, such as national vaccination programs, aim to prevent disease cases in the vaccinated but also act indirectly in the unvaccinated community via herd protection.

Years of research have been undertaken to elucidate the mechanisms of immune stimulation and protection to allow rational vaccine design. Even with evidence such as this it is not possible to make exact predictions regarding the level of effectiveness of vaccination within a population. Particularly in respect to any consequent changes in the epidemiology of colonisation and disease. Therefore it is essential to monitor any impact, positive or negative to allow the pharmaceutical industry, policy makers and health service providers to respond to unforeseen or exaggerated effects.

The human nasopharynx is potentially inhabited by multiple bacterial species, strains and other microorganisms. Finite nutrients and space are thought to result in competition for colonisation (Murphy *et al.* 2009). Intra species competition and niche specificities occur for pneumococci; serotypes and even STs are known to differ in their association with colonisation and disease (Lipsitch 1997; Rajam *et al.* 2007; Auranen *et al.* 2010). Any intervention that is likely to result in removal of constituents of a niche will affect the complex balance and dynamics resulting in reconstitution of the new natural equilibrium. Therefore targeting a subpopulation of pneumococcal types with conjugate vaccines has the potential to alter the epidemiology of colonisation, both pneumococcal and total species diversity.

The present study was undertaken to characterise pneumococci in the nasopharyngeal niche during conjugate vaccine introduction. Determining the prevalence of pneumococcal nasopharyngeal carriage, and monitoring serotype and genotype distribution in healthy children, during conjugate vaccine introduction allowed this evaluation. The surveillance was conducted over a five-year period from the introduction of the first pneumococcal conjugate vaccine in the UK in 2006. The study aimed to detect changes in the circulating pneumococcal population and provide data that could inform current and future vaccine policy.

Next generation sequencing technologies that allow whole bacterial genome sequencing are available and increasingly affordable. The work described in this thesis seeks to validate and utilise such technology for comprehensive typing and analysis of

the carriage study isolate collection, as a cost and time efficient alternative to traditional methods for epidemiologists.

To date, a small number of completed pneumococcal genomes have been published (Donati *et al.* 2010; Camilli *et al.* 2011) and many more draft genome sequences are at various stages of completion or are part of on-going sequencing projects (Institute 2011; NCBI 2011; The Sanger Institute 2011). These genomes predominantly represent laboratory strains and IPD clinical isolates (Dopazo *et al.* 2001; Hoskins *et al.* 2001; Tettelin *et al.* 2001; Lanie *et al.* 2007; Ding *et al.* 2009).

The determination of whole genome sequences described in this thesis will add value to current publically available genomes; providing the genomic sequence of a large number of carried pneumococci that are currently under-represented. This will include a diversity of serotypes and genotypes associated with colonisation that could be used as comparators to invasive isolates.

1.7 Hypothesis

The hypothesis of this thesis was that during PCV implementation there would be a shift in pneumococcal carriage epidemiology. This was predicted to occur through vaccine serotype and associated genotypes decreases with subsequent replacement by non-vaccine serotype and associated genotypes

1.7.1 Aims

- To infer how PCV implementation has impacted pneumococcal carriage epidemiology for extrapolation to disease and contribute explanation for current disease epidemiology.
- Deduce epidemiologically relevant data from whole genome sequences for surveillance purposes.

1.7.2 Objectives

- To determine the pneumococcal carriage rate, characterise the carried pneumococci and detect and temporal changes in the carriage rate and serotype and genotype prevalence.
- To validate the robust inference of serotype and determination of genotype from whole genome sequences by comparing to data generated by traditional methods.

2. METHODS

2.1 Samples

Bachelor of Medicine (BM) and subsequently BMedSci students, were trained to perform nasopharyngeal swabbing by qualified research nurses. They then collected nasopharyngeal swabs from children aged four years and under attending the outpatient department of Southampton General Hospital, as part of their 4th year BM/BMedSci research project. Written parental/guardian informed consent was obtained with ethical approval for this procedure from Southampton and South West Hampshire Research Ethics Committee 'B' (REC 06/Q1704/105). Age was the primary exclusion criteria, additionally only one member of an approached family was swabbed with one exception in year 2. The child swabbed was not necessarily the child attending the clinic. Rayon tipped, Transwabs (Medical wire, Corsham, UK) with flexible shafts were used to obtain samples and were immediately immersed into the charcoal Amies media. Swabs were taken during the winter months of October to March of 2006/7, 2007/8, 2008/9, 2009/10 and 2010/11 until ≥100 *S. pneumoniae* were isolated within each sampling period.

In period 2010/11 a questionnaire was introduced for the parent/guardian to complete after the informed consent process. This questionnaire allowed metadata, of relevance to potential pneumococcal carriage, to be collected on the participants including age, recollected vaccination history, and recent antibiotic usage and respiratory illness which could be extrapolated to the study population for the entire study period. The primary aim of the carriage study was to build a collection of pneumococcal carriage isolates from which any changes in pneumococcal carriage epidemiology could be detected and this data used to help understand any change in the epidemiology of pneumococcal disease with respect to the introduction of PCV7. The ability to detect a change in the carriage rate was achieved by collecting a sample size of ≥100 pneumococci positive swabs per study year. This was based on the lowest expected carriage rate of 10% and would provide 80% power to detect an estimated 50% relative reduction in pneumococcal carriage after the introduction of PCV-7 at a 5% significance level.

2.1.1 Microbiological Processing

State registered biomedical scientists at the Health Protection Agency Microbiology Services, Southampton, carried out initial microbiological processing of swabs. Swabs were processed within 9 hours of collection by plating onto Columbia Colistin Naladixic Acid agar (CNA, Oxoid, Basingstoke, UK) to screen for *S. pneumoniae* each sampling period. Presumptive *S. pneumoniae* were plated on Columbia blood agar (CBA, Oxoid) with an optochin disc (Oxoid). Ten well-isolated colonies of pneumococci from each patient were used to inoculate individual Microbank vials of beads (Pro-Lab Diagnostics, Neston, UK) as master stocks. Isolates were collected at this stage from the HPA and stored at –70°C.

2.1.2 Culture and storage

All strains were initially grown from the 1st master stock vial by inoculating Columbia blood horse agar (Oxoid) with a single bead. Where growth was not observed for the first master stock the second vial was used to inoculate, this was necessary for <5% of samples. When growth could not be achieved in this manner one bead was enriched for 4 hours in 100µl BHI at 37ºC in 5% CO_2. 75µl of the enriched BHI was streak-plated onto agar. An optochin disc (Mast Group Ltd, Merseyside, UK) was placed on all agar plates to determine optochin susceptibility. Isolates with a >14mm zone of inhibition around the disc were classed at sensitive. Isolates that had any colonies within this inhibition zone were classed as intermediately resistant and growth that showed no zone or sign of inhibition was classed as resistant. Plates were incubated upside down at 37ºC in 5% CO_2 overnight (~18 hours).

Working stocks were created from pure cultures of the master stock, sub-culturing of one colony onto a second CBA plate was performed if necessary in cases of mixed or scanty growth, observed for up to 20% of isolates particular in samples stored for longer periods of time, years 1-3. Approximately half a plate of growth was collected and used to inoculate 1ml sterile 50% glycerol (Fisher Scientific UK Ltd, Loughborough, UK), Brain Heart Infusion (BHI) broth (Oxoid) and repeated with the second half of the growth to create duplicate glycerol stocks. One of the glycerol stocks was used as the working stock whilst the other was stored separately as a backup. Master and glycerol stocks were stored at -70ºC.

2.1.3 Strain exclusion and inclusion

A previous temporal analysis of the initial years (1-3) of this pneumococcal carriage study (Tocheva *et al.* 2011) excluded one isolate of *S. pneumoniae*, strain 2337, from year 2. Isolates 2336 and 2337 were collected sequentially from siblings and were likely to be colonised by similar organisms and strains. Therefore isolate 2337 was previously excluded to remove bias from the calculation of carriage rates and subsequent analysis. The isolate 2337 was included for whole genome sequencing. Two isolates from year 5 of the carriage study, 5265 and 5267 were excluded from sequencing as low growth did not permit extraction of high yield, high quality DNA (2.2.2) for sequencing. Sequencing data that suggested that DNA extractions were not pure –heterologous SNPs, resulted in the exclusion of four isolates, 0316, 2336, 5202 and 5251. Final and definitive strain data e.g. serotype and sequence type included in the analysis was derived from whole genome data. Where there was disagreement between previously generated data and WGS resultant data, WGS data was taken as the primary data provided that all other WGS data supported it. This decision was made as any one sample may potentially contain multiple pneumococcal serotypes/strains. This study does not aim to nor has the design to properly investigate this phenomenon, due to the culture based collection methodology, particularly in the early years, 1-4 for which the protocol did not include storage of the swab and relied solely on culture. Non-typeable presumptive non-pneumococcal streptococci, 5006, 5047, 5163 and 5170 were excluded from temporal analysis of pneumococcal epidemiology. Therefore, of the 527 samples sent for sequencing, data for 519 were included in the core analysis.

2.2 DNA Extraction

DNA was extracted from cultured isolates to allow their molecular characterisation. The extraction method differed depending on the downstream molecular analysis required.

2.2.1 For PCR or MLST

Samples were grown by plating 20µl of the glycerol/BHI working stock on CBA (Oxoid) and incubating overnight at 37ºC in 5% CO_2 with an optochin disc (Mast Group). DNA extraction was initiated using an in-house lysis step. A sweep of 5 colonies was transferred from an agar plate into 200µl of lysis buffer (10mM Tris pH8, 100mM EDTA pH8, 0.5% w/v SDS) in a micro-centrifuge tube. Vigorous vortexing was used to resuspend the bacteria. The solution was incubated at 37 ^{o}C for 1 hour. 20µl proteinase K (Qiagen, Crawley, UK) was added and vortexed, after which a further incubation at 56 ^{o}C for 1 hour was carried out. The QIAamp DNA mini and Blood mini kit protocol (Qiagen) was then followed as per manufacturer's instructions from step 5b of the DNA purification from tissue protocol (Qiagen 2010). Briefly, 200µl of the Qiagen lysis buffer AL was added and mixed by pulse vortexing before 10 minutes incubation at 70ºC. Following this step 200µl ethanol was added then vortexed. The sample was then added to the QIAamp mini spin column and centrifuged at 6000 *x g* (8000rpm) for 1 minute in a bench-top centrifuge (Sanyo MSE, Micro Centaur). The filtrate was discarded and 500µl of buffer AW1 added before centrifuging at 6000 *x g* (8000rpm) for a further minute, the filtrate was discarded. Five hundred microlitres of buffer AW2 was then added and centrifuged at 20,000 *x g* (14, 000rpm) for 3 minutes. The column was moved into a clean 1.5ml micro-centrifuge tube and 200µl of the elution buffer AE added to the column. This was spun 6000 *x g* (8000rpm) for 1 minute and the elution kept, this step was repeated once more. The two resulting DNA elutions were stored at less than 4ºC.

2.2.2 For whole genome sequencing

Twenty microlitres of working stock was used to inoculate a CBA plate (Oxoid) and grown overnight. The plate was sub cultured from one colony the following day and allowed to grow overnight with an optochin disc (Mast Group) to confirm susceptibility. The resulting growth was then scraped into 1ml pre-warmed BHI (Oxoid) and spun at 4000rpm for 1 minute. 0.75ml of the supernatant was added to 9ml pre-warmed BHI (Oxoid). Approximate cell viability and culture incubation time of 6 hours was calculated

from viable counts of three representative pneumococcal liquid cultures at 4, 5 and 6 hours, where 100µl was taken at each time point and serially diluted down to 10^8. To calculate viable bacteria per ml of culture 10µl of each dilution was spotted in triplicate onto CBA (Oxoid) and incubated overnight. Colonies were counted at a dilution that had between 10-50 colonies and the number of cells in 10ml was calculated in order to produce a pellet of approximately (but no more than) 2.2 x 10^{10} colony forming units/ml to avoid blocking the column. Seeded liquid cultures were incubated at 37ºC for 6 hours to achieve this. The resultant bacterial suspension was centrifuged at 5500rpm for 15 minutes, and the supernatant discarded. The resulting pellet was frozen for future use with the Qiagen 100/G genomic tip protocol.

DNA was extracted for whole genome sequencing using the Qiagen 100/G genomic tips as per manufacturer's instructions for gram-positive bacteria from liquid culture (Qiagen 2001). In brief, the bacterial pellet was re-suspended in 3.5ml buffer B1 containing 10mg/ml RNase A by thorough vortexing. Eighty microlitres of 100mg/ml Lysozyme (Sigma-Aldrich Company Ltd, Gillingham, UK) and 100µl Proteinase K (Qiagen) were added to the sample which was then incubated at 37ºC for 30 minutes. One point two ml of buffer B2 was added and mixed by inverting three times before incubation at 50ºC for 30 minutes. Each Qiagen Genomic-tip 100/G was equilibrated using 4ml of buffer QBT and allowed to empty by gravity flow between each step. The sample was vortexed and added to the column and washed twice with 7.5ml buffer QC. The column was moved to a 15ml collection tube before addition of 5ml of the elution buffer QF, pre-warmed to 50ºC. The eluted DNA was precipitated using 3.5ml isopropanol, inverted several times to mix and centrifuged at 5500 *x g* for 15 minutes at 4 ºC. The supernatant was discarded and replaced with 2ml of 70% ethanol chilled to 4 ºC, vortexed briefly and centrifuged at 5500 *x g* for 10 minutes at 4 º C in a Biofuge Primo R centrifuge (Thermo Scientific, Heraeus, Loughborough, UK). The pellet was air dried and re-suspended in 0.1-0.4µl of TE (10mM Tris, 1mM EDTA, pH 8.0 with HCl), dependant on the visible size of the DNA pellet. The pellet was dissolved overnight on a shaker (Luckham, Sussex, UK). All buffers as required by the Qiagen genomic DNA handbook were prepared in-house as per manufactures instructions outlined in the manual.

2.2.2.1 Genomic DNA quality

The Qubit™ fluorometer (Invitrogen, Paisley, UK) was used to measure the concentration of DNA in each sample. Samples were diluted to approximately 20µg/µl with a minimum volume of 100µl in TE buffer as per instructions from the Wellcome

Trust Sanger Institute. Ten microlitres of neat DNA sample was added to 1µl of loading buffer (Bioline, London, UK) and mixed, samples were run on a 1% w/v agarose (Geneflow, Fradley, UK) gel in 1 x TAE buffer (121g Tris base, 28.6ml glacial acetic acid, 50ml 0.5M EDTA). Ten microlitres of hyperladder III (Bioline) was used to determine the approximate average length of DNA fragments. Gels were run for 1 hour at 150 volts before staining in 100m1 x TAE plus 10µl/l ethidium bromide for 20 minutes on a Rotartest shaker (Luckman Ltd, Sussex, UK). The gels were viewed in a 2UV^{TM} trans-illuminator (UVP, Cambridge, UK) under UV light (302nm) and a digital image taken to confirm that bands were distinct and not fragmented and that the predominance of the DNA was >1kb in size.

2.3 Typing

Serotyping, multi-locus sequence typing (MLST) and genomic sequence analysis was performed to allow phenotypic and genetic characterisation of isolates.

2.3.1 Capsular typing

The pneumococcal capsular type can be determined or inferred by a number of different methods a combination of which was used in this thesis.

2.3.1.1 PCR serotyping

S. pneumoniae isolates were serotyped by PCR using the CDC protocol for deducing pneumococcal serotype designed for USA clinical samples (Pai *et al.* 2006; Centers for Disease Control and prevention 2012). DNA was extracted as per 2.2.1. The CDC protocol is split into pools of multiplex reactions of three to six serotypes for which the primers have been carefully selected to produce different stepwise bands sizes, each pool also includes the primers for the pneumococcal capsular polysaccharide synthesis gene A (*cpsA*) as a positive pneumococcal control. The reaction volume used was 12.5µl, for all of the pneumococcal pools. Each reaction mixture contained 0.5µl 50mM $MgCl_2$ (Bioline), 6.5µl Biomix ^{TM} red (Bioline), 1µl of DNA extract, the varying quantities of forward and reverse primers for each of the serotypes in the pool and primers for *cpsA*. The appropriate volume of nuclease-free water was used to make up the reaction to 12.5µl. DNA extracts from isolates of known serotype were used as positive controls and water was used as a negative control.

Serotype was deduced from the PCR products using gel electrophoresis. The full reaction volume that includes a red loading dye was loaded into a 2% w/v agarose (Geneflow) gel in 1 x TAE buffer. Ten microlitres of a 100bp hyperladder IV (Bioline) was used to determine band size. The cycling parameters, reaction mixtures and serotype pools are described in the appendices Table 24, Table 25 and Table 26 respectively. Gels were run for 1-2hours at 150 volts before staining in a 10µl/100ml ethidium bromide in 1 x TAE for 20 minutes on a shaker. The gels were viewed in a trans-illuminator (UVP) under UV light (302nm) and a digital image taken. Serotype/group was then deduced when all positive and negative controls were satisfactory and a band was present in a sample that corresponds exactly to an expected serotype band size.

2.3.1.2 Phenotypic serotyping

The CDC protocol for deduction of serotype can only positively identify 40 serotypes out of the 94 known. Additionally for a number of reactions it cannot discriminate to the serotype level but only to the serogroup level (a serogroup can represent a number of serotypes). Therefore further serotyping is required. Where necessary, isolates were either sent to Statens Serum Institut (Copenhagen, Denmark) for Quellung (Neufeld 1902) to be performed or serotyped in-house using the slide agglutination method described below.

2.3.1.2.1 Slide agglutination method

An unpublished standard operating procedure obtained through collaboration with Public Health England (PHE) was used to determine pneumococcal capsule type via slide agglutination (S. Martin 2012). A CBA plate was inoculated from the working stock. One ¼ or ½ plate of pure growth was transferred into 5ml of Todd Hewitt Broth (Oxoid) to seed a liquid culture. These liquid cultures were incubated with loose lids, to allow gas exchange, at 35^{o}C in 4% CO_2 incubator overnight. After overnight culture the tube was spun down at 1500rpm for 30 minutes and the supernatant removed leaving ~200ul and bacterial pellet. The pellet was re-suspended by vortexing or pipetting mixing this suspension was used to perform slide agglutination using appropriate Statens Serum institute antisera or latex factors raised against pneumococcal polysaccharide capsular antigens. Targeted slide agglutination based on the results of PCR serotyping was performed.

A suspension of 10µl was dispensed onto a glass slide, and then 10ul of antisera or 1ul of latex factor was added. The slide was then rocked to mix the reagent and culture. Agglutination (clumping with clearing background) of the suspension was recorded as a positive result. If no agglutination was seen, gentle rocking of the slide was continued for up to 30 seconds to observe a positive result. Interpretation of the agglutination reactions seen with the factor antisera follows (Table 3). A single reaction or a pattern of reactions may have been necessary to determine the final serotype of an isolate. If there was weak/possible or no visible agglutination a further aliquot antiserum and/or culture suspension was added to the 'drop' and mixed. If there was still no agglutination, the test was repeated with 10µl omniserum (polyvalent antiserum containing antibodies to ~90 known serotypes of pneumococcal capsular polysaccharide), a positive reaction confirmed the presence of pneumococci expressing one of the ~90 capsules for which antibodies are included. If there was no agglutination with omniserum the isolate was noted potentially acapsular or closely

related streptococcal species to be confirmed with a combination of phenotypic and genotypic data.

Table 3. Pneumococcal factor sera reactivity profiles

Serotype	Reaction factors			
	6b	6c	6d	
6A	+	-	-	
6B	-	+	-	
6C	-	-	+	
	7b	7c	7e	7f
7F	+	-	-	-
7A	+	+	-	-
7B	-	-	+	-
7C	-	-	-	+
	9b	9d	9e	9g
9A	-	+	-	-
9L	+	-	-	-
9N	+	-	+	-
9V	-	+	-	+
	10b	10d	10f	
10F	+	-	-	
10A	-	+	-	
10B	+	+	-	
10C	+	-	+	
	11b	11c	11f	11g
11F	+	-	-	+
11A	-	+	-	-
11B	+	-	+	+
11C	+	+	+	-
11D	+	+	-	-
	12b	12c	12e	
12F	+	-	-	
12A	-	+	-	
12B	+	+	+	
	15b	15c	15e	15h
15F	+	+	-	-
15A	-	+	-	-
15B	+	-	+	+
15C	-	-	+	-
	16b	16c		
16F	+	-		
16A	-	+		
	17b	17c		
17F	+	-		
17A	-	+		
	18c	18d	18e	18f
18F	+	-	+	+
18A	-	+	-	-
18B	-	-	+	-
18C	+	-	+	-

Serotype	Reaction factors				
	19b	19c	19f	7h	
19F	+	-	-	-	
19A	-	+	-	-	
19B	-	-	-	+	
19C	-	-	+	+	
	22b	22c			
22F	+	-			
22A	-	+			
	23b	23c	23d		
23F	+	-	-		
23A	-	+	-		
23B	-	-	+		
	24c	24d	24e		
24F	-	+	-		
24A	+	+	-		
24B	-	-	+		
	25b	25c			
25F	+	-			
25A	-	+			
	28b	28c			
28F	+	-			
28A	-	+			
	32a	32b			
32F	+	-			
32A	+	+			
	33b	33e	33f	6a	20b
33F	+	-	-	-	-
33A	+	-	-	-	+
33B	-	-	+	-	-
33C	-	+	(+)	-	-
33D	-	-	+	+	-
	35a	35b	35c	29b	42a
35F	+	+	-	-	-
35A	+	-	+	-	-
35B	+	-	+	+	-
35C	+	-	+	-	+
	41a	41b			
41F	+	+			
41A	+	-			
	47a	43b			
47F	+	-			
47A	+	+			

Vaccine serotypes indicated in **bold** (PCV7/13 blue, PCV13 only red**)**. (+) Weak positive reaction. Data from (Statens Serum Institut 2012)

2.3.2 Multi Locus Sequence Typing

Fragments of the following seven house-keeping genes were sequenced to determine the pneumococcal MLST by Qiagen Genomic Services for isolates collected in years 1 through to 4 of the study (Hilden, Germany); *aroE* (shikimate dehydrogenase), *ddl* (D-alanine-D-alanine ligase), *gdh* (glucose-6-phosphate dehydrogenase), *gki* (glucose kinase), *recP* (transketolase), *spi* (signal peptidase I) and *xpt* (xanthine phosphoribosyltransferase) (Enright & Spratt 1998). MLST DNA sequence data (raw data and text files) and ST assignments were received by our laboratory. MLST for year 5 isolates were derived from whole genome data using a method described in section 2.3.4. Further analysis on all ST data was performed using the pneumococcal MLST website (MLST.net 2012). New alleles or combinations of alleles were submitted to the curated online database (MLST.net 2012), new allele or ST numbers were assigned by the curator following quality checks on the submitted data. eBURST was utilised produced graphical representations of the MLST distribution of the carriage isolates (Feil *et al.* 2004). Clonal complexes (CC) were defined as a network of STs that share 6 out of 7 alleles for the 7 loci with at least one other ST in the network termed as single locus variants (SLVs). The founded of the clonal complex is defined as the ST connected with the largest number of other STs that differ from itself only by one allele (SLVs). A subgroup founder is a ST that appears to have diversified to produce multiple SLVs.

2.3.3 Species discrimination

To discriminate between strains of capsular deficient non-typeable pneumococci and other closely related streptococci that lack the pneumococcal specific capsular loci, an extended set of MLST genes were analysed. Capsular negative isolates that exhibited ≥1 unknown MLST allele, n=5, had each of the seven MLST alleles submitted for online analysis using the nucleotide basic local alignment search tool (BLASTn) (National Centre for Biotechnology Information 2013) to determine the top species match. All capsular negative isolates, n=12, had eMLSA (electronic Multi locus sequence analysis) alleles for the genes *map* (Methionine aminopeptidase), *pfl* (pyruvate formate lysase), *ppaC* (Inorganic pyrophosphatase), *pyk* (Pyruvate kinase), *rpoB* (RNA polymerase beta subunit), *sodA* (Superoxide dismutase), *and tuf* (Elongation factor Tu) submitted for online analysis using BLASTn. Presumptive species assignment was made from the most commonly occurring species found as top match for a single strain. The designation of non-pneumococcal streptococci was additional confirmed

with consistent phylogenetic separation from the non-capsular pneumococci for each of the eMLSA genes individually and their concatenation.

2.3.4 Whole genome sequencing

Library preparation and sequencing was performed solely by staff at the Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton in collaboration with Dr Stephen Bentley. Quantified DNA extraction tubes of the required >100µl of 20ng/ml for sequencing were barcoded as part of this body of work for tracing within the Sanger Institute prior to being couriered. Sequencing was performed on multiple Illumina HiSeq 2000 instruments in seven individual HiSeq runs. The sequencing utilised v3 reagents in multiplexed pools of 24 per lane. Paired end libraries were prepared with 150-250bp inserts. Read length parameters were set to 2 x 75bp with 8 cycles for the sequencing of the multiplex IDs, consequently 158 cycles were performed.

2.3.4.1 Quality, mapping and assembly

FASTQ read data from the Illumina sequencers entered the automated qc_grind pipeline for quality control (QC) including both the standard Illumina QC and in house post sequencing analysis using NPG: Short read Sequencing pipeline (The Sanger Institute 2012). This included assignment of quality scores to each base position within reads, distribution of percentage GC content, mean insert size falls within requested size range and % mapping to the designated reference genome S*treptococcus_pneumoniae*_ATCC_700669 of serotype 23F, ST81 (Croucher *et al.* 2009) using SMALT version 0.5.3 and generating BAM files (The Sanger Institute 2012). The publically available Integrative Genomics Viewer (IGV) from the Broad Institute was used for visualisation of the bam alignments to the reference (Broad Institute 2012).

FASTQ reads were manually submitted for *de novo* assembly using Velvet assembler (European Bioinformatics Institute 2012) to generate FASTA contig files on the Sanger Institute's high performance cluster computing facility. This utilised an in-house Velvet optimiser script for bacterial genomes that adjusts parameters such as K-mer length for each assembly.

2.3.4.2 *In silico* analysis

In silico analysis for the purpose of this thesis will be defined as the bioinformatics analysis of whole genome data to obtain results that would traditionally be gathered by alternative methods including MLST and PCR.

2.3.4.2.1 *In silico* MLST

An in-house Sanger Institute script was utilised to submit batches of sample data to determine the seven, pneumococcal, MLST alleleic sequences from each sample's FASTQ reads. This script requires a concatenated MLST allele reference, then the Sanger Institute software SSAHA-sequence to which sequence search and alignment by hashing algorithm SSAHA (The Sanger Institute 2012) maps reads to and uses iCORN- iterative correction of reference nucleotides (Otto *et al.* 2010) to morph the reference into the concatenated sample ST. The output in xml format contained the extracted seven alleles for each sample that could then be submitted to MLST.net for allele and ST assignment.

2.3.4.2.2 *In silico* PCR

An existing Sanger Institute script that extracts a set of query genes using primer sequences was used to perform *in silico* PCR (Sanger institute, unpublished method). A tab delimited 5 column input file of primer name, forward primer sequence, reverse primer sequence, maximum product length, and minimum product length was required. Each new line could contain additional primer specifications.

The script searched a specified FASTA contig file for the forward and reverse primers with a tolerated 3bp mismatch per primer. To yield the extracted FASTA sequence and output it to file primers had to be observed in the expected orientation within the specified maximum and minimum size range. This required both primers to be located within a single contig. The option to include primer in the output sequence was utilised where primers were located at the termini of the sequence of interest. The locations of potential primer regions that did not satisfy the stated requirements were printed to file to allow manual interrogation of why a sample might be unexpectedly negative. The *in silico* PCRs were performed on genome assemblies using primers listed in Appendix Table 27.

2.3.4.2.3 *In silico* Pneumolysin PCR

Pneumolysin sequences derived from assemblies were translated and compared to published alleles (Kirkham *et al.* 2006; Jefferies *et al.* 2010b) to allow designation of allele type or identification of novel sequences (2.3.4.2.5). Where data was available from previous work the two data source results were compared to allow the validity of designation from whole genome sequence to be assessed.

2.3.4.2.4 *In silico* PCR capsular typing

In silico capsular typing offered a method to infer capsular type by replicating PCR (Centers for Disease Control and prevention 2012) in a fast and highly parallel manner from assemblies with the added advantage of extracting the complete sequence of the genes targeted. As serotypes within a serogroup can differ in a single SNP, the *in silico* capsular sequence output was used to offer increased discriminatory power over conventional PCR.

2.3.4.2.5 Allele designator

A Perl script was designed for comparisons of alleles outputted from *in silico* PCR (1.3.4.2), this was formulated and designed by Rebecca Gladstone with assistance from a Perl scripter at the Sanger Institute. The script (Figure 3) compared submitted alleles in FASTA format to known alleles contained in an accessory file. When an allele matched a known allele the appropriate allele assignment was made and the observation recorded to file. If the allele was novel it was added to the supplementary list of alleles and assigned the next novel allele number. This allowed the interrogation of large numbers of alleles, identifying novel alleles and assigning an allele number to the allele for each sample. Finally, the output was designed to list all the allele names and their sequences observed, the number of times an individual allele was seen and a list of samples the specific allele was observed in. This script could be adapted to any gene allele comparison by replacement of all references to the four-character gene name with the next gene of interest, which crucially would be in the name of the input FASTA files to be interrogated. Manual exclusion of alleles that contain paired end gaps consisting of unknown base (N) inserts were excluded from new allele designation.

Figure 3. Allelic designator

```
#!/usr/bin/perl
use strict;
use warnings;

### Script finds and counts occurence of xxxX alleles both existing and novel

my (%hash, %newhash, %filehash);  #       Initialise hashes for use later

open(xxxX, "/directory_file_path/xxxX.fa");      #       Open file that contains known alleles
while (<xxxX>) {
        if ($_ =~ /^>xxxX/) {   #       If line begins with ">xxxX"
#               print "$_\n";
                my @line = split(/>/,$_);       #       Split the line containing sequence and name info for xxxX alleles
                $hash{$line[2]} = $line[1];     #       Store sequence of known xxxX allele as key ($line[2]), and name of the xxxX allele as the value ($line[1])
        }
}
close(xxxX);

open(MASTER, "/directory_file_path/xxxXdone.txt");       # File list for strains that have undergone insilico pcr for xxxX and returned a xxxX fasta file
#if (-e "/directory_file_path/xxxXmatch.txt") { # Checks if file xxxXmatch.txt exists
#       `rm /directory_file_path/xxxXmatch.txt`;        # If above true then removes previous files to prevent rewriting results to the files multiple times
#       `rm /directory_file_path/xxxXempty.txt`;
#       `rm /directory_file_path/xxxXnovel.txt`;
#}
open(OUT, ">/directory_file_path/xxxXmatch.txt");        #opens/creates files that now def dont exist
open(OUT2, ">/directory_file_path/xxxXempty.txt");
open(OUT3, ">/directory_file_path/xxxXnovel.txt");

while (<MASTER>) {
        my $file = $_;  # assigns the $_ (line) to $file
        $file = "\/lustre\/scratch108\/bacteria\/rg9\/cdcpcr\/" . $file;        #rename $file, concatenate to file path so that script has absolute location of file
        chomp($file);
        open(IN, "$file");      #opens xxxX.fasta file
        my $size = -s $file;    #stores size of file in $size
#       print "$size\n";
        if ($size == 0) {       #if file empty
                print OUT2 "$file\n"; #prints out to file xxxXempty.txt
                next;   # then go to next file
        }
        while (<IN>) {  #reads xxxX.fasta file
                if ($_ =~ /^>NODE/) {   #skips header line
                        next;           #cont. of above
                }
                my $line = $_;  #default assigned to $line
                $newhash{$line}++;
#assigning seq to key, value is assigned 1 (seen once so far) if not yet seen and incrementing the value each subsequent time it is seen
                $filehash{$file}=$line;         #seq is assigned to value in %filehash key is file path ($file)
                if (grep {$_ eq $line} keys %hash) {
# eq is == equivalent for 'strings', finds seq ($line) in each of the keys in %hash which contain the seq of known xxxX alleles IF IT exists in this set of X alleles
                        chomp($file);
                        print OUT "$file\t$hash{$line}\n";      #if found prints path to file and seq match in xxxXmatch.txt
                } else {
                        print OUT3 "$file\n$line\n\n";          #if not found i.e. novel allele, prints path and seq to xxxXnovel.txt
                }
        }
        close(IN);
}
```

Figure 3. Continued: Allelic designator

```
open(ALLELES,">/directory_file_path/xxxX_variants.txt");        #open new output file handles
open(NOVEL,">/directory_file_path/Novel_xxxXalleles.txt");      #as above
my @keys = keys %newhash;                #array, list of keys of %newhash sequences that have been seen at least once
my @sorted = sort (keys %filehash);      #array, sorts filepaths ABCD
my $number = 100;                        #set incremental start value for new xxxX alleles
print ALLELES "There are " . scalar(@keys) . " different alleles.\n\n\n\n";     #prints to xxxX_variants.txt 'there are x number of unique (differemt) alleles
foreach my $allele (keys %newhash) {     #loops through unique alleles seen and assigns current one to $allele
        if ($hash{$allele}) {            #is the unique sequence one of the known xxxX alleles
                print ALLELES "\n$hash{$allele}\t$newhash{$allele}\n";  #if true print cpsA allele number, number of times seen
                foreach my $filepath2 (@sorted) {                      #stores the current file path which have been sorted, in $filepath
                        if ($filehash{$filepath2} eq $allele) { #if the sequence of the filepath matches the seq of the unique $allele
                                print ALLELES "$filepath2\n";   #prints files associated with each unique allele
                        }
                }
        } else {                                                       #if seq is not a known allele, novel ply alleles added to xxxXalleles.fa on 030512
                print NOVEL ">xxxX" . $number . "\n$allele\n";  #print incremented from 19 >Ply assigns plynumber to a novel ply seq
                print ALLELES "\nSequence (xxxX" . $number . ") below is seen ". $newhash{$allele} . " times\n$allele\n";
# prints number of times seq found followed by that seq
                $number++;                                             #increment action
                foreach my $filepath (@sorted) {                       #stores the current file path which have been sorted, in $filepath
                        if ($filehash{$filepath} eq $allele) {  #if the sequence of the filepath matches the seq of the unique and novel $allele
                                print ALLELES "$filepath\n";    #prints files associated with unique novel allele
                        }
                }
        }
}
close(NOVEL);
close(ALLELES);
close(MASTER);
close(OUT);
close(OUT2);
close(OUT3);

exit;
```

Written in the programming language Perl for this Thesis by a bioinformatician of The Sanger Institute with guidance and complete (line by line) biological and functional instruction from Rebecca Gladstone. Identifies all unique alleles and assigns allele designation.

2.3.4.2.6 Capsular mapping

Utilising a script written by Nick Croucher at the Sanger Institute, capsular grouping could be inferred from Fastq reads. Reads were mapped against the concatenated capsular locus sequence, *aliA* through to *dexB*, of 91 serotypes. Where the coverage was greater than one, a match was called. Multiple matches could occur as serotypes, particularly within a serogroup, can have closely related sequences in the capsular region, in this case all matches were reported and ordered by their mapping score (% coverage of the CPS locus), The top matches infer capsular grouping and were used to support other findings.

2.3.4.3 Allelic diversity

A number of web-based tools were utilised for comparison of the sequence diversity within a gene using the alleles outputted from the allelic designator (Methods 1.3.4.3). ClustalW2 was used for the multiple alignment of DNA or amino acid sequence inputted in FASTA format for SNP/non-synonymous change detection (EMBL-EBI 2012) whilst phylogeny.fr allowed maximum likelihood phylogenetic tree generation from FASTA sequences in ‘one click’ mode using MUSCLE, Gblocks, PhyML and Treedyn (Dereeper *et al.* 2008). A phylogram was chosen to represent the phylogenetic tree that has branch spans proportional to the amount of character change. Translation of DNA to AA was performed using the online ExPASy tool (Swiss Institute of Bioinformatics 2013).

2.4 Statistics

The data in the study was descriptive. Data is collected as counts of observed colonisation. Therefore the data set was categorical with the frequencies of events recorded and percentages calculated. Confidence intervals (CI) were calculated at 95% confidence for incidence values, unless otherwise stated. Statistically significant changes in proportions between two groups were determined using Fisher's exact test to produce a two-tailed *p*-value. The statistically significant changes between the means of two groups were determined using an independent sample *t*-test for differences in means to produce a two-tailed *p*-value. To determine whether there was an association between two factors the Fisher's exact with Monte Carlo estimation and 99% confidence intervals was used.

Odds ratios for examining invasive disease potential of serotypes were calculated as OR=(ad)/(bc) were a) is the number of IPD isolates of serotype X b) is the number of carriage isolates of serotype X c) is the number of IPD isolates that are not serotype X and d) is the number of carriage isolates that are not serotype X (Brueggemann *et al.* 2003), an OR >1 indicates increased invasive disease potential, and an OR <1 indicates decreased invasive disease potential. 95% confidence intervals were calculated using CIA, epidemiological studies, and the unmatched case control study option. CIs that did not cross 1 indicate significance ($p<0.05$). For serotypes that were either not observed in carriage or IPD the ORs and their CIs could not be calculated. CIA, proportions and their differences, unpaired, were therefore used, with 95% confidence intervals for the differences. CIs that did not cross 0 indicate significance ($p<0.05$)

The Simpson's Diversity Index was used to measure the sequence type and serotype diversity of the nasopharyngeal habitat. It takes into account the number of types present, as well as the abundance of each type. The Simpson's diversity index 1-D was used and calculated, $D=\sum n(n-1)/N(N-1)$. Where n is the total number of organisms of a particular sub population and N is the total number of organisms of all sub populations. The 1-D Simpson's diversity index is consequently more intuitive as 0 indicates no diversity through to 1 where all isolates are different.

3. RESULTS - WHOLE GENOME DATA

3.1 Quality Control

It was necessary to assess the quality of the sequencing procedure, the sequences produced and the *in silico* methods used to derive data from the sequences, before analysis of epidemiologically relevant data could be performed.

3.1.1 Sequencing statistics

The results of the quality control of whole genome sequencing performed by qc_grind are presented to the researcher in the form of a Sanger internal web interface, per sample, to be referred to but non exportable, an example from this data set, typical of the majority of sample output graphs, is used to illustrate the quality of the sequencing, with a GC plot that follows the reference for both the forward and reverse reads, average Q scores of >30 for each base along the length of the read in both orientations and ~150bp inserts (Figure 4). This data confirms that sequencing process completed as anticipated and the genome has the expected GC profile (Figure 4).

The average sequencing depth for the 527 samples submitted for sequencing was 411 times coverage per base (411X). The minimum observed depth of coverage of 68X was well above the sequencing gold standard of 20X (Table 4). Only 12/527 isolates submitted for sequencing fell below the 80% mapping QC threshold. Two of these isolates, 0008 and 2070, had abnormal GC plots indicative of low-level non-streptococcal contamination (Figure 5). These isolates had 79.2%, 76.5% mapping and 334X, 74X sequencing depth respectively.

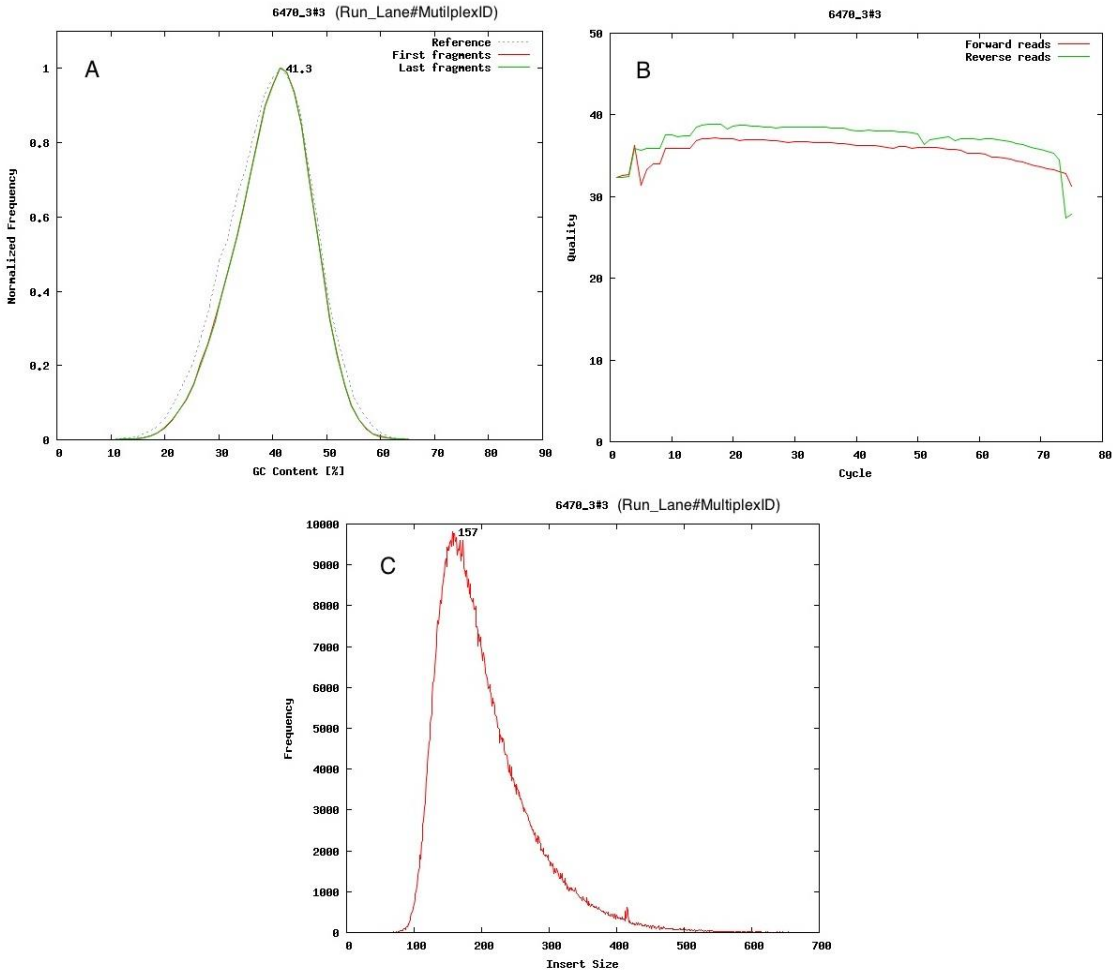
In silico MLST and serotype results for these two isolates exactly matched wet-lab results due to the high level sequencing depth, therefore their serotype and MLST data was included in the overall analysis but any novel alleles derived from the genome data underwent more rigorous manual inspection. Four of the 12 isolates had >70% mapping and were classed as non-typeable (NT) by both the analysis performed within this thesis and by previous investigation (Tocheva *et al.* 2011). One sample (0011) with 75.5% coverage similar to that of NT isolates was previously classed as 22F, the ST449 (designation from previous analysis (Tocheva *et al.* 2011) and this analysis) was only observed as NT in the MLST database thus also supported that this was indeed a NT isolate as the 22F capsular genes could not be observed in the whole genome data. The remaining five isolates were otherwise excluded from the core data

set: as they were non-pneumococcal isolates n=4 or mixed population sequencing n=1 (2.1.3 Strain exclusion and inclusion, 2.3.3 Discriminating non-typeable pneumococci and non-pneumococci).

3.1.2 *In silico* PCR validation

PCR results for the pneumolysin gene were used to validate the *in silico* PCR and allele designator script. The pneumolysin gene was selected as there are a limited set of described protein alleles (Ply), and historical data existed for early isolates of the carriage study on Ply allele type (Kirkham *et al.* 2006; Jefferies *et al.* 2010b). A total of 192 isolates from years 1 and 2 had Ply alleles assigned historically by PCR and Sanger sequencing (Jefferies *et al.* 2010b) that could be compared to the Ply alleles assigned from whole genome data. There was 100% concordance between the two data sets therefore the *in silico* PCR and allele designator were declared fit for purpose to replace conventional gene detection and designation of alleles.

Figure 4. Example QC output



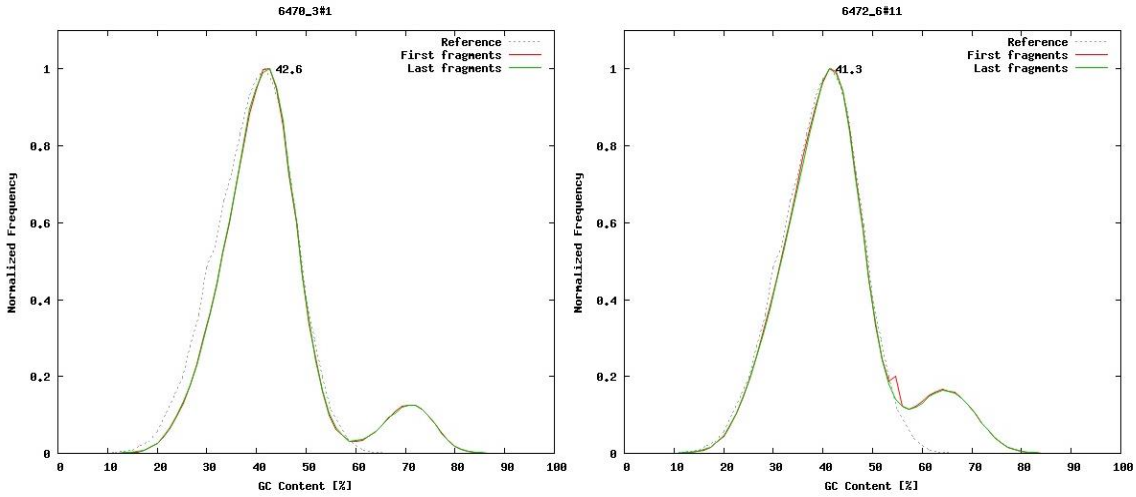
Qc_grind output graphs for sample 0014 as an example of sequencing data quality for the data set

Table 4. Estimated base coverage

Per Sample	Min	Max	Mean	Median
Coverage depth (X)	68	1307	411	402
*Mapping against 23F reference (%)	74%	99%	89%	89%
Yield (Kb)	171089	3087239	1021514	1004885

*Excludes isolates excluded from core dataset in Methods strain exclusion and inclusion

Figure 5. GC indication of non-streptococcal contamination



Qc_grind GC graphs for samples 0008 and 2070 showing low level contamination of sequence data with a species with a higher GC content of ~65-70% (small peak to the right hand of the main peak).

3.1.3 *In silico* MLST validation

MLST was determined by conventional methods (Qiagen genomic services) for isolates from years 1-4. This data was used to validate MLST assignation from whole genome data. Comparable data was initially available for 420 isolates. There was disagreement between the two methods for four isolates (Table 5). The conventionally assigned ST differed from the *in silico* ST at all seven loci, additionally for three out of the four isolates the corresponding serotype also differed between the two data sets. To assess whether the sequencing chemistry could introduce errors in typing exact isolate pairs were needed for the comparisons, the contradictory profiles for the four isolates were distinct but previously observed profiles that could not have been produced by random sequencing errors and are a result of obtaining data from distinct strains from within a single sample. Therefore these four isolates were excluded from the validation but included in the final dataset. The remaining 416 strains achieved 100% concordance in ST designation therefore *in silico* MLST was declared fit for purpose to assign STs, specifically for those isolates for which MLST had not been determined by conventional methods.

3.1.4 Data consolidation

When comparing the serotype and MLST data sets determined by the traditional methods and results generated *in silico*, there were a number of conflicting results to be resolved.

3.1.4.1 MLST designation

There was a single instance of matching serotype data with disagreement between generated STs for strain 2082 (11A), the original designation of ST4695 was not otherwise observed in the entire MLST database as serotype 11A (MLST.net 2012). Additionally serotype 11A is commonly seen as ST62 as generated by WGS. ST62 was subsequently deemed to be the true ST (Table 5) and the original designation ST4695 belongs to a strain, likely not serotype 11A that may have been a result of multiple serotype colonisation or human error. Three strains 0232, 2173 and 3226 were in disagreement between the two data sets for both serotype and ST. In all cases the whole genome data was used as the primary data for subsequent analysis after it was revealed that sequencing of a secondary strain of *S. pneumoniae* obtained from single patient sample had likely occurred. Investigation into sequential pneumococcal positive isolates 2255 and 2258 revealed that a labelling/multiplexing error was detected for

which the exact combination of expected ST, serotype and Ply allele were found in the other, this error was easily corrected and data included.

3.1.4.2 Serotype designation

The following section is split into discussing isolates from years 1-3 and years 4-5, as the former includes serotype determination as part of previous work and in silico from genome data generated as part of this thesis whilst the later was determined in the laboratory and derived from genome data both as part to this thesis.

3.1.4.2.1 Years 1-3

The *in silico* PCR output for capsular type was compared to final serotype designation from the first three study years (Tocheva *et al.* 2011). This comparison highlighted nine isolates (year 1 n=0, year 2 n=3, year 3 n=6) previously designated serotype 14 using serotype 14 primers (Pai *et al.* 2006) which were determined to be 15B/15C *in silico* (Table 5) using the CDC primer recommendations of 2012 (Centers for Disease Control and prevention 2012). The isolates belonged to ST199 n=6, ST2220 n=1 and ST411 n=1, none of which have ever been recorded as occurring in a serotype 14 background within the MLST database up to 01/06/2013 (MLST.net 2012). This conflicting serogroup 15(B/C) and serotype 14 designations were confirmed to be serogroup 15(B/C) by the Statens Serum Institute using the Quellung reaction. The serotype 14 primers previously used (Pai *et al.* 2006) have subsequently been replaced by primers that do not cross react with serotypes 15B and 15C in the CDC protocol (Dias *et al.* 2007; Centers for Disease Control and prevention 2012), a repeat of the wet lab PCR with updated primers again confirmed 15B/15C designation.

Twelve additional isolates from years 1-3 with conflicting serotype information were noted. These designations were re-assigned using Quellung and/or MLST data to determine if A: a distinct strain had been sequenced or B: the original designation was a false positive (Table 5).

3.1.4.2.2 Years 4 and 5

For years 4 and 5, 216 presumptive pneumococcal isolates were available for comparison of the two methods. Multiple non-specific PCR products were observed within the serotyping pools for 35 samples with the potential to be interpreted as specific capsular gene product dependent on band size. For the majority (n=32), one of the bands corresponded to the single capsular type indicated by the *in silico* analysis. Furthermore, for isolates with the final designation serotype 21, 8/22 isolates from

years 4 and 5 had a PCR band additional to (7/8) or instead of (1/8) the serotype 21 specific band, which appeared to correspond to capsular PCR grouping 35A_C_42. In all cases the *in silico* PCR only identified capsular type 21 and MLST data was in support of the *in silico* designation.

Despite the dependency on good assembly there were no observed *in silico* false negatives due to contig breaks in the capsular region. All isolates for which there was no match with the *in silico* PCR, where either a serotype not included within the CDC protocol (33D, n=1) or did not have a capsule, were confirmed with Quellung. There were however PCR false positives where non-specific bands for 4/12 non-capsular, *cpsA* negative, isolates were observed. Two PCR false negatives were also observed where the genes for 15B_C and 6A_B_C_D were detected *in silico* and confirmed by Quellung.

Overall, the *in silico* capsular typing method, when whole genome data already existed, was extremely rapid and results were reproducible, highly sensitive and specific, not subject to interpretation and incurred only cost associated with personnel time. Therefore for capsular typing of diverse carriage isolates with the *in silico* method was superior to the CDC PCR (Centers for Disease Control and prevention 2012) which was designed for well-characterised invasive isolates with limited diversity. *In silico* capsular typing was declared fit for purpose to assign capsular type, specifically for those isolates for which there were ambiguous results by conventional methods. Capsular designation was further supported by isolate specific known MLST-serotype associations and confirmed where necessary with Quellung to the serotype level for those isolates grouped but not typed and those that produced negative results.

The *in silico* capsular analysis also highlighted the sequencing data from DNA extractions were not pure for three isolates, (0316, 2336, 5251) as they were positive for two distinct serotypes (not observed in-vivo), the whole genome data for these isolates could then be excluded from further analysis as described in 2.1.3 Strain exclusion and inclusion.

There were 522 pneumococcal positive isolates available for analysis. Of these, 519 isolates were successfully processed for whole genome sequencing (2.1.3 Strain exclusion and inclusion), the data for which was used for all subsequent analysis below the level of species described in this thesis.

Table 5. Resolution of discordant results years 1-3

	Discordant results		**Final designation**		
Strain ID	Serotype (Tocheva *et al.* 2011)/**Rebecca Gladstone**	MLST (Tocheva *et al.* 2011)/**Rebecca Gladstone**	Serotype	MLST	Comments
0011	22F/**NT**	449/**449**	**NT**	**449**	Misattributed serotype
0048	11A/**10A**	1497/**1497**	**10A**	**1497**	Misattributed serotype
0232	15B/**19A**	1262/**276**	**19A**	**276**	Different strain sequenced
0305	15A/**15B**	199/**199**	**15B**	**199**	Misattributed serotype
0311	15B/**15A**	63/**63**	**15A**	**63**	Misattributed serotype
2076	14/**15B**	199/**199**	**15B**	**199**	14-15B/C cross reaction with (Pai *et al.* 2006) primers
2082	11A/**11A**	4695/**62**	**11A**	**62**	Distinct strain previously submitted for MLST
2142	14/**15C**	199/**199**	**15C**	**199**	14-15B/C cross reaction with (Pai *et al.* 2006) primers
2173	11A/**35F**	1760/**1635**	**35F**	**1635**	Different strain sequenced
2227	17F**/31**	568/**568**	**31**	**568**	Misattributed serotype
2287	14/**15C**	199/**199**	**15C**	**199**	14-15B/C cross reaction with (Pai *et al.* 2006) primers
3079	14/**15C**	2220/**2220**	**15C**	**2220**	14-15B/C cross reaction with (Pai *et al.* 2006) primers
3165	4/**15B**	1262/**1262**	**15B**	**1262**	Misattributed serotype
3226	33F/**22F**	100/**433**	**22F**	**433**	Different strain sequenced
3234	15A/**38**	393/**393**	**38**	**393**	Misattributed serotype
3250	14/**15B**	411/**411**	**15B**	**411**	14-15B/C cross reaction with (Pai *et al.* 2006) primers
3251	14/**15B**	199/**199**	**15B**	**199**	14-15B/C cross reaction with (Pai *et al.* 2006) primers
3257	14/**15B**	199/**199**	**15B**	**199**	14-15B/C cross reaction with (Pai *et al.* 2006) primers
3259	7C/**20**	235/**235**	**20**	**235**	Misattributed serotype
3271	14/**15C**	199/**199**	**15C**	**199**	14-15B/C cross reaction with (Pai *et al.* 2006) primers
3315	14/**15B**	199/**199**	**15B**	**199**	14-15B/C cross reaction with (Pai *et al.* 2006) primers

3.1.5 Discriminating non-typeable pneumococci and non-pneumococci

There were 11 isolates (sequenced and passed quality control measures) that could not be assigned a serotype and were designated non-typeable in terms of capsule. This group could consist of both traditional pneumococci that are a-capsular and closely related pneumococcal species that lack a capsule. It was important to distinguish which isolates were true pneumococci for the epidemiological analysis of pneumococcal carriage rates. All non-typeable isolates from years 1-3 (n=5) had been previously been confirmed by Quellung and non-pneumococci already excluded before isolates were used for this thesis. All 11 isolates were subjected to the data analysis described below to distinguish between non-pneumococcal or non-typeable pneumococci.

Isolates exhibited a dry colony texture in line with a lack of capsule expression. All five isolates from years 1-3 were optochin sensitive as were the two isolates from year 4. Of the year five isolates two were clearly resistant and two sensitive to optochin in a 5% CO_2 atmosphere which can be used to distinguish pneumococci from *S. pseudomoniae* which can display a sensitive phenotype in this environment but pneumococci stored at -70°C can exhibit resistance (Arbique *et al.* 2004; Robson *et al.* 2007).

The PCR assay for deducing capsular type were all *cpsA* negative for these 12 isolates as well as being negative for all serotype specific products. This was reproduced *in silico* with no match for serotype specific primers or *cpsA*. Mapping of reads for these isolates against known capsular loci all had a top match to capsule 25F but of less than 50% identity with the exception of 4165 which had a 87% match to the serotype 21 loci but still serotype 21 *wzy* negative.

Pneumolysin could not be detected in assemblies from isolates 0011, 3140, 3143, 3267, 5163 and 5170. Novel Ply alleles Ply-21, Ply-26, the homologue Ply-20 and previously described Ply-17 were detected in the isolates and could not be used to distinguish non-pneumococcal and non-typeable isolates.

MLST assigned STs to all year 1-3 isolates (ST448, 449). New ST allelic combinations were observed for two isolates from year four whilst four isolates, all collected in year 5, exhibited new allelic sequences for the majority (≥5/7 genes n=3), or all seven of the MLST genes (n=1).

The phenotypic and genotypic descriptions above did not conclusively determine whether isolates were non-typeable or non-pneumococci. Therefore all eleven 'non-typeable' isolates had the sequence determined for the seven MLSA (Bishop *et al.* 2009) typing genes which are additional to the MLST genes and can be used to differentiate between Streptococcal species. Each MLSA gene sequence and all novel MLST gene sequences were submitted to BLASTn to determine the top species hits in online databases.

Seven isolates had known MLST alleles for all genes and had a ≥99% match to *S. pneumoniae* for all seven MLSA genes and were concluded to be non-typeable pneumococci. The remaining four capsular negative isolates (5006, 5047, 5163, 5170) top BLASTn species matches were almost exclusively *Streptococcus mitis* or *Streptococcus pseudopneumoniae* and not *S. pneumoniae*, with high percentage max identities of ≥96% (Table 6). The combined phenotypic and genotypic data presented above allowed the inclusion of 7 non-typeable pneumococci and the exclusion of four isolates (5006, 5047, 5163, 5170) from the pneumococcal epidemiological analysis as evidence suggests they are not pneumococci but other very closely related species.

Table 6. BLAST top species percentage match max identity for extended MLST gene panel for potential non-pneumococcal streptococcal isolates

	Strain ID			
Gene	**5006**	**5047**	**5163**	**5170**
*aro*E	Spp 99%	Sp 97%	Sm/Sp 97%	Sp allele 1
gdh	Sp allele 166	Sm 98%	Sm 98%	Spp/Sm 99%
gki	Sm 100%	Sm 96%	Sp/Sm 98%	Sp 100%
*rec*P	Sp/Sm 98%	Sm 97%	Sm 96%	Spp 100%
spi	Sp 100%	Sm 97%	Sm 96%	Spp 99%
xpt	Sp allele 1	Sp 97%	Sp 98%	Sp 97%
ddl	Spp 100%	Sp 97%	Sp allele 300	Spp 100%
map	Spp 100%	Sm 100%	Sm 96%	Spp 100%
pfl	Spp 100%	Sm 99%	Sm 99%	Spp 100%
*ppa*C	Sm 96%	Sm 98%	Sm 97%	Sm 96%
pyk	Spp 99%	Sm 100%	Sm 99%	Spp 99%
*rpo*B	Sm 99%	Sm 98%	Sm 98%	Spp 100%
*sod*A	Spp 100%	Sm 98%	Sm 100%	Sp/Spp 100%
tuf	Spp 100%	Sm 99%	Sm 100%	Spp 100%
Consensus Streptococcal Species	Ssp	Sm	Sm	Ssp

Sp= *Streptococcus pneumoniae*, Spp=*Streptococcus pseudopneumoniae*, Sm=*Streptococcus mitis,* multiple top matches listed if shared percentage max identity.

3.1.6 Discussion of results

Data derived from next generation sequencing platforms was used to provide epidemiological inputs for analysis of the pneumococcal carriage study isolates. Therefore it is important to discuss the validity and potential weaknesses of the data and any changes in the method for obtaining traditional typing data. SNP based differentiation is a core use of whole genome data however next generation sequencing platforms are associated with their own inherent errors in base calling. For Illumina specifically this is associated with de-phasing – a sequencing lag in some fragments (n+1 or n-1) which is routinely corrected for during the Illumina base calling process (Allhoff *et al.* 2013). Additionally there are known to be higher error rates towards the ends of reads, with an error rate of 0.3% at the start of a read and 3.8% at the tail end of the read previously reported (Dohm *et al.* 2008; Allhoff *et al.* 2013). Particular motifs are known to trigger de-phasing base call errors, these are sequence depend errors and therefore only found in that orientation/direction of sequencing (Dohm *et al.* 2008; Allhoff *et al.* 2013). All of these errors have the potential to introduce SNPs impacting genotyping including MLST and serotype inference however paired read technology goes some way to compensate for errors at the end of reads or in particular orientations by providing complementary data, this is further combated when there is high coverage, as with the sequences in this thesis, as a consensus sequence is made during assembly from the sequencing reads for which the majority will mask errors occurring at a lower frequency. The observation made in thesis, when using exact pairs of data from WGS data and traditional methods, that pneumolysin alleles and MLST profiles could be exactly reproduced when a single base pair difference would have prevented this observation indicates an undetectable error rate when the consensus assembly data was used.

The CDC protocol for deducing pneumococcal capsular type was developed based on the serotype distribution and rank order in the Active Bacterial Core surveillance (ABCs) and validated with disease isolates during 2002-2003 (Pai *et al.* 2006). The serotype specific primers were developed using the cps loci data published from The Sanger Institute for which only single representatives of each serotype were included (Bentley *et al.* 2006). The wet laboratory PCR may not have functioned as well as the *in silico* adaption due to the number of factors stated above; the isolates in question were carriage isolates from a latter period of time during an epidemiological shift which was changing the rank prevalence of serotypes and increasing previously rare types. Any of these factors could result in sequence diversity in the primer region and a

subsequent reduction in the biological specificity and sensitivity of the primers and introduced unforeseen cross-reactions. The pool ordering was designed as a screening process identifying the most common types first and only continuing through further pools if not yet identified so that large numbers of isolates could be screened in the most time efficient manner preventing the need to run unnecessary PCR pools. However cross reactions appearing in the early pools for serotypes not screened for until later pools result in mistyping, with an increased number of isolates with serotypes screened for in the later pool due in changes in prevalence. Additionally atypical product sizes would be mistyped, as this is central to assigning serotype in this method. This is where the *in silico* adaption has an advantage, all primers are easily screened for in all isolates and furthermore the sequence is obtained to confirm that the product is specific regardless of the product size.

3.2 Whole genome sequence analysis

Whole genome data was used to determine the sequence of a number of genes encoding virulence factors, antigens important to pneumococcal vaccines- current and experimental, and genes used for molecular typing. Characterisation of the diversity within these genes and their resultant proteins highlights the degree of heterogeneity, which may have the potential to impact vaccine effectiveness through vaccine type replacement, reduction in efficacy or allow vaccine escape.

3.2.1 Pneumolysin alleles

All 519 pneumococcal whole genome sequences and the four presumptive non-pneumococcal sequences were included in this analysis as species closely related to pneumococci have been documented to contain pneumolysin homologues (Jefferies *et al.* 2007; Johnston *et al.* 2010). Pneumolysin is a pore-forming toxin of the cholesterol dependant cytolysin family. Associated with pneumococcal virulence it is a component of a number of experimental serotype independent vaccines that are likely to provide the next generation of vaccines for protection against pneumococcal disease (Ljutic *et al.* 2012; Hu *et al.* 2013; Kamtchoua *et al.* 2013). Pneumolysin sequence heterogeneity and the presence of divergent alleles could affect antigenic epitopes, it important for this to be documented before implementation so that the potential for replacement or decreased efficacy can be considered and inform policy.

A pneumolysin allele could not be extracted using this assembly-based method for only seven isolates; four non-typeables, two non-pneumococci and one serotype 21 isolate. Fifty-two unique pneumolysin DNA sequences (*ply)* were observed in the carriage study isolates. A further eight *ply* sequences previously deposited on Genbank (Kirkham *et al.* 2006; Jefferies *et al.* 2007) were not observed in this population.

Translation of the fifty-two nucleotide alleles detected in this dataset resulted in 20 unique protein alleles (Ply), of which ten were novel. The eight previously recognised nucleotide alleles that were not observed in this dataset represented a further eight known protein alleles. Therefore a total of 28 Ply alleles have now been reported to date. However the two potential non-pneumococcal isolates for which the Ply sequence could be determined accounted for one novel pneumolysin allele P20 (n=1) and the only observation of the previously documented pneumolysin sequence P17 allele (n=1), (Kirkham *et al.* 2006; Jefferies *et al.* 2010b). Despite being previously documented in pneumococci Ply 17 had a 100% BLASTP top match of the extracted

and translated protein sequence with the NCBI reference sequence (WP_001284352.1) of pneumolysin from a *Streptococcus pseudopneumoniae* strain, but a 99% 2nd match with GenBank AB021358.1 a pneumococcal strain. The novel Ply 20 allele had a 100% BLASTP top match with GenBank pseudopneumolysin ACJ76900.1 from a *Streptococcus pseudopneumoniae* strain and a further six 99% matches with Mitilysin and other pseudopneumolysins. The lengths of the Ply alleles observed were 469 AA (Ply 5 and Ply 28) and 471 AA with majority of alleles being the latter (Ply 1, 2, 8, 9, 11, 15-27).

3.2.1.1 Gene phylogeny constructed from Ply

A gene tree was created, based on the pneumolysin protein alleles, to describe the diversity of the protein sequences and their relation. DNA alleles were not used as multiple alleles coded for the proteins and protein alignments have 21 characters reducing multiple sequence misalignment when compared to the four 4 bases of DNA, therefore the phylogenies can be more robust. All novel Ply alleles grouped with other previously documented alleles in multiple alignments in terms of length and non-synonymous SNP/INDEL pattern (Figure 6) and the gene tree (Figure 7).

The gene tree phylogram (Figure 7) includes soft polytomies, when branches with very low (<10%) support are collapsed, that are a result from the reduced amount of data in the protein alleles for resolving the evolutionary relationship between closely related alleles. The predominance of strains carried the Ply-2 (51%) or Ply-1 (36%) alleles, when the Ply-1 and 2 clades were taken into account they accounted for 94% of the isolates analysed (41% and 53% respectively), only a small minority of strains (<6%) carried alleles that were not part of clades that included Ply-1 or Ply-2. Ply-28, Ply-23 and Ply-21 were the most commonly occurring novel alleles (n=5, 3 and 2 respectively) with all other novel alleles observed only in one isolate. The clade that includes Ply-2 shares a most recent common ancestor (MRCA) before it shares a MRCA with the clade that includes Ply-1. Within the Ply-1 clade, Ply-9 and novel Ply-27 are more closely related to each other than to the rest of the clade. For the Ply-2 clade, Ply-10 and novel Ply-28 are more closely related to each other than they are to the rest of the clade.

3.2.1.1.1 Clustering of novel Ply alleles

Novel Ply-26 observed in a NT strain in the dataset clusters with Ply-17 a pneumolysin allele observed in a non-pneumococcal strain in this dataset. Novel Ply-21 that was observed in NT strains 4096 and 4165, clusters with known Ply-16, which, was

observed in this dataset in the serotype 6D isolate 2029. The remaining novel alleles all belonged to one of the polytomous clades that include Ply1 (n=3) or Ply2 (n=3).

Importantly the placement in the gene tree of the ‘pneumolysin’ protein allele Ply-20 of isolate 5006, which is excluded from the pneumococcal carriage analysis as it was not identified as a pneumococcus using the criteria described in section 2.3.3 Species discrimination, supports the non-pneumococcal designation. Ply-20 from 5006 shares a MRCA with the pneumolysin homologue mitilysin (Jefferies *et al.* 2007) before it shares one with the known pneumolysin genes, therefore allele Ply-20 is not a novel pneumolysin but a homologue and a total of 27 genuine pneumococcal protein alleles have now been described to date.

Figure 6. Multiple alignments of pneumolysin protein alleles, known and novel, AA 1-244

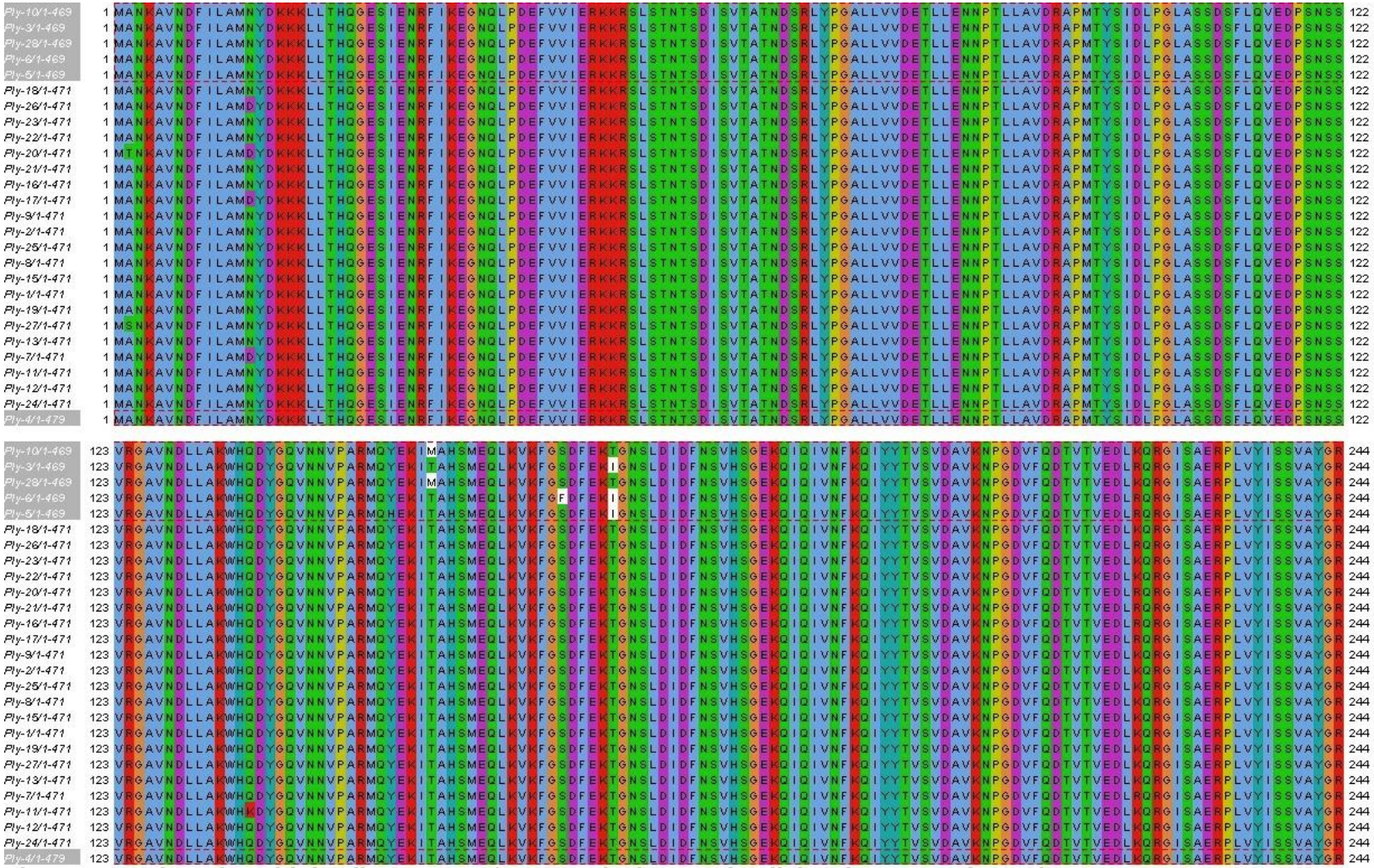
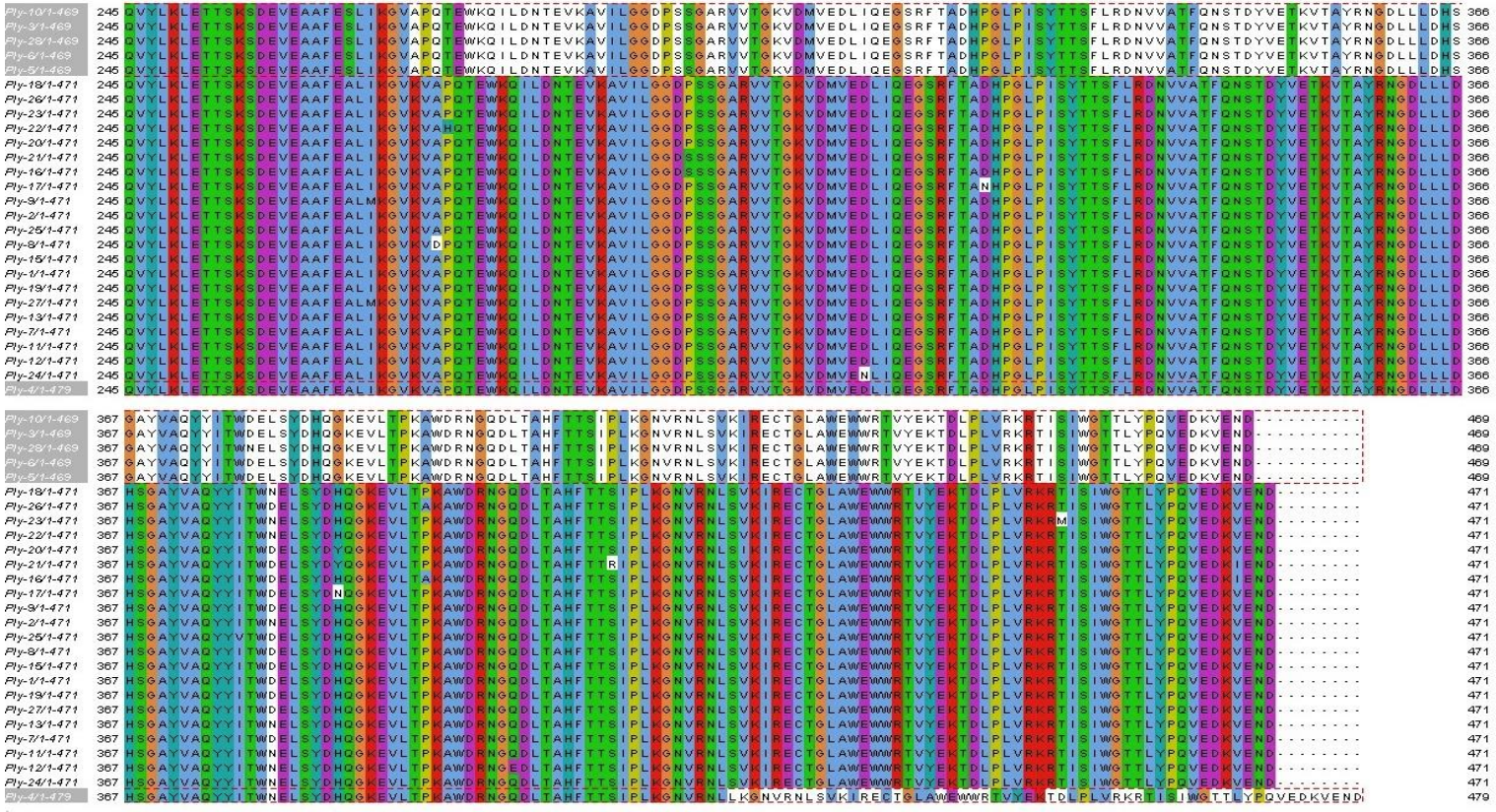
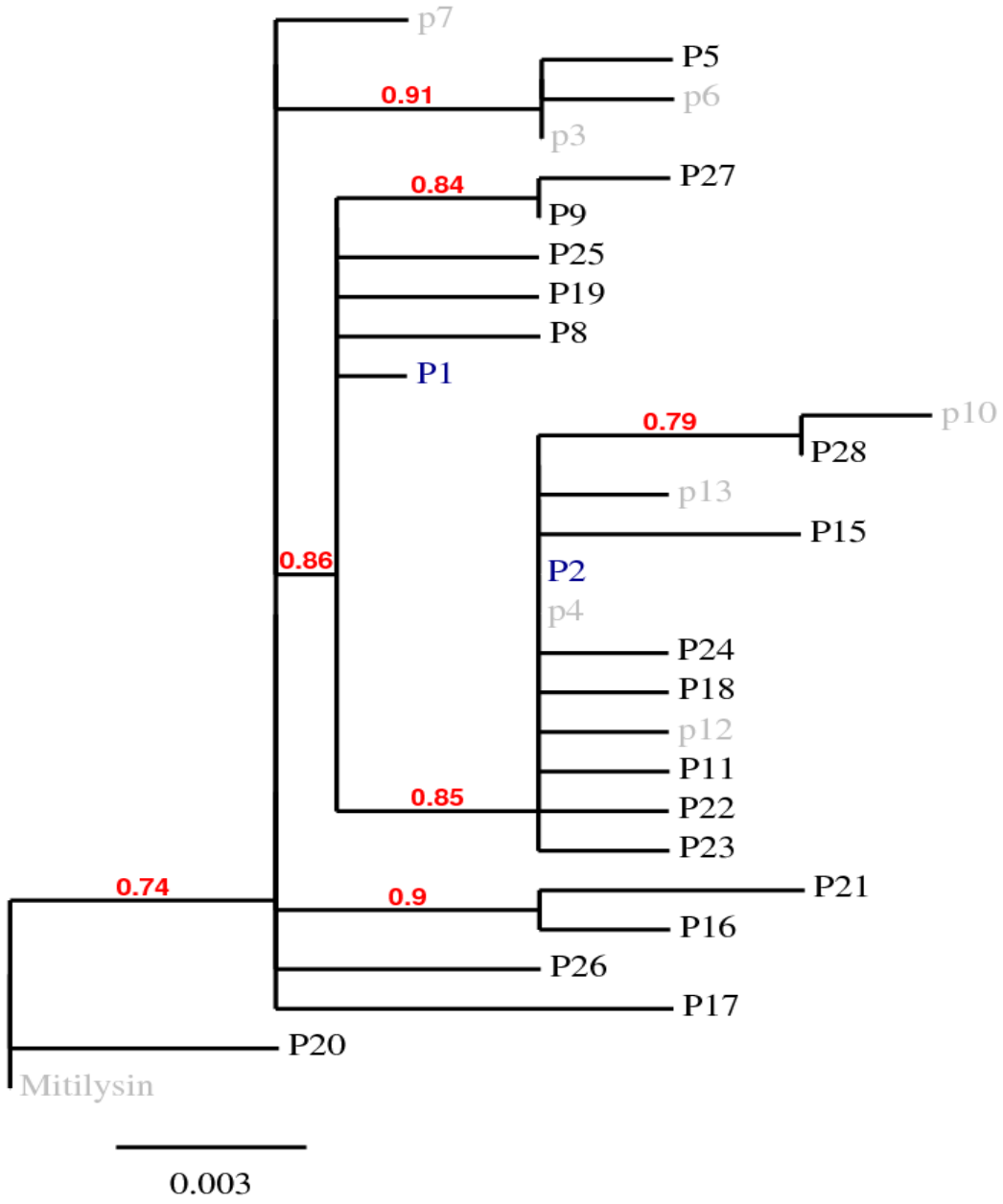


Figure 6 continued. Multiple alignments of pneumolysin protein alleles, known and novel, AA 245-471



Ply 14 allele, which has a large insertion (Jefferies *et al.* 2007) was not observed in this dataset, and was not included in alignment for visualisation purposes in SeaView (Gouy *et al.* 2010). Greyed out allele names indicate groups of alleles differing in length from 471AA.

Figure 7. Gene phylogeny of pneumolysin protein alleles previously documented or observed in 519 genomes



Maximum likelihood phylogram of protein sequences for known pneumolysin alleles denoted P if present in the dataset and p (grayed text) if not observed in this carriage dataset. Branch support values are given in red. Translated mitilysin gene (GenBank EF066519.1) used as an out-group, to root the tree. Blue font highlights the dominant protein alleles Ply 1 and Ply 2. Branches with support values of less than 10% were collapsed. Ply 14 allele, which has a large insertion (Jefferies *et al.* 2007) and was not observed in this dataset, was not included in phylogram for visualisation purposes.

3.2.2 Capsular gene sequences

The genetic diversity within capsule defining genes has the potential to affect capsular structure, immunogenic profile and therefore the efficacy of vaccines against different subsets of a serotype (as defined at present) or recognition of a novel serotype. Therefore the sequence of capsule defining genes was used to definitively designate serotype for serotypes for which the genetic basis is known but are not all readily distinguished by traditional methods. This analysis was performed for serotypes within the serogroups 6 (Park *et al.* 2007a; Bratcher *et al.* 2010), 11 (Oliver *et al.* 2013a), 15 (van Selm *et al.* 2003), 19 (Morona *et al.* 1999) and 22 (Salter *et al.* 2012) as they accounted for 310 (60%) of the isolates, the genetic basis was known and the serogroup components could not be otherwise be fully resolved using other genotypic methods. Serotype 21 *wzy* was also characterised as PCR based techniques gave ambiguous results (3.1.4.2 Serotype designation). The remaining serotypes within a serogroup for which the genetic basis is not known or were not common in the dataset were not analysed in this manner.

The increased serotype designation allows epidemiological analysis of serotype level fluctuations. Additionally determination of the sequence of these genes allowed characterisation of the diversity of those capsular defining genes within a single serotype that have previously been assumed to be highly conserved (Bentley *et al.* 2006).

3.2.2.1 Serogroup 6

Serogroup 6 currently includes four distinct serotypes, 6A-D, which were all observed in this data set (6 n=130, 6A n=39, 6B n=46, 6C n=44, 6D n=1). A single non synonymous SNP (within the *wci*P gene encoding a putative rhamnosyl transferase produces the immunological difference between serotype 6A and 6B (G584A, S195N) (Mavroidi *et al* 2004), whilst serotypes 6C and 6D have a 1,029bp sequence that replaces a 1,222bp region in *wci*N (314 AA), termed *wci*Nbeta (374 AA) both of which encode glycosyltransferases (Park *et al.* 2007a; Bratcher *et al.* 2010). The 6C/D *wci*Nbeta gene occurs in the 6A or 6B *wci*P SNP background respectively. These differences mean that serotypes in serogroup 6 can be easily deduced from the sequences of these two genes. The combination of a *wci*P allele containing the 6B non-synonymous SNP (Figure 8) and a 1455bp *wci*Nbeta allele (Figure 9), isolate 2029 was determined to be the only serotype 6D observation in this dataset. There are not currently PCR or antibody based techniques available to definitively assign 6D and this

isolate 2029 was previously designated 6C within this carriage study (Tocheva *et al.* 2011).

3.2.2.1.1 wciN/beta alleles

Within-serotype variation was observed for the *wci*N and *wci*Nbeta alleles that contribute to the 6A/6B 6C/6D capsular types respectively. There were six *wci*N 6A alleles that translated into three protein alleles that occurred in 35, 5 and 1 isolates, with differences at two positions (Ala52Ser, Ala150Thr). The Ala150Thr change in a 6A background has been recently reported and suggested to result in an alteration in the serological and chemical profiles and has been putatively assigned the novel serotype 6F (Oliver *et al.* 2013b), as this report comes after the formation of this thesis this single isolate 4237 from year 4 remains to be considered serotype 6A for epidemiological analysis.

The dominant 6B *wci*N allele (n=43) was identical to the dominant 6A *wci*N allele (n=35, Figure 9) but two additional alleles (DNA and subsequent protein) also existed each found in one isolate. One of these two proteins designated *wci*N 6B_B differed by only one amino acid (Gly121Asp) whilst the second allele in isolate 0041 designated 6B by Quellung had a much larger DNA allele; 2029bp rather than 1689bp and a 314AA coding sequence which differed dramatically in the 247-305 positions of the protein with changes at 24 different positions when compared to 6B_A (Figure 10). For 6C there were five alleles that translated to four protein sequences differing at three positions in the 374AA sequence for *wci*Nbeta (Tyr22Phe, Ala281Thr, Asn371Asp). The single representative of the 6D *wci*Nbeta allele (Figure 9) differed from the dominant 6C protein sequence by only one AA (Glu368Gly).

3.2.2.1.2 wciP alleles

The consistently distinguishing SNP of 6A from 6B is in *wci*P, however additional variation within this gene and its resultant protein, which potentially could also affect the immunogenic profile of the capsule, was observed. There were four distinct *wci*P alleles observed for 6A, one dominant allele *wci*P_6A_1 and three rare alleles *wci*P_6A_2, *wci*P_6A_3 and *wci*P_6A_4, for which there was some association with STs (Table 7). These alleles were all 987bp in length but differed at three nucleotide positions, all three were non-synonymous resulting in AA substitutions in the resultant 328AA protein in the open reading frame (Table 8A). The AAs involved in the substitutions have differing properties of hydropathy, polarity and charge of side chains (Table 9) and could therefore alter protein structure of the putative rhamnosyl

transferase encoded by the *wci*P gene that is involved in sugar linkage for which a single non-synonymous SNP differentiates 6A and 6B.

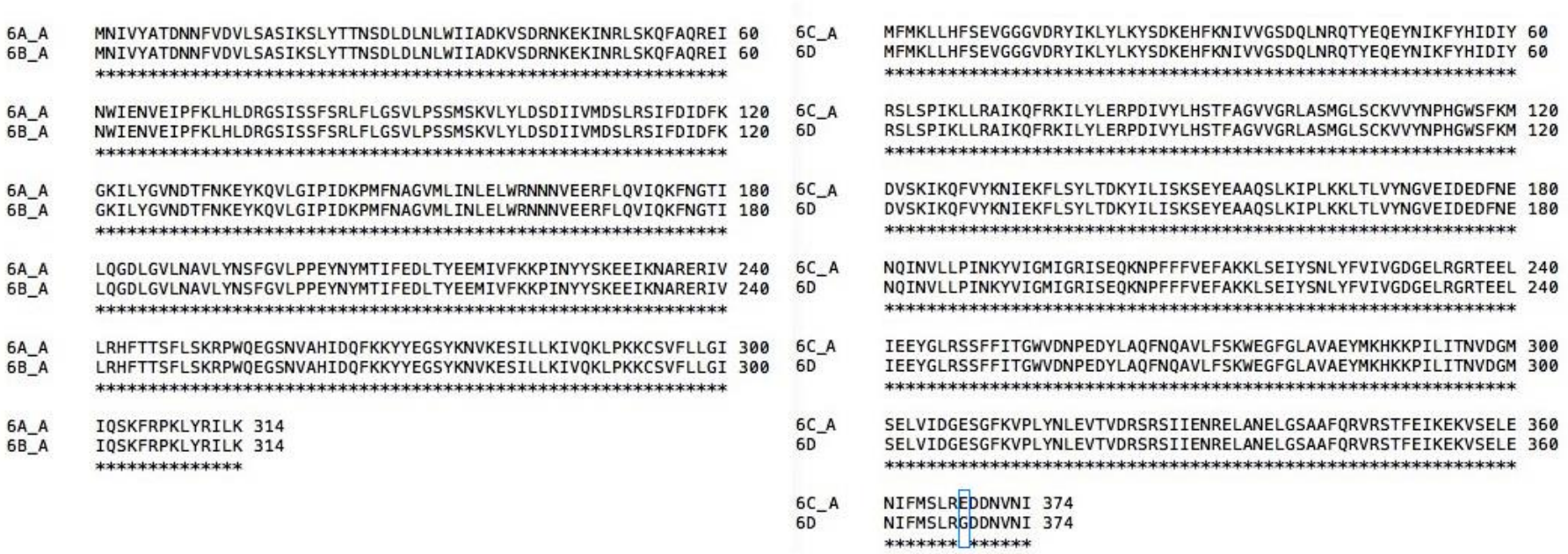
Similarly for the *wci*P gene in serotype 6B strains there were 5 distinct alleles with one dominant allele *wci*P 6B_1 and four rare alleles *wci*P 6B_2 *to wci*P 6B_5. All five alleles were 987bp in length but with a greater number of variable sites than for the *wci*P 6A alleles, which result in five non-synonymous changes and four distinct protein alleles (Table 8). Translations of *wci*P_6B_1 and *wci*P_6B_4 have only synonymous differences and the protein termed wciP_6B_A. The AAs involved in the substitutions that result in the four distinct protein sequences have differing properties of hydropathy, polarity and charge of side chains (Table 9) although for residue 62 this is only a small change in hydrophilicity from substitution of Leucine for Phenylalanine in 6B_D. All five sites are different to the three sites of non-synonymous variance in *wci*P 6A alleles. The *wci*P gene for 6C isolates had four distinct alleles with four variable sites in the 987bp sequence of which three were non-synonymous (Table 8C). Again *t*he AAs involved in the substitutions have differing properties of hydropathy, polarity and charge of side chains although for residue 129 this is only a small change in hydrophilicity from substitution of Leucine for Methionine in 6C_A and 6C_D (Table 9)

The relationship of the serogroup 6 alleles is inferred in Figure 11. It is clear that the 6D wciP allele is most closely related to serotype 6B as it shares a MRCA with a number of 6B alleles before it shares a MRCA with the other serotypes. It is also clear than the *wci*P alleles in 6C isolates are most closely related to serotype 6A alleles as they share a MRCA with all 6A alleles before they share a MRCA with the other serotypes. This supports the theory that 6C descended from 6A and suggests that 6D has arisen from 6B with the recombinogenic replacement of *wci*N with *wci*Nbeta. Conversely the *wci*Nbeta allele has a single non-synonymous SNP from the dominant 6C *wci*Nbeta allele, supporting the hypothesis that the origins of the *wci*Nbeta allele of the 6D isolate is serotype 6C or a common source (Figure 9).

Figure 8. Protein alignment, of dominant, serogroup 6 *wci*P alleles

The defining 6A/6C and 6B/6D amino acid change, Serine (S) to Asparagine (N) respectively, the consequence of a single SNP

Figure 9. Protein alignment, of dominant, serogroup 6 *wci*N and wciNbeta alleles



The only non-synonymous SNP difference between the dominant 6C wciNbeta allele and the observed 6D allele

Figure 10. *wci*N protein alleles from serotype 6B isolates

Blue box highlights single AA change between wciN protein alleles 6B_A and 6B_B.

Red box highlights region of high AA change in protein allele 6B_C.

Table 7. Protein allele frequency and associated ST for serogroup 6 *wci*P gene

	Count	Year/s	ST/s
wciP_6A_A	31	1-5	65, 460
wciP_6A_B	6	1,2,4 & 5	42, 327, 1862
wciP_6A_C	2	1 & 4	1876, 3240
wciP_6A_D	1	2	1862
wciP_6B_A	42	1-5	138, **176**, 402, 469, 2457, 2496, 4020
wciP_6B_B	1	1	**176**
wciP_6B_C	1	4	**176**
wciP_6B_D	1	2	242
wciP_6C_A	35	1-5	395, 1692, 1714, Novel ST
wciP_6C_B	7	2,3 & 5	398, 1150, 1390, 4581, 8028
wciP_6C_C	1	3	1600
wciP_6C_D	1	2	1692

Bold type ST indicates STs for which multiple wciP alleles were observed

Table 8 A, B & C. Substitutions in protein alleles for serogroup 6 *wci*P gene

A. (6A)	Position of 328		
Protein	**128**	**195**	**255**
wciP_6A_A	**M**	S	R
wciP_6A_B	T	S	**M**
wciP_6A_C	T	**N**	R
wciP_6A_D	T	S	R

B. (6B)	Position of 328				
Protein	**62**	**134**	**244**	**282**	**289**
wciP_6B_A	**F**	N	G	I	G
wciP_6B_B	F	N	G	I	**S**
wciP_6B_C	F	N	G	**S**	G
wciP_6B_D	**L**	K	**R**	I	G

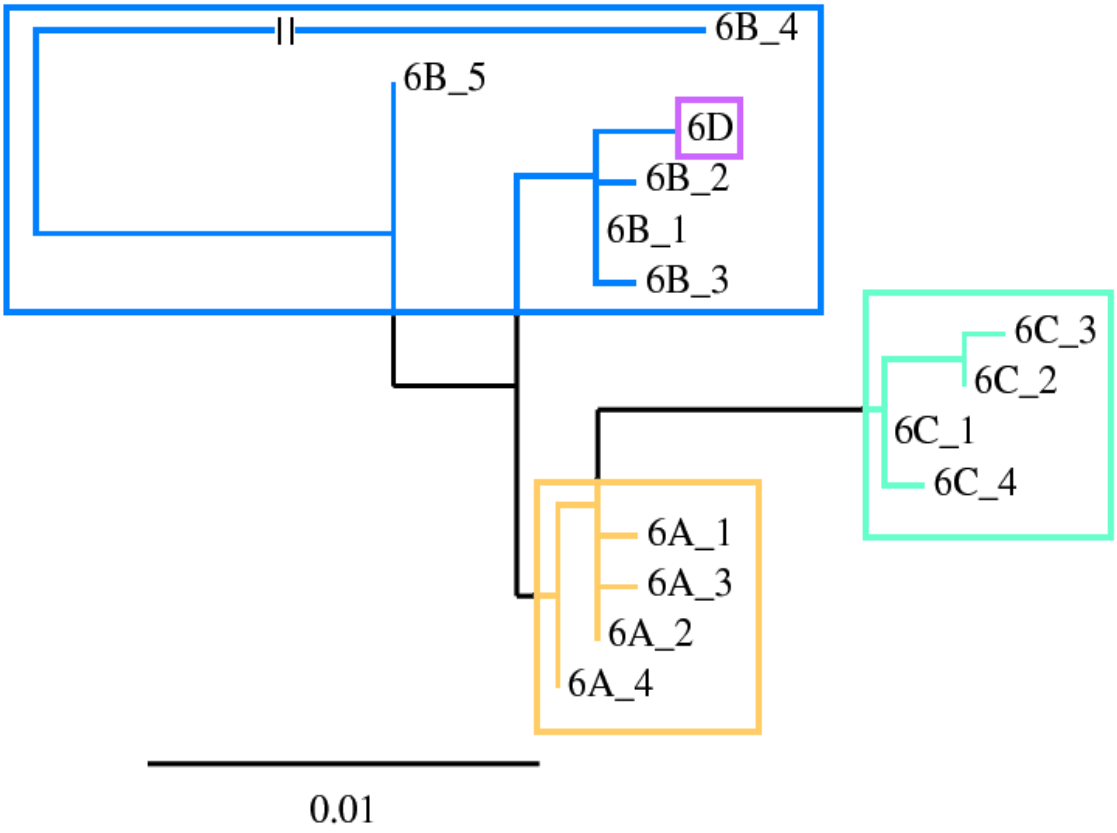
C. (6C)	Position of 328		
Protein	**44**	**126**	**313**
wciP_6C_A	D	**M**	I
wciP_6C_B	D	L	I
wciP_6C_C	**Y**	L	I
wciP_6C_D	D	**M**	**N**

Table 9. Amino acids and their properties

Amino Acid	3 letter code	1 letter code	Side-chain polarity	Side-chain charge	**Hydropathy index**
Alanine	Ala	A	Non-polar	Neutral	1.8
Arginine	Arg	R	Basic polar	Positive	-4.5
Asparagine	Asn	N	Polar	Neutral	-3.5
Aspartic acid	Asp	D	Acidic polar	Negative	-3.5
Glutamic acid	Glu	E	Acidic polar	Negative	-3.5
Glycine	Gly	G	Non-polar	Neutral	-0.4
Isoleucine	Ile	I	Non-polar	Neutral	4.5
Leucine	Leu	L	Non-polar	Neutral	3.8
Lysine	Lys	K	Basic polar	Positive	-3.9
Methionine	Met	M	Non-polar	Neutral	1.9
Phenylalanine	Phe	F	Non-polar	Neutral	2.8
Serine	Ser	S	Polar	Neutral	-0.8
Threonine	Thr	T	Polar	Neutral	-0.7
Tyrosine	Tyr	Y	Polar	Neutral	-1.3
Valine	Val	V	Non-polar	Neutral	4.2

Adapted from (Kyte & Doolittle 1982; Cooper & Hausman 2004), only AAs relevant to results included

Figure 11. Phylogram of *wci*P DNA alleles for serogroup 6



A phylogram, gene tree of serogroup 6 wciP using mid-point rooting. Coloured blocks highlight different serotypes. Broken line indicates shortened branch for visualisation purposes.

3.2.2.2 Serogroup 11

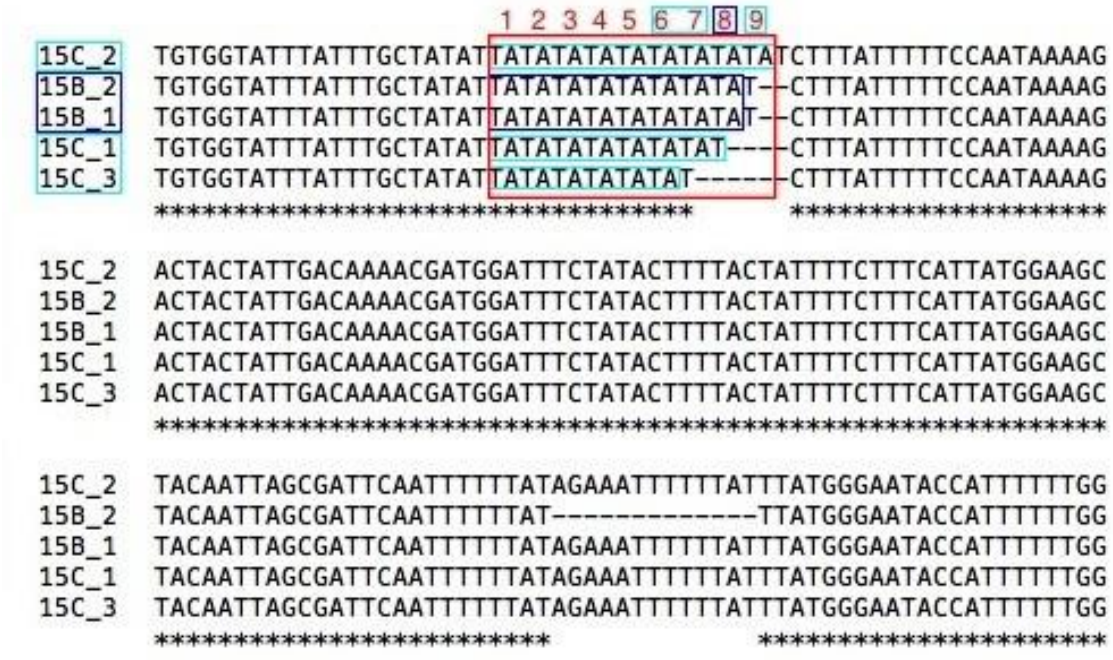
Serogroup 11 currently includes six serotypes, 11A-E, of which only 11A were observed in this dataset (11A n=37). Serotypes 11A and 11D were indistinguishably typed as 11A/D in the CDC PCR methodology (Centers for Disease Control and prevention 2012). Recently the genetic basis for distinguishing 11A and 11D has been determined as a non-synonymous SNP in the glycosyl-transferase encoding gene *wcr*L. This results in the replacement of Asparagine with Serine in 11D at AA112 (Oliver *et al.* 2013a). The presence of serine AA122 in the allele translations was therefore used to designate the serotype 11A to all 11A/D PCR typed isolates (n=37) four of which gave subjectively ambiguous results for slide agglutination of factor 11b which should be negative for 11A and positive in 11D and appeared to exhibit weak positivity. All isolates had an identical nucleotide and consequent protein allele for *wcr*L except for one isolate (5113) that had two SNPs resulting in non-synonymous changes in the 240AA protein at positions 153 and 159 the former being a change from Threonine, a AA with a polar hydrophobic side chain, to Isoleucine with a non-polar hydrophilic side chain and the latter Leucine to Phenylalanine (Table 9).

3.2.2.3 Serogroup 15

Serogroup 15 currently includes four serotypes, 15A-C, F, of which only 15A-C were observed in this dataset (15 n=50, 15A n=13, 15B n=23, 15C n=14). The wzy gene is used to differentiate 15A/F from 15B/C using PCR (Centers for Disease Control and prevention 2012). Quellung was used to further differentiate 15A from 15F. The genetic and biochemical basis is known for the differences between 15B and 15C, the *wciZ* gene (synonym cps15bM) allows differentiation of these serotypes from the sequence. 15B contains an eight unit tandem repeat of TA, 15C has a more variable number of TA tandem repeat units which result in a non-functional O-acetyltransferase gene and differences in O-acetylation between the two serotypes (van Selm *et al.* 2003). The tandem repeat was used to serotype 15B/C isolates in this study (Figure 12). For isolates of serotypes 15B (n=23) and 15C (n=14) two and three DNA alleles were observed respectively. Previously van Selm *et al,* reported that 15C isolates contained 7 or 9 TA repeats, this observation was repeated in this dataset, additionally two isolates of 15C (0145, 0310) were observed to have only 6 TA repeats.

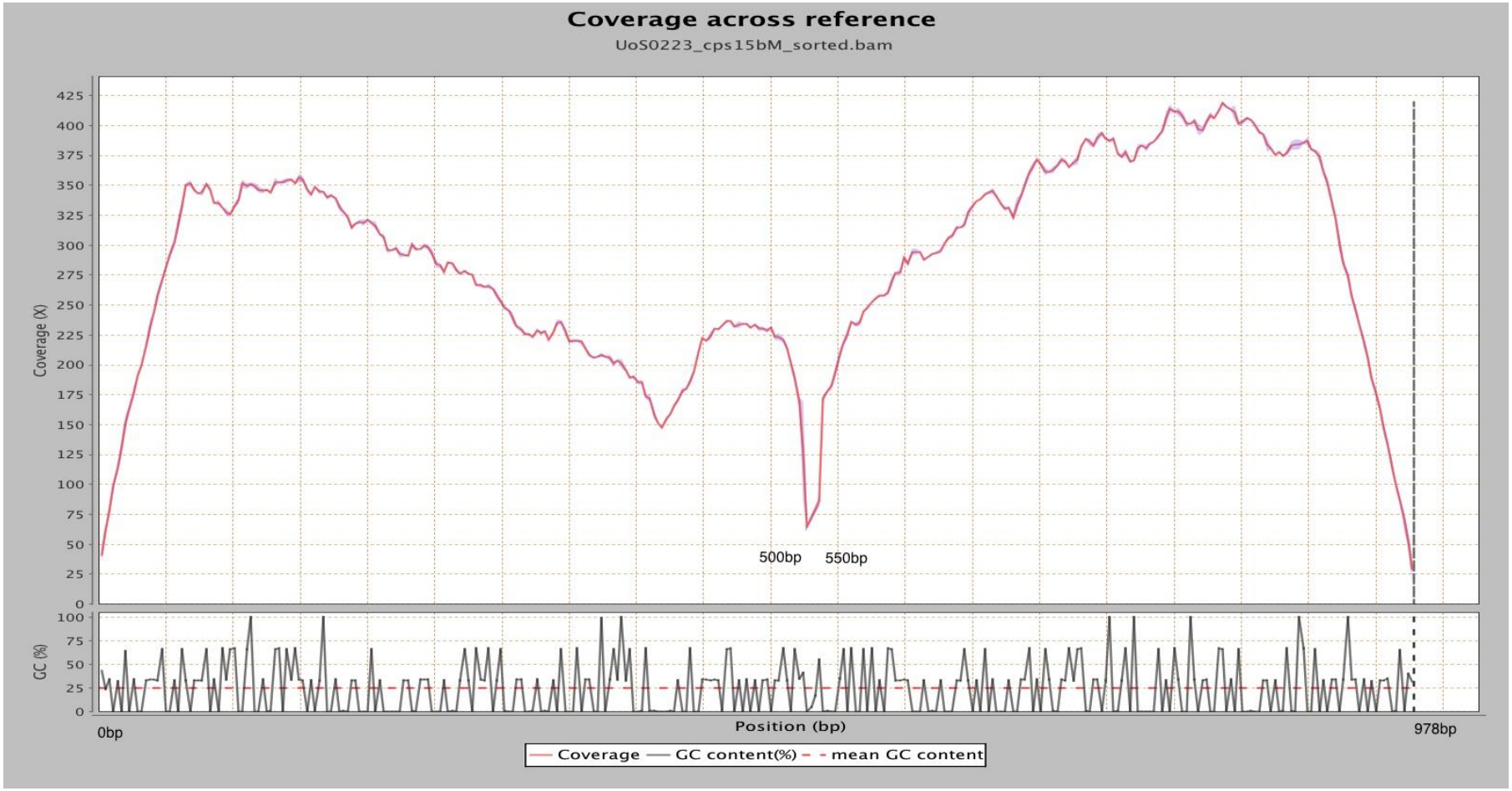
All alleles for *wciZ* for 15B and 15C alike differed only in the tandem repeat region with the exception of 15B_2 which was only observed once in isolate 0223, and despite having the 15B eight-unit tandem repeat had an additional putative deletion of 13 bp in the assembly (Figure 12), which, when translated would induce a premature stop in the protein. However mapping against a wciZ reference sequence (GenBank AY250187.1) revealed complete coverage of the *wciZ* gene, with a drop in coverage at the region in question, 535-547 of 978bp (Figure 13). Viewing the BAM file in Artemis revealed that a large number but not all reads contained the deletion (Figure 14). With previous work documenting this isolate as 15B by Quellung combined with mapping evidence the observations for this isolate are best explained by a mixed sample with two separate coding sequences for wciZ, one with a deletion which would likely result in 15C like capsular expression and one without that would account for the 15B positive Quellung reaction.

Figure 12. Partial sequence DNA Alignment of the *wciZ* gene tandem repeat region



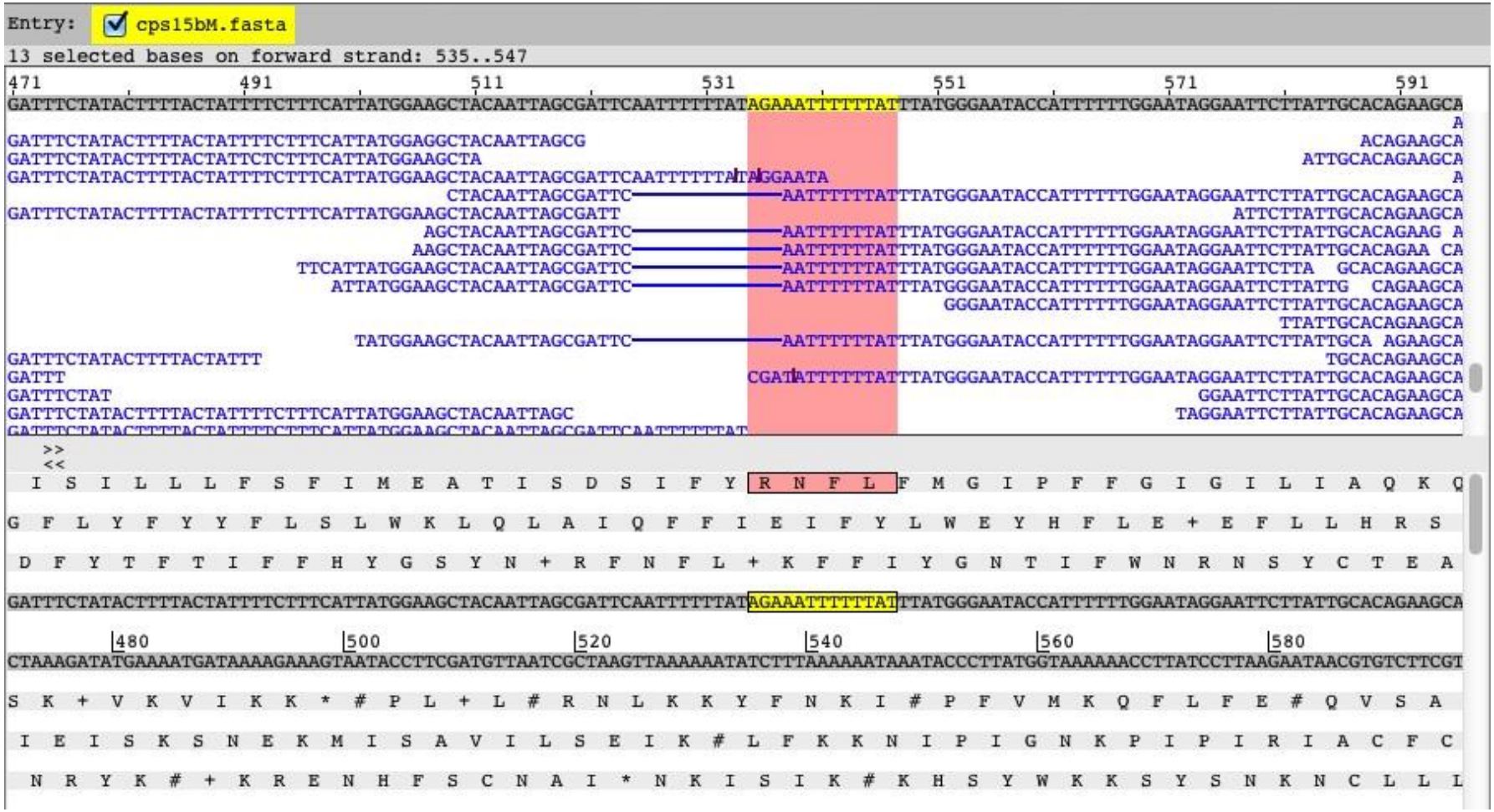
Tandem TA repeat region highlighted in red. Eight TA units, which infer 15B capsule type highlighted in dark blue. Six, seven and nine TA units indicative of 15C capsule.

Figure 13. Coverage across wciZ reference of isolate 0223



Flanking base positions either side of region missing from assembly show a clear drop in coverage at this location

Figure 14. Section of BAM read pileup for isolate 0223 against reference for *wciZ*



The region missing from the assembly is highlighted in red (reads) and yellow (reference sequence).

3.2.2.4 Serogroup 19

There are four recognised members of serogroup 19, 19A, B, C and 19F, of which only 19A and 19F were observed in this dataset (19A n=31, 19F n=20). Serotypes 19A and 19F are reported to have syntenic (same arrangement of genetic loci) capsular loci and structurally differ only in linkage between glucose and rhamnose during polymerization of the polysaccharide capsule; this is thought to be conferred by the action of the *wzy* gene a putative oligosaccharide repeat unit polymerase (Morona *et al.* 1999; Mavroidi *et al.* 2007). Within the dataset described in this thesis, the alleles for 19A and 19F were clearly specific for the serotype with a large number of SNPs and a ClustalW score of only 78.43 for the two dominant alleles confirming the serotype designation by PCR methods.

Four distinct alleles were observed the for the 19A-specific 1335bp *wzy* gene with four variable base positions. When translated to a 444AA protein there were only three alleles that varied in frequency in the population (Table 10A). Interestingly two of the *wzy* protein alleles (19A_A,19A_B) that differ by two AAs are observed within a single ST 199. The position 109 Isoleucine-Valine change and position 88 Valine-alanine change resulted in a minor and larger reduction in hydrophilicity respectively. The Isoleucine-serine change at position 81 could have a more substantial effect on the resultant protein with changes in side chain polarity and a switch between being hydrophobic and hydrophilic (Table 9).

Three distinct alleles were observed the for 19F-specific 1338bp *wzy* gene with three variable base positions. When translated to a 445AA protein there were only two alleles circulating the dataset with similar incidence (Table 10B). The AA change only resulted in a change in hydrophilicity (Table 9).

Table 10 A & B. Substitutions in protein alleles for serogroup 19 *wzy* gene

A. (19A)	Position				
Protein	**81**	**88**	**109**	Count	STs
wzy_19A_A	I	V	I	22	199, 276, 416
wzy_19A_B	**S**	V	**V**	8	162, 193, 199, 201, 482
wzy_19A_C	I	**A**	I	1	172

B. (19F)			
Protein	Position 72	Count	STs
wzy_19F_A	I	12	162, 309, 422, 4696
wzy_19F_B	**L**	7	177, 179, 311

3.2.2.5 Serotype 21

Twenty-eight isolates were determined to be serotype 21. The serotype 21 capsular locus has been shown to cluster with the biosynthetic loci of *Streptococcus mitis* and *oralis* whilst the CPS structure is unknown (Mavroidi *et al* 2007). Therefore the key genetic components of the serotype 21 expression are still to be determined. The *wzx* gene, which encodes a flippase involved in translocation of the CPS, has however been deemed suitably divergent to *wzx* found in other serotypes to allow the capsular designation of serotype 21 based on PCR amplification of a band of 192bp from this gene (Centers for Disease Control and prevention 2012). The exact sequence of this gene was therefore selected to assess the sequence diversity of this gene. This serotype was displaying multiple bands PCR positive for alternative serotypes also typed using the *wzx* gene (23B and 35A/35C/42) using the CDC methodology (Centers for Disease Control and prevention 2012). There were two alleles observed, both 1425bp long with only a single bp difference, the C-T SNP in the dominant allele (n=19) also differed from the gi|68643497_13131-14555 GenBank reference. The 2nd allele (n=9) was identical to the reference. The two resulting protein sequences differed at only one position (Phe445Ser).

3.2.2.6 Serogroup 22

Sergroup 22 consists of serotypes 22A and 22F of which only 22F was observed in this dataset (22F n=27). The genetic basis for differentiation of 22A and 22F, based on the highly distinct and *wcw*C genes (Salter *et al.* 2012), was delayed due to the sequencing of an atypical 22F in a previous study (Mavroidi *et al.* 2007). The 22A/F designation of the CDC PCR was supported by the sequences of the *wcw*A and *wcw*C genes and allowed a designation of 22F, which was also further confirmed by slide agglutination as part of this thesis for isolates from years 4-5 and by Statens Serum Insititut for all other isolates.

The *wcw*A and *wcw*C genes encode a glycosyltransferase and O-acetyltransferase both involved in polysaccharide biosynthesis. Two 1270bp alleles of 22F *wcw*A were observed with an insertion in both alleles when compared with the 22F *wcw*A reference HE651300.1 out with the coding sequence when translated and not affecting the protein sequence. An additional SNP in the was present within the protein coding sequence in the rare allele (n=1), which was non-synonymous at AA position 144 of 402 (Alanine to Valine) when compared to the dominant allele (n=27) with an increase in hydrophilicity. Furthermore there was only one 705bp allele for *wcw*C, which differed

by only one non-synonymous SNP from the 22F *wcw*C reference HE651300.1 with an Alanine to Valine difference from the reference at AA164 of 234.

3.2.3 Discussion of results

In recent years pneumococcal studies additional to that described in this thesis have been published that have deduced MLST and serotype data from WGS primarily using a mapping approach for deriving this data (Croucher *et al.* 2011; Everett *et al.* 2012). The extra resolution and complete repertoire that whole genome data offers allowed the designation of serotypes and identification of heterogeneous alleles for further analysis from an epidemiological perspective within this thesis. The characterisation of a number genes involved in pneumococcal virulence, that code for current and experimental vaccine targets was necessary as this thesis hypothesized that carriage isolates would represent a more diverse collection, than the IPD isolates that are traditionally studied, with implications for vaccine usage and design. The observation of novel pneumolysin alleles in the carried pneumococcal population and the presence of these alleles and homologues in other streptococcal species (Jefferies *et al.* 2007; Johnston *et al.* 2010) has importance for the design and development of future vaccines using a pneumolysin component. The inclusion of pneumolysin is designed to provide broad serotype independent protection therefore the specificity of an immune response is important, it could preferentially act upon particular pneumolysin protein variants potentially leading to vaccine type replacement or conversely have broad activity and also apply immunological pressure to non-pneumococcal strain containing pneumolysin homologues potentially leading to species replacement.

This thesis also documents that there is protein sequence diversity for a number of serotypes -in the genes that determine capsular type- that has not been well described to date. Until recently, only one representative cps locus sequence was published for 90 of the >95 serotypes (Bentley *et al.* 2006) but as PCVs have been introduced a number of studies have looked in more detail at the diversity of cps genes within a serogroup - primarily serogroup 6 and 19 with multiple isolates for each serotype (Elberse *et al.* 2011a; Song *et al.* 2011; Oliver *et al.* 2013b). However the number of isolates representing each serotype was still less than 14 for each of these studies compared to the larger number of strains in this thesis ranging from 14 strains per serotype (15C) up to 46 isolates for 6B (with the exception of 6D, n=1 and putative 6F, n=1).

For serogroup 6, Elberse *et al* reported three 6A *wci*N alleles with differences in two positions Ala52Ser also observed in this dataset and Glu77Gly only observed as Glutamic acid. Two alleles were observed for 6A *wci*P, Met128Thr only observed as Methionine in this thesis. For serotype 6B there were no non-synonymous SNPs reported in the *wci*N or *wci*P alleles by Elberse *et al* contrary to the four observed in this thesis. Finally for 6C they observed the Leu126Met but not the additional Asp44Tyr, Ile313Asn changes identified in this thesis. This thesis and Elberse *et al* identified 3 variable protein positions for *wci*Nbeta in 6C but only one of these was present in both datasets Try22Phe. The serogroup *wci*N protein substitution at AA121 has also been observed, with additional changes at site 137 reported but not observed in this dataset (Song *et al.* 2011). Finally the single AA sequence polymorphism was also observed between 6C and 6D isolates at Glu368 outside this thesis (Song *et al.* 2011). There have been additional investigations into the evolution of serogroup 6 that report multiple independent recombinogenic events that result in the serotype 6D, in this thesis the 6D data from *wci*Nbeta and *wci*P suggests a similar event to that which Bratcher *et al* describe with 6C being the donor and 6B the recipient (Bratcher *et al.* 2011; Otsuka *et al.* 2013).

Elberse *et al* additionally looked into serotypes 19A/F where they identified a single non-synonymous SNP in 19A *wzy* and “a few non-synonymous substitutions” (Elberse *et al.* 2011a) in 19F *wzy* with one strain reported to have 2 SNPs and a second to have 4 when compared to a reference but it was not disclosed which nature these were or whether some of the SNPs were shared. This suggests that they observed more diversity in 19F than 19A in *wzy* unlike the observation made that 19A was more diverse in this thesis. Their data also suggested that these variants were circulating in the same population simultaneously with multiple representatives of the different *wzy* alleles for 19A and 19F (Elberse *et al.* 2011a).

As perhaps expected when sampling larger number of isolates novel amino acid substitutions were observed in capsular defining genes within this dataset that were not observed in other studies to date, a number of non-synonymous changes were identified by both recent work by others and this thesis. However studies also identified additional protein sequences not seen in this dataset highlighting that the current sampling depth is not sufficient to allow the identification of all variants in the diverse pneumococcal population (Elberse *et al.* 2011a; Song *et al.* 2011). Whilst this thesis investigated the diversity in the capsular defining genes of particular serogroups others have looked at diversity in a number of capsular genes (Elberse *et al.* 2011a; Song *et*

al. 2011), there is potential for changes in any cps loci gene to result in novel serotypes that would not have been characterised in this analysis.

The investigation by Oliver *et al* into the serological and chemical profiles of serogroup 6 isolates resulted in the indication of two novel serotypes and further demonstrates the power of single non-synonymous SNPs on immunogenic epitopes and the importance of investigating within serotype genetic diversity. Any of the amino acid substitutions have the potential to be a subclass of a serotype and affect affinity or have a different immunogenic profile and be a novel serotype belonging to that serogroup. The ability to screen genome data real time as information on the genetic basis is published for novel serotypes or other new discoveries is a strong advantage of working with whole genome sequences and allowed the quick identification of the single isolates of serotype 6D and 6F in the carriage dataset.

4. CHAPTER 4 RESULTS – EPIDEMIOLOGY

It was hypothesised that during PCV implementation there would be a shift in pneumococcal carriage epidemiology as the vaccines target a minority of the existing serotypes. This was hypothesised to occur through vaccine serotype and associated genotypes decreases with subsequent replacement by non-vaccine serotype and associated genotypes. Comparisons between years of the study were made to identify any such changes with statistical methods.

4.1 Carriage epidemiology

A total of 1,712 swabs were collected between winters 2006/7 and 2010/11 with 287-399 collected in a single year to obtain ~100 pneumococcal isolates in each study period (Table 11).

4.1.1 Isolation rates

Only isolates successfully sequenced and unequivocally determined to be *S. pneumoniae* within this study were used to calculate carriage rates. Pneumococci were isolated from 519 of the 1,712 swabs (Table 11) giving an overall carriage rate of 30.32% for the five years with 95% confidence of the true carriage rate being within 2.17% of this value (28.15 to 32.49%). The base line carriage rate in year one was 31.79% (CI, 26.72 to 36.86%). The lowest and highest carriage rates of 27.82% and 34.49% occurred in years four (CI, 23.42 to 32.22%) and five (CI, 28.99 to 39.99%) respectively with a non-significant difference in carriage rates between these two years based on the Fishers exact test which produced a two tailed p-value (p=0.065), combined with considerable overlap in their confidence intervals. There were no significant differences between the base line carriage in year 1 and any other year, with p-values of >0.1 (Table 11). The carriage rate has therefore been observed to remain statistically stable throughout the study period.

Table 11. *Streptococcus pneumoniae* carriage

		S. pneumoniae			
Study year (Winter)	Number of swabs	Isolated	% carriage	95% CI	p-value
1 (2006/7)	324	103	31.79	26.72 to 36.86	Base line
2 (2007/8)	373	104	27.88	23.33 to 32.43	0.28
3 (2008/9)	328	102	31.40	26.38 to 36.42	0.87
4 (2009/10)	399	111	27.82	23.42 to 32.22	0.25
5 (2010/11)	287	99	34.49	28.99 to 39.99	0.49
Combined	1712	519	30.32	28.15 to 32.49	

4.1.2 Participant metadata

Participant demographic and clinical information was available from a questionnaire completed by the parents/guardians of participants in the 5th study year. Comparisons were made between children from whom pneumococci were isolated (carriers) and those children whose nasopharyngeal swabs did not grow pneumococci (non-carriers). The mean age of carriers and non-carriers were 23.28 and 18.66 months old respectively, there was a statistically significant difference between the mean ages for these two groups (p=0.019) using a two-tailed independent samples T-Test for differences in means, with the pneumococcal carriers being statistically older. Of the carriers and non-carriers 54.55% and 65.08%, respectively, were under 2 years old. There were no statistically significant differences detected between carriers and non-carriers, for the age categories 0-23, 24-35, 36-47 or 48-60 months of age (Table 12). There was weak evidence for a difference in pneumococcal carriage status in those of less than 1 year old (p=0.0542) with a greater proportion of non-carriers than carriers belonging to this age group. However the 95% confidence intervals overlapped (non-carriers 34.25% to 48.29%, carriers 20.33% to 38.25%).

There was no significant association or disassociation between antibiotic usage in the past month and pneumococcal carriage (Table 12). A statistically significant difference between carriers and non-carriers was observed when the occurrence of any upper respiratory tract illness (URTI) in the last month was reported (p=<0.0001) with more URTIs reported for the pneumococcal carriers (Table 12). The appropriate doses of PCV, according to their guardian, had been received for over 91% of non-carriers and 88.79% of carriers, with no statistically significant difference between the two groups (p=1) (Table 12).

Table 12. Year 5 metadata analysis of pneumococcal carriers versus non-carriers

Age category in months	Pneumococcal negative participants (N)	Proportion of Pneumococcal negative participants (%)	CI 95%	Pneumococcal positive participants (N)	Proportion of Pneumococcal positive participants (%)	CI 95%	P-value
0-23	123	65.08%	61.43% to 74.73%	54	54.55%	44.74% to 64.36%	0.0975
24-35	27	14.29%	9.3% to 19.28%	17	17.17%	9.74% to 24.6%	0.6052
36-47	20	10.58%	6.19% to 14.97%	13	13.13%	6.48% to 19.78%	0.561
48-60	16	8.47%	4.5% to 12.44%	13	13.13%	6.48% to 19.78%	0.2215
Antibiotics in past month	44	23.28%	14.96% to 31.6%	26	26.26%	17.59% to 34.93%	0.567
Respiratory illness in past month	122	64.55%	55.13% to 73.97%	84	84.85%	77.79% to 91.91%	0.0003*
Received PCV doses as required	172	91.01%	85.38% to 96.64%	90	90.91%	85.25% to 96.57%	1

4.1.3 Serotypes

Thirty-eight different serotypes were observed during the study. Annually this ranged from 23 to 29 distinct serotypes, with eight serotypes (21% of all observed serotypes) observed at only one time point. Thirteen serotypes were observed in all time points (Figure 15), nine of which were NVT, three were VTs unique to PCV13 (6A, 19A and 3) and only one PCV7 VT was present in all years, serotype 6B. A single year represented a minimum of 59% of the serotypes seen in the entire study period. Eleven presumptive pneumococcal, *cpsA* negative, isolates could not be typed using PCR (traditional and *in silico*) or Quellung. Five of these isolates, all collected in year 5, exhibited features atypical of pneumococci and these were therefore excluded and analysed separately (2.1.3 Strain inclusion and exclusion). The seven remaining non-typeable pneumococcal isolates were observed in year 1 n=1, year 3 n=4 and year 4 n=2, these NT isolates were included as 7 examples of one NVT 'serotype' entity in the following calculations.

The Simpson's diversity index (1-D) was used to assess the serotype diversity of the population taking into account the number of types and their abundance, where 0 is equal to no diversity and 1, infinite diversity. A serotype diversity of 0.95 was calculated for the total study period. The indexes calculated for each study year remained stable and reflected the index for the total study period with a range of 0.91 (Year1) to 0.95 (Year 3 and 4). This describes a consistently highly diverse population with regards to serotype.

The most common serotype was 6B with 46 isolates closely followed by 6C (n=44) and 6A (n=39) representing 8.8% (6.36% to 11.24%), 8.5% (6.1% to 10.9%) and 7.5% (5.23% to 9.77%) of the complete isolate set respectively. The distribution of the study isolates across the serotypes observed when viewed in rank order of prevalence (Figure 16) is a continuous sliding scale from those serotypes which are very common through to those that are rare, with an absence of any steps separating the most common and rarest serotypes. This representation of the serotype makeup also makes clear that a large number of rarer serotypes contribute less to the total isolate set than a small number of very common serotypes with the first 8 serotypes accounting for >50% of the isolates and last 19 serotypes accounting for <10% of the isolates (Figure 16).

The 38 serotypes represented 26 serogroups, when this is accounted for in a Simpson's diversity index (1-D) calculation the population diversity is reduced to 0.89

but still describes a highly diverse serotype distribution. The predominant serogroup was serogroup 6 with 130 isolates representing over 25% of the complete isolate set (21.32% to 28.78%).

Figure 15. Distinct serotypes observed by time point

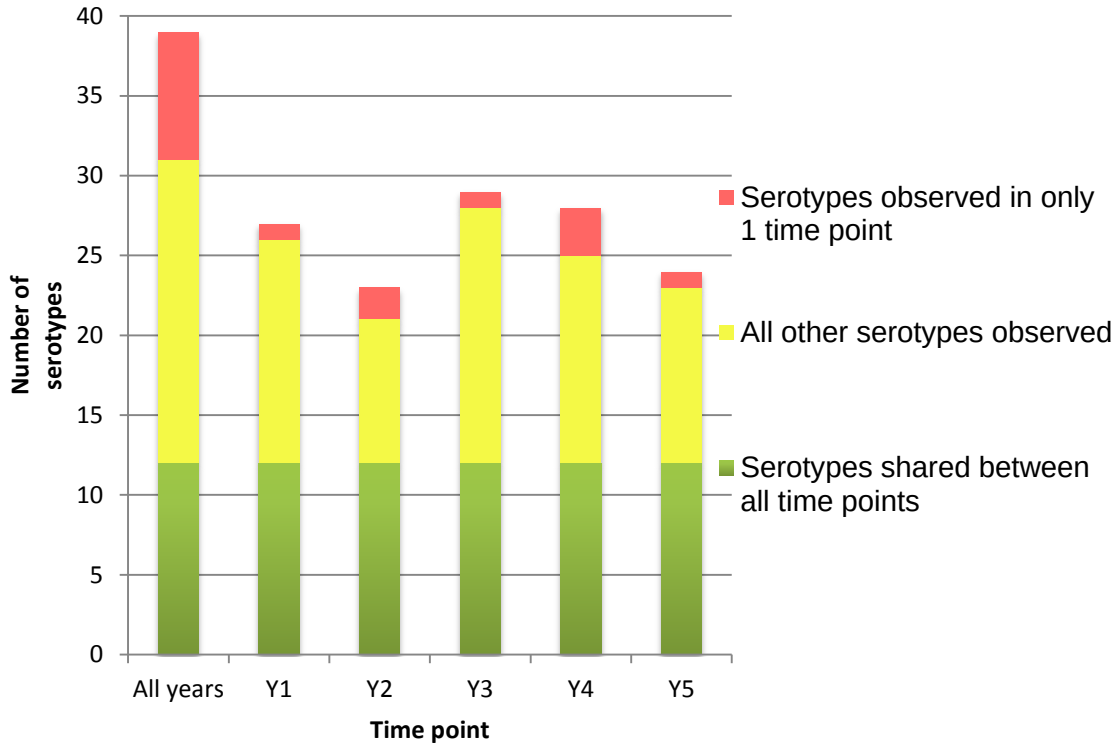
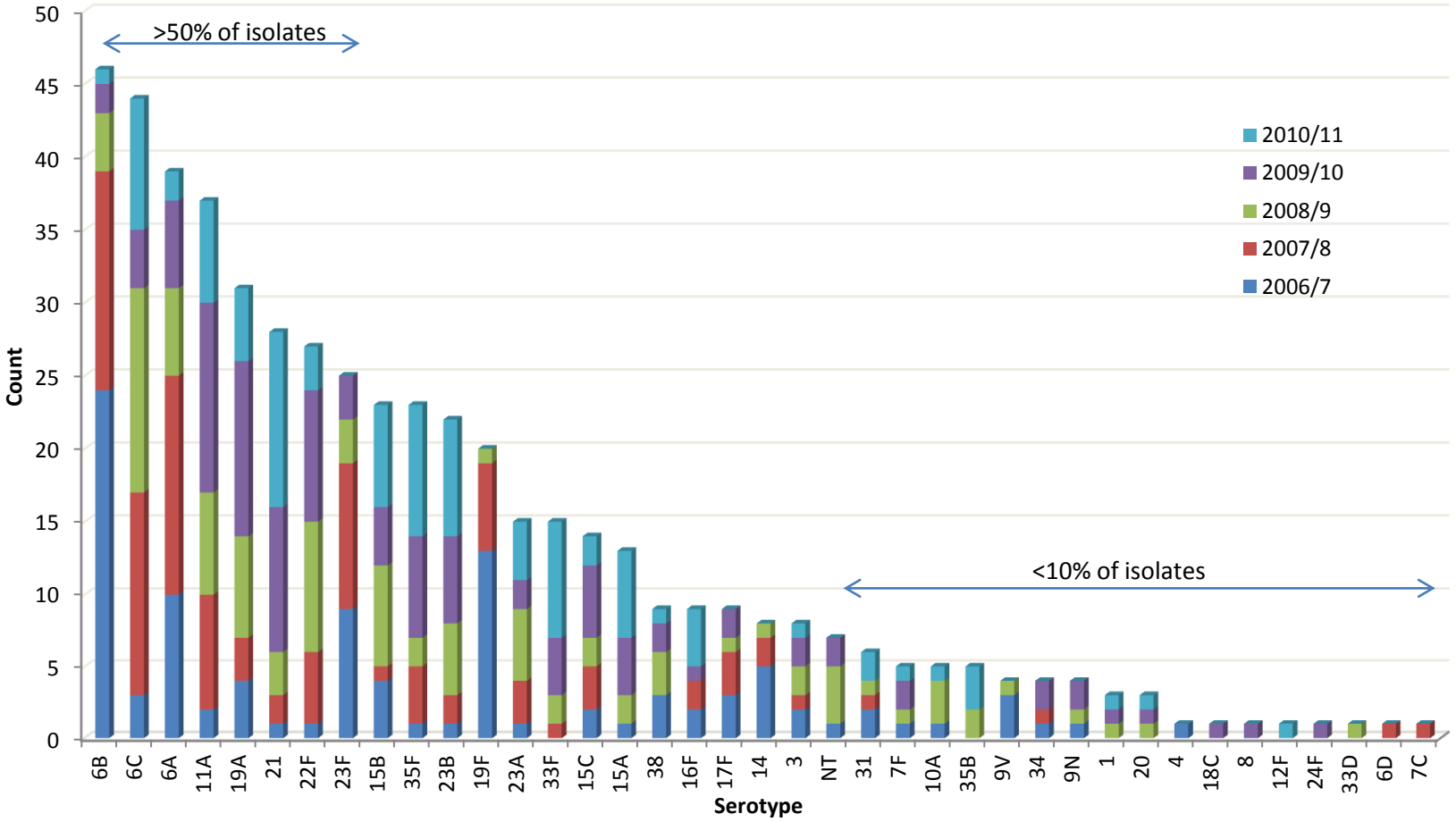


Figure 16. Incidence of serotypes for the study period in rank order

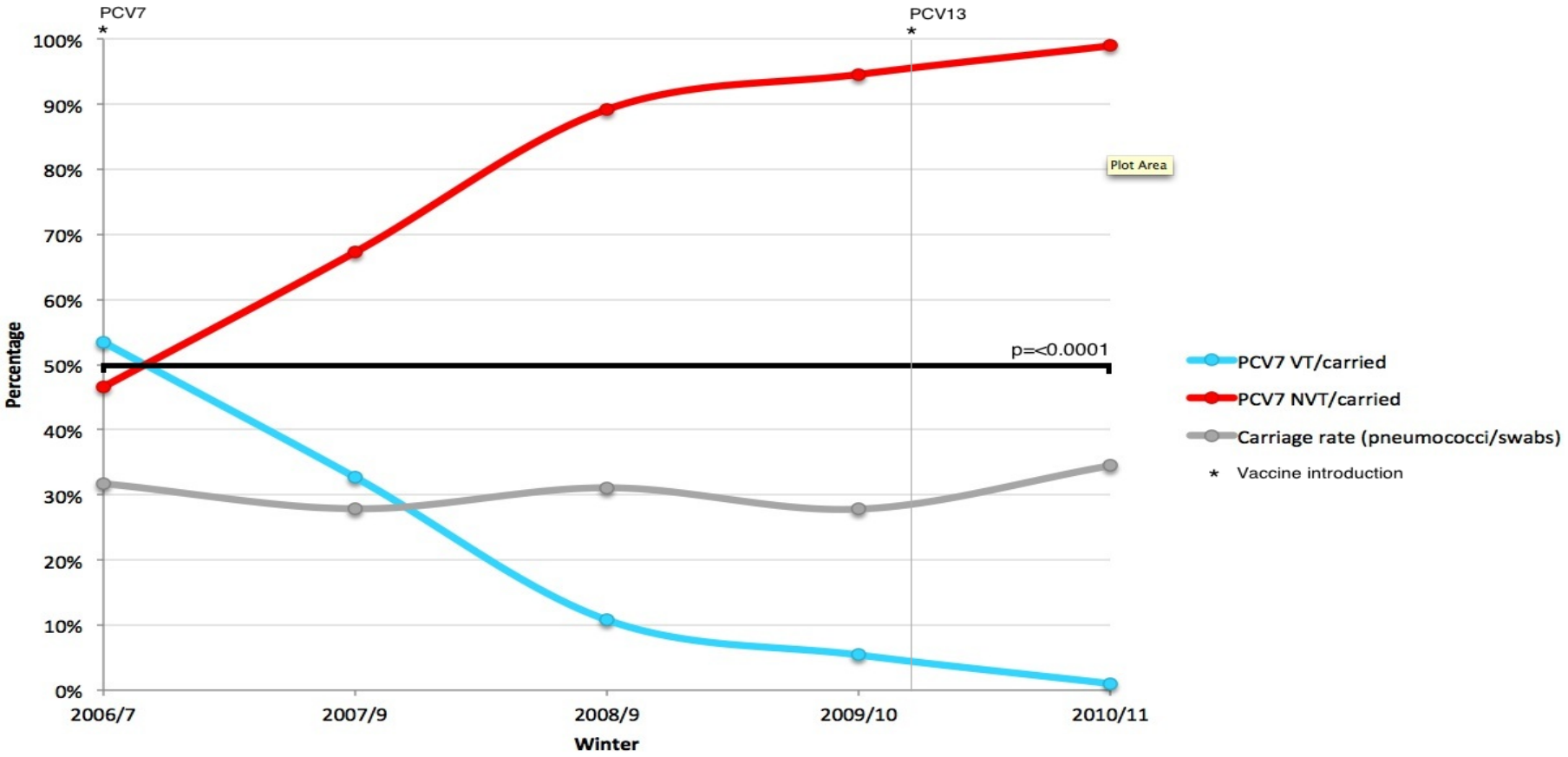


4.1.3.1 PCV7 Vaccine and non-vaccine serotype prevalence

Whilst the carriage rate of pneumococci remained stable, the relative proportions of vaccine serotype to non-vaccine serotypes did not (Figure 17).

PCV7 vaccine types decreased significantly (p=<0.0001) between the baseline, year 1 (53%, CI 43.36 to 62.64%) and year 5 (1%, -0.96 to 2.96%). Conversely PCV7 non-vaccine types were observed to significantly increase (p=<0.0001) between the baseline year 1 (47%, 37.36 to 56.64%) and year 5 (99%, 97.04 to 100.96%). Significant changes to the mutually exclusive vaccine serotype and non-vaccine serotype groups occurred between all alternate year combinations and concurrent years 1 and 2 and year 2 and 3 (Table 13,Table 14). At the year five time-point, near complete replacement of PCV7 vaccine types with non-vaccine types had occurred with only a single isolate of 6B observed (Figure 17).

Figure 17. Carriage rate of total pneumococci, PCV7 vaccine types and non-vaccine types



NB a smoothed line is used to connect the changes between data points and only gives theoretical, over simplified, shape to the graph, refer to data points to determine values.

Table 13. P values, all against all years for PCV7 and PCV13 vaccine serotype, non-vaccine serotype prevalence.

			PCV7		
Year	1	2	3	4	5
1		0.0032*	<0.0001*	<0.0001*	<0.0001*
2	0.0161*		0.0002*	<0.0001*	<0.0001*
3	<0.0001*	0.0006*		0.2058	0.005*
4	<0.0001*	0.0002*	0.8775		0.123
5	<0.0001*	0.0001*	0.0041*	0.0078*	
			PCV13		

*Statistically significant, orange cells PCV7 data, blue cells PCV13 data

Table 14. Annual rates and confidence intervals for PCV7 vaccine and non-vaccine serotypes

Year	PCV7 VT n(%)	CI	PCV7 NVT n(%)	95% CI
1	55 (53%)	43.36 to 62.64%	48 (47%)	37.36 to 56.64%
2	34 (33%)	23.96 to 42.04%	70 (67%)	57.96 to 76.04%
3	11 (11%)	4.93 to 17.07%	91 (89%)	82.93 to 95.07%
4	6 (5%)	0.95 to 9.05%	105 (95%)	90.95 to 99.05%
5	1 (1%)	-0.96 to 2.96%	98 (99%)	97.04 to 100.96%

P-values calculated with respect to the year 1 baseline data.

4.1.3.2 PCV13 vaccine and non-vaccine serotype prevalence

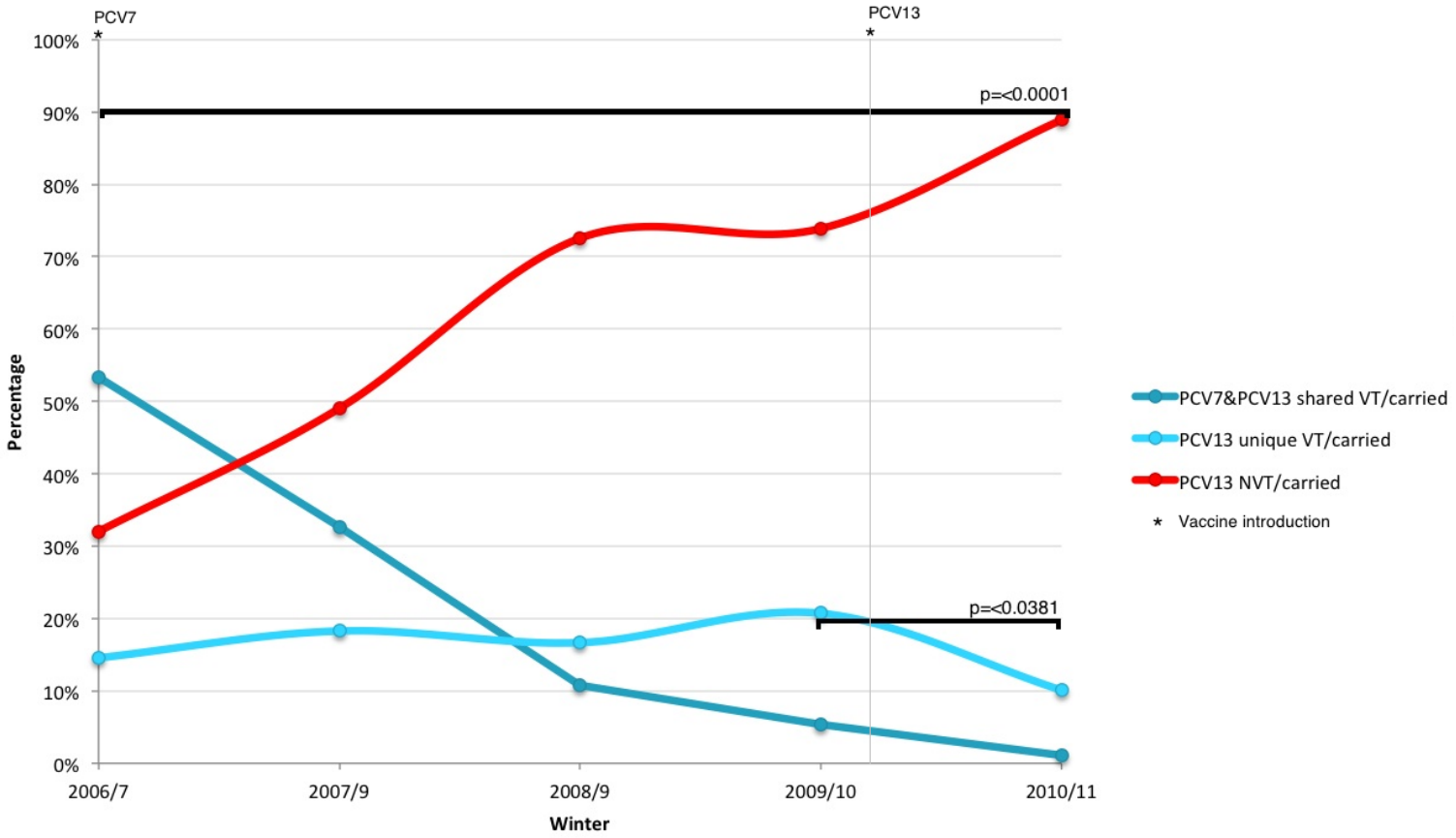
A statistically significant change ($p<0.0001$) in the relative proportion of PCV13 vaccine to non-vaccine types was observed between year 1 and year 5 (Table 15). Additionally a significant change to the relative proportions of vaccine serotype to non-vaccine serotypes occurred between concurrent years 4 and 5 ($p<0.01$) not observed when analysed in the context of PCV7 serotype (Table 15).

As PCV13 shares 7 serotypes with PCV7 it was necessary to separately analyse the 6 serotypes unique to PCV13 (Table 2) to distinguish the potential impact of PCV7 and PCV13 on carriage (Table 2). A significant decrease ($p<0.05$) in vaccine serotypes unique to PCV13 occurred only between years 4 and 5, which relates with PCV13 implementation between these two time points (Table 15,Table 16)

4.1.3.3 Vaccine serotype coverage

By year 5 the PCV7 coverage of the serotypes in carriage was only 1% compared to the base line year coverage of 53%. The baseline year coverage of PCV13 before its implementation was 68% (95% CI, 58.99 to 77%), which dropped to 26% (95% CI, 17.84-35.16%) due to the shared PCV7 serotypes during PCV7 introduction. In the year 5 winter (2010/11), immediately following PCV13 introduction in April of 2010, the coverage of PCV13 dropped further to only 11% (95% CI, 4.84 to 17.16%).

Figure 18. Carriage of shared PCV7 and PCV13 vaccine types, PCV13 unique vaccine types and PCV13 non-vaccine types



NB a smoothed line is used to connect the changes between data points and only gives theoretical, over simplified, shape to the graph, refer to data points to determine values.

Table 15. P values, all against all years for unique PCV13 vaccine serotypes versus all other serotypes

Year	2	3	4	5
1	0.5742	0.7033	0.2835	0.3974
2		0.855	0.7316	0.111
3			0.4862	0.2157
4				0.0381*

Table 16. Annual rates and confidence intervals for the six vaccine serotypes unique to PCV13

Year	Unique PCV13 VT n(%)	95% CI
1	15 (15%)	8.1-21.9%
2	19 (18%)	10.62-25.38%
3	17 (17%)	9.71-24.29%
4	23 (21%)	13.42-28.58%
5	10 (10%)	4.09-15.91%

4.1.3.4 Specific vaccine serotype fluctuations

Statistically significant decreases in serotype prevalence between the baseline year 1 and the final study year 5 were observed for PCV7 serotypes 6B (p=<0.0001), 19F (p=0.0002), 23F (p=0.0033) and PCV7 VRT/PCV13 VT 6A (p=0.0335). Weak evidence for an increase (p=0.0692) was observed for PCV13 VT 19A between year 1 (0.15% to 7.61%) and year 4 (5.03% to 16.59%) before the introduction of PCV13 to the routine childhood immunisation schedule. A decrease in PCV7 serotype 14 between years 1 and 5 was approaching significance (p=0.0597), however prevalence was low (year 1 n=5, year 5 n=0). As the 15B/C isolates originally misidentified as serotype 14 (n=9) were only from years 2 and 3, correction of this data did not therefore affect this analysis. PCV13 vaccine serotype 5 was not observed in the data set. All fluctuations observed for the remaining six PCV serotypes 1, 3, 4, 7F, 9V and 18C were not statistically significant (p>0.4).

4.1.3.5 Specific non-vaccine serotype fluctuations

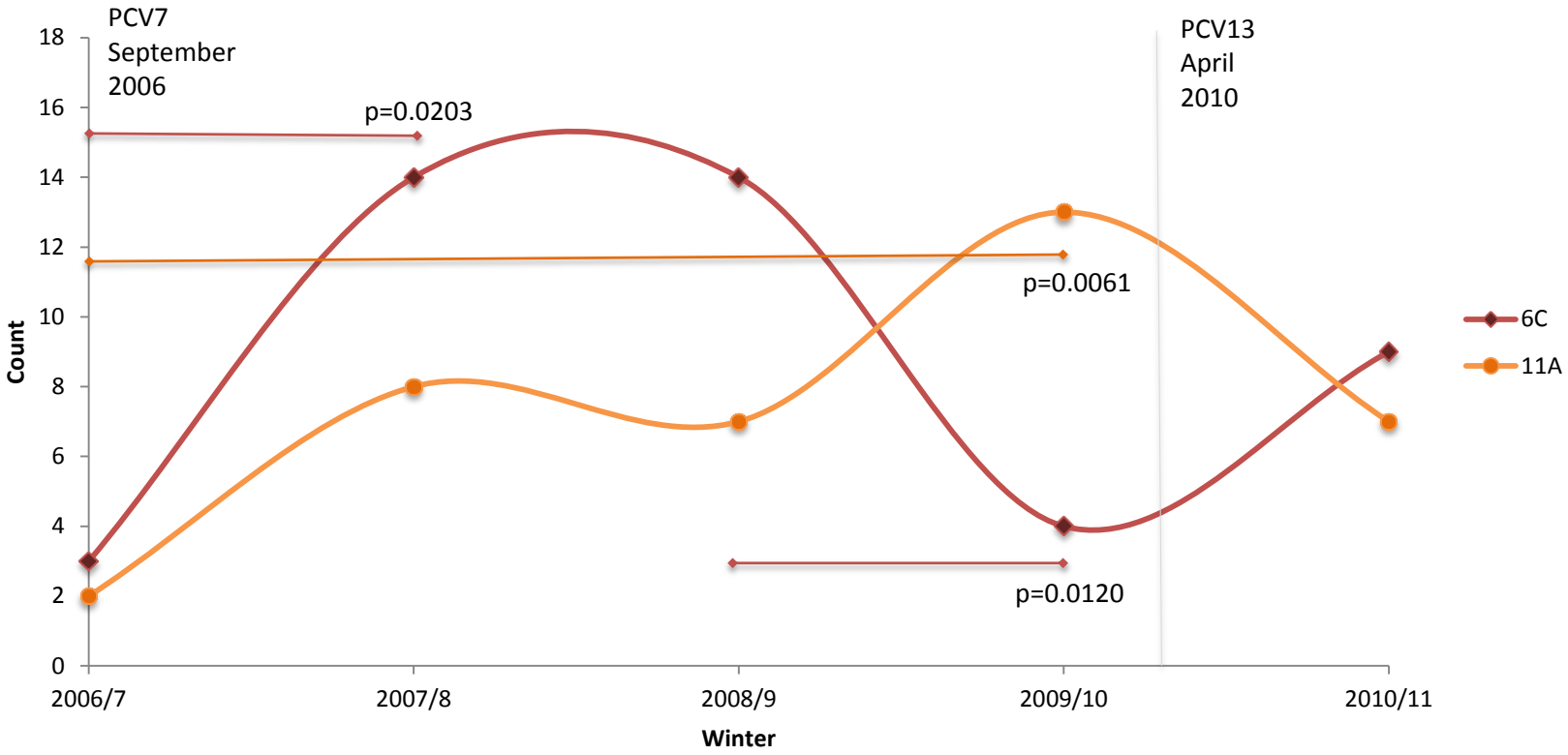
Twenty-seven serotypes were observed during the course of the study that are not included in PCV13. Of the 27 non-vaccine serotypes significant increases between year 1 and year 5 were observed for serotypes 21, 23B, 33F and 35F (Table 17). Weak evidence for increases was observed for serotype 15A and 6C (p=0.0601, p=0.0778) between year 1 and year 5. Serotypes 11A and 6C did however have statistically significant fluctuations within the study period not detectable between years 1 and 5 (Figure 19). The PCV vaccine related serotype 6C was previously reported to significantly increase in prevalence during the introduction of PCV7 between years 1 and 2 and years 1 and 3, primarily through clonal expansion (Tocheva *et al.* 2010; Loman NJ *et al.* 2013), however a significant decrease (p=0.012) was also observed between years 3 (7.05% to 20.41%) and 4 (0.13% to 7.07%) (Figure 19).

Table 17. Serotypes for which significant changes were observed between year 1 and 5

Vaccine	Serotype	Year 1 n (%)	CI	Year 5 n (%)	CI	P-value
PCV7/13	19F	13 (12.62%)	6.21% to 19.03%	0 (0%)	-	0.0002
PCV7/13	23F	9 (8.74%)	3.29% to 14.19%	0 (0%)	-	0.0033
PCV7/13	6B	24 (23.30%)	15.14% to 31.46%	1 (1.01%)	-0.96% to 2.96%	<0.0001
PCV13	6A	10 (9.71%)	3.99% to 15.43%	2 (2.02%)	-0.76% to 4.76%	0.0335
NVT	6C	3 (2.91%)	-0.34% to 6.16%	9 (9.09%)	3.43% to 14.75%	#0.0778
NVT	15A	1 (0.97%)	-0.92% to 2.86%	6 (6.06%)	1.36% to 10.76%	#0.0601
NVT	21	1 (0.97%)	-0.92% to 2.86%	12 (12.12%)	5.69% to 18.55%	0.0011
NVT	23B	1 (0.97%)	-0.92% to 2.86%	8 (8.08%)	2.71% to 13.45%	0.0170
NVT	33F	0 (0%)	-	8 (8.08%)	2.71% to 13.45%	0.0029
NVT	35B	0 (0%)	-	3 (3.03%)	-0.35% to 6.41%	0.0268
NVT	35F	1 (0.97%)	-0.92% to 2.86%	9 (9.09%)	3.43% to 14.75%	0.0087
		Year 1		**Year 4**		
NVT	11A	2 (1.94%)	-0.72% to 4.6%	13 (11.7%)	-5.77 to 29.17%	0.0061
		Year 1		**Year 2**		
NVT	6C	3 (2.91%)	-0.34% to 6.16%	14 (13.46%)	6.9% to 20.02%	0.0203
		Year 3		**Year 4**		
NVT	6C	14 (13.73%)	7.05% to 20.41%	4 (3.60%)	0.13% to 7.07%	0.0120

Above significance threshold, weak evidence

Figure 19. Temporal fluctuation of serotypes 6C and 11A



NB a smoothed line is used to connect the changes between data points and only gives theoretical, over simplified, shape to the graph, refer to data points to determine values.

4.1.3.6 Invasive disease potential

The invasive disease potential of a serotype relates to how often it is found to cause IPD and how often it occurs in carriage. Odds ratios, i.e. the odds of the serotype being isolated from IPD relative to the odds of the same serotype being isolated from carriage, can be used to estimate invasive disease potential. An OR >1 indicates increased invasive disease potential, and an OR <1 indicates decreased invasive disease potential (Table 18). This can be interpreted as the likelihood of a serotype occurring in IPD relative to carriage, for example, an OR of eight would suggest that the serotype is eight times more likely to occur in IPD than carriage. Odds of isolation from carriage identified in this study were calculated relative to odds of isolation from IPD as calculated from a study carried out in England and Wales by Miller *et al* using adjusted data from children <5 years of age in 2008-2010 (Miller *et al.* 2011b). The serotypes that were most prevalent in year 5 (6C, 21, 23B, 33F, 35F) had statistically significant ORs of <1 except for serotype 33F (n=8, 8%, 2.71% to 13.45%), which had a non-significant OR of 1.25. The non-vaccine serotype 8 was not common in this carriage dataset (n=1) and had statistically significant OR of 15.15. Vaccine types 3, 7F and 19A had statistically significant ORs of >1 and were observed in carriage in year 5 (n=1, n=1, n=5 respectively). Of the 24 serotypes observed in year 5, four were observed to have a statistically significant OR of >1 (1, 3, 7F and 19A) all four are targeted by PCV13. Four of the top five serotypes observed in year five had statistically significant ORs of <1.

A comparison of the statistically significant ORs reported from an Oxford, UK study (Brueggemann *et al.* 2003), with the ORs calculated from this carriage data and PHE data (Miller *et al.* 2011b) was made. Statistically significant ORs were observed for five serotypes in the Oxford study, three >1 (1,18C and 4) and two <1 (14, 23F). For the carriage data set there were 13 serotypes for which statistically significant ORs could be calculated, five of which were ORs >1 and eight ORs <1 (Table 18). Ten of these serotypes were significant where the Oxford study was not, conversely significance for serotypes 4 and 14 could not be observed in this dataset unlike the Oxford dataset (Table 18).

Table 18. Odds ratios of invasive disease potential

Serotype	**Thesis Carriage isolates (n)**	**IPD (n)** Miller *et al* 2011	**OR**	**95% CI**	OR Brueggemann *et al* 2003	CI Brueggemann *et al* 2003
1	**3**	**68**	**25.43**	**7.95 to 81.36**	**9.6**	**1.1 to 86.5**
7F	**5**	**104**	**25.22**	**10.18 to 62.43**	4.7	0.4 to 52.6
8	**1**	**15**	**15.15**	**1.99 to 115.09**	0.9	0.2 to 4.9
12F	1	8	7.97	0.99 to 63.95	0.0	CNC
18C	1	7	6.96	0.85 to 56.77	**5.8**	**2.6 to 13.2**
3	**8**	**30**	**3.85**	**1.75 to 8.48**	0.3	0.1 to 1.1
19A	**31**	**100**	**3.68**	**2.41 to 5.62**	1.1	0.4 to 3.1
24F	1	3	2.96	0.31 to 28.55	0	CNC
4	1	3	2.96	0.31 to 28.55	**12.1**	**1.4 to 104.2**
33F*	15	19	1.25	0.63 to 2.5	0.5	0.1 to 4
10A	5	6	1.18	0.36 to 3.9	0	CNC
22F	27	30	1.10	0.64 to 1.87	0	CNC
15C	14	13	0.91	0.42 to 1.96	0.1 (B/C)	0 to 1.1
9N	4	3	0.74	0.16 to 3.3	0.4	0 to 3.2
19F	20	15	0.73	0.37 to 1.44	0.6	0.3 to 1.1
20	3	2	0.65	0.11 to 3.93	0	N/A
38	9	6	0.65	0.23 to 1.84	0.6	0.1 to 5.3
35B	5	3	0.59	0.14 to 2.47	0	CNC
15B	23	13	0.54	0.27 to 1.09	0.1 (B/C)	0 to 1.1
9V	4	2	0.49	0.09 to 2.68	1.5	0.6 to 4
15A	13	5	0.37	0.13 to 1.05	0	CNC
14	8	3	0.37	0.1 to 1.38	**8.8**	**5.1 to 15.4**
16F	9	3	0.32	0.09 to 1.2	0	CNC
17F	9	3	0.32	0.09 to 1.2	0	CNC
23B*	**22**	**7**	**0.30**	**0.13 to 0.72**	0	CNC
23A	**15**	**4**	**0.26**	**0.08 to 0.78**	0	CNC
11A	**37**	**8**	**0.20**	**0.09 to 0.43**	0	CNC
21*	**28**	**6**	**0.20**	**0.08 to 0.49**	0.2	0 to 1.6
6A/C*	**83**	**19**	**0.20**	**0.12 to 0.33**	0.7 (6A)	0.3 to 1.7
6B/D	**47**	**8**	**0.15**	**0.07 to 0.33**	0.6 (6B)	0.3 to 1
23F	**25**	**4**	**0.15**	**0.05 to 0.44**	**0.4**	**0.2 to 0.8**
35F*	**23**	**3**	**0.12**	**0.04 to 0.41**	0	CNC

Serotype	Thesis carriage isolates (n)	IPD (n) Miller *et al* 2011	**Difference in proportions**	**95% CI of difference**	OR Brueggemann *et al* 2003	95% CI of difference
5	0	5	0.01	0 to 0.02	DNC	DNC
7C	1	0	0.00	-0.01 to 0.01	DNC	DNC
33D	1	0	0.00	-0.01 to 0.01	DNC	DNC
NT	7	0	-0.1	-0.03 to 0	0	DNC
31	6	0	-0.1	0.02 to 0	0	DNC
34	4	0	-0.1	0.02 to 0	DNC	DNC

PCV7/13	PCV13 only	NVT			CNC= Could not calculate	DNC= Did not calculate

Adjusted IPD data for children >5 years old between 2008-10 (Miller *et al.* 2011b). Historical OR (Brueggemann *et al.* 2003). * Top 5 most common serotypes in year 5 carriage data set. **Bold type-** significant difference (p=<0.05) in likelihood, between occurrence in carriage and disease.

4.1.4 Multi locus sequence types

One hundred and twenty seven distinct ST's were observed during the study. Annually this ranged from 43 to 51 distinct ST's, with 78 ST's (61% of all observed STs) observed at only one time point (Figure 20). Any single year was a minority representation (33.86% to 40.15%) of the STs seen in the entire study period. Of the 78 STs observed at a single time point, 70 of these were observed as a consequence of a single isolate, whilst the additional eight were a consequence of pairs of isolates within a single year. Eight ST's were observed at all time points.

When clonal complexes (CC) of SLVs (identified by this dataset using eBURST (Feil *et al.* 2004) Methods 2.3.2) were accounted for there were only 78 ST units comprising of 56 single STs and the 22 clonal complexes. Of the twenty-two clonal complexes noted within the dataset half of these were pairs of STs differing by a single allele. When clonal complexes were taken into account 39 of the 78 ST units (49% of all observed ST units) were observed at a single time point. Additionally 9 of the 78 ST units were observed at all time points.

Statistically significant changes in prevalence of STs between year 1 and year 5 were observed for 7 individual ST's, with decreases for ST's 138, 162, 176 and increases for ST's 100, 432, 439 and 1635 (Table 19). There was weak evidence for a decrease of ST 65 ($p=0.0653$) between year 1 and 5, however a significant change ($p=0.0026$) could be observed between year 2 (11.54%, 5.4% to 17.68%) and year 5 (1.01%, -0.96% to 2.98%). All other fluctuations observed between year 1 and 5 for the remaining 119 sequence types did not reach statistical significance ($p>0.1$), only 15 STs were shared between the two years.

When clonal complexes were considered the number of shared ST units increased to 17, between year 1 and year 5, these 17 units encompassed 54 different STs. All significant changes for STs remained significant when clonal complexes were taken into account except ST 439, as part of CC42, for which there was no longer a significant difference ($p=0.1603$).

Figure 20. Distinct sequence types (ST's) observed by time point

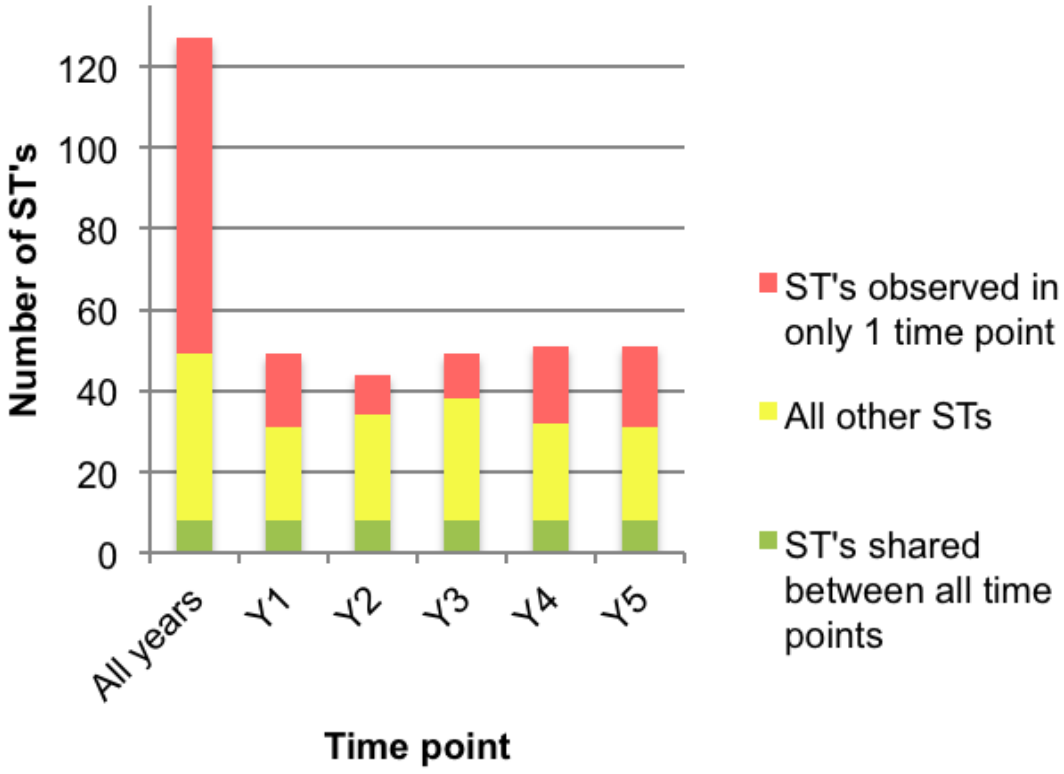


Table 19. Significant changes to individual ST prevalence between year 1 and year 5

ST	Year 1 n(%)	CI	Year 5 (N,%)	CI	p-value	Change
65	7(6.8%)	1.93% to 11.65%	1(1.0%)	-0.96% to 2.98%	#0.0653	Decrease
100	0(0%)	-	5(5.0%)	0.71 to 9.29%	0.0268	Increase
138	6(5.8%)	1.29 to 10.31%	0(0%)	-	0.0291	Decrease
162	9(8.7%)	3.26 to 14.14%	1(1.0%)	-0.96 to 2.96%	0.0187	Decrease
176	10(9.7%)	3.98 to 15.42%	1(1.0%)	-0.96 to 2.96%	0.0098	Decrease
432	0(0%)	-	6(6.0%)	1.32 to 10.68%	0.0128	Increase
439	1(0.97%)	-0.92 to 2.86%	7(7.0%)	1.97 to 12.03%	0.0325	Increase
1635	0(0%)	-	6(6.0%)	1.32 to 10.68%	0.0128	Increase

Not significant, weak evidence for change

The Simpson's diversity index (1-D) for MLST diversity was 0.97 for the total study period, reduced only to 0.96 when accounting for clonal complexes. Each year the index remained stable with a range of 0.96 (Year 2) to 0.98 (Year 5), and describes a highly diverse population throughout the study period with greater diversity described by ST than serotype.

4.1.5 MLST and serotype associations

A highly significant statistical bi-directional association between ST and serotype, was observed for the 519 isolate data set with a Fisher's exact p-value of <0.0001. This observation was replicated with data from years 1 and 5 individually with all Fisher's exact p-values <0.0001. However a Chi-Squared p-value of <0.0001 and <0.001 was calculated using the entire data set for the association of serotypes with STs, and the associated of ST with serotypes respectively. This highlights the greater diversity of the serotype in terms of ST than the ST in terms of serotype.

To explore this further the Simpson's diversity index (1-D) was calculated for all serotypes with >15 isolates (n=13) and all STs with >10 isolates (n=13), (Table 20). Whilst STs are highly associated with a single or very limited number of serotypes, serotypes are associated with STs in a more variable manner with as many as 10 STs associated with serotype 6C (n=44) and only one for serotype 3 (n=8). It could be observed that the average Simpson's diversity index (1-D) for serotype and STs was higher for the serotypes than for the STs, (0.53, 0.11 respectively). Eight STs were entirely clonal in terms of serotype; serotype 11A had the least diversity observed for the serotypes, with only two associated ST's among the 37 isolates giving an index of 0.05, a highly conserved population. Conversely serotype 19F had the highest genetic diversity with seven different ST's within 20 isolates and ST 199 had an index of 0.62.

As Simpson's index of diversity does not take into account the relationship between ST's or the relationship within a serogroup the analysis was repeated after the clonal complex grouping of ST's within a serotype and serogrouping of serotypes within an ST was performed in order to better represent the diversity. When clonal complexes and serogroups were taken into account the Simpson's diversity index (1-D) decreased in both cases to 0.36 and 0.08 respectively. Whilst collapsing the CC and serogroups had little effect on 19F and 11A and STs 199 and 162 this analysis highlights the clonality of serotypes 35F and 23F and ST 62 for which the diversity index decreased to 0 when CC or serogroup were taken into account (Table 20). Additionally this analysis shows that the exact level of diversity is variable and specific to the ST or serotype in

question. When observing the ST's for which significant changes were observed between year 1 and 5 the majority (6/7) were associated with a single serotype (Table 21). For the 57 STs represented in the dataset by more than one isolate only 9 of these were associated with more than one serotype and only three STs with the maximum of three serotypes. When serogroups were taken into account, only 7 STs were associated with more than one serogroup and not a single ST was associated with >2 serogroups.

When observing dominant STs for which there was a significant change in prevalence between year 1 and 5, of the associated serotype, all fluctuations in the STs mirrored the serotypes in terms of direction (increase/decrease). Furthermore for the 9 serotypes with significant changes between years 1 and 5 there was also a significant change observed in the dominant ST for 7 of the serotypes (Table 22) with weak evidence for ST65 the dominant ST for serotype 6A (p=0.0653).

Table 20. Simpson's Diversity index (1-D) for serotypes and STs

Serotype	Isolates (n)	ST's (n)	Simpson's 1-D	CC (n)	Simpson's 1-D CC
6B	46	9	0.71	5	0.47
6C	44	10	0.65	5	0.33
11A	37	2	0.08*	1	0.05
6A	39	7	0.41	6	0.37
19A	31	8	0.45	8	0.45
21	28	6	0.72	3	0.64
22F	27	5	0.67	2	0.48
15B	23	5	0.58	2	0.40
35F	23	4	0.44	1	0.00*
23B	22	4	0.26	3	0.18
19F	20	7	0.79**	6	0.74**
23A	15	5	0.56	1	0.00*
33F	15	4	0.60	2	0.51
ST	**Isolates (n)**	**Serotype's (n)**	**Simpson's 1-D**	**Serogroups (n)**	**Simpson's 1-D Serogroup**
199	45	3	0.62**	2	0.51**
62	35	2	0.16	1	0.00*
65	30	1	0.00*	1	0.00*
1692	25	1	0.00*	1	0.00*
176	22	1	0.00*	1	0.00*
439	19	1	0.00*	1	0.00*
1635	17	1	0.00*	1	0.00*
36	15	1	0.00*	1	0.00*
193	13	2	0.15	2	0.15
433	13	1	0.00*	1	0.00*
162	12	3	0.53	2	0.41
138	11	1	0.00*	1	0.00*
438	10	1	0.00*	1	0.00*

*Lowest diversity **highest diversity

Table 21. STs for which significant changes were observed and their associated serotypes

ST	Count	Associated serotype (n,%)
65	30	6A (**30**,100%)
100	9	33F (**9**,100%)
138	11	6B (**11**,100%)
162	12	19F (**8**,67%) 9V (**3**,25%), 19A (**1**,8%)
176	22	6B (**22**,100%)
432	8	21 (**8**,100%)
439	19	23B (**19**,100%)
1635	17	35F (**17**,100%)

Table 22. Significant changes for both serotype and primary associated ST

ST	Change	P-value	Associated Serotype	Vaccine status	Change	P-value
36	Decrease	#0.1216	23F	PCV7/13	Decrease	0.0033
65	Decrease	# 0.0653	6A	PCV13	Decrease	0.0335
100	Increase	0.0268	33F	NVT	Increase	0.0029
138	Decrease	0.0291	6B	PCV7/13	Decrease	<0.0001
162	Decrease	0.0187	19F	PCV7/13	Decrease	0.0002
176	Decrease	0.0098	6B	PCV7/13	Decrease	<0.0001
432	Increase	0.0128	21	NVT	Increase	0.0011
439	Increase	0.0325	23B	NVT	Increase	0.017
1635	Increase	0.0128	35F	NVT	Increase	0.0087

Non statistically significant change, weak evidence

4.1.5.1 Capsular switches

The presence of multiple serotypes belonging to a single ST or *vice versa* is an indication that a capsular switch event is likely to have occurred at some point during pneumococcal evolution. If it is assumed for simplicity and to prevent false positive indications that, any capsular switch events involving a specific serotype and a specific genetic lineage did not occur multiple times then the presence of multiple STs belonging to a clonal complex associated with that serotype can be explained by evolution of the MLST through mutation. If it is also assumed that the presence of multiple serotypes belonging to a single serogroup sharing an ST is due to evolution of the capsule after the establishment of the ST then a screen for the most likely capsular switch events can be made that involves independent genetic lineages and independent capsules. NT isolates were excluded from this analysis. There were 10 indications of capsular switch events observed where multiple serogroups were associated with a single ST unit and a further 57 where multiple ST units were observed for a single serogroup (Table 23). Serogroups 6 and 19 were observed with the largest number of ST units, 15 and 13 respectively. Clonal complexes CC33, CC172 and CC193 were seen in association with largest number of serogroups, with 3 different serogroups each. The direction, donor or recipient, or timing of these events cannot be inferred from MLST and serotype data alone.

Table 23. Indications of capsular switch events

	Serotype/group																		
ST units	1	6	7	9	10A	11	14	15	16F	19	20	21	22F	23	31	33	34	35	Count
9, 15							*												1
30, 414	*								*										2
CC42		*								*				*					3
36														*					1
58								*											1
62, 10002						*													1
63								*											1
65, 97, 460		*			*														2
66				*															1
73								*											1
81														*					1
100																*			1
124							*												1
138		*																	1
162				*						*									2
172, 1150, 1373		*								*				*					3
CC176		*																	1
177, 179										*									1
191			*																1
CC193								*		*		*							3
198																		*	1
CC199								*		*									2
227	*																		1
235											*								1
242		*																	1
276										*									1

Table 23. Indications of capsular switch events continued

	Serotype/group																		
ST units	1	6	7	9	10A	11	14	15	16F	19	20	21	22F	23	31	33	34	35	Count
306	*																		1
309										*									1
CC395		*																	1
398		*																	1
405				*															1
416										*									1
422										*									1
432, 6190												*							1
433, 819, 4325													*						1
444															*				1
CC1635																		*	1
452																		*	1
461, 861					*														1
482										*									1
568															*				1
673, 1012, 3916																*			1
698, 3734													*						1
1201										*									1
1262, 8711								*											1
1390		*																	1
1497					*														1
1600, 4581, 8028		*																	1
1766															*				1
1797			*																1

Table 23. Indications of capsular switch events continued

	Serotype/group																		
ST units	1	6	7	9	10A	11	14	15	16F	19	20	21	22F	23	31	33	34	35	Count
1862		*																	1
1876		*																	1
1877												*							1
1884																	*		1
2042									*										1
2457		*																	1
2631														*					1
3240		*																	1
3460		*																	1
4019						*													1
4423				*															1
4424				*															1
4696										*									1
4697																*			1
6187								*											1
6188											*								1
8027																	*		1
Count	3	15	2	5	3	2	2	7	2	13	2	3	2	3	3	3	2	3	

* Represents observation of the specific combination of ST and serotype. Count of specific observations does not relate to the number of isolates involved.

4.1.5.2 eBURST analysis of MLSTs

The ST's observed in this study represented the preponderance of the major clonal complexes observed in the MLST database (MLST.net 2012) with the exceptions of CC276 and CC320 (Figure 21) which are highly associated with serotypes 19A/F submitted predominantly from mainland European locations, in this study 19A and 19F were most commonly found to be ST199 and ST162/179 respectively. The carriage study STs are spread throughout the database diagram and the novel STs observed (green halos,Figure 21) are within established CCs also represented elsewhere in the carriage dataset; CCs 62, 199, 1692 associated with serogroups 11, 19 and 6 which corresponded to the serotypes of the isolates with the novel alleles. The major clonal complexes present in the carriage dataset and the percentage contribution of the CCs to the complete data set can be observed in the eBURST representation (Figure 22). The variable number of associated serotypes of the CCs demonstrates that the degree to which clonality in terms of serotype is unique to the clonal complex (Figure 22). A comparison of the STs and CCs unique to and shared between the first and last year of the study largely recreates the representation of the whole carriage dataset, the major clonal complexes exist in a combination of shared STs and those unique to both years (Figure 23). Thirteen clonal links of the 22 CCs can be observed in the comparison of these two years alone indicative of the existence of stable clonal lineages throughout the study.

Figure 21. eBURST diagram of carriage MLST dataset as a query of the entire *S. pneumoniae* MLST database as of 06/2013

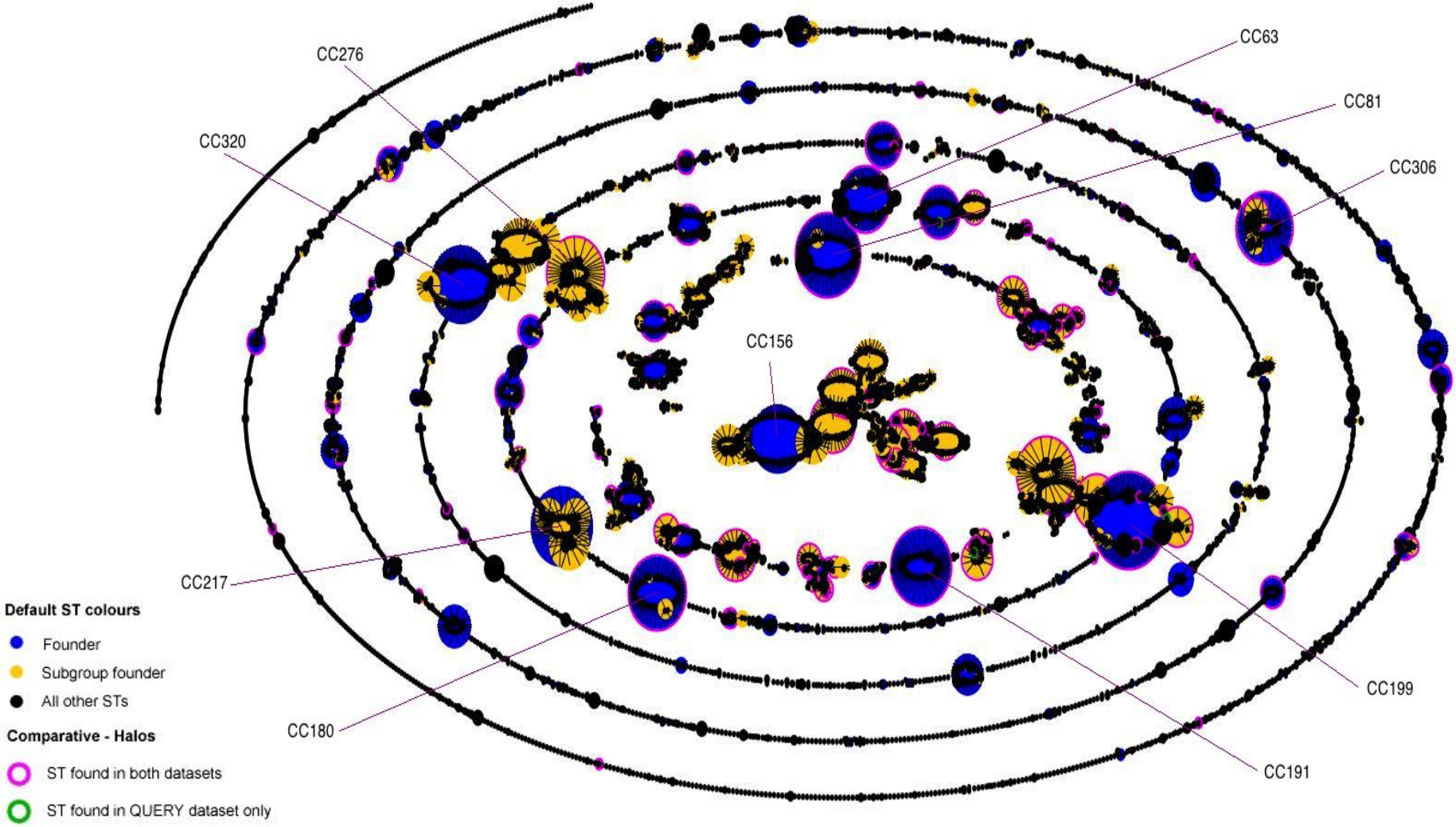


Figure 22. eBURST diagram of carriage MLST dataset

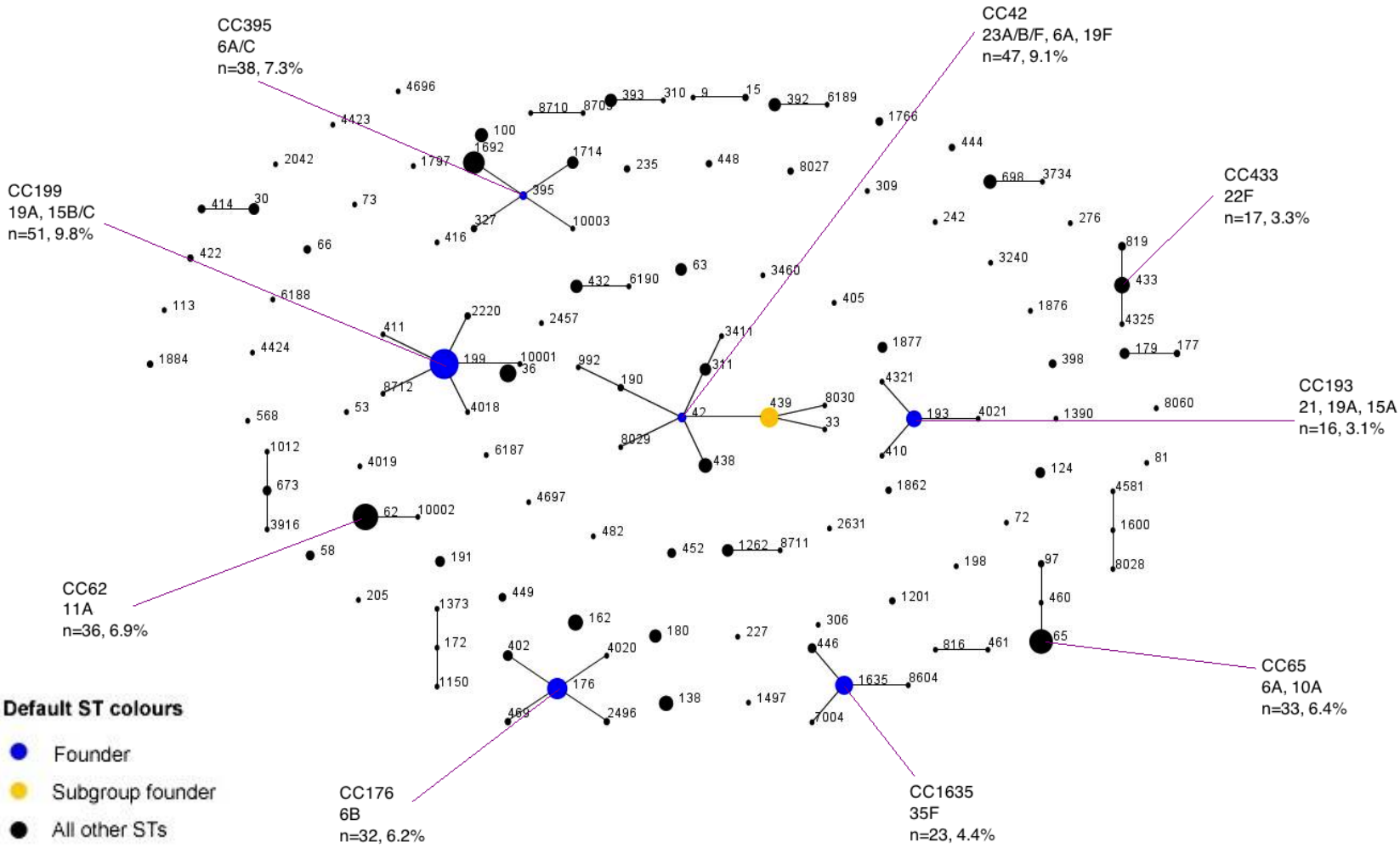
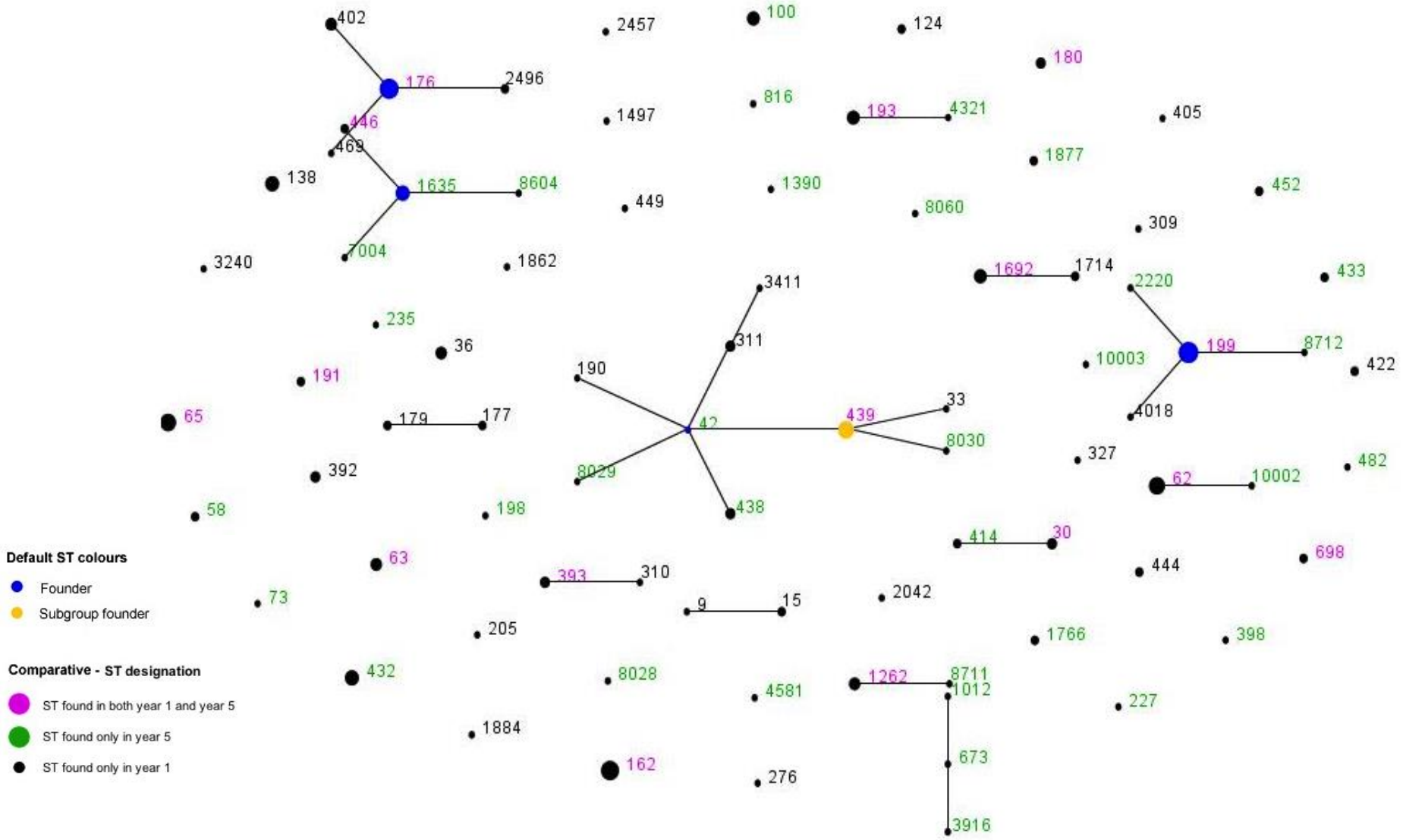


Figure 23. eBURST diagram of carriage MLST datasets year 1 and year 5 comparison



4.1.6 Discussion of results

It was interesting to note that a smaller percentage of carriers than non-carriers respectively, were under 2 years old although there were no statistically significant differences between the two groups for age detected. Carriage in the pre PCV era has been documented to be highest in the under twos not observed in this study (Hussain *et al.* 2005) but a Spanish study also observed carriage to be lower in this age group in the PCV era (Hernandez-Bou *et al.* 2012). Upper respiratory tract infections (URTI) identified as being associated with pneumococcal carriage in this thesis have previously been identified as a risk factor for children and association with influenza like illness in adults (Ansaldi *et al.* 2012b; Hsieh *et al.* 2012). The interactions between colonising bacteria and their relationship with URTI is a not yet well understood and interesting area of investigation outside of the remit of this thesis (Pettigrew *et al.* 2008; van den Bergh *et al.* 2012; Bosch *et al.* 2013; Ruohola *et al.* 2013).

Although the pneumococcal population remained stable over the introduction of PCVs, in terms of the carriage rate and the diversity as a function of the number of serotypes and genotypes observed year by year, the exact make-up of the population changed dramatically over time. The replacement of vaccine types with non-vaccine types and the consequent genotypic replacement of VT associated STs with STs associated with NVTs contributed to this stable state during PCV introduction. The presence of the major clonal complexes STs in both years 1 and 5 suggests that all though the exact genotypes varied over the course of the study there were still a continuation of specific genetic lineages. The significant changes in STs associated with specific VT or NVT illustrates further the strong relationship between serotype and genotype. This replacement with genotypes already existing in the pre PCV era was also observed in the US, which also suggested that the capsular switch variants also existed in the population (Croucher *et al.* 2013). Evidence for serotype switch events was common in this data set although they were not investigated at the genome level for the homologous exchange of the capsular loci, work by others suggests that capsular switch events occur regularly in the pneumococcal population but any that have been involved in ‘vaccine escape’ are historical events that have been selected for (Pai *et al.* 2005; Hanage *et al.* 2011; Croucher *et al.* 2012).

The 5 year data analysis mirrored and continued the overarching trends of vaccine types being replaced by non-vaccine types observed for the first three years of the study reported previously, with VT and associated genotype replacement with the sustained significant decrease for STs 162 and observed changes that then reached

statistical significance in the 5 year data set for serotypes 6A, 6B, 21, 23B and STs 176, 138 (Tocheva *et al.* 2011; Tocheva *et al.* 2013). Conversely the statistically significant change for 6C and associated ST 1692 (Tocheva *et al.* 2010) was not sustained and disappeared by year 4 before the introduction of PCV13 which, included 6A reported to stimulate cross reactive antibodies for 6C (Cooper *et al.* 2010). This observation could be explained by the hypothesis made in 1997 on the vaccination against bacteria with multiple serotypes that there may be increases in serotype with partial protection (PCV7 6B cross reactivity for 6C) followed by a decrease as a function of vaccine coverage (Lipsitch 1997).

5. DISCUSSION, CONCLUSIONS AND FUTURE WORK

This thesis hypothesised that during PCV implementation there would be a shift in pneumococcal carriage epidemiology. This was predicted to occur through vaccine serotype and associated genotypes decreases, with subsequent replacement by non-vaccine serotypes and the associated genotypes. Secondly this thesis hypothesised that data required to make the above epidemiological observations could be robustly derived from whole genome sequences (WGS). Furthermore the whole genome data could provide additional data of relevance to epidemiological temporal surveillance of circulating pneumococci, during PCV introduction and on-going delivery. The implications of the findings of this thesis in context will now be discussed.

5.1 Whole genome data

New sequencing technologies are set to revolutionise the field of molecular and clinical microbiology and many others (Chan *et al.* 2012; Didelot *et al.* 2012; Koser *et al.* 2012a; Reuter *et al.* 2013). The current era is one in which sequencing capabilities have been increasing faster than Moore's law (Moore 1965), suggested to be doubling every 6-9 months (Loman *et al.* 2012). Meanwhile the cost per isolate through increasing data yields and run throughput is reducing to the extent where it is no longer prohibitive to produce the raw data. Additionally companies provide sequencing services making whole genome data (WGD) accessible outside of dedicated sequencing centres, which have a large amount of sequencing expertise and investment in sequencing platforms for their own purposes.

There is growing interest for the integration of next generation sequencing in the clinical setting with proof of principal studies fast emerging. These include *Mycobacterium abscessus* transmission in cystic fibrosis, retrospective *Mycobacterium tuberculosis* outbreak transmission networks, MRSA outbreak investigations and genome derived antimicrobial resistance profiles (Koser *et al.* 2012b; Bryant *et al.* 2013; Harris *et al.* 2013; Koser *et al.* 2013; Walker *et al.* 2013). The utilisation of these technologies for research purposes and clinical use, needs to be fully validated and incorporated into general practice alongside the more traditional sequencing projects, to further advance science and clinical practice.

The instruments and protocols themselves are already becoming more user friendly, and could be introduced to diagnostic laboratories in a comparable manner to the introduction of PCR into routine microbiology. However there are still hurdles yet to overcome and barriers to the integration of WGS into clinical practice. One major consideration for the successful introduction of the technology is the selection of, the most appropriate applications of the technology, for practical and beneficial routine clinical use. Moreover, thought needs to be given to how the data needed will be derived and what standardised procedures there will be for doing so.

An important decision point and logistical issue is the question of who will routinely analyse clinical WGD. Currently dedicated specialist personnel are required for the bioinformatics, as there are features of the analyses that inhibit full automation at present. Such features include the complexity of the raw data generated and the sheer number of approaches and questions that could be asked of the data. Additionally the storage of the consequent data is also of paramount importance. Whilst data volume is perhaps an obvious problem, the storage of data in an accessible manner of benefit to the whole clinical community is perhaps a more important one. This would be essential for purposes such as strain comparison, outbreak surveillance and tracing, and will likely be harder to achieve than the physical storage space and require global collaborative databases. Finally the translation of data, to clinical action should require an understanding of the technology and analysis, by those involved in the clinical use of whole genome data. This would include the registrar or consultant microbiologists making the care decisions for individual patients, epidemiologists and communicable disease control consultants for public health measures. It would also require those in direct contact with patients to be confident and capable of imparting the meaning and importance of results derived from whole genome sequencing. This will likely require raising awareness, engagement, and training in readiness for the introduction, to ensure the change is managed successfully, and should not be underestimated in the importance or the size of the task. One approach could be to form dedicated specialist teams to work with the data with a high level of training, rather than full integration into routine practice, at present specialist academic and clinical partnerships are leading the way (Chan *et al.* 2012; Didelot *et al.* 2012; Koser *et al.* 2012a; Torok & Peacock 2012; Reuter *et al.* 2013).

Whole genome sequencing will allow powerful, clinically relevant and previously infeasible analysis. Examples of such research in the pneumococcal field include; the investigation into the evolution of pneumococci in the context of PCVs, and identifying and understanding pneumococcal vaccine escape strains, both providing data on the

evolutionary response to clinical intervention (Croucher *et al.* 2011; Golubchik *et al.* 2012). One additional application of whole genome sequences to clinical science will be to allow comprehensive genotyping of isolates through a single method, whilst generating the whole sequence, which can then be stored and further interrogated after the initial typing.

5.1.1 Comprehensive whole genome genotyping

Genotyping is a central part of characterising a population of microorganisms and can be used to clearly distinguish or cluster isolates and monitor their prevalence over time. MLST has been widely used for genotyping *S. pneumoniae*, both from carriage and disease, as a simple proxy for genetic groups. This allows monitoring of the prevalence of similar isolates in terms of genetic background and emergence of new groups or clonal expansion (Enright & Spratt 1998; Maiden *et al.* 1998; Hanage *et al.* 2005a; Tocheva *et al.* 2010; Pichon *et al.* 2013; Tocheva *et al.* 2013). MLST is usually determined using Sanger sequencing but can equally be determined using alternative sequencing methods. MLST is of limited resolution and will not always be able to distinguish subgroups of a clonal population whereas whole genome sequencing is of the highest possible resolution. As I have demonstrated in this work MLST and other genotypic methods can be deduced from the data with vast additional discriminatory power. The pneumococcal MLST scheme utilises seven housekeeping genes as do a number of bacterial MLST schemes. In the advent of MLST from whole genome sequences, the panel of genes used to differentiate strains need not be limited to seven and can be extended considerably. The Bacterial isolate genome sequence database (BIGSdb) is one example of extending the MLST to a gene-by-gene approach of whole genomes. It is designed to be scalable for both the number of strains in the collection and the number of loci you wish to interrogate. As with the analysis performed in this thesis BIGSdb relies upon the assembly of contiguous regions of sequence from WGS reads (Jolley & Maiden 2010; Maiden *et al.* 2013).

5.1.2 *In silico* MLST

Determination of MLST from whole genome sequences is now becoming widely accepted within the bioinformatics community. A number of publically available tools now exist including web based platforms that can be used for this purpose utilising both assembly based and mapping based approaches (Jolley & Maiden 2010; Otto *et al.* 2010; Inouye *et al.* 2012; Larsen *et al.* 2012).

Since the inception of the pneumococcal MLST database, new alleles have been assigned after quality checking of sequence data by the database curator(s). In the past this data was in the form of trace files from Sanger sequencing. It is uncertain whether a defined process for submission and quality checking of MLST alleles extracted from whole genome data is yet in place. The submission of WGS from this study to the curators of the pneumococcal MLST database for allele assignment prompted internal discussion regarding development of protocols for the curation of this type of data. Following some delay, it was suggested that submission of VCF or FastQ files trimmed to the allele region was the preferred approach. In principle these formats mimic the role of the trace file in conventional curation and would require different but similar expertise to curate.

MLST from WGS may become the norm in the future, with large genome pneumococcal projects looking to characterise population diversity determining MLST in this manner already occurring (Everett *et al.* 2012; Croucher *et al.* 2013). These studies, particularly those characterising isolates from previously underrepresented geographical locations, have the potential to identify large numbers of new allelic sequences in more divergent lineages, which would increase the need to submit WGD for curation.

5.1.3 Defining the pneumococcus

The differentiation of pneumococci, most easily classified by the expression of a capsule, from closely related streptococcal species is not as clear-cut for non-typeable isolates. Two classes of non-typeable pneumococci have been described previously, the classical non-typeable lineage clearly phylogenetically distinct from other pneumococci, and other more sporadic losses of the capsule spread through different lineages (Croucher *et al.* 2013). The latter have typical capsular pneumococcal genomes lacking predominantly only a fully functional capsular loci resulting in low-level expression not easily detectable or no capsule expression. Isolates that are phenotypically non-typeable pneumococci can be distinguished from non-pneumococcal species by detection of the remnants of pneumococcal capsular biosynthesis genes which can elucidate the serotype background before loss of function as has been described using microarray (Scott *et al.* 2012).

The profile of the classical non-typeable lineage is harder to distinguish from non-pneumococcal species using phenotypic and genotypic methods and has resulted in the mistaken disease aetiology of pneumococci rather than the non-pneumococcal

streptococci. Only three quarters of 132 atypical pneumococci from IPD in Spain actually determined to be pneumococci after extensive genotyping and phenotyping (Rolo *et al.* 2013). The pneumococcal population is highly diverse and recombinogenic not only within the species but also the Streptococcal genus (Donati *et al.* 2010; Croucher *et al.* 2012). No known single species-defining feature can be used which will not be missing in the minority or occasionally present in other species. This includes the specificity of optochin susceptibility, bile solubility and even the detection of 'pneumococcal like' pneumolysin, autolysin *lyt*A and capsular biosynthesis genes in other streptococcal species (Whatmore *et al.* 2000; Arbique *et al.* 2004; Robson *et al.* 2007; Carvalho *et al.* 2012; Carvalho Mda *et al.* 2013).

The distributed genome hypothesis (Ehrlich *et al.* 2005; Donati *et al.* 2010) suggests that pneumococci have genomic plasticity, whereby isolates have a unique subset of genes from the pneumococcal 'pan-genome' or 'supragenome' which is the full complement of genes of the species. The core genome shared between all isolates has been estimated, using clusters of orthologous genes, to represent between 1.14 and 1.5Mb of the ~1.98-2.19Mb genomes (Hiller *et al.* 2007; Donati *et al.* 2010; Croucher *et al.* 2013). It has been reported than on average 74% of the genome is core, the remainder of the genome, termed the accessory genome, is more variably distributed in strains. There are two theories for the pan-genome, the finite (closed) and unlimited (open) pan-genome, whereby horizontal gene transfer is thought to contribute new genes to an open pneumococcal pan-genome, with up to 30% of sequence diversity in genes shared with *S. mitis* estimated to be from horizontal transfer events (Donati *et al.* 2010; Croucher *et al.* 2012).

The complete genetic resolution of whole genomes provides the data and a means of characterising isolates and differentiating non-typeable pneumococci and other streptococcal species (Salter *et al.* 2012). However, as more isolates are sequenced and the pan genome further sampled, particularly rare and more distantly related divergent strains, the concept of species maybe further tested as a genetic continuum is unveiled (Arbique *et al.* 2004; Donati *et al.* 2010). Perhaps then the use of the term species would be better complimented with a classification based on functional context, with the clinical perspective of disease potential. Furthermore the ideas of the commensal, opportunistic and pathogenic bacteria could contribute to the clinically relevant grouping of streptococci if the key differing virulence determinants were further understood.

5.1.4 Sequencing the serotype

With computing power and scripts available and after the initial set up the process of deducing serotype from whole genome data was quick, easy and useful. However utilising sequence data from multiple isolates to characterise the diversity of capsule-defining genes of serotypes has not been fully exploited to date particular outside of serogroups 6 and 19 for which a number of studies exist (Morona *et al.* 1999; Bratcher *et al.* 2011; Elberse *et al.* 2011b; Song *et al.* 2011; Otsuka *et al.* 2013). There has only been one major genotypic study into all pneumococcal serotypes, which only had single or a limited number of representatives of each serotype, documenting the diversity within a serotype may allow the quick identification of sub-types for which phenotypic and biochemical investigations can then be made (Bentley *et al.* 2006; Mavroidi *et al.* 2007). Predictions have been made on the functions and specificities of the capsular loci genes but full investigation into the relationship between sequence and capsular expression has been limited to a few serotypes (Aanensen *et al.* 2007; Calix *et al.* 2011; Oliver *et al.* 2013a; Oliver *et al.* 2013b), many genes are documented as putative.

Data from this thesis confirms that there is protein sequence diversity in the capsular defining genes of all serotypes for which this analysis was performed. The genes encoded were involved in a number of key stages of capsular expression including an oligosaccharide polymerase, flippase, glycosyltransferases, O-acetyltransferases and a rhamosyltranferase. This diversity has the potential to alter the capsular structure and immunogenic properties and therefore may represent previously unrecognised serotypes that could affect vaccine effectiveness through reduced efficacy, coverage and replacement. If serotypes with differing immunological properties are not completely resolved, estimates of vaccine serotype coverage, cost effectiveness and efficacy against any current definition of a serotype not yet fully discriminated, could be an inaccurate estimation of the potential of the vaccine to prevent disease cases. High throughput sequencing offers a method to generate large quantities of data on genes of the pneumococcal capsular loci and cost no longer prohibits the sequencing and analysis of multiple strains representing a serotype. It is important to investigate the diversity and identify potential sub-strains that could affect vaccine effectiveness. It is also of importance to follow these discoveries with phenotypic confirmation of expression and biochemical analysis to determine if there are any substantial immunological differences of clinical relevance, the number of serotypes that exist however is large and therefore focus should be placed on those causing disease and targeted by current and investigational vaccines.

5.2 Carriage epidemiology

Previously in this carriage dataset the phenomenon of serotype replacement was observed for PCV7 implementation and vaccine type associated genotype replacement (Tocheva *et al.* 2011; Tocheva *et al.* 2013). The data presented in this thesis adds to this observation, revealing that further replacement occurred for PCV7 serotypes and associated lineages. Additionally replacement continued to be observed after PCV13 introduction including for the additional six PCV13 vaccine types and is also likely to continue until near complete replacement has occurred as has been observed for PCV7 in this thesis and in the US in carriage (Hanage *et al.* 2010). Similar observations have been made elsewhere in studies around the world have also reported the loss of VT in the carried pneumococcal population and rapid replacement with NVT suggesting this is not a local geographical phenomenon (Ansaldi *et al.* 2012a; Cho *et al.* 2012; Croucher *et al.* 2013; Parra *et al.* 2013).

The significant reduction in serotypes unique to PCV13 in the winter immediately following the April 2010 introduction into the UKs routine childhood immunisation schedule illustrates the speed at which clinical interventions can impact pneumococcal epidemiology. This further demonstrates the value of carriage studies for rapidly detecting these epidemiological shifts. Weinberger *et al* have also recently presented evidence that carriage data can be used to model PCV impact on IPD (Weinberger *et al.* 2013). Furthermore, carriage is establishing itself as an endpoint for assessing vaccine efficacy because of the important contribution of carriage to indirect vaccine effects -herd protection, and serotype replacement. Methodologies for vaccine trials now exist, which, allow estimation of vaccine efficacy from cross sectional measurements of carriage taking into account prevalence, acquisition rate and duration of carriage (Rinta-Kokko *et al.* 2009; Auranen *et al.* 2013). Surveillance of carriage over the course of both PCV7 and PCV13 implementation and routine usage provides us with data that describes how the pneumococcal population, or indeed any microbial population with only a subset of the existing subtypes targeted by an intervention, could respond to future interventions. This allows the formulation of the most suitable approach for sustaining control of disease. Serotype replacement of *H. influenzae* after the introduction the serotype B (Hib) vaccine or meningococci after the meningitis C vaccine has not been observed to the same extent as for pneumococci despite the existence of multiple serotypes (Borrow 2013; Menzies *et al.* 2013). This has been hypothesised to be due to the initial low prevalence of *H. influenzae* type B (Hib) carriage making changes harder to detect (Lipsitch 1997) and this explanation could

ultimately apply to *N. meningitidis* as well (Trotter *et al.* 2005). Ultimately pneumococcal carriage study data needs to be related to pneumococcal disease epidemiology to make informed decisions and clinically relevant response to carriage replacement that has the potential to impact disease.

5.2.1 Does carriage extrapolate to disease?

A crude way to assess potential vaccine serotype coverage is to extrapolate from the serotype prevalence in carriage, as colonisation is a precursor to disease. However, this doesn't take in to account the differing potential for serotypes to cause a disease case (Brueggemann *et al.* 2003) and over simplifies the prediction. Nonetheless, observed PCV7 vaccine serotype coverage was 1% in the 5th year of study and it could be deduced that the majority of colonisation events, a precursor to infection, would not be PCV7 preventable if they progressed to disease cases. Therefore PCV7 would have very little capacity to sustain control of pneumococcal disease. This supports the decision to replace PCV7 with PCV13 (Department of Health 2010) in response to other replacement observations, primarily in IPD surveillance from the then Health Protection Agency (Public Health England 2009). Furthermore in year 5 PCV13 vaccine serotype coverage in carriage was already only 11% with complete carriage replacement as seen for PCV7 in this thesis likely in the following years.

The invasive disease potential of the serotypes that replace vaccine types is a fundamental influence on the level to which replacement in carriage results in disease replacement, and the extent to which the overall burden of disease impacted. It was observed in this thesis that the predictions of the invasive disease potential of a number of serotypes, using this carriage data set and UK post PCV IPD cases, differ from those previously generated in the pre PCV era in Oxford (Brueggemann *et al.* 2003). However the Oxford study generated ORs from a smaller carriage data set n=351 of children <5 years and limited set of IPD isolates n=150 rather than the PHE surveillance used in this thesis n=528 and 519 carriage isolates. The only non-vaccine type that was found in this thesis to have a significant difference between occurrence in carriage and disease with a propensity for disease was serotype 8. The most prevalent serotypes circulating in the carriage population observed in this thesis were more highly associated with carriage than disease, but the frequency at which these are carried can still result in substantial number of disease cases. Targeting of serotypes that are of high prevalence in carriage with low disease potential can result in more disease replacement than the targeting of serotypes of low prevalence in carriage with high propensities for disease (Flasche *et al.* 2011).

5.2.1.1 UK pneumococcal IPD epidemiology

Epidemiological shifts have also been observed in the UK for IPD. Furthermore the observations for carriage made in this thesis can be related to changes in IPD epidemiology. National surveillance of IPD in England and Wales allowed PCV7 serotype replacement in disease to be assessed (Miller *et al.* 2011b). When data from 2009-10 was compared with 2000-06 there was a staggering 98% reduction in PCV7 VT IPD in children of less than 2 years old which relates to the observation of near removal of PCV7 VT from carriage in this thesis. Additionally there was an 81% reduction in IPD in the elderly (>65 years) reported. This elderly population is not routinely vaccinated with PCVs and the observed reduction in disease is a likely consequence of herd protection conveyed through the reduced carriage of VT observed. However in 2009-2010 there were still 5809 cases of IPD in all age groups with an annual incidence of 10.6 per 100,000 compared with an adjusted incidence (for missing data and increases in ascertainment) of 16.1 per 100,000. This was only a 34% reduction in the annual incidence (Miller *et al.* 2011b).

NVT replacement in carriage observed in this thesis also occurred in disease, a 68% increase in NVT IPD carriage was observed in children of less than 2 years of age. The key difference between the observations in carriage and disease is the extent of replacement. In carriage, replacement was near complete and the overall carriage rate remained unchanged whilst for disease the reduction in PCV7 VT was offset by increases in NVT disease but overall there was still a reduction in disease. One possible reason for this observation is that the replacing NVTs in carriage in this instance have a greater propensity for the carriage state and a reduced invasive disease potential than the VT. These VT were specifically targeted because of their association with disease cases. The key serotypes involved in PCV7 disease replacement were by 2010 were reported to be 7F, and 19A both included in PCV13 and 22F which is included in Merck's PCV15 (Miller *et al.* 2011b; Skinner *et al.* 2011), the latter not yet in routine clinical use. Changes in these serotypes in carriage did not reach statistical significance, which may be a factor of sample size. For 19A and 7F it could be explained by their statistically significant Odds Ratio being greater than 1, used as a measure of the invasiveness and showing higher association with disease. The OR value for 7F, calculated from the PHE IPD data and this thesis's carriage data was comparable to the OR of serotype 1 established to be a particularly invasive serotype (Miller *et al.* 2011b).

In this thesis a significant reduction in PCV13 unique serotypes (n=6) was observed, which was associated in time with a 50% reduction in PCV13 specific IPD. This reduction in IPD was observed for the serotypes additional to PCV7 in children under 2 years between 2009/2010 (n=175) and 2010/2011 (n=89), by a comparison of pre and post PCV13 implementation IPD cases. This further illustrates the speed at which vaccination impacts disease epidemiology as with carriage (Miller *et al.* 2011a).

Reports from the US suggested that serotype replacement has not featured as heavily in the epidemiology of disease as the UK with a maximum estimate of 5 additional cases per 100,000 annually reported by one study, particularly involving 19A (Rosen *et al.* 2011). However Miller *et al* suggest that when like comparisons are made with paediatric hospital admissions, PCV7 serotype replacement in the US and England and wales is comparable to that observed in the US (Miller *et al.* 2011b).

The observations discussed above illustrate that PCVs have a rapid ability to eliminate the targeted serotypes from the nasopharyngeal niche and subsequently disease, but that they allow serotype replacement and therefore have a limited life span to continue preventing disease cases. Prevention of colonisation rather than only prevention of invasion is a key feature of PCVs that contributes to their cost effectiveness by adding to herd protection (Melegaro & Edmunds 2004b).

5.2.2 Continuing control of pneumococcal disease

Serotype switches have been documented to occur in the pneumococcal population. The identification of numerous serotypes found with a single ST and larger numbers of STs found within a single serotype within this thesis further demonstrates how common the artefacts of serotype switches are in the population (Pai *et al.* 2005; Temime *et al.* 2008; Hanage *et al.* 2011; Croucher *et al.* 2013). The importance of these extant switches in vaccine effectiveness comes into play when targeting specific serotypes of high disease potential to control particularly virulent genotypes. These genotypes may be contributing to the invasive disease potential of a serotype because of the association between serotypes and genotypes. Antibiotic resistant lineages are an important example of this where the presence of these lineages with alternate serotypes prevents the control of hard to treat disease cases previously associated primarily with a vaccine type.

With an increased serotype valency PCV from Merck in development but not yet licensed (Skinner *et al.* 2011) and a number of experimental protein based or whole cell vaccine in experimental stages, the future of vaccine policy progression and

immediate continued control of pneumococcal disease is unclear. It is likely that a decision will need to be made on how to tackle serotype replacement, either with a vaccine that does not promote serotype replacement or a vaccine policy that responds continuously to it. Vaccine cost effectiveness relies in part on the elimination of carriage which results in replacement if this is combined with broad spectrum pneumococcal targets this could well avoid serotype replacement but result in species replacement. This could have a detrimental affect on controlling broader infectious diseases that would perhaps be more difficult to quantify or attribute to a vaccination schedule.

Continuing control with conjugate vaccines could involve further increasing the valency of vaccines with a 15-valent in trials and anecdotal suggestions of ~20 valent in the pipeline from pharmaceuticals. An alternative strategy would be to react to serotype replacement in an influenza-like periodical arms race. Conjugate vaccines could not be manufactured in seasonal timeframes but over 5-year periods, which would fit with the timeframe and observation of this study. To enable this to happen there would likely need to be changes to regulatory and licensing requirements with smaller safety and immunogenicity trials to make this approach to controlling specific serotypes with limited valency conjugates feasible.

5.3 Limitations

A limitation of this study is that it utilised children attending an outpatient clinic, which may not be a true representation of pneumococcal epidemiology in healthy individuals. However this population is a proxy for the wider community that will contain individuals in various states of health and allowed the timely collection of strains in a time sensitive study. Furthermore subjects were not attending clinics, which would likely contain immunosuppressed, or immunocompromised individuals with no intentional bias towards any particular clinic. Wherever possible the sibling of the outpatient was swabbed to reduce this potential bias. The observed carriage rate was comparable with other UK and developed world studies between 25-50%, which targeted populations classically considered to be ‘healthy’ (Pebody *et al.* 2009; Hanage *et al.* 2011; Wroe *et al.* 2012), but it is important to note that all populations will have a distribution of health statuses. This population did however have a high vaccine uptake so it adequately represents the vaccinated population. Additionally others have also used the outpatient clinic to provide timely access to children for pneumococcal carriage studies (Hernandez-Bou *et al.* 2012).

It has to be considered that the surveillance of carriage in this study as with other culture dependant studies will likely be underestimating the level of pneumococcal carriage as discussed in 1.3.1 -Detection of carriage, and the full review of pneumococcal carriage study protocol (Gladstone *et al.* 2012b). Furthermore this study did not set out to identify multiple serotype colonisation events and therefore only includes data for a single strain per positive participant. A previous culture based pilot study, as part of a BM student study in depth, did not identify multiple serotypes (unpublished data). However differences identified between participant isolates cultured for this thesis and those previously isolated from an individual participant samples (3.1.4 –Data consolidation) and the identification of mixed serotype sequences in WGD (2.1.3 –Strain exclusion and inclusion) suggests this did occur in the study population. This phenomenon would be better investigated in a culture independent fashion (Turner *et al.* 2013) or with more sensitive methods to detect the minority serotype including broth enrichment and microarray (Carvalho *et al.* 2010; Turner *et al.* 2011).

The impact of PCVs on the nasopharyngeal niche inclusive of all microbial inhabitants was not the focus of this thesis. There is potential however for PCV related species replacement given the documented specific interactions between constituent species of the nasopharynx and the removal of pneumococcal VT. The carriage study has collected key nasopharyngeal bacteria additional to pneumococci since year three of the study. The three years of additional species available at the formulation of this thesis project were not considered sufficient for the deduction of PCV impact due to the lower isolation rates and is now being analysed by others as part of the growing data set.

A limitation of working with genotypic data to describe and explain pneumococcal epidemiology is that genotype does not always relate to phenotype. For example, any functional or structural repercussions of novel sequence, that may affect vaccine efficacy, cannot be identified from genotype alone; this will require complimentary phenotypic, biochemical and immunological work.

Finally the gene-by-gene approach that thesis took to utilising whole genome data also does not fully explore all the data generated.

5.4 Future directions

This thesis generated new hypotheses to be tested including the potential for protein sequence variation to have structural implications and consequently distinct immunological profiles for which protein structures could be modelled and immunogenicity profiles investigated. A huge amount of data was generated by whole genome sequencing the >500 isolates into which many more exploratory lines of investigations could be made. This thesis made a gene-by-gene investigation for typing purposes. However, it would be of interest to investigate further utilising the complete compliment of sequence data to describe the phylogenetic structure of the population over PCV7 introduction, and determine both the accessory and core genome size for the population and further years by year to reveal further the impact of PCV7 on the genetic diversity of the population over time. Finally the collection of pneumococcal isolates used in this thesis continues with seven years of isolates already collected and plans in place for reaching the 10-year time point. This growing collection will need to be further characterised to determine the continued impact of PCV13 and assess when equilibrium is reached.

5.5 Concluding remarks

The data in this thesis has embraced cutting edge technology and provides a novel application of next generation sequencing to a new medical public health intervention; the surveillance of pneumococcal carriage epidemiology over the course of PCV implementation in the UK.

During the introduction of PCV7 and PCV13 the pneumococcal carriage population shifted rapidly with vaccine serotypes and associated genotypes replaced by non-vaccine serotypes and associated genotypes, this was mirrored to a lesser extent in disease in the UK. Additionally, variation in the protein sequences of future and experimental vaccine targets was observed previously undocumented with the potential to affect vaccine serotype coverage or further implications for replacement in the pneumococcal population.

The data generated adds to the characterisation of the pneumococcal population and current understanding of pneumococcal response to clinical interventions that can help to inform future vaccine policy and progression and the design of experimental vaccines.

APPENDICES

CDC PCR for deduction of pneumococcal serotype

Table 24. Serotype PCR cycling parameters

Thermal Cycling Parameters	
1 x	95ºC for 15min
35 x	94ºC for 30 sec
	54ºC for 90 sec
	72ºC for 60 sec
1 x	72ºC for 10min
∞	10ºC

Table 25. Reaction mixture

Reagent	**Concentration**	**(µL)/rxt**
MgCl2	50mM	0.5
Fwd Primers	100µM	Varies*
Rev Primers	100µM	Varies*
Red PCR Mix	2x	6.25
DNA Extract	Neat	1
H_2O	Neat	Make up to 12.5

Table 26. Primers concentrations multiplexing and product size

Primer Info		To get final concentration:			
Reaction 1	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**	**Conc. (µM) 1 rxn**	**Band (bp)**
cpsA -F	100	0.0125	1.875	0.1	160
cpsA -R	100	0.0125	1.875	0.1	
6(6A/6B/6C)-F	100	0.0375	5.625	0.3	250
6(6A/6B/6C)-R	100	0.0375	5.625	0.3	
3-F	100	0.0375	5.625	0.3	371
3-R	100	0.0375	5.625	0.3	
19A-F	100	0.0375	5.625	0.3	566
19A-R	100	0.0375	5.625	0.3	
22F/22A-F	100	0.0625	9.375	0.5	643
22F/22A-R	100	0.0625	9.375	0.5	
16F-F	100	0.05	7.5	0.4	717
16F-R	100	0.05	7.5	0.4	
Primer Info		**To get final concentration:**			
Reaction 2	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**	**Conc. (µM) 1 rxn**	**Band (bp)**
cpsA -F	100	0.0125	1.875	0.1	160
cpsA -R	100	0.0125	1.875	0.1	
8-F	100	0.025	3.75	0.2	201
8-R	100	0.025	3.75	0.2	
33F/33A/37-F	100	0.0375	5.625	0.3	338
33F/33A/37-R	100	0.0375	5.625	0.3	
15A/15F-F	100	0.0375	5.625	0.3	434
15A/15F-R	100	0.0375	5.625	0.3	
7F/7A-F	100	0.05	7.5	0.4	599
7F/7A-R	100	0.05	7.5	0.4	
23A-F	100	0.0625	9.375	0.5	722
23A-R	100	0.0625	9.375	0.5	
Primer Info		**To get final concentration:**			
Reaction 3	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**	**Conc. (µM) 1 rxn**	**Band (bp)**
cpsA-F	100	0.0125	1.875	0.1	160
*cps*A-R	100	0.0125	1.875	0.1	
19F-F	100	0.0625	9.375	0.5	304
19F-R	100	0.0625	9.375	0.5	
12F/12A/44/46-F	100	0.0625	9.375	0.5	376
12F/12A/44/46-R	100	0.0625	9.375	0.5	
11A/11D-F	100	0.0375	5.625	0.3	463
11A/11D-R	100	0.0375	5.625	0.3	

38/25F-F	100	0.0375	5.625	0.3	574
38/25F-R	100	0.0375	5.625	0.3	
35B-F	100	0.0625	9.375	0.5	677
35B-R	100	0.0625	9.375	0.5	
Primer Info		**To get final concentration:**		**Conc. (µM) 1 rxn**	**Band (bp)**
Reaction 4	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**		
24(24A/24B/24F)-F	100	0.0125	1.875	0.1	99
24(24A/24B/24F)-R	100	0.0125	1.875	0.1	
*cps*A -F	100	0.0125	1.875	0.1	160
*cps*A -R	100	0.0125	1.875	0.1	
7C/7B/40-F	100	0.0375	5.625	0.3	260
7C/7B/40-R	100	0.0375	5.625	0.3	
4-F	100	0.0375	5.625	0.3	430
4-R	100	0.0375	5.625	0.3	
18(18A/18B/18/C/18F)-F	100	0.0375	5.625	0.3	573
18(18A/18B/18/C/18F)-R	100	0.0375	5.625	0.3	
9V/9A-F	100	0.0625	9.375	0.5	816
9V/9A-R	100	0.0625	9.375	0.5	
Primer Info		**To get final concentration:**		**Conc. (µM) 1 rxn**	**Band (bp)**
Reaction 5	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**		
*cps*A -F	100	0.0125	1.875	0.1	160
*cps*A -R	100	0.0125	1.875	0.1	
14-F	100	0.0375	5.625	0.3	189
14-R	100	0.0375	5.625	0.3	
1-F	100	0.0375	5.625	0.3	280
1-R	100	0.0375	5.625	0.3	
23F-F	100	0.0625	9.375	0.5	384
23F-R	100	0.0625	9.375	0.5	
15B/15C-F	100	0.0375	5.625	0.3	496
15B/15C-R	100	0.0375	5.625	0.3	
10A-F	100	0.0625	9.375	0.5	628
10A-R	100	0.0625	9.375	0.5	
Primer Info		**To get final concentration:**		**Conc. (µM) 1 rxn**	**Band (bp)**
Reaction 6	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**		
39-F	100	0.025	3.75	0.2	98
39-R	100	0.025	3.75	0.2	

*cps*A -F	100	0.0125	1.875	0.1	160
*cps*A -R	100	0.0125	1.875	0.1	
10F/10C/33C-F	100	0.0375	5.625	0.3	248
10F/10C/33C-R	100	0.0375	5.625	0.3	
5-F	100	0.0375	5.625	0.3	362
5-R	100	0.0375	5.625	0.3	
35F/47F-F	100	0.0375	5.625	0.3	517
35F/47F-R	100	0.0375	5.625	0.3	
17F-F	100	0.0625	9.375	0.5	693
17F-R	100	0.0625	9.375	0.5	
Primer Info		**To get final concentration:**			
Reaction 7	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**	**Conc. (µM) 1 rxn**	**Band (bp)**
*cps*A -F	100	0.0125	1.875	0.1	160
*cps*A -R	100	0.0125	1.875	0.1	
23B-F	100	0.025	3.75	0.2	199
23B-R	100	0.025	3.75	0.2	
35A/35C/42-F	100	0.0375	5.625	0.3	280
35A/35C/42-R	100	0.0375	5.625	0.3	
34-F	100	0.0375	5.625	0.3	408
34-R	100	0.0375	5.625	0.3	
9N/9L-F	100	0.0625	9.375	0.5	516
9N/9L-R	100	0.0625	9.375	0.5	
31-F	100	0.0625	9.375	0.5	701
31-R	100	0.0625	9.375	0.5	
Primer Info		**To get final concentration:**			
Reaction 8	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**	**Conc. (µM) 1 rxn**	**Band (bp)**
*cps*A -F	100	0.0125	1.875	0.1	160
*cps*A -R	100	0.0125	1.875	0.1	
21-F	100	0.025	3.75	0.2	192
21-R	100	0.025	3.75	0.2	
2-F	100	0.0375	5.625	0.3	290
2-R	100	0.0375	5.625	0.3	
20-F	100	0.0375	5.625	0.3	514
20-R	100	0.0375	5.625	0.3	
13-F	100	0.05	7.5	0.4	655
13-R	100	0.05	7.5	0.4	
Primer Info		**To get final concentration:**			
Reaction 6C	**Original Conc. (µM)**	**Primer (µL)**	**(µL)/Mmix**	**Conc. (µM) 1 rxn**	**Band (bp)**

*cps*A -F	100	0.0125	1.875	0.1	160
*cps*A -R	100	0.0125	1.875	0.1	
6A/6B/6C-F	100	0.0375	5.625	0.3	250
6A/6B/6C-R	100	0.0375	5.625	0.3	
6C-F	100	0.0625	9.375	0.5	727
6C-R	100	0.0625	9.375	0.5	

Table 27. Primer sequences

Target	Gene	Forward	Reverse	Reference/sequence from	Use
Serotype 1	*wzy*	CTCTATAGAATGGAGTATATAAACTATGGTTA	CCAAAGAAAATACTAACATTATCACAATATTGGC	(Centers for Disease Control and prevention, 2012)	PCR, *in silico*
Serotype 2	*wzy*	TATCCCAGTTCAATATTTCTCCACTACACC	ACACAAAATATAGGCAGAGAGAGACTACT		PCR, *in silico*
Serotype 3	*galU*	ATGGTGTGATTTCTCCTAGATTGGAAAGTAG	CTTCTCCAATTGCTTACCAAGTGCAATAACG		PCR, *in silico*
Serotype 4	*wzy*	CTGTTACTTGTTCTGGACTCTCGATAATTGG	GCCCACTCCTGTTAAAATCCTACCCGCATTG		PCR, *in silico*
Serotype 5	*wzy*	ATACCTACACAACTTCTGATTATGCCTTTGTG	GCTCGATAAACATAATCAATATTTGAAAAAGTATG		PCR, *in silico*
Serotypes 6A_6B_6C_6D	*wciP*	AATTTGTATTTTATTCATGCCTATATCTGG	TTAGCGGAGATAATTTAAAATGATGACTA		PCR, *in silico*
Serotypes 6C_6D	*wciNbeta*	CATTTTAGTGAAGTTGGCGGTGGAGTT	AGCTTCGAAGCCCATACTCTTCAATTA		PCR, *in silico*
Serotypes 7C_7B_40	*wcwL*	CTATCTCAGTCATCTATTGTTAAAGTTTACGACGGGA	GAACATAGATGTTGAGACATCTTTTGTAATTTC		PCR, *in silico*
Serotypes 7F_7A	*wzy*	TCCAAACTATTACAGTGGGAATTACGG	ATAGGAATTGAGATTGCCAAAGCGAC		PCR, *in silico*
Serotype 8	*wzy*	GAAGAAACGAAACTGTCAGAGCATTTACAT	CTATAGATACTAGTAGAGCTGTTCTAGTCT		PCR, *in silico*
Serotypes 9N_9L	*wzx*	GAACTGAATAAGTCAGATTTAATCAGC	ACCAAGATCTGACGGGCTAATCAAT		PCR, *in silico*
Serotypes	*wzy*	GGGTTCAAAGTCAGACAGTGA	CCATGAATGAAATCAACATTGT		PCR, *in silico*

Target	**Gene**	**Forward**	**Reverse**	**Reference/sequence from**	**Use**
9V_9A		ATCTTAA	CAGTAGC		
Serotype10A	*wcrG*	GGTGTAGATTTACCATTAGTGTCGGCAGAC	GAATTTCTTCTTTAAGATTCGGATATTTCTC		PCR, *in silico*
Serotypes 10F_10C_33C	*wzx*	GGAGTTTATCGGTAGTGCTCATTTTAGCA	CTAACAAATTCGCAACACGAGGCAACA		PCR, *in silico*
Serotypes 11A_11D	*wzy*	GGACATGTTCAGGTGATTTCCCAATATAGTG	GATTATGAGTGTAATTTATTCCAACTTCTCCC		PCR, *in silico*
Serotypes 12F_12A_44_46	*wzx*	GCAACAAACGGCGTGAAAGTAGTTG	CAAGATGAATATCACTACCAATAACAAAAC		PCR, *in silico*
Serotype 3	*wzx*	TACTAAGGTAATCTCTGGAAATCGAAAGG	CTCATGCATTTTATTAACCGCTTTTTGTTC		PCR, *in silico*
Serotype 14	*wzy*	GAAATGTTACTTGGCGCAGGTGTCAGAATT	GCCAATACTTCTTAGTCTCTCAGATGAAT		PCR, *in silico*
Serotypes 15A_15F	*wzy*	ATTAGTACAGCTGCTGGAATATCTCTTC	GATCTAGTGAACGTACTATTCCAAAC		PCR, *in silico*

Target	**Gene**	**Forward**	**Reverse**	**Reference/sequence from**	**Use**
Serotypes 15B_15C	*wzy*	TTGGAATTTTTTAATTAGTGGCTTACCTA	CATCCGCTTATTAATTGAAGTAATCTGAACC	(Centers for Disease Control and prevention, 2012)	PCR, *in silico*
Serotype 16F	*wzy*	GAATTTTTCAGGCGTGGGTGTTAAAAG	CAGCATATAGCACCGCTAAGCAAATA		PCR, *in silico*
Serotype 17F	*wci*P	TTCGTGATGATAATTCCAATGATCAAACAAGAG	GATGTAACAAATTTGTAGCGACTAAGGTCTGC		PCR, *in silico*
Serotypes 18A_18B_18C_18F	*wzy*	CTTAATAGCTCTCATTATTCTTTTTTTAAGCC	TTATCTGTAAACCATATCAGCATCTGAAAC		PCR, *in silico*
Serotype19A	*wzy*	GAGAGATTCATAATCTTGCACTTAGCCA	CATAATAGCTACAAATGACTCATCGCC		PCR, *in silico*
Serotype 19F	*wzy*	GTTAAGATTGCTGATCGATTAATTGATATCC	GTAATATGTCTTTAGGGCGTTTATGGCGATAG		PCR, *in silico*
Serotype 20	*wci*L	GAGCAAGAGTTTTTCACCTGACAGCGAGAAG	CTAAATTCCTGTAATTTAGCTAAAACTCTTATC		PCR, *in silico*
Serotype 21	*wzx*	CTATGGTTATTTCAACTCAATCGTCACC	GGCAAACTCAGACATAGTATAGCATAG		PCR, *in silico*
Serotypes 22F_22A	*wcw*V	GAGTATAGCCAGATTATGGCAGTTTTATTGTC	CTCCAGCACTTGCGCTGGAAACAACAGACAAC		PCR, *in silico*
Serotype 23A	*wzy*	TATTCTAGCAAGTGACGAAGATGCG	CCAACATGCTTAAAAACGCTGCTTTAC		PCR, *in silico*
Serotype 23B	*wzx*	CCACAATTAGCGCTATATTCATTC	GTCCACGCTGAATAAAATGAA		PCR, *in silico*

Target	Gene	Forward	Reverse	Reference/sequence from	Use
		AATCG	GCTCCG		
Serotype 23F	*wzy*	GTAACAGTTGCTGTAGAGGGAATTGGCTTTTC	CACAACACCTAACACTCGATGGCTATATGATTC		PCR, *in silico*
Serotypes 24A_24B_24F	*wzy*	GCTCCCTGCTATTGTAATCTTTAAAGAG	GTGTCTTTTATTGACTTTATCATAGGTCGG		PCR, *in silico*
Serotype 31	*wzy*	GGAAGTTTTCAAGGATATGATAGTGGTGGTGC	CCGAATAATATATTCAATATATTCCTACTC		PCR, *in silico*
Serotypes 33F_33A_37	*wzy*	GAAGGCAATCAATGTGATTGTGTCGCG	CTTCAAAATGAAGATTATAGTACCCTTCTAC		PCR, *in silico*
Serotype 34	*wzy*	GCTTTTGTAAGAGGAGATTATTTTCACCCAAC	CAATCCGACTAAGTCTTCAGTAAAAAACTTTAC		PCR, *in silico*
Serotypes 35A_35C_42	*wzx*	ATTACGACTCCTTATGTGACGCGCATA	CCAATCCCAAGATATATGCAACTAGGTT		PCR, *in silico*
Serotype 35B	*wcr*H	GATAAGTCTGTTGTGGAGACTTAAAAAGAATG	CTTTCCAGATAATTACAGGTATTCCTGAAGCAAG		PCR, *in silico*
Serotypes 35F_47F	*wzy*	GAACATAGTCGCTATTGTATTTTATTTAAAGCAA	GACTAGGAGCATTATTCCTAGAGCGAGTAAACC		PCR, *in silico*

Target	**Gene**	**Forward**	**Reverse**	**Reference/sequence from**	**Use**
Serotype 38_25F_25A	*Wzy*	CGTTCTTTTATCTCACTGTATAGTATCTTTATG	ATGTTTGAATTAAAGCTAACGTAACAATCC	(Centers for Disease Control and prevention, 2012)	PCR, *in silico*
Serotype 39	*Wzy*	TCATTGTATTAACCCTATGCTTTATTGGTG	GAGTATCTCCATTGTATTGAAATCTACCAA		PCR, *in silico*
cpsA	*cpsA*	GCAGTACAGCAGTTTGTTGGACTGACC	GAATATTTTCATTATCAGTCCCAGTC		PCR, *in silico*
Methionine aminopeptidase	*map*	GCWGACTCWTGTTGGGCWTATGC	TTARTAAGTTCYTTCTTCDCCTTG	(Bishop et al., 2009)	*in silico*
Pyruvate formate lyase	*Pfl*	AACGTTGCTTACTCTAAACAAACTGG	ACTTCRTGGAAGACACGTTGWGTC		*in silico*
Inorganic pyrophosphatase	*ppa*C	GACCAYAATGAATTYCARCAATC	TGAGGNACMACTTGTTTSTTACG		*in silico*
Pyruvate kinase	*Pyk*	GCGGTWGAAWTCCGTGGTG	GCAAGWGCTGGGAAAGGAAT		*in silico*
RNA polymerase beta subunit	*rpo*B	AARYTIGGMCCTGAAGAAAT	TGIARTTTRTCATCAACCATGTG		*in silico*
Superoxide dismutase	*sod*A	TRCAYCATGAYAARCACCAT	ARRTARTAMGCRTGYTCCCARACRTC		*in silico*
Elongation factor Tu	*tuf*	GTTGAAATGGAAATCCGTGACC	GTTGAAGAATGGAGTGTGACG		*in silico*

Target	Gene	Forward	Reverse	Reference/sequence from	Use
Pneumolysin	*Ply*	ATGGCAAATAAAGCAGTAAA	CTAGTCATTTTCTACCTTATCTTCTA	FQ312041.2	*in silico*
Serogroup 15	*wciZ*	ATGGTTACTAAAGATAAAGGATTAACA	TTAACTAAACCTTTTTAATAGTCTGTTT	CR931663	*in silico*
Serotype19A	Wzy	ATGACTTATTTATTTTTACTCTGCCTGACCTTAT	TCAGTTTTCCGATTTAATACTTGTGTCAGAG	CR931675	*in silico*
Serotype 19F	Wzy	ATGAGTTATTTATTTTTACTTTGCCTTACAT	TTATTCTTCTTTTAATTTGATACTTGAGTCAG	CR931678	*in silico*
Serogroup 6	*wci*P	ATGGGAAAGTCAGTTGCAATTTTAATG	TTATAAGAGTTCTTTCATAATATATTTTTT	CR931638	*in silico*
Serogroup 6	*wci*N beta	TTTTATAATGTTCTAAAGGGAGATATGAGT	GCTTTTTTAGCAGGCGACATAGTTTTTCTT	EF538718.1	*in silico*
Serotype 21	*wzx*	ATGAAAGTTCTTAAAAATTACGCCTATAAT	TTAAATCTCCCTATTTTTTAATAGTTCCT	CR931680	*in silico*
Serotype 22F	*wcw*A	TTATACTTTAGGCTTTCACCATCACGTA	TTAATTAATCTTTTGATTATCATTTGTAATC	HE651300	*in silico*
Serotype 22F	*wcw*C	ATGAAAGGATTACAAATGATAATCAAA	TTTAATACTACTTTCTTTCCATTAAAAT	HE651300	*in silico*

GLOSSARY

Term	Definition
Contig	A contig (from contiguous) is a set of overlapping DNA segments that together represent a consensus region of DNA
In silico	Analysis performed using computational methods

Abbreviation	Detail
AOM	Acute otitis media
*aro*E	Encodes shikimate dehydrogenase
BHI	Brain heart infusion
bp	Base pair
CC	Clonal complex
CDC	Centers for Disease Control and Prevention
CDS	Coding sequence
CI	Confidence interval
ClpP	Caseinolytic protease
CNA	Colistin naladixic acid agar
cps15bM/wciZ	Encodes capsular polysaccharide O-acetyltransferase
cpsA	Encodes capsular polysaccharide synthesis gene A
ddl	Encodes D-alanine-D-alanine ligase
(d)dNTPs	(Di)-deoxynucleoside triphosphates
eBURST	Electronic-based on related sequence types
*erm*B	Encodes erythromycin ribosomal methylase
ETA	Egg-thioglycolate-antibiotic medium

Glossary

GAVI	Formerly The Global Alliance for Vaccines and Immunisation
gdh	Encodes glucose-6-phosphate dehydrogenase
gki	Encodes glucose kinase
IPD	Invasive pneumococcal disease
Mb	Megabase
MLST	Multi locus sequence type/typing
NS	Non synonymous
NTHi	Non typeable *Haemophilus influenzae*
NVT	Non vaccine serotype
PCV	Pneumococcal conjugate vaccine
PCV-13	Thirteen valent pneumococcal conjugate vaccine
PCV-7	Seven valent pneumococcal conjugate vaccine
PHE	Public Health England
PHiD-CV	Pneumococcal, with *Haemophilus influenzae* protein D, conjugate vaccine
Ply/*ply*	Encodes pneumolysin Protein/DNA
PPV	Pneumococcal polysaccharide vaccine
PspA	Pneumococcal surface protein A
PspC	Pneumococcal surface protein C
PTP	Pico titre plate
*rec*P	Encodes transketolase
SLV	Single locus variant
SNP	Single nucleotide polymorphism
spi	Encodes signal peptidase I

ST	Sequence type
STGG	Skim milk-tryptone-glucose-glycerin
*tet*M	Encodes tetracycline-resistant ribosome protection protein
VRT	Vaccine related serotype
VT	Vaccine serotype
WCV	Whole cell vaccine
*wci*N	Encodes capsular polysaccharide glycosyltransferase
*wci*Nbeta	Encodes capsular polysaccharide glycosyltransferase
*wci*P	Encodes capsular polysaccharide rhamnosyl transferase
*wcr*L	Encodes capsular polysaccharide glycosyltransferase
*wcw*A	Encodes capsular polysaccharide glycosyltransferase
*wcw*C	Encodes capsular polysaccharide O-acetyltransferase
WGS	Whole genome sequence
WGD	Whole genome data
WHO	World Health Organisation
wzx	Encodes capsular polysaccharide translocation flippase
wzy	Encodes capsular oligosaccharide repeat unit polymerase
xpt	Encodes xanthine phosphoribosyltransferase

References

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

Abdullahi, O., E. Wanjiru, R. Musyimi, N. Glass & J. A. G. Scott (2007). "Validation of nasopharyngeal sampling and culture techniques for detection of Streptococcus pneumoniae in children in Kenya." *J. Clin. Microbiol* **45**(10): 3408-3410.

Allhoff, M., A. Schonhuth, M. Martin, I. G. Costa, S. Rahmann & T. Marschall (2013). "Discovering motifs that induce sequencing errors." *BMC Bioinformatics* **14 Suppl 5**, S1.

Anevlavis, S. & D. Bouros (2010). "Community acquired bacterial pneumonia." *Expert Opinion on Pharmacotherapy* **11**(3), 361-374.

Ansaldi, F., D. de Florentiis, P. Canepa, M. Zancolli, M. Martini, A. Orsi, P. Durando & G. Icardi (2012a). "Carriage of Streptoccoccus pneumoniae 7 years after implementation of vaccination program in a population with very high and long-lasting coverage, Italy." *Vaccine* **30**(13), 2288-2294.

Ansaldi, F., D. De Florentiis, V. Parodi, E. Rappazzo, M. Coppelli, M. Martini, C. Alicino, P. Durando & G. Icardi (2012b). "Bacterial carriage and respiratory tract infections in subjects > or = 60 years during an influenza season: implications for the epidemiology of community acquired pneumonia and influenza vaccine effectiveness." *J Prev Med Hyg* **53**(2), 94-97.

Ansorge, W. J. (2009). "Next-generation DNA sequencing techniques." *N Biotechnol* **25**(4), 195-203.

Arbique, J. C., C. Poyart, P. Trieu-Cuot, G. Quesne, G. Carvalho Mda, A. G. Steigerwalt, R. E. Morey, D. Jackson, R. J. Davidson & R. R. Facklam (2004). "Accuracy of phenotypic and genotypic testing for identification of Streptococcus pneumoniae and description of Streptococcus pseudopneumoniae sp. nov." *J Clin Microbiol* **42**(10), 4686-4696.

Auranen, K., J. Mehtala, A. Tanskanen & M. S. Kaltoft (2010). "Between-Strain Competition in Acquisition and Clearance of Pneumococcal Carriage--Epidemiologic Evidence From a Longitudinal Study of Day-Care Children." *Am. J. Epidemiol.* **171**(2), 169-176.

Auranen, K., H. Rinta-Kokko, D. Goldblatt, H. Nohynek, K. L. O'Brien, C. Satzke, B. Simell, A. Tanskanen, H. Kayhty & Pneumococcal Carriage Group (2013). "Colonisation endpoints in Streptococcus pneumoniae vaccine trials." *Vaccine* **32**(1): 153-158.

Avery, O. T. & W. F. Goebel (1929). "Chemo-Immunological Studies on Conjugated Carbohydrate-Proteins : Ii. Immunological Specificity of Synthetic Sugar-Protein Antigens." *J Exp Med* **50**(4), 533-550.

Avery, O. T., MacLeod, C. M., & McCarty, M. (1944). "Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation desoxyribonucleic acid fraction isolated from pneumococcus type III." *Journal of Experimental Medicine* **79**, 137-157.

Avni, T., N. Mansur, L. Leibovici & M. Paul (2010). "PCR using blood for diagnosis of invasive pneumococcal disease: systematic review and meta-analysis." *J Clin Microbiol* **48**(2), 489-496.

Balmer, P., R. Borrow & P. D. Arkwright (2007). "The 23-valent pneumococcal polysaccharide vaccine does not provide additional serotype antibody protection in children who have been primed with two doses of heptavalent pneumococcal conjugate vaccine." *Vaccine* **25**(34), 6321-6325.

Bättig, P., L. J. Hathaway, S. Hofer & K. Mühlemann (2006). "Serotype-specific invasiveness and colonization prevalence in Streptococcus pneumoniae correlate with the lag phase during in vitro growth." *Microbes and Infection* **8**(11), 2612-2617.

Bentley, S. D., D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabbinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail other authors (2006). "Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes." *PLoS Genet* **2**(3), e31.

Bishop, C. J., D. M. Aanensen, G. E. Jordan, M. Kilian, W. P. Hanage & B. G. Spratt (2009). "Assigning strains to bacterial species via the internet." *BMC Biol* **7**, 3.

Black, S., H. Shinefield, R. Baxter, R. Austrian, L. Bracken, J. Hansen, E. Lewis & B. Fireman (2004). "Postlicensure surveillance for pneumococcal invasive disease after use of heptavalent pneumococcal conjugate vaccine in Northern California Kaiser Permanente." *Pediatr Infect Dis J* **23**(6), 485-489.

Black, S. B. M., H. R. M. Shinefield, S. M. M. Ling, J. Hansen, B. M. Fireman, D. M. Spring, J. M. Noyes, E. M. Lewis, P. M. Ray, J. D. Lee other authors (2002). "Effectiveness of heptavalent pneumococcal conjugate vaccine in children younger than five years of age for prevention of pneumonia." *Pediatric Infectious Disease Journal* **21**(9), 810-815.

Block, S. L., J. Hedrick, C. J. Harrison, R. Tyler, A. Smith, R. Findlay & E. Keegan (2004). "Community-Wide Vaccination with the Heptavalent Pneumococcal Conjugate Significantly Alters the Microbiology of Acute Otitis Media." *The Pediatric Infectious Disease Journal* **23**(9), 829-833.

Bluestone, C. D., J. S. Stephenson & L. M. Martin (1992). "Ten-year review of otitis media pathogens." *The Pediatric Infectious Disease Journal* **11**(8), S7-11.

Boersma, W. G., A. Lowenberg, Y. Holloway, H. Kuttschrutter, J. A. M. Snijder & G. H. Koeter (1993). "The Role of Antigen Detection in Pneumococcal Carriers: A Comparison between Cultures and Capsular Antigen Detection in Upper Respiratory Tract Secretions." *Scandinavian Journal of Infectious Diseases* **25**(1), 51-56.

Bogaert, D., R. de Groot & P. W. M. Hermans (2004a). "Streptococcus pneumoniae colonisation: the key to pneumococcal disease." *The Lancet Infectious Diseases* **4**(3), 144-154.

Bogaert, D., B. Keijser, S. Huse, J. Rossen, R. Veenhoven, E. van Gils, J. Bruin, R. Montijn, M. Bonten & E. Sanders (2011). "Variability and Diversity of Nasopharyngeal Microbiota in Children: A Metagenomic Analysis." *PLoS ONE* **6**(2), e17035.

Bogaert, D., A. van Belkum, M. Sluijter, A. Luijendijk, R. de Groot, H. C. Rümke, H. A. Verbrugh & P. W. M. Hermans (2004b). "Colonisation by Streptococcus pneumoniae and Staphylococcus aureus in healthy children." *The Lancet* **363**(9424), 1871-1872.

Borrow, R. (2013). "The evidence for use of pneumococcal conjugate over polysaccharide in children." *Hum Vaccin Immunother* **9**(6).

Bosch, A. A., G. Biesbroek, K. Trzcinski, E. A. Sanders & D. Bogaert (2013). "Viral and bacterial interactions in the upper respiratory tract." *PLoS Pathog* **9**(1), e1003057.

Bowen, M. K., L. C. Thiele, B. D. Stearman & I. G. Schaub (1957). "The Optochin sensitivity test: a reliable method for identification of pneumococci." *J Lab Clin Med* **49**(4), 641-642.

Bratcher, P. E., K. H. Kim, J. H. Kang, J. Y. Hong & M. H. Nahm (2010). "Identification of natural pneumococcal isolates expressing serotype 6D by genetic, biochemical and serological characterization." *Microbiology* **156**(Pt 2), 555-560.

Bratcher, P. E., I. H. Park, M. B. Oliver, M. Hortal, R. Camilli, S. K. Hollingshead, T. Camou & M. H. Nahm (2011). "Evolution of the capsular gene locus of Streptococcus pneumoniae serogroup 6." *Microbiology* **157**(Pt 1), 189-198.

Broad Institute (2012). "IGV. http://www.broadinstitute.org/igv/." **Accessed 01/12/12**.

Brueggemann, A., R. Pai, D. Crook & B. Beall (2007). "Vaccine escape recombinants emerge after pneumococcal vaccination in the United States." *PLoS Pathog* **3**, e168.

Brueggemann, A. B., D. T. Griffiths, E. Meats, T. E. Peto, D. W. Crook & B. G. Spratt (2003). "Clonal Relationships between Invasive and Carriage Streptococcus pneumoniae and Serotype- and Clone- Specific Differences in Invasive Disease Potential." *The Journal of Infectious Diseases* **187**(9), 1424-1432.

Brueggemann, A. B., T. E. Peto, D. W. Crook, J. C. Butler, K. G. Kristinsson & B. G. Spratt (2004). "Temporal and Geographic Stability of the Serogroup-Specific Invasive Disease Potential of Streptococcus pneumoniae in Children." *The Journal of Infectious Diseases* **190**(7), 1203-1211.

Brugger, S., L. Hathaway & K. Muhlemann (2009). "Streptococcus pneumoniae strain co-colonization in the nasopharynx." *J Clin Microbiol* **47**, 1750 - 1756.

Brugger, S. D., P. Frey, S. Aebi, J. Hinds & K. Muhlemann (2010). "Multiple Colonization with S. pneumoniae before and after Introduction of the Seven-Valent Conjugated Pneumococcal Polysaccharide Vaccine." *PLoS ONE* **5**(7), e11638.

Bryant, J. M., D. M. Grogono, D. Greaves, J. Foweraker, I. Roddick, T. Inns, M. Reacher, C. S. Haworth, M. D. Curran, S. R. Harris other authors (2013). "Whole-genome sequencing to identify transmission of Mycobacterium abscessus between patients with cystic fibrosis: a retrospective cohort study." *Lancet* **381**(9877), 1551-1560.

Burman, L. A., R. Norrby & B. Trollfors (1985). "Invasive pneumococcal infections: incidence, predisposing factors, and prognosis." *Rev Infect Dis* **7**(2), 133-142.

Calix, J. J. & M. H. Nahm (2010). "A New Pneumococcal Serotype, 11E, Has a Variably Inactivated wcjE Gene." *Journal of infectious diseases* **202**(1), 29-38.

Calix, J. J., M. H. Nahm & E. R. Zartler (2011). "Elucidation of structural and antigenic properties of pneumococcal serotype 11A, 11B, 11C, and 11F polysaccharide capsules." *J Bacteriol* **193**(19), 5271-5278.

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

Camilli, R., R. Bonnal, M. Del Grosso, M. Iacono, G. Corti, E. Rizzi, M. Marchetti, L. Mulas, F. Iannelli, F. Superti other authors (2011). "Complete genome sequence of a serotype 11A, ST62 Streptococcus pneumoniae invasive isolate." *BMC Microbiology* **11**(1), 25.

Cao, J., D. Chen, W. Xu, T. Chen, S. Xu, J. Luo, Q. Zhao, B. Liu, D. Wang, X. Zhang other authors (2007). "Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP." *Vaccine* **25**(27), 4996-5005.

Cao, J., D. Li, Y. Gong, N. Yin, T. Chen, C. K. Wong, W. Xu, J. Luo, X. Zhang, C. W. K. Lam other authors (2009). "Caseinolytic protease: a protein vaccine which could elicit serotype-independent protection against invasive pneumococcal infection." *Clinical & Experimental Immunology* **156**(1), 52-60.

Capeding, M., H. Nohynek, L. Sombrero, L. Pascual, E. Sunico, G. Esparar, E. Esko, M. Leinonen & P. Ruutu (1995). "Evaluation of sampling sites for detection of upper respiratory tract carriage of Streptococcus pneumoniae and

Haemophilus influenzae among healthy Filipino infants." *J. Clin. Microbiol.* **33**(11), 3077-3079.

Carter, R. J. F. (2006). "Infants too young to receive pneumococcal conjugate vaccine benefit from herd immunity." *Thorax* **61**(7), 610-.

Carvalho, M. D., G. Jagero, G. M. Bigogo, M. Junghae, F. C. Pimenta, I. Moura, A. Roundtree, Z. Li, L. Conklin, D. R. Feikin other authors (2012). "Potential non-pneumococcal confounding of PCR-based determination of serotype in carriage." *J Clin Microbiol* **50**(9): 3146-3147.

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

Carvalho M. D., F. C. Pimenta, I. Moura, A. Roundtree, R. E. Gertz, Jr., Z. Li, G. Jagero, G. Bigogo, M. Junghae, L. Conklin other authors (2013). "Non-pneumococcal mitis-group streptococci confound detection of pneumococcal capsular serotype-specific loci in upper respiratory tract." *PeerJ* **1**, e97.

Casey, J. R., D. G. Adlowitz & M. E. Pichichero (2009). "New Patterns in the Otopathogens Causing Acute Otitis Media Six to Eight Years After Introduction of Pneumococcal Conjugate Vaccine." *Pediatr Infect Dis J* **29**(4): 304-309.

Centers for Disease Control and Prevention (2005). "Direct and indirect effects of routine vaccination of children with 7-valent pneumococcal conjugate vaccine on incidence of invasive pneumococcal disease--United States, 1998-2003." *MMWR Morb Mortal Wkly Rep* **54**(36), 893-897.

Centers for Disease Control and prevention (2007). "Emergence of antimicrobial-resistant serotype 19A Streptococcus pneumoniae--Massachusetts, 2001-2006." *MMWR Morb Mortal Wkly Rep* **56**(41), 1077-1080.

Centers for Disease Control and Prevention (2009). "Morbidity and Mortality Weekly Report. Pneumonia Hospitalizations Among Young Children Before and After Introduction of Pneumococcal Conjugate Vaccine -United States, 1997-2006." *JAMA* **301**(12), 1220-1222.

Centers for Disease Control and prevention (2010). "Licensure of a 13-Valent pneumococcal vaccine and reccomendations for use among children." *Advisory Committee on Immunization Practices (ACIP), Morbidity and Mortality Weekly Report 2010* **59**, 258-261.

Centers for Disease Control and prevention (2011). "Invasive Pneumococcal Disease and 13-Valent Pneumococcal Conjugate Vaccine (PCV13) Coverage Among Children Aged </=59 Months --- Selected U.S. Regions, 2010--2011." *MMWR Morb Mortal Wkly Rep* **60**, 1477-1481.

Centers for Disease Control and prevention (2012). "PCR deduction of pneumococcal serotypes. http://www.cdc.gov/ncidod/biotech/strep/pcr.htm " **Accessed 01/12/2012**.

Chan, J. Z., M. J. Pallen, B. Oppenheim & C. Constantinidou (2012). "Genome sequencing in clinical microbiology." *Nat Biotechnol* **30**(11), 1068-1071.

Charalambous, B. M., S. L. Batt, A. C. Peek, H. Mwerinde, N. Sam & S. H. Gillespie (2003). "Quantitative Validation of Media for Transportation and Storage of Streptococcus pneumoniae." *J. Clin. Microbiol.* **41**(12), 5551-5556.

Cherian, T., E. K. Mulholland, J. B. Carlin, H. Ostensen, R. Amin, M. d. Campo, D. Greenberg, R. Lagos, M. Lucero, S. A. Madhi other authors (2005). "Standardized interpretation of paediatric chest radiographs for the diagnosis of pneumonia in epidemiological studies." *Bulletin of the World Health Organization* **83**, 353-359.

Chester, F. D. (1901). *A manual of determinative bacteriology*. New York The Macmillan Company; London, Macmillan & Co., Ltd.

Cho, E. Y., H. M. Kang, J. Lee, J. H. Kang, E. H. Choi & H. J. Lee (2012). "Changes in serotype distribution and antibiotic resistance of nasopharyngeal isolates of Streptococcus pneumoniae from children in Korea, after optional use of the 7-valent conjugate vaccine." *J Korean Med Sci* **27**(7), 716-722.

Christensen, H., M. May, L. Bowen, M. Hickman & C. L. Trotter (2010). "Meningococcal carriage by age: a systematic review and meta-analysis." *The Lancet Infectious Diseases* **10**(12), 853-861.

Clark, T. A., I. A. Murray, R. D. Morgan, A. O. Kislyuk, K. E. Spittle, M. Boitano, A. Fomenkov, R. J. Roberts & J. Korlach (2012). "Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing." *Nucleic Acids Res* **40**(4), e29.

Clarke, S. C., J. M. Jefferies, A. J. Smith, J. McMenamin, T. J. Mitchell & G. F. S. Edwards (2006). "Potential Impact of Conjugate Vaccine on the Incidence of Invasive Pneumococcal Disease among Children in Scotland." *J. Clin. Microbiol.* **44**(4), 1224-1228.

Coffey, T. J., C. G. Dowson, M. Daniels, J. Zhou, C. Martin, B. G. Spratt & J. M. Musser (1991). "Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of Streptococcus pneumoniae." *Molecular Microbiology* **5**(9), 2255-2260.

Coffey, T. J., M. C. Enright, M. Daniels, J. K. Morona, R. Morona, W. Hryniewicz, J. C. Paton & B. G. Spratt (1998). "Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of Streptococcus pneumoniae." *Molecular Microbiology* **27**(1), 73-83.

Cooper, D., X. Yu, M. Sidhu, M. H. Nahm, P. Fernsten & K. U. Jansen (2010). "Development of an opsonophagocytic assay to Streptococcus pneumoniae serotype 6C: Demonstration of cross-functional responses to 6C in Prevenar-13 immune sera." *ESPID* **2010** European Society for Paediatric Infectious Disease Meeting, Nice.

Cooper, D., X. Yu, M. Sidhu, M. H. Nahm, P. Fernsten & K. U. Jansen (2011). "The 13-valent pneumococcal conjugate vaccine (PCV13) elicits cross-functional opsonophagocytic killing responses in humans to Streptococcus pneumoniae serotypes 6C and 7A." *Vaccine* **29**(41), 7207-7211.

Cooper, G. M. & R. E. Hausman (2004). *The cell : a molecular approach*. Washington, D.C. Sunderland, Mass., ASM Press ; Sinauer Associates.

Croucher, N. J., J. A. Finkelstein, S. I. Pelton, P. K. Mitchell, G. M. Lee, J. Parkhill, S. D. Bentley, W. P. Hanage & M. Lipsitch (2013). "Population genomics of post-vaccine changes in pneumococcal epidemiology." *Nat Genet* **45**(6), 656-663.

Croucher, N. J., S. R. Harris, L. Barquist, J. Parkhill & S. D. Bentley (2012). "A high-resolution view of genome-wide pneumococcal transformation." *PLoS Pathog* **8**(6), e1002745.

Croucher, N. J., S. R. Harris, C. Fraser, M. A. Quail, J. Burton, M. van der Linden, L. McGee, A. von Gottberg, J. H. Song, K. S. Ko other authors (2011). "Rapid Pneumococcal Evolution in Response to Clinical Interventions." *Science* **331**(6016), 430-434.

Croucher, N. J., D. Walker, P. Romero, N. Lennard, G. K. Paterson, N. C. Bason, A. M. Mitchell, M. A. Quail, P. W. Andrew, J. Parkhill other authors (2009). "Role of Conjugative Elements in the Evolution of the Multidrug-Resistant Pandemic Clone Streptococcus pneumoniaeSpain23F ST81." *J. Bacteriol.* **191**(5), 1480-1489.

de Roux, A., B. Schmole-Thoma, G. R. Siber, J. G. Hackell, A. Kuhnke, N. Ahlers, S. A. Baker, A. Razmpour, E. A. Emini, P. D. Fernsten other authors (2008). "Comparison of pneumococcal conjugate polysaccharide and free polysaccharide vaccines in elderly adults: conjugate vaccine elicits improved

antibacterial immune responses and immunological memory." *Clin Infect Dis* **46**(7), 1015-1023.

De Schutter, I., A. Malfroot, D. Piérard & S. Lauwers (2006). "Pneumococcal serogroups and serotypes in severe pneumococcal pneumonia in Belgian children: Theoretical coverage of the 7-valent and 9-valent pneumococcal conjugate vaccines." *Pediatric Pulmonology* **41**(8), 765-770.

De Silva, S., G. Wood, T. Quek, C. Parrott & C. M. Bennett (2010). "Comparison of flocked and rayon swabs for detection of nasal carriage of Staphylococcus aureus among pathology staff members." *J Clin Microbiol* **48**(8), 2963-2964.

Department of Health (2006). Important changes to the childhood immunisation programme, PL/CMO/2006/1.

Department of Health (2010). "Introduction of Prevenar 13® into the Childhood Immunisation Programme. Department of Health, Gateway reference: 13581."

Department of Health (2011a). "Joint Committee on Vaccination and Immunisation. Advice on the pneumococcal vaccination programme for people aged 65 years and older. Department of Health, Gateway reference: 15777."

Department of Health (2011b). "Joint Committee on Vaccination and Immunisation. Advice on the pneumococcal vaccination programme for people aged 65 years and older. Department of Health, Gateway reference: 16380."

Department of Health (2011c). "Joint Committee on Vaccination and Immunisation. Statement on the routine pneumococcal vaccination programme for adults aged 65 years and older. 20 July 2011 JCVI minutes."

Department of Health (2012). "Chapter 25 Pneumococcal. Immunisation against infectious disease - 'The Green Book' - 2006 updated edition, 295-313, Gateway reference: 7523 ".

Dereeper, A., V. Guignon, G. Blanc, S. Audic, S. Buffet, F. Chevenet, J. F. Dufayard, S. Guindon, V. Lefort, M. Lescot other authors (2008). "Phylogeny.fr: robust phylogenetic analysis for the non-specialist." *Nucleic Acids Res* **36**(Web Server issue), W465-469.

Dias, C. A., L. M. Teixeira, G. Carvalho Mda & B. Beall (2007). "Sequential multiplex PCR for determining capsular serotypes of pneumococci recovered from Brazilian children." *J Med Microbiol* **56**(Pt 9), 1185-1188.

Didelot, X., R. Bowden, D. J. Wilson, T. E. Peto & D. W. Crook "Transforming clinical microbiology with bacterial genome sequencing." *Nat Rev Genet* **13**(9), 601-612.

Didelot, X., R. Bowden, D. J. Wilson, T. E. Peto & D. W. Crook (2012). "Transforming clinical microbiology with bacterial genome sequencing." *Nat Rev Genet* **13**(9), 601-612.

Ding, F., P. Tang, M.-H. Hsu, P. Cui, S. Hu, J. Yu & C.-H. Chiu (2009). "Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant Streptococcus pneumoniae serotype 14." *BMC Genomics* **10**(1), 158.

Dohm, J. C., C. Lottaz, T. Borodina & H. Himmelbauer (2008). "Substantial biases in ultra-short read data sets from high-throughput DNA sequencing." *Nucleic Acids Res* **36**(16), e105.

Donati, C., N. L. Hiller, H. Tettelin, A. Muzzi, N. Croucher, S. Angiuoli, M. Oggioni, J. Dunning Hotopp, F. Hu, D. Riley other authors (2010). "Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species." *Genome Biology* **11**(10), R107.

Dopazo, J. n., A. Mendoza, J. Herrero, F. Caldara, Y. Humbert, L. Friedli, M. Guerrier, E. Grand-Schenk, C. Gandin, M. de Francesco other authors (2001). "Annotated Draft Genomic Sequence from a Streptococcus pneumoniae Type 19F Clinical Isolate." *Microbial Drug Resistance* **7**(2), 99-125.

Douce, G., K. Ross, G. Cowan, J. Ma & T. J. Mitchell (2010). "Novel mucosal vaccines generated by genetic conjugation of heterologous proteins to

pneumolysin (PLY) from Streptococcus pneumoniae." *Vaccine* **28**(18), 3231-3237.

Droege, M. & B. Hill (2008). "The Genome Sequencer FLX System--longer reads, more applications, straight forward bioinformatics and more complete data sets." *J Biotechnol* **136**(1-2), 3-10.

Eastham, K. M., R. Freeman, A. M. Kearns, G. Eltringham, J. Clark, J. Leeming & D. A. Spencer (2004). "Clinical features, aetiology and outcome of empyema in children in the north east of England." *Thorax* **59**(6), 522-525.

Ehrlich, G. D., F. Z. Hu, K. Shen, P. Stoodley & J. C. Post (2005). "Bacterial plurality as a general mechanism driving persistence in chronic infections." *Clin Orthop Relat Res*(437), 20-24.

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

Elberse, K. E. M., I. van de Pol, S. Witteveen, H. G. J. van der Heide, C. S. Schot, A. van Dijk, A. van der Ende & L. M. Schouls (2011b). "Population Structure of Invasive *Streptococcus pneumoniae* in the Netherlands in the Pre-Vaccination Era Assessed by MLVA and Capsular Sequence Typing." *PLoS ONE* **6**(5), e20390.

Elemraid, M. A., A. D. Sails, G. J. Eltringham, J. D. Perry, S. P. Rushton, D. A. Spencer, M. F. Thomas, K. M. Eastham, F. Hampton, A. R. Gennery other authors (2013). "Aetiology of paediatric pneumonia after the introduction of pneumococcal conjugate vaccine." *Eur Respir J* **42**(6): 1595-1603.

EMBL-EBI (2012). "ClustalW multiple sequence alignment. http://www.ebi.ac.uk/Tools/msa/clustalw2/." **Accessed 01/12/12**.

Emlickov, A., H, K. Urb, P, M. Ad, V, J. Motlov & P. Lebedov, B (2006). "Characteristics of Streptococcus pneumoniae, Haemophilus influenzae, Moraxella catarrhalis and Staphylococcus aureus isolated from the nasopharynx of healthy children attending day-care centres in the Czech Republic." *Epidemiology and Infection* **134**(06), 1179-1187.

Enright, M. C. & B. G. Spratt (1998). "A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease." *Microbiology* **144**(11), 3049-3060.

Eskola, J., T. Kilpi, A. Palmu, J. Jokinen, J. Haapakoski, E. Herva, A. Takala, H. Kayhty, P. Karma, R. Kohberger other authors (2001). "Efficacy of a Pneumococcal Conjugate Vaccine against Acute Otitis Media." *N Engl J Med* **344**(6), 403-409.

European Bioinformatics Institute (2012). "Velvet assembler. http://www.ebi.ac.uk/~zerbino/velvet/." **Accessed 01/12/12**.

Everett, D. B., J. Cornick, B. Denis, C. Chewapreecha, N. Croucher, S. Harris, J. Parkhill, S. Gordon, E. D. Carrol, N. French other authors (2012). "Genetic characterisation of Malawian pneumococci prior to the roll-out of the PCV13 vaccine using a high-throughput whole genome sequencing approach." *PLoS One* **7**(9), e44250.

Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage & B. G. Spratt (2004). "eBURST: Inferring Patterns of Evolutionary Descent among Clusters of Related Bacterial Genotypes from Multilocus Sequence Typing Data." *J. Bacteriol.* **186**(5), 1518-1530.

Flasche, S., A. J. Van Hoek, E. Sheasby, P. Waight, N. Andrews, C. Sheppard, R. George & E. Miller (2011). "Effect of Pneumococcal Conjugate Vaccination on Serotype-Specific Carriage and Invasive Disease in England: A Cross-Sectional Study." *PLoS Med* **8**(4), e1001017.

Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick other

authors (1995). "Whole-genome random sequencing and assembly of Haemophilus influenzae Rd." *Science* **269**(5223), 496-512.

GAVI (2011). "Accelerated Vaccine Introduction - progress report. Report to the GAVI Alliance Board. Personal communication."

Gertz, R. E., Jr., Z. Li, F. C. Pimenta, D. Jackson, B. A. Juni, R. Lynfield, J. H. Jorgensen, G. Carvalho Mda, B. W. Beall & T. Active Bacterial Core Surveillance (2010). "Increased penicillin nonsusceptibility of nonvaccine-serotype invasive pneumococci other than serotypes 19A and 6A in post-7-valent conjugate vaccine era." *J Infect Dis* **201**(5), 770-775.

Gibson, L. F. & J. T. Khoury (1986). "Storage and survival of bacteria by ultra-freeze." *Letters in Applied Microbiology* **3**(6), 127-129.

Giglio, N., P. Micone & A. Gentile (2011). "The pharmacoeconomics of pneumococcal conjugate vaccines in Latin America." *Vaccine* **29 Suppl 3**, C35-42.

Gladstone, R. A., J. M. Jefferies, S. N. Faust & S. C. Clarke (2011). "Continued control of pneumococcal disease in the UK - the impact of vaccination." *J Med Microbiol* **60**(Pt 1), 1-8.

Gladstone, R. A., J. M. Jefferies, S. N. Faust & S. C. Clarke (2012a). "Pneumococcal 13-valent conjugate vaccine for the prevention of invasive pneumococcal disease in children and adults." *Expert Rev Vaccines* **11**(8), 889-902.

Gladstone, R. A., J. M. Jefferies, S. N. Faust & S. C. Clarke (2012b). "Sampling methods for the study of pneumococcal carriage: a systematic review." *Vaccine* **30**(48), 6738-6744.

Gladstone, R. A., A. S. Tocheva, G. Afimeke, J. Garland, J. M. Jefferies, S. F. Faust & S. C. Clarke (2010). *Polymicrobial colonisation of the nasopharynx in UK children*. International symposium for pneumococci and pneumococcal diseases, Tel Aviv.

Glennie, S. J., D. Banda, K. Gould, J. Hinds, A. Kamngona, D. D. Everett, N. A. Williams & R. S. Heyderman (2013). "Defective pneumococcal-specific Th1 responses in HIV-infected adults precedes a loss of control of pneumococcal colonization." *Clin Infect Dis* **56**(2), 291-299.

Golubchik, T., A. B. Brueggemann, T. Street, R. E. Gertz, Jr., C. C. Spencer, T. Ho, E. Giannoulatou, R. Link-Gelles, R. M. Harding, B. Beall other authors (2012). "Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event." *Nat Genet* **44**(3), 352-355.

Gouy, M., S. Guindon & O. Gascuel (2010). "SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building." *Mol Biol Evol* **27**(2), 221-224.

Grabenstein, J. D. & K. P. Klugman (2012). "A century of pneumococcal vaccination research in humans." *Clin Microbiol Infect* **18 Suppl 5**, 15-24.

Gray, B. (2002). "Egg-Based Media for Delayed Processing of Nasopharyngeal Swabs in Colonization Studies of Streptococcus pneumoniae." *European Journal of Clinical Microbiology & Infectious Diseases* **21**(9), 666-670.

Green, M. C., E. O. Mason, S. L. Kaplan, L. B. Lamberth, S. H. Stovall, L. B. Givner, J. S. Bradley, T. Q. Tan, W. J. Barson, J. A. Hoffman other authors (2011). "Increase in Prevalence of Streptococcus pneumoniae Serotype 6C at Eight Children's Hospitals in the United States from 1993 to 2009." *J. Clin. Microbiol.* **49**(6), 2097-2101.

Greenberg, D., N. Givon-Lavi, N. Newman, J. Bar-Ziv & R. Dagan (2011). "Nasopharyngeal carriage of individual Streptococcus pneumoniae serotypes during pediatric pneumonia as a means to estimate serotype disease potential." *Pediatr Infect Dis J* **30**(3), 227-233.

Gregson, A. L. & R. Edelman (2003). "Does antigenic overload exist? The role of multiple immunizations in infants." *Immunol Allergy Clin North Am* **23**(4), 649-664.

Griffith, F. (1928). "The Significance of Pneumococcal Types." *J Hyg (Lond)* **27**(2), 113-159.

Grijalva, C. G., J. P. Nuorti, P. G. Arbogast, S. W. Martin, K. M. Edwards & M. R. Griffin (2007). "Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis." *The Lancet* **369**(9568), 1179-1186.

Grimwood, K., P. Anderson, V. Anderson, L. Tan & T. Nolan (2000). "Twelve year outcomes following bacterial meningitis: further evidence for persisting effects." *Archives of Disease in Childhood* **83**(2), 111-116.

Haber, M., A. Barskey, W. Baughman, L. Barker, C. G. Whitney, K. M. Shaw, W. Orenstein & D. S. Stephens (2007). "Herd immunity and pneumococcal conjugate vaccine: A quantitative model." *Vaccine* **25**(29), 5390-5398.

Hamel, J., N. Charland, I. Pineau, C. Ouellet, S. Rioux, D. Martin & B. R. Brodeur (2004). "Prevention of Pneumococcal Disease in Mice Immunized with Conserved Surface-Accessible Proteins." *Infect. Immun.* **72**(5), 2659-2670.

Hanage, W. P. (2008). "Serotype-specific problems associated with pneumococcal conjugate vaccination." *Future Microbiology* **3**(1), 23-30.

Hanage, W. P., C. J. Bishop, S. S. Huang, A. E. Stevenson, S. I. Pelton, M. Lipsitch & J. A. Finkelstein (2011). "Carried pneumococci in Massachusetts children: the contribution of clonal expansion and serotype switching." *Pediatr Infect Dis J* **30**(4), 302-308.

Hanage, W. P., J. A. Finkelstein, S. S. Huang, S. I. Pelton, A. E. Stevenson, K. Kleinman, V. L. Hinrichsen & C. Fraser (2010). "Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete." *Epidemics* **2**(2), 80-84.

Hanage, W. P., T. Kaijalainen, E. Herva, A. Saukkoriipi, R. Syrjanen & B. G. Spratt (2005a). "Using multilocus sequence data to define the pneumococcus." *J Bacteriol* **187**(17), 6223-6230.

Hanage, W. P., T. H. Kaijalainen, R. K. Syrjanen, K. Auranen, M. Leinonen, P. H. Makela & B. G. Spratt (2005b). "Invasiveness of Serotypes and Clones of Streptococcus pneumoniae among Children in Finland." *Infect. Immun.* **73**(1), 431-435.

Hansman, D. & M. M. Bullen (1967). "A RESISTANT PNEUMOCOCCUS." *The Lancet* **290**(7509), 264-265.

Harboe, Z. B., T. L. Benfield, P. Valentiner-Branth, T. Hjuler, L. Lambertsen, M. Kaltoft, K. Krogfelt, H. C. Slotved, J. J. Christensen & H. B. Konradsen (2010). "Temporal trends in invasive pneumococcal disease and pneumococcal serotypes over 7 decades." *Clin Infect Dis* **50**(3), 329-337.

Harris, S. R., E. J. Cartwright, M. E. Torok, M. T. Holden, N. M. Brown, A. L. Ogilvy-Stuart, M. J. Ellington, M. A. Quail, S. D. Bentley, J. Parkhill other authors (2013). "Whole-genome sequencing for analysis of an outbreak of meticillin-resistant Staphylococcus aureus: a descriptive study." *Lancet Infect Dis* **13**(2), 130-136.

Hausdorff, W. P., J. Bryant, P. R. Paradiso & G. R. Siber (2000). "Which Pneumococcal Serogroups Cause the Most Invasive Disease: Implications for Conjugate Vaccine Formulation and Use, Part I." *Clinical Infectious Diseases* **30**(1), 100-121.

Hausdorff, W. P., R. Dagan, F. Beckers & L. Schuerman (2009). "Estimating the direct impact of new conjugate vaccines against invasive pneumococcal disease." *Vaccine* **27**(52), 7257-7269

Hausdorff, W. P., D. R. Feikin & K. P. Klugman (2005). "Epidemiological differences among pneumococcal serotypes." *The Lancet Infectious Diseases* **5**(2), 83-93.

Helicos Biosciences (2012). "Heliscope single molecule sequencer. http://www.helicosbio.com." **Accessed 01/12/12**.

Hendley, J. O., M. A. Sande, P. M. Stewart & J. M. G. Jr (1975). "Spread of Streptococcus pneumoniae in Families. I. Carriage Rates and Distribution of Types." *The Journal of Infectious Diseases* **132**(1), 55-61.

Hernandez-Bou, S., J. J. Garcia-Garcia, A. Gene, C. Esteva, E. del Amo & C. Munoz-Almagro (2012). "Pneumococcal carriage in children attending a hospital outpatient clinic in the era of pneumococcal conjugate vaccines in Barcelona." *Diagn Microbiol Infect Dis* **74**(3), 258-262.

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

Hiller, N., B. Janto, J. Hogg, R. Boissy, S. Yu, E. Powell, R. Keefe, N. Ehrlich, K. Shen & J. Hayes (2007). "Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome." *J Bacteriol* **189**(22), 8186 - 8195.

Hoberman, A., J. L. Paradise, H. E. Rockette, N. Shaikh, E. R. Wald, D. H. Kearney, D. K. Colborn, M. Kurs-Lasky, S. Bhatnagar, M. A. Haralam other authors (2011). "Treatment of acute otitis media in children under 2 years of age." *N Engl J Med* **364**(2), 105-115.

Hoge, C. W., M. R. Reichler, E. A. Dominguez, J. C. Bremer, T. D. Mastro, K. A. Hendricks, D. M. Musher, J. A. Elliott, R. R. Facklam & R. F. Breiman (1994). "An Epidemic of Pneumococcal Disease in an Overcrowded, Inadequately Ventilated Jail." *New England Journal of Medicine* **331**(10), 643-648.

Hogenesch, H., A. Dunham, B. Hansen, K. Anderson, J. F. Maisonneuve & S. L. Hem (2011). "Formulation of a killed whole cell pneumococcus vaccine - effect of aluminum adjuvants on the antibody and IL-17 response." *J Immune Based Ther Vaccines* **9**, 5.

Holden, M. T., L. Y. Hsu, K. Kurt, L. A. Weinert, A. E. Mather, S. R. Harris, B. Strommenger, F. Layer, W. Witte, H. de Lencastre other authors (2013). "A genomic portrait of the emergence, evolution, and global spread of a methicillin-resistant Staphylococcus aureus pandemic." *Genome Res* **23**(4), 653-664.

Hoskins, J., W. E. Alborn, Jr., J. Arnold, L. C. Blaszczak, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D. J. Fu, W. Fuller other authors (2001). "Genome of the bacterium Streptococcus pneumoniae strain R6." *J Bacteriol* **183**(19), 5709-5717.

Hsieh, Y. C., C. H. Chiu, K. Y. Chang, Y. C. Huang, C. J. Chen, C. Y. Kuo, P. Y. Chen, K. P. Hwang & T. Y. Lin (2012). "The impact of the heptavalent pneumococcal conjugate vaccine on risk factors for Streptococcus pneumoniae carriage in children." *Pediatr Infect Dis J* **31**(9), e163-168.

Hu, L., S. B. Joshi, M. R. Liyanage, M. Pansalawatta, M. R. Alderson, A. Tate, G. Robertson, J. Maisonneuve, D. B. Volkin & C. R. Middaugh (2013). "Physical characterization and formulation development of a recombinant pneumolysoid protein-based pneumococcal vaccine." *J Pharm Sci* **102**(2), 387-400.

Huss, A., PhD, P. Scott, MSc, A. E. Stuck, MD, C. Trotter, PhD & M. Egger, MD MSc (2009). "Efficacy of pneumococcal vaccination in adults: a meta-analysis." *CMAJ* **180**(1), 48-58.

Hussain, M., A. Melegaro, R. G. Pebody, R. George, W. J. Edmunds, R. Talukdar, S. A. Martin, A. Efstratiou & E. Miller (2005). "A longitudinal household study of Streptococcus pneumoniae nasopharyngeal carriage in a UK setting." *Epidemiol Infect* **133**(5), 891-898.

Hyman, E. D. (1988). "A new method of sequencing DNA." *Anal Biochem* **174**(2), 423-436.

Inouye, M., T. C. Conway, J. Zobel & K. E. Holt (2012). "Short read sequence typing (SRST): multi-locus sequence types from short reads." *BMC Genomics* **13**, 338.

J. Craig Venter Institute (2011). "Comprehensive microbial resource." http://cmr.tigr.org **Accessed 06/2011**.

International Human Genome Sequencing Consortium (2004). "Finishing the euchromatic sequence of the human genome." *Nature* **431**(7011), 931-945.

Inverarity D, Diggle M, Edwards G & T. Mitchell (2006). "An Evaluation of Media Suitable for the Transportation by Air of Streptococcus pneumoniae." *Journal of infection* **55,** e65-e66.

Jackson, L. A., A. Gurtman, K. Rice, K. Pauksens, R. N. Greenberg, T. R. Jones, D. A. Scott, E. A. Emini, W. C. Gruber & B. Schmoele-Thoma (2013). "Immunogenicity and safety of a 13-valent pneumococcal conjugate vaccine in adults 70 years of age and older previously vaccinated with 23-valent pneumococcal polysaccharide vaccine." *Vaccine* **31**(35), 3585-3593.

Jacobs, M. R., C. E. Good, S. Bajaksouzian & A. R. Windau (2008). "Emergence of Streptococcus pneumoniae serotypes 19A, 6C, and 22F and serogroup 15 in Cleveland, Ohio, in relation to introduction of the protein-conjugated pneumococcal vaccine." *Clin Infect Dis* **47**(11), 1388-1395.

Jacups, S. P. & A. Cheng (2011). "The epidemiology of community acquired bacteremic pneumonia, due to Streptococcus pneumoniae, in the Top End of the Northern Territory, Australia--Over 22 years." *Vaccine* **29**(33): 5386-5392.

Jefferies, J., L. Nieminen, L.-A. Kirkham, C. Johnston, A. Smith & T. J. Mitchell (2007). "Identification of a Secreted Cholesterol-Dependent Cytolysin (Mitilysin) from Streptococcus mitis." *J. Bacteriol.* **189**(2), 627-632.

Jefferies, J. M., A. J. Smith, G. F. S. Edwards, J. McMenamin, T. J. Mitchell & S. C. Clarke (2010a). "Temporal analysis of invasive pneumococcal clones from Scotland illustrates fluctuations in diversity of serotype and genotype in the absence of pneumococcal conjugate vaccine." *J. Clin. Microbiol.* **48**(1), 87-96.

Jefferies, J. M., A. S. Tocheva, H. Rubery, J. Bennett, J. Garland, M. Christodoulides, S. Faust, A. J. Smith, T. J. Mitchell & S. C. Clarke (2010b). "Identification of novel pneumolysin (Ply) alleles from paediatric carriage isolates of Streptococcus pneumoniae." *J Med Microbiol* **59**(Pt 7): 808-814.

Jefferies, J. M. C., A. Smith, S. C. Clarke, C. Dowson & T. J. Mitchell (2004). "Genetic Analysis of Diverse Disease-Causing Pneumococci Indicates High Levels of Diversity within Serotypes and Capsule Switching." *J. Clin. Microbiol.* **42**(12), 5681-5688.

Jin, P., F. Kong, M. Xiao, S. Oftadeh, F. Zhou, C. Liu, F. Russell & Gwendolyn L. Gilbert (2009). "First Report of Putative Streptococcus pneumoniae Serotype 6D among Nasopharyngeal Isolates from Fijian Children." *The Journal of Infectious Diseases* **200**(9), 1375-1380.

Johnson, A. P., P. Waight, N. Andrews, R. Pebody, R. C. George & E. Miller (2007). "Morbidity and mortality of pneumococcal meningitis and serotypes of causative strains prior to introduction of the 7-valent conjugant pneumococcal vaccine in England." *Journal of Infection* **55**(5), 394-399.

Johnson, H. L., M. Deloria-Knoll, O. S. Levine, S. K. Stoszek, L. Freimanis Hance, R. Reithinger, L. R. Muenz & K. L. O'Brien (2010). "Systematic Evaluation of Serotypes Causing Invasive Pneumococcal Disease among Children Under Five: The Pneumococcal Global Serotype Project." *PLoS Med* **7**(10), e1000348.

Johnston, C., J. Hinds, A. Smith, M. van der Linden, J. Van Eldere & T. J. Mitchell (2010). "Detection of Large Numbers of Pneumococcal Virulence Genes in Streptococci of the Mitis Group." *J. Clin. Microbiol.* **48**(8), 2762-2769.

Jolley, K. A. & M. C. Maiden (2010). "BIGSdb: Scalable analysis of bacterial genome variation at the population level." *BMC Bioinformatics* **11**, 595.

Jones, R. N., M. R. Jacobs & H. S. Sader (2010). "Evolving trends in Streptococcus pneumoniae resistance: implications for therapy of community-acquired bacterial pneumonia." *International Journal of Antimicrobial Agents* **36**(3), 197-204.

Kaltoft, M. S., U. B. Skov Sørensen, H.-C. Slotved & H. B. Konradsen (2008). "An easy method for detection of nasopharyngeal carriage of multiple Streptococcus pneumoniae serotypes." *Journal of Microbiological Methods* **75**(3), 540-544.

Kamtchoua, T., M. Bologa, R. Hopfer, D. Neveu, B. Hu, X. Sheng, N. Corde, C. Pouzet, G. Zimmermann & S. Gurunathan (2013). "Safety and immunogenicity of the pneumococcal pneumolysin derivative PlyD1 in a single-antigen protein vaccine candidate in adults." *Vaccine* **31**(2), 327-333.

Kastenbauer, S. & H. W. Pfister (2003). "Pneumococcal meningitis in adults." *Brain* **126**(5), 1015-1025.

Kedes, L. & G. Campany (2011). "The new date, new format, new goals and new sponsor of the Archon Genomics X PRIZE competition." *Nat Genet* **43**(11), 1055-1058.

Kieninger D. M, Kueper K, Steul K, Juergens C, Ahlers N, Baker S, Giardina P, Gruber W & Scott D (2008). Safety and immunologic non-inferiority of 13-valent pneumococcal conjugate vaccine compared to 7-valent pneumococcal conjugate vaccine given as a 4-dose series with routine vaccines in healthy infants and toddlers, Presented at the 48th Annual ICAAC/IDSA 46th Annual Meeting. Washington DC, October 25-28, 2008

Kim, S. Y., G. Lee & S. J. Goldie (2010). "Economic evaluation of pneumococcal conjugate vaccination in The Gambia." *BMC Infect Dis* **10**, 260.

Kirkham, L. S., J. M. C. Jefferies, A. R. Kerr, Y. Jing, S. C. Clarke, A. Smith & T. J. Mitchell (2006). "Identification of Invasive Serotype 1 Pneumococcal Isolates That Express Nonhemolytic Pneumolysin." *J. Clin. Microbiol.* **44**(1), 151-159.

Kononen, E., H. Jousimies-Somer, A. Bryk, T. Kilpi & M. Kilian (2002). "Establishment of streptococci in the upper respiratory tract: longitudinal changes in the mouth and nasopharynx up to 2 years of age." *J Med Microbiol* **51**(9), 723-730.

Koser, C. U., J. M. Bryant, J. Becq, M. E. Torok, M. J. Ellington, M. A. Marti-Renom, A. J. Carmichael, J. Parkhill, G. P. Smith & S. J. Peacock (2013). "Whole-genome sequencing for rapid susceptibility testing of M. tuberculosis." *N Engl J Med* **369**(3), 290-292.

Koser, C. U., M. J. Ellington, E. J. Cartwright, S. H. Gillespie, N. M. Brown, M. Farrington, M. T. Holden, G. Dougan, S. D. Bentley, J. Parkhill other authors (2012a). "Routine use of microbial whole genome sequencing in diagnostic and public health microbiology." *PLoS Pathog* **8**(8), e1002824.

Koser, C. U., M. T. Holden, M. J. Ellington, E. J. Cartwright, N. M. Brown, A. L. Ogilvy-Stuart, L. Y. Hsu, C. Chewapreecha, N. J. Croucher, S. R. Harris other authors (2012b). "Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak." *N Engl J Med* **366**(24), 2267-2275.

Koshy, E., J. Murray, A. Bottle, M. Sharland & S. Saxena (2010). "Impact of the seven-valent pneumococcal conjugate vaccination (PCV7) programme on childhood hospital admissions for bacterial pneumonia and empyema in England: national time-trends study, 1997-2008." *Thorax* **65**(9), 770-774.

Kronvall, G. (1973). "A rapid slide-agglutination method for typing pneumococci by means of specific antibody adsorbed to protein A-containing staphylococci." *J Med Microbiol* **6**(2), 187-190.

Kyaw, M. H., P. Christie, S. C. Clarke, J. D. Mooney, S. Ahmed, I. G. Jones & H. Campbell (2003). "Invasive pneumococcal disease in Scotland, 1999-2001:

use of record linkage to explore associations between patients and disease in relation to future vaccination policy." *Clin Infect Dis* **37**(10), 1283-1291.

Kyte, J. & R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein." *J Mol Biol* **157**(1), 105-132.

Ladhani, S., M. E. Ramsay, M. Chandra & M. P. Slack (2008). "No evidence for Haemophilus influenzae serotype replacement in Europe after introduction of the Hib conjugate vaccine." *The Lancet Infectious Diseases* **8**(5), 275-276.

Ladhani, S. N., M. P. Slack, N. J. Andrews, P. A. Waight, R. Borrow & E. Miller (2013). "Invasive pneumococcal disease after routine pneumococcal conjugate vaccination in children, England and Wales." *Emerg Infect Dis* **19**(1), 61-68.

Lafong, A. C. & E. Crothers (1988). "Simple latex agglutination method for typing pneumococci." *J Clin Pathol* **41**(2), 230-231.

Lahti, E., J. Mertsola, T. Kontiokari, E. Eerola, O. Ruuskanen & J. Jalava (2006). "Pneumolysin polymerase chain reaction for diagnosis of pneumococcal pneumonia and empyema in children." *European Journal of Clinical Microbiology & Infectious Diseases* **25**(12), 783-789.

Lambertsen, L., M. Brendstrup, H. Friis & J. J. Christensen (2010). "Molecular characterization of invasive penicillin non-susceptible Streptococcus pneumoniae from Denmark, 2001 to 2005." *Scand J Infect Dis* **42**(5), 333-340.

Lammers, A. J., A. P. de Porto, S. Florquin, O. J. de Boer, H. J. Bootsma, P. W. Hermans & T. van der Poll (2011). "Enhanced vulnerability for Streptococcus pneumoniae sepsis during asplenia is determined by the bacterial capsule." *Immunobiology* **216**(8): 863-870.

Lanie, J. A., W.-L. Ng, K. M. Kazmierczak, T. M. Andrzejewski, T. M. Davidsen, K. J. Wayne, H. Tettelin, J. I. Glass & M. E. Winkler (2007). "Genome Sequence of Avery's Virulent Serotype 2 Strain D39 of Streptococcus pneumoniae and Comparison with That of Unencapsulated Laboratory Strain R6." *J. Bacteriol.* **189**(1), 38-51.

Larsen, M. V., S. Cosentino, S. Rasmussen, C. Friis, H. Hasman, R. L. Marvig, L. Jelsbak, T. Sicheritz-Ponten, D. W. Ussery, F. M. Aarestrup other authors (2012). "Multilocus sequence typing of total-genome-sequenced bacteria." *J Clin Microbiol* **50**(4), 1355-1361.

Leach, A. J., E. Stubbs, K. Hare, J. Beissbarth & P. S. Morris (2008). "Comparison of nasal swabs with nose blowing for community-based pneumococcal surveillance of healthy children." *J Clin Microbiol* **46**(6), 2081-2082.

Lebon, A., N. J. Verkaik, C. P. de Vogel, H. Hooijkaas, H. A. Verbrugh, W. J. B. van Wamel, V. W. V. Jaddoe, A. Hofman, P. W. M. Hermans, T. J. Mitchell other authors (2011). "The Inverse Correlation between Staphylococcus aureus and Streptococcus pneumoniae Colonization in Infants Is Not Explained by Differences in Serum Antibody Levels in the Generation R Study." *Clin. Vaccine Immunol.* **18**(1), 180-183.

Lee, C.-C., N. A. Middaugh, S. R. C. Howie & M. Ezzati (2010). "Association of Secondhand Smoke Exposure with Pediatric Invasive Bacterial Disease and Bacterial Carriage: A Systematic Review and Meta-analysis." *PLoS Med* **7**(12), e1000374.

Lee, H., M. H. Nahm, R. Burton & K. Kim (2009). "Immune Response in Infants to the Heptavalent Pneumococcal Conjugate Vaccine against Vaccine-Related Serotypes 6A and 19A." *Clin. Vaccine Immunol.* **16**(3), 376-381.

Levene, M. J., J. Korlach, S. W. Turner, M. Foquet, H. G. Craighead & W. W. Webb (2003). "Zero-mode waveguides for single-molecule analysis at high concentrations." *Science* **299**(5607), 682-686.

Lewis, T., N. J. Loman, L. Bingle, P. Jumaa, G. M. Weinstock, D. Mortiboy & M. J. Pallen (2010). "High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak." *Journal of Hospital Infection* **75**(1), 37-41.

Lijek, R. S., S. L. Luque, Q. Liu, D. Parker, T. Bae & J. N. Weiser (2012). "Protection from the acquisition of Staphylococcus aureus nasal carriage by cross-reactive antibody to a pneumococcal dehydrogenase." *Proc Natl Acad Sci U S A* **109**(34), 13823-13828.

Lipsitch, M. (1997). "Vaccination against colonizing bacteria with multiple serotypes." *Proc Natl Acad Sci U S A* **94**(12), 6571-6576.

Ljutic, B., M. Ochs, B. Messham, M. Ming, A. Dookie, K. Harper & S. F. Ausar (2012). "Formulation, stability and immunogenicity of a trivalent pneumococcal protein vaccine formulated with aluminum salt adjuvants." *Vaccine* **30**(19), 2981-2988.

Lloyd-Evans, N., T. J. O'Dempsey, I. Baldeh, O. Secka, E. Demba, J. E. Todd, T. F. McArdle, W. S. Banya & B. M. Greenwood (1996). "Nasopharyngeal carriage of pneumococci in Gambian children and in their families." *Pediatr Infect Dis J* **15**(10), 866-871.

Loman NJ, Gladstone RA, Constantinidou C, Tocheva AS & J. JMC (2013). "Clonal Expansion within Pneumococcal Serotype 6C after Use of Seven-Valent Vaccine." *PLoS ONE* **8**(5), e64731.

Loman, N. J., C. Constantinidou, J. Z. Chan, M. Halachev, M. Sergeant, C. W. Penn, E. R. Robinson & M. J. Pallen (2012). "High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity." *Nat Rev Microbiol* **10**(9), 599-606.

Lu, Y.-J., L. Leite, V. M. Gonçalves, W. d. O. Dias, C. Liberman, F. Fratelli, M. Alderson, A. Tate, J.-F. Maisonneuve, G. Robertson other authors (2011). "GMP-grade pneumococcal whole-cell vaccine injected subcutaneously protects mice from nasopharyngeal colonization and fatal aspiration-sepsis." *Vaccine* **28**(47), 7468-7475.

Lucero, M. G., V. E. Dulalia, L. T. Nillos, G. Williams, R. A. Parreno, H. Nohynek, I. D. Riley & H. Makela (2009). "Pneumococcal conjugate vaccines for preventing vaccine-type invasive pneumococcal disease and X-ray defined pneumonia in children less than two years of age." *Cochrane Database Syst Rev*(4), CD004977.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant other authors (1998). "Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms." *Proc Natl Acad Sci U S A* **95**(6), 3140-3145.

Maiden, M. C., M. J. van Rensburg, J. E. Bray, S. G. Earle, S. A. Ford, K. A. Jolley & N. D. McCarthy (2013). "MLST revisited: the gene-by-gene approach to bacterial genomics." *Nat Rev Microbiol* **11**(10), 728-736.

Malley, R., M. Lipsitch, A. Stack, R. Saladino, G. Fleisher, S. Pelton, C. Thompson, D. Briles & P. Anderson (2001). "Intranasal Immunization with Killed Unencapsulated Whole Cells Prevents Colonization and Invasive Disease by Capsulated Pneumococci." *Infect. Immun.* **69**(8), 4870-4873.

Mardis, E. R. (2008). "The impact of next-generation sequencing technology on genetics." *Trends Genet* **24**(3), 133-141.

Mardis, E. R. (2010). "The $1,000 genome, the $100,000 analysis?" *Genome Med* **2**(11), 84.

Margolis, E., A. Yates & B. R. Levin (2010). "The ecology of nasal colonization of Streptococcus pneumoniae, Haemophilus influenzae and Staphylococcus aureus: the role of competition and interactions with host's immune response." *BMC Microbiol* **10**, 59.

Mavroidi, A., Godoy, D., Aanensen, D. M., Robinson, D. A., Hollingshead, S. K., Spratt, B. G. (2004). "Evolutionary genetics of the capsular locus of serogroup 6 pneumococci." J Bacteriol **186**(24): 8181-8192.

Mavroidi, A., D. M. Aanensen, D. Godoy, I. C. Skovsted, M. S. Kaltoft, P. R. Reeves, S. D. Bentley & B. G. Spratt (2007). "Genetic relatedness of the Streptococcus pneumoniae capsular biosynthetic loci." *J Bacteriol* **189**(21), 7841-7855.

Maxam, A. M. & W. Gilbert (1977). "A new method for sequencing DNA." *Proc Natl Acad Sci U S A* **74**(2), 560-564.

McChlery, S. M., K. J. Scott & S. C. Clarke (2005). "Clonal analysis of invasive pneumococcal isolates in Scotland and coverage of serotypes by the licensed conjugate polysaccharide pneumococcal vaccine: possible implications for UK vaccine policy." *European Journal of Clinical Microbiology & Infectious Diseases* **24**(4), 262-267.

McEllistrem, M. C., J. Adams, Edward O. Mason & Ellen R. Wald (2003). "Epidemiology of Acute Otitis Media Caused by Streptococcus pneumoniae Before and After Licensure of the 7-Valent Pneumococcal Protein Conjugate Vaccine." *The Journal of Infectious Diseases* **188**(11), 1679-1684.

McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefevre, A. Tomasz other authors (2001). "Nomenclature of Major Antimicrobial-Resistant Clones of Streptococcus pneumoniae Defined by the Pneumococcal Molecular Epidemiology Network." *J. Clin. Microbiol.* **39**(7), 2565-2571.

Melegaro, A. & W. J. Edmunds (2004a). "The 23-valent pneumococcal polysaccharide vaccine. Part II. A cost-effectiveness analysis for invasive disease in the elderly in England and Wales." *Eur J Epidemiol* **19**(4), 365-375.

Melegaro, A. & W. J. Edmunds (2004b). "Cost-effectiveness analysis of pneumococcal conjugate vaccination in England and Wales." *Vaccine* **22**(31-32), 4203-4214.

Melegaro, A., W. J. Edmunds, R. Pebody, E. Miller & R. George (2006). "The current burden of pneumococcal disease in England and Wales." *Journal of Infection* **52**(1), 37-48.

Melin, M., K. Trzcinski, M. Antonio, S. Meri, R. Adegbola, T. Kaijalainen, H. Kayhty & M. Vakevainen (2010a). "Serotype-Related Variation in Susceptibility to Complement Deposition and Opsonophagocytosis among Clinical Isolates of Streptococcus pneumoniae." *Infect. Immun.* **78**(12), 5252-5261.

Melin, M., K. Trzcinski, S. Meri, H. Kayhty & M. Vakevainen (2010b). "The Capsular Serotype of Streptococcus pneumoniae Is More Important than the Genetic Background for Resistance to Complement." *Infect. Immun.* **78**(12), 5262-5270.

Meningitis Research Foundation (2010). "Pneumococcal meningitis http://www.meningitis.org/facts." *MRF* **Accessed 28/08/13**.

Menzies, R. I., P. Markey, R. Boyd, A. P. Koehler & P. B. McIntyre (2013). "No evidence of increasing Haemophilus influenzae non-b infection in Australian Aboriginal children." *Int J Circumpolar Health* **72**.

Metzker, M. L. (2009). "Sequencing technologies - the next generation." *Nat Rev Genet* **11**(1), 31-46.

Millar, E. V., K. L. O'Brien, E. R. Zell, M. A. Bronsdon, R. Reid & M. Santosham (2009). "Nasopharyngeal Carriage of Streptococcus pneumoniae in Navajo and White Mountain Apache Children Before the Introduction of Pneumococcal Conjugate Vaccine." *Pediatric Infectious Disease Journal* **28**(8), 711-716.

Miller, E., N. J. Andrews, P. A. Waight, M. P. Slack & R. C. George (2011a). "Effectiveness of the new serotypes in the 13-valent pneumococcal conjugate vaccine." *Vaccine* **29**(49), 9127-9131.

Miller, E., N. J. Andrews, P. A. Waight, M. P. Slack & R. C. George (2011b). "Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study." *Lancet Infect Dis* **11**(10), 760-768.

MLST.net (2012). "Multi-locus sequence typing. http://www.MLST.net." **Accessed 01/12/12**.

Monasta, L., L. Ronfani, F. Marchetti, M. Montico, L. Vecchi Brumatti, A. Bavcar, D. Grasso, C. Barbiero & G. Tamburlini (2012). "Burden of disease caused by otitis media: systematic review and global estimates." *PLoS One* **7**(4), e36226.

Moore, G. E. (1965). "Cramming More Components onto Integrated Circuits." *Electronics* **38**(8), 114-117.

Morona, J. K., R. Morona & J. C. Paton (1999). "Comparative genetics of capsular polysaccharide biosynthesis in Streptococcus pneumoniae types belonging to serogroup 19." *J Bacteriol* **181**(17), 5355-5364.

Morsczeck, C., T. Prokhorova, J. Sigh, M. Pfeiffer, M. Bille-Nielsen, J. Petersen, A. Boysen, T. Kofoed, N. Frimodt-Møller, P. Nyborg-Nielsen other authors (2008). "Streptococcus pneumoniae: proteomics of surface proteins for vaccine development." *Clinical Microbiology and Infection* **14**(1), 74-81.

Mullick, J., S. S. Cherian, V. A. Potdar, M. S. Chadha & A. C. Mishra (2011). "Evolutionary dynamics of the influenza A pandemic (H1N1) 2009 virus with emphasis on Indian isolates: Evidence for adaptive evolution in the HA gene." *Infection, Genetics and Evolution* **In Press, Corrected Proof**.

Murphy, T. F., L. O. Bakaletz & P. R. Smeesters (2009). "Microbial Interactions in the Respiratory Tract." *The Pediatric Infectious Disease Journal* **28**(10), S121-S126 110.1097/INF.1090b1013e3181b1096d1097ec.

Musher, D. M., A. M. Rueda, M. H. Nahm, E. A. Graviss & M. C. Rodriguez-Barradas (2008). "Initial and subsequent response to pneumococcal polysaccharide and protein-conjugate vaccines administered sequentially to adults who have recovered from pneumococcal pneumonia." *J Infect Dis* **198**(7), 1019-1027.

Mutreja, A., D. W. Kim, N. R. Thomson, T. R. Connor, J. H. Lee, S. Kariuki, N. J. Croucher, S. Y. Choi, S. R. Harris, M. Lebens other authors (2011). "Evidence for several waves of global transmission in the seventh cholera pandemic." *Nature* **477**(7365), 462-465.

Nahm, Moon H., J. Lin, Jonathan A. Finkelstein & Stephen I. Pelton (2009). "Increase in the Prevalence of the Newly Discovered Pneumococcal Serotype 6C in the Nasopharynx after Introduction of Pneumococcal Conjugate Vaccine." *The Journal of Infectious Diseases* **199**(3), 320-325.

National Centre for Biotechnology Information (2013). "Standard nucleotide BLAST. http://blast.ncbi.nlm.nih.gov/Blast.cgi." **Acessed 01/01/13**.

NCBI (2011). "Genome information by genome sequence." http://www.ncbi.nlm.nih.gov/genome/ **Accessed 06/2010**.

Neufeld, F. (1902). "Ueber die Agglutination der Pneumokokken und uber die Theorien der Agglutination." *Zeitschrift fur Hygiene Infektionskrankheiten* **34**, pp. 54–72.

Newton, R., J. Hinds & L. Wernisch (2011). "Empirical Bayesian models for analysing molecular serotyping microarrays." *BMC Bioinformatics* **12**, 88.

Nyren, P. & A. Lundin (1985). "Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis." *Anal Biochem* **151**(2), 504-509.

O'Brien, K., L. , L. Wolfson, J. , J. Watt, P., E. Henkle, M. Deloria-Knoll, N. McCall, E. Lee, K. Mulhollland, C. Charles, S. L. Orin other authors (2009). "Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates." *The Lancet* **374**(9693), 893-902.

O'Brien, K. L., M. A. Bronsdon, R. Dagan, P. Yagupsky, J. Janco, J. Elliott, C. G. Whitney, Y.-H. Yang, L.-G. E. Robinson, B. Schwartz other authors (2001). "Evaluation of a Medium (STGG) for Transport and Optimal Recovery of Streptococcus pneumoniae from Nasopharyngeal Secretions Collected during Field Studies." *J. Clin. Microbiol.* **39**(3), 1021-1024.

O'Brien, K. L. M. M., H. M. Nohynek & G. The Who Pneumococcal Vaccine Trials Carriage Working (2003). "Report from a WHO Working Group: standard method for detecting upper respiratory carriage of Streptococcus pneumoniae." *Pediatric Infectious Disease Journal* **22**(2), e1-e11.

Ogunniyi, A. D., M. Grabowicz, D. E. Briles, J. Cook & J. C. Paton (2007). "Development of a Vaccine against Invasive Pneumococcal Disease Based on Combinations of Virulence Proteins of Streptococcus pneumoniae." *Infect. Immun.* **75**(1), 350-357.

Okoniewski, M. J., J. Meienberg, A. Patrignani, A. Szabelska, G. Matyas & R. Schlapbach (2013). "Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers." *Biotechniques* **54**(2), 98-100.

Olafsdottir, T. A., K. Lingnau, E. Nagy & I. Jonsdottir (2011). "Novel Protein-Based Pneumococcal Vaccines Administered with the Th1-Promoting Adjuvant IC31 Induce Protective Immunity against Pneumococcal Disease in Neonatal Mice." *Infect Immun* **80**(1), 461-468.

Oliver, M. B., C. Jones, T. R. Larson, J. J. Calix, E. R. Zartler, J. Yother & M. H. Nahm (2013a). "Streptococcus pneumoniae serotype 11D has a bi-specific glycosyltransferase and expresses two different capsular polysaccharide repeating units." *J Biol Chem*.

Oliver, M. B., M. P. van der Linden, S. A. Kuntzel, J. S. Saad & M. H. Nahm (2013b). "Discovery of Streptococcus pneumoniae serotype 6 variants with glycosyltransferases synthesizing two differing repeating units." *J Biol Chem*.

Otsuka, T., B. Chang, A. Wada & M. Okazaki (2013). "Molecular epidemiology and serogroup 6 capsular gene evolution of pneumococcal carriage in a Japanese birth cohort study." *J Med Microbiol*.

Otto, T. D., M. Sanders, M. Berriman & C. Newbold (2010). "Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology." *Bioinformatics* **26**(14), 1704-1707.

Oxford Nanopore Technologies (2012). "Oxford Nanopore Tehcnologies. http://www.nanoporetech.com/." **Accessed 01/12/12**.

Pai, R., R. E. Gertz & B. Beall (2006). "Sequential Multiplex PCR Approach for Determining Capsular Serotypes of Streptococcus pneumoniae Isolates." *J. Clin. Microbiol.* **44**(1), 124-131.

Pai, R., M. R. Moore, T. Pilishvili, R. E. Gertz, C. G. Whitney, B. Beall & T. Active Bacterial Core Surveillance (2005). "Postvaccine genetic structure of Streptococcus pneumoniae serotype 19A from children in the United States." *J Infect Dis* **192**(11), 1988-1995.

Park, I. H., S. Park, S. K. Hollingshead & M. H. Nahm (2007a). "Genetic basis for the new pneumococcal serotype, 6C." *Infect Immun* **75**(9), 4482-4489.

Park, I. H., D. G. Pritchard, R. Cartee, A. Brandao, M. C. C. Brandileone & M. H. Nahm (2007b). "Discovery of a New Capsular Serotype (6C) within Serogroup 6 of Streptococcus pneumoniae." *J. Clin. Microbiol.* **45**(4), 1225-1233.

Parra, E. L., F. De La Hoz, P. L. Diaz, O. Sanabria, M. E. Realpe & J. Moreno (2013). "Changes in Streptococcus pneumoniae serotype distribution in invasive disease and nasopharyngeal carriage after the heptavalent pneumococcal conjugate vaccine introduction in Bogota, Colombia." *Vaccine* **31**(37), 4033-4038.

Pasteur, Chamberland & Roux (1881). "Sur une maladie nouvelle provoquee par la salive d'un enfant mort de la rage." *Compt. rend* **92** (4), 159-165.

Pebody, R. G., O. Morgan, Y. Choi, R. George, M. Hussain & N. Andrews (2009). "Use of antibiotics and risk factors for carriage of Streptococcus pneumoniae: a longitudinal household study in the United Kingdom." *Epidemiol Infect* **137**(4), 555-561.

Pettigrew, M. M., J. F. Gent, K. Revai, J. A. Patel & T. Chonmaitree (2008). "Microbial interactions during upper respiratory tract infections." *Emerg Infect Dis* **14**(10), 1584-1591.

Pichon, B., S. N. Ladhani, M. P. Slack, A. Segonds-Pichon, N. J. Andrews, P. A. Waight, E. Miller & R. George (2013). "Changes in molecular epidemiology of streptococcus pneumoniae causing meningitis following introduction of pneumococcal conjugate vaccination in England and Wales." *J Clin Microbiol* **51**(3), 820-827.

Pickering, J. W., T. B. Martins, R. W. Greer, M. C. Schroder, M. E. Astill, C. M. Litwin, S. W. Hildreth & H. R. Hill (2002). "A multiplexed fluorescent microsphere immunoassay for antibodies to pneumococcal capsular polysaccharides." *Am J Clin Pathol* **117**(4), 589-596.

Pikis, A., J. M. Campos, W. J. Rodriguez & J. M. Keith (2001). "Optochin Resistance in Streptococcus pneumoniae: Mechanism, Significance, and Clinical Implications." *Journal of infectious diseases* **184**(5), 582-590.

Pilishvili, T., C. Lexau, Monica M. Farley, J. Hadler, Lee H. Harrison, Nancy M. Bennett, A. Reingold, A. Thomas, W. Schaffner, Allen S. Craig other authors (2010). "Sustained Reductions in Invasive Pneumococcal Disease in the Era of Conjugate Vaccine." *The Journal of Infectious Diseases* **201**(1), 32-41.

Pillai, D., D. Shahinas, A. Buzina, R. Pollock, R. Lau, K. Khairnar, A. Wong, D. Farrell, K. Green, A. McGeer other authors (2009). "Genome-wide dissection of globally emergent multi-drug resistant serotype 19A Streptococcus pneumoniae." *BMC Genomics* **10**(1), 642.

Pletz, M. W., U. Maus, N. Krug, T. Welte & H. Lode (2008). "Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species." *International Journal of Antimicrobial Agents* **32**(3), 199-206.

Poddar, S. K., M. H. Sawyer & J. D. Connor (1998). "Effect of inhibitors in clinical specimens on Taq and Tth DNA polymerase-based PCR amplification of influenza A virus." *J Med Microbiol* **47**(12), 1131-1135.

Pollock, M. R. (1948). "Unsaturated fatty acids in cotton wool plugs." *Nature* **161**(4100), 853.

Pop, M. & S. L. Salzberg (2008). "Bioinformatics challenges of new sequencing technology." *Trends in Genetics* **24**(3), 142-149.

Prymula, R., P. Kriz, E. Kaliskova, T. Pascal, J. Poolman & L. Schuerman (2009). "Effect of vaccination with pneumococcal capsular polysaccharides conjugated to Haemophilus influenzae-derived protein D on nasopharyngeal carriage of Streptococcus pneumoniae and H. influenzae in children under 2 years of age." *Vaccine* **28**(1), 71-78.

Prymula, R., P. Peeters, V. Chrobok, P. Kriz, E. Novakova, E. Kaliskova, I. Kohl, P. Lommel, J. Poolman, J.-P. Prieels other authors (2006). "Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both Streptococcus pneumoniae and non-typable Haemophilus influenzae: a randomised double-blind efficacy study." *The Lancet* **367**(9512), 740-748.

Public Health England (2009). "Invasive Pneumococcal Disease (IPD) in England & Wales after 7 Valent Conjugate Vaccine (PCV7). Potential impact of 10 and 13 valent vaccines. http://www.hpa.org.uk/webc/HPAwebFile/HPAweb_C/1245581527892." **Accessed 01/06/11**.

Public Health England (2012). "Current Epidemiology of Invasive Pneumococcal Disease (IPD). http://www.hpa.org.uk." **Accessed 01/01/12**.

Qiagen (2001). "QIAGEN Genomic DNA Handbook." *Qiagen Handbook* **August 2001**.

Qiagen (2010). "QIAamp DNA Mini and Blood Mini Handbook." *Qiagen Handbook* **Third edition**.

Rajam, G., D. Jackson, T. Pilishvili, C. G. Whitney, R. R. Facklam, G. M. Carlone & S. Romero-Steiner (2007). "An in vitro model to assess pneumococcal adherence to nasopharyngeal cells under competition conditions." *Journal of Microbiological Methods* **70**(2), 219-226.

Rapola, S., E. Salo, P. Kiiski, M. Leinonen & A. K. Takala (1997). "Comparison of four different sampling methods for detecting pharyngeal carriage of Streptococcus pneumoniae and Haemophilus influenzae in children." *J Clin Microbiol* **35**(5), 1077-1079.

Reuter, S., M. J. Ellington, E. J. Cartwright, C. U. Koser, M. E. Torok, T. Gouliouris, S. R. Harris, N. M. Brown, M. T. Holden, M. Quail other authors (2013). "Rapid Bacterial Whole-Genome Sequencing to Enhance Diagnostic and Public Health Microbiology." *JAMA Intern Med* **173**(15): 1397-1404.

Rinta-Kokko, H., R. Dagan, N. Givon-Lavi & K. Auranen (2009). "Estimation of vaccine efficacy against acquisition of pneumococcal carriage." *Vaccine* **27**(29), 3831-3837.

Robberstad, B., C. R. Frostad, P. E. Akselsen, K. J. Kvaerner & A. K. Berstad (2011). "Economic evaluation of second generation pneumococcal conjugate vaccines in Norway." *Vaccine* **29**(47), 8564-8574.

Robson, R. L., S. Essengue, N. A. Reed & R. T. Horvat (2007). "Optochin resistance in Streptococcus pneumoniae induced by frozen storage in glycerol." *Diagn Microbiol Infect Dis* **58**(2), 185-190.

Roche, A., P. T. Heath, M. Sharland, D. Strachan, A. Breathnach, J. Haigh & Y. Young (2007). "Prevalence of nasopharyngeal carriage of pneumococcus in preschool children attending day care in London." *Arch Dis Child* **92**(12), 1073-1076.

Rodgers, G. L., A. Arguedas, R. Cohen & R. Dagan (2009). "Global serotype distribution among Streptococcus pneumoniae isolates causing otitis media in children: Potential implications for pneumococcal conjugate vaccines." *Vaccine* **27**(29), 3802-3810.

Rohde, H., J. Qin, Y. Cui, D. Li, N. J. Loman, M. Hentschke, W. Chen, F. Pu, Y. Peng, J. Li other authors (2011). "Open-source genomic analysis of Shiga-toxin-producing E. coli O104:H4." *N Engl J Med* **365**(8), 718-724.

Rolo, D., S. S. A, A. Domenech, A. Fenoll, J. Linares, H. de Lencastre, C. Ardanuy & R. Sa-Leao (2013). "Disease isolates of Streptococcus pseudopneumoniae and non-typeable S. pneumoniae presumptively identified as atypical S. pneumoniae in Spain." *PLoS One* **8**(2), e57047.

Romero, P., E. Garcia & T. Mitchell (2009). "Development of a Prophage Typing System and Analysis of Prophage Carriage in Streptococcus pneumoniae." *Appl Environ Microbiol* **75**(6), 1642 - 1649.

Rosen, J. B., A. R. Thomas, C. A. Lexau, A. Reingold, J. L. Hadler, L. H. Harrison, N. M. Bennett, W. Schaffner, M. M. Farley, B. W. Beall other authors (2011). "Geographic variation in invasive pneumococcal disease following pneumococcal conjugate vaccine introduction in the United States." *Clin Infect Dis* **53**(2), 137-143.

Rothberg, J. M., W. Hinz, T. M. Rearick, J. Schultz, W. Mileski, M. Davey, J. H. Leamon, K. Johnson, M. J. Milgrew, M. Edwards other authors (2011). "An integrated semiconductor device enabling non-optical genome sequencing." *Nature* **475**(7356), 348-352.

Rovers, M. M., A. G. Schilder, G. A. Zielhuis & R. M. Rosenfeld (2004). "Otitis media." *Lancet* **363**(9407), 465-473.

Rubin, L. G., A. Rizvi & A. Baer (2008). "Effect of Swab Composition and Use of Swabs versus Swab-Containing Skim Milk-Tryptone-Glucose-Glycerol (STGG) on Culture- or PCR-Based Detection of Streptococcus pneumoniae in Simulated and Clinical Respiratory Specimens in STGG Transport Medium." *J. Clin. Microbiol.* **46**(8), 2635-2640.

Ruohola, A., M. M. Pettigrew, L. Lindholm, J. Jalava, K. S. Raisanen, R. Vainionpaa, M. Waris, P. A. Tahtinen, M. K. Laine, E. Lahti other authors (2013). "Bacterial and viral interactions within the nasopharynx contribute to the risk of acute otitis media." *J Infect* **66**(3), 247-254.

S. Martin, K. B., C. Sheppard (2012). "Laboratory typing methods for *Streptococcus pneumoniae* " *Public Health England, Standard operating procedure* **R-6404/07-12**.

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

Sanger, F., G. M. Air, B. G. Barrell, N. L. Brown, A. R. Coulson, C. A. Fiddes, C. A. Hutchison, P. M. Slocombe & M. Smith (1977a). "Nucleotide sequence of bacteriophage phi X174 DNA." *Nature* **265**(5596), 687-695.

Sanger, F. & A. R. Coulson (1975). "A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase." *J Mol Biol* **94**(3), 441-448.

Sanger, F., S. Nicklen & A. R. Coulson (1977b). "DNA sequencing with chain-terminating inhibitors." *Proc Natl Acad Sci U S A* **74**(12), 5463-5467.

Scott, D. A., S. F. Komjathy, B. T. Hu, S. Baker, L. A. Supan, C. A. Monahan, W. Gruber, G. R. Siber & S. P. Lockhart (2007). "Phase 1 trial of a 13-valent pneumococcal conjugate vaccine in healthy adults." *Vaccine* **25**(33), 6164-6166.

Scott, J. R., J. Hinds, K. A. Gould, E. V. Millar, R. Reid, M. Santosham, K. L. O'Brien & W. P. Hanage (2012). "Nontypeable pneumococcal isolates among navajo and white mountain apache communities: are these really a cause of invasive disease?" *J Infect Dis* **206**(1), 73-80.

Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour & T. S. Whittam (1986). "Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics." *Appl Environ Microbiol* **51**(5), 873-884.

Shakir, E., Cameron, C., Denham, B., McMenamin, J. (2009). "Respiratory and immunisation quarterly report." *HPS Wkly Rep* **43**, 407-408.

Shi, Y., P. C. Simpson, J. R. Scherer, D. Wexler, C. Skibola, M. T. Smith & R. A. Mathies (1999). "Radial capillary array electrophoresis microplate and scanner for high-performance nucleic acid analysis." *Anal Chem* **71**(23), 5354-5361.

Siberry, G., K. N. Brahmadathan, R. Pandian, M. K. Lalitha, M. C. Steinhoff & T. J. John (2001). "Comparison of different culture media and storage temperatures for the long-term preservation of Streptococcus pneumoniae in the tropics." *Bull World Health Organ* **79**(1), 43-47.

Silva, N. A., J. McCluskey, J. M. C. Jefferies, J. Hinds, A. Smith, S. C. Clarke, T. J. Mitchell & G. K. Paterson (2006). "Genomic Diversity between Strains of the Same Serotype and Multilocus Sequence Type among Pneumococcal Clinical Isolates." *Infect. Immun.* **74**(6), 3513-3518.

Sjostrom, K., C. Spindler, A. Ortqvist, M. Kalin, A. Sandgren, S. Kuhlmann-Berenzon & B. Henriques-Normark (2006). "Clonal and Capsular Types Decide Whether Pneumococci Will Act as a Primary or Opportunistic Pathogen." *Clinical Infectious Diseases* **42**(4), 451-459.

Skinner, J. M., L. Indrawati, J. Cannon, J. Blue, M. Winters, J. Macnair, N. Pujar, W. Manger, Y. Zhang, J. Antonello other authors (2011). "Pre-clinical evaluation of a 15-valent pneumococcal conjugate vaccine (PCV15-CRM197) in an infant-rhesus monkey immunogenicity model." *Vaccine* **29**(48), 8870-8876.

Smith, L. M., S. Fung, M. W. Hunkapiller, T. J. Hunkapiller & L. E. Hood (1985). "The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis." *Nucleic Acids Res* **13**(7), 2399-2412.

Smith, T., D. Lehmann, J. Montgomery, M. Gratten, I. D. Riley & M. P. Alpers (1993). "Acquisition and Invasiveness of Different Serotypes of Streptococcus pneumoniae in Young Children." *Epidemiology and Infection* **111**(1), 27-39.

Song, J. H., J. Y. Baek & K. S. Ko (2011). "Comparison of capsular genes of Streptococcus pneumoniae serotype 6A, 6B, 6C, and 6D isolates." *J Clin Microbiol* **49**(5), 1758-1764.

Song, J. H., R. Dagan, K. P. Klugman & B. Fritzell (2012). "The relationship between pneumococcal serotypes and antibiotic resistance." *Vaccine* **30**(17), 2728-2737.

Spijkerman, J., S. M. Prevaes, E. J. van Gils, R. H. Veenhoven, J. P. Bruin, D. Bogaert, A. J. Wijmenga-Monsuur, G. P. van den Dobbelsteen & E. A. Sanders (2012). "Long-term effects of pneumococcal conjugate vaccine on nasopharyngeal carriage of S. pneumoniae, S. aureus, H. influenzae and M. catarrhalis." *PLoS ONE* **7**(6), e39730.

Spratt, B. G. & B. M. Greenwood (2000). "Prevention of pneumococcal disease by vaccination: does serotype replacement matter?" *The Lancet* **356**(9237), 1210-1211.

Statens Serum Institut (2012). "Pneumcoccal antisera. http://www.ssi.dk/English/SSI Diagnostica/Products from SSI Diagnostica/Antisera antibodies/Pneumococcus antisera/Neufeld.aspx." **Accessed 01/12/12**.

Stein, K. E. (1992). "Thymus-Independent and Thymus-Dependent Responses to Polysaccharide Antigens." *The Journal of Infectious Diseases* **165**, S49-S52.

Sternburg (1881). "A fatal form of septicemia in the rabbit produced by the subcutaneous injection of human saliva." *Natl. Bd. Health Bull.* **2**(44), 781-783,.

Swiss Institute of Bioinformatics (2013). "ExPASy Bioinformatics resource portal. http://web.expasy.org/translate/." **Accessed 01/07/2013**.

Tahtinen, P. A., M. K. Laine, P. Huovinen, J. Jalava, O. Ruuskanen & A. Ruohola (2011). "A placebo-controlled trial of antimicrobial treatment for acute otitis media." *N Engl J Med* **364**(2), 116-126.

Teele, D. W., J. O. Klein, C. Chase, P. Menyuk & B. A. Rosner (1990). "Otitis media in infancy and intellectual ability, school achievement, speech, and language at age 7 years. Greater Boston Otitis Media Study Group." *J Infect Dis* **162**(3), 685-694.

Temime, L., P. Boelle, L. Opatowski & D. Guillemot (2008). "Impact of capsular switch on invasive pneumococcal disease incidence in a vaccinated population." *PLoS ONE* **3**, e3244.

Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson other authors (2001). "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae." *Science* **293**(5529), 498-506.

The Sanger Institute (2011). "*Streptococcus pneumoniae* http://www.sanger.ac.uk/Projects/S_pneumoniae/ " **Accessed 01/12/2012**.

The Sanger Institute (2012). "Software. http://www.sanger.ac.uk/resources/software/." **Accessed 01/12/12**.

Tillett, W. S. & T. Francis (1930). "Serological Reactions in Pneumonia with a Non-Protein Somatic Fraction of Pneumococcus." *J Exp Med* **52**(4), 561-571.

Tocheva, A., J. M. Jefferies, M. Christodoulides, S. N. Faust & S. C. Clarke (2010). "Increase in serotype 6C pneumococcal carriage during the implementation of the heptavalent pneumococcal conjugate vaccine." *Emerging Infectious Diseases* **16**(1).

Tocheva, A. S., J. M. Jefferies, M. Christodoulides, S. N. Faust & S. C. Clarke (2013). "Distribution of carried pneumococcal clones in UK children following the introduction of the 7-valent pneumococcal conjugate vaccine: a 3-year cross-sectional population based analysis." *Vaccine* **31**(31), 3187-3190.

Tocheva, A. S., J. M. Jefferies, H. Rubery, J. Bennett, G. Afimeke, J. Garland, M. Christodoulides, S. N. Faust & S. C. Clarke (2011). "Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of Streptococcus pneumoniae in young children." *Vaccine* **29**(26), 4400-4404.

Torok, M. E. & S. J. Peacock (2012). "Rapid whole-genome sequencing of bacterial pathogens in the clinical microbiology laboratory--pipe dream or reality?" *J Antimicrob Chemother* **67**(10), 2307-2308.

Trotter, C. L., N. J. Gay & W. J. Edmunds (2005). "Dynamic Models of Meningococcal Carriage, Disease, and the Impact of Serogroup C Conjugate Vaccination." *Am. J. Epidemiol.* **162**(1), 89-100.

Turner, P., J. Hinds, C. Turner, A. Jankhot, K. Gould, S. D. Bentley, F. Nosten & D. Goldblatt (2011). "Improved detection of nasopharyngeal cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray." *J Clin Microbiol* **49**(5), 1784-1789.

Turner, P., C. Turner, A. Jankhot, K. Phakaudom, F. Nosten & D. Goldblatt (2013). "Field evaluation of culture plus latex sweep serotyping for detection of multiple pneumococcal serotype colonisation in infants and young children." *PLoS One* **8**(7), e67933.

U.S National institutes of Health (2012). "Clincial Trials. http://www.clinicaltrials.gov " **Accessed 01/11/12**.

Uddin, S., R. Borrow, M. R. Haeney, A. Moran, R. Warrington, P. Balmer & P. D. Arkwright (2006). "Total and serotype-specific pneumococcal antibody titres in children with normal and abnormal humoral immunity." *Vaccine* **24**(27-28), 5637-5644.

University of Southampton (2012). "Optoelectronics Research Centre. http://www.orc.soton.ac.uk/dna_sequencing.html " **Accessed 01/12/12**.

Vainio, A., O. LyytikÄlnen, R. Sihvonen, T. Kaijalainen, L. TeirilÄ, M. Rantala, P. Lehtinen, P. Ruuska & A. Virolainen (2009). "An outbreak of pneumonia associated with S. pneumoniae at a military training facility in Finland in 2006." *APMIS* **117**(7), 488-491.

Vakevainen, M., C. Eklund, J. Eskola & H. Kayhty (2001). "Cross-Reactivity of Antibodies to Type 6B and 6A Polysaccharides of Streptococcus pneumoniae, Evoked by Pneumococcal Conjugate Vaccines, in Infants." *The Journal of Infectious Diseases* **184**(6), 789-793.

Valouev, A., J. Ichikawa, T. Tonthat, J. Stuart, S. Ranade, H. Peckham, K. Zeng, J. A. Malek, G. Costa, K. McKernan other authors (2008). "A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning." *Genome Res* **18**(7), 1051-1063.

van de Sande-Bruinsma, N., H. Grundmann, D. Verloo, E. Tiemersma, J. Monen, H. Goossens, M. Ferech, G. European Antimicrobial Resistance Surveillance System & G. European Surveillance of Antimicrobial Consumption Project (2008). "Antimicrobial drug use and resistance in Europe." *Emerg Infect Dis* **14**(11), 1722-1730.

van den Bergh, M. R., G. Biesbroek, J. W. Rossen, W. A. de Steenhuijsen Piters, A. A. Bosch, E. J. van Gils, X. Wang, C. W. Boonacker, R. H. Veenhoven, J. P. Bruin other authors (2012). "Associations between pathogens in the upper respiratory tract of young children: interplay between viruses and bacteria." *PLoS One* **7**(10), e47711.

van der Linden, M., A. Al-Lahham, W. Nicklas & R. R. Reinert (2009). "Molecular Characterization of Pneumococcal Isolates from Pets and Laboratory Animals." *PLoS ONE* **4**(12), e8286.

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

Wade, N. (2006). "The quest for the $1,000 human genome: DNA sequencing in the doctor's office? At birth? It may be coming closer." *N Y Times Web*, F1, F3.

Wadowsky, R. M., S. Laus, T. Libert, S. J. States & G. D. Ehrlich (1994). "Inhibition of PCR-based assay for Bordetella pertussis by using calcium alginate fiber and aluminum shaft components of a nasopharyngeal swab." *J Clin Microbiol* **32**(4), 1054-1057.

Walker, T. M., C. L. Ip, R. H. Harrell, J. T. Evans, G. Kapatai, M. J. Dedicoat, D. W. Eyre, D. J. Wilson, P. M. Hawkey, D. W. Crook other authors (2013). "Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study." *Lancet Infect Dis* **13**(2), 137-146.

Watson, J. D. & R. M. Cook-Deegan (1991). "Origins of the Human Genome Project." *FASEB J* **5**(1), 8-11.

Watson, K., K. Carville, J. Bowman, P. Jacoby, T. V. Riley, A. J. Leach & D. Lehmann (2006). "Upper respiratory tract bacterial carriage in Aboriginal and non-Aboriginal children in a semi-arid area of Western Australia." *Pediatr Infect Dis J* **25**(9), 782-790.

Watt, J. P., K. L. O'Brien, S. Katz, M. A. Bronsdon, J. Elliott, J. Dallas, M. J. Perilla, R. Reid, L. Murrow, R. Facklam other authors (2004). "Nasopharyngeal versus oropharyngeal sampling for detection of pneumococcal carriage in adults." *J Clin Microbiol* **42**(11), 4974-4976.

Weinberger, D. M., D. T. Bruden, L. R. Grant, M. Lipsitch, K. L. O'Brien, S. I. Pelton, E. A. Sanders & D. R. Feikin (2013). "Using Pneumococcal Carriage Data to Monitor Postvaccination Changes in Invasive Disease." *Am J Epidemiol* **178**(9): 1488-1495.

Whatmore, A. M., A. Efstratiou, A. P. Pickerill, K. Broughton, G. Woodard, D. Sturgeon, R. George & C. G. Dowson (2000). "Genetic relationships between clinical isolates of Streptococcus pneumoniae, Streptococcus oralis, and Streptococcus mitis: characterization of "Atypical" pneumococci and organisms allied to S. mitis harboring S. pneumoniae virulence factor-encoding genes." *Infect Immun* **68**(3), 1374-1382.

Woolfson, A., R. Huebner, A. Wasas, S. Chola, P. Godfrey-Faussett & K. Klugman (1997). "Nasopharyngeal carriage of community-acquired, antibiotic-resistant Streptococcus pneumoniae in a Zambian paediatric population." *Bull World Health Organ* **75**(5), 453-462.

World Health Organisation (2004). "Chronic suppurative otitis media." *Burden of illness and management options*

World Health Organisation (2008). "WHO/Health Canada consultation on serological criteria for evaluation and licensing of new pneumococcal vaccines. Executive summary." *WHO*.

Wright, A., W. Parry Morgan, L. Colebrook & R. W. Dodgson (1914). "Observations on prophylatic inoculation against pneumococccus infections, and on the results which have been achieved by it." *The Lancet* **183**(4715), 87-95.

Wright, H. D. (1934). "A substance in cotton-wool inhibitory to the growth of the pneumococcus." *J Pathol Bacteriol* **38**, 499-501.

Wroe, P. C., G. M. Lee, J. A. Finkelstein, S. I. Pelton, W. P. Hanage, M. Lipsitch, A. E. Stevenson, S. L. Rifas-Shiman, K. Kleinman, M. M. Dutta-Linn other authors (2012). "Pneumococcal carriage and antibiotic resistance in young children before 13-valent conjugate vaccine." *Pediatr Infect Dis J* **31**(3), 249-254.

Wysocki, J., J. C. Tejedor, D. Grunert, R. Konior, J. Garcia-Sicilia, M. Knuf, L. Bernard, I. Dieussaert & L. Schuerman (2009). "Immunogenicity of the 10-Valent Pneumococcal Non-typeable Haemophilus influenzae Protein D Conjugate Vaccine (PHiD-CV) When Coadministered With Different Neisseria

meningitidis Serogroup C Conjugate Vaccines." *Pediatric Infectious Disease Journal* **28**(4), S77-S88.

Yeh, S. H., A. Gurtman, D. C. Hurley, S. L. Block, R. H. Schwartz, S. Patterson, K. U. Jansen, J. Love, W. C. Gruber, E. A. Emini other authors (2010). "Immunogenicity and safety of 13-valent pneumococcal conjugate vaccine in infants and toddlers." *Pediatrics* **126**(3), e493-505.

Yu, J., J. Lin, K. H. Kim, W. H. Benjamin, Jr. & M. H. Nahm (2011). "Development of an automated and multiplexed serotyping assay for Streptococcus pneumoniae." *Clin Vaccine Immunol* **18**(11), 1900-1907.

Zhang, Q., S. Choo & A. Finn (2002). "Immune responses to novel pneumococcal proteins pneumolysin, PspA, PsaA, and CbpA in adenoidal B cells from children." *Infect Immun* **70**(10), 5363-5369.

Zhang, Q., Q. Ma, Q. Li, W. Yao & C. Wang (2011). "Enhanced protection against nasopharyngeal carriage of Streptococcus pneumoniae elicited by oral multiantigen DNA vaccines delivered in attenuated Salmonella typhimurium." *Mol Biol Rep* **38**(2), 1209-1217.

BIBLIOGRAPHY

Papers published during the course of this thesis candiature are listed below with the full texts immediately following:

Gladstone, R. A., J. M. Jefferies, S. N. Faust & S. C. Clarke (2011). "Continued control of pneumococcal disease in the UK - the impact of vaccination." *J Med Microbiol* **60**(Pt 1), 1-8.

Gladstone, R. A., J. M. Jefferies, S. N. Faust & S. C. Clarke (2012a). "Pneumococcal 13-valent conjugate vaccine for the prevention of invasive pneumococcal disease in children and adults." *Expert Rev Vaccines* **11**(8), 889-902.

Gladstone, R. A., J. M. Jefferies, S. N. Faust & S. C. Clarke (2012b). "Sampling methods for the study of pneumococcal carriage: a systematic review." *Vaccine* **30**(48), 6738-6744.

Loman NJ, Gladstone RA, Constantinidou C, Tocheva AS & J. JMC (2013). "Clonal Expansion within Pneumococcal Serotype 6C after Use of Seven-Valent Vaccine." *PLoS ONE* **8**(5), e64731.

Review

Continued control of pneumococcal disease in the UK – the impact of vaccination

R. A. Gladstone,[1] J. M. Jefferies,[1,2,3] S. N. Faust[1,3,4] and S. C. Clarke[1,2,3]

Correspondence
S. C. Clarke
S.C.Clarke@soton.ac.uk

[1]Division of Infection, Inflammation and Immunity, University of Southampton School of Medicine, UK

[2]Health Protection Agency, Southampton, UK

[3]Southampton NIHR Respiratory Biomedical Research Unit, Division of Infection, Inflammation and Immunity, Southampton University School of Medicine, UK

[4]Wellcome Trust Clinical Research Facility, Southampton University Hospitals Trust, Southampton, UK

Streptococcus pneumoniae, also known as the pneumococcus, is an important cause of morbidity and mortality in the developed and developing world. Pneumococcal conjugate vaccines were first introduced for routine use in the USA in 2000, although the seven-valent pneumococcal conjugate vaccine (PCV7) was not introduced into the UK's routine childhood immunization programme until September 2006. After its introduction, a marked decrease in the incidence of pneumococcal disease was observed, both in the vaccinated and unvaccinated UK populations. However, pneumococci are highly diverse and serotype prevalence is dynamic. Conversely, PCV7 targets only a limited number of capsular types, which appears to confer a limited lifespan to the observed beneficial effects. Shifts in serotype distribution have been detected for both non-invasive and invasive disease reported since PCV7 introduction, both in the UK and elsewhere. The pneumococcal *Haemophilus influenzae* protein D conjugate vaccine (PHiD-CV, Synflorix; GlaxoSmithKline) and 13-valent pneumococcal conjugate vaccine (PCV13, Prevenar 13; Pfizer) have been newly licensed. The potential coverage of the 10- and 13-valent conjugate vaccines has also altered alongside serotype shifts. Nonetheless, the mechanism of how PCV7 has influenced serotype shift is not clear-cut as the epidemiology of serotype prevalence is complex. Other factors also influence prevalence and incidence of pneumococcal carriage and disease, such as pneumococcal diversity, levels of antibiotic use and the presence of risk groups. Continued surveillance and identification of factors influencing serotype distribution are essential to allow rational vaccine design, implementation and continued effective control of pneumococcal disease.

Introduction

Streptococcus pneumoniae is a Gram-positive encapsulated bacterium. Currently 46 serogroups and 93 serotypes have been documented, the latest additions being serotype 6C (Park *et al.*, 2007b), serotype 6D (Jin *et al.*, 2009) and serotype 11E (Calix & Nahm, 2010). Capsular polysaccharides are highly immunogenic and are the main target for pneumococcal vaccines. However, the bacterium is capable of transformation, the horizontal exchange of genetic information, at both the intra- and interspecies level. This recombination of genetic material can result in subtle changes, which impact on the disease biology of strains and can also allow capsular switch to occur (Silva *et al.*, 2006). This phenomenon results from recombination of heterologous DNA at the capsular locus. As a consequence, clonal isolates, as determined by multilocus sequence typing, can express different polysaccharide capsules (serotype). Isolates of the same serotype can also be of different sequence type (ST) (Coffey *et al.*, 1998). The polysaccharide capsule is a key component of virulence, and serotypes differ in their association with invasive disease, antibiotic resistance and outbreak potential (Brueggemann *et al.*, 2003; Magee & Yother, 2001; Weinberger *et al.*, 2009).

The pneumococcal niche

Humans are the major reservoir of *S. pneumoniae*, carrying the bacteria asymptomatically in the nasopharynx (Hussain *et al.*, 2005). Within the human population, young children are the key source of pneumococci. Prior to the seven-valent polysaccharide conjugate vaccine (PCV7) introduction in the UK, a carriage rate of 45 % had been reported in children under 2 years old, compared to only 8 % in those older than 18 (Hussain *et al.*, 2005).

Despite being part of the respiratory commensal flora, the pneumococcus is also responsible for significant morbidity

and mortality in the UK and worldwide (Melegaro *et al.*, 2006; Mulholland, 2007). Pneumococcal disease ranges from acute otitis media (AOM) through to pneumonia and invasive disease (IPD) such as meningitis and septicaemia. Certain populations are at high risk of IPD and other pneumococcal diseases. These include infants under 2 years old, for whom a UK pre-PCV7 study estimated 15 pneumococcal meningitis cases per 100 000 (Melegaro *et al.*, 2006). The elderly are also at risk, with approximately 45 cases of IPD per 100 000 occurring pre-PCV7 in persons over 65 years in Scotland (Kyaw *et al.*, 2003). Additional risk groups for serious pneumococcal infection include children between 2 and 5 years old and the immunocompromised (Burman *et al.*, 1985; Kyaw *et al.*, 2003).

Control of pneumococcal disease

The 23-valent polysaccharide vaccine (PPV, Pneumovax; Merck) has been available for over 25 years. This vaccine is used today to vaccinate at-risk adults and the elderly. Unfortunately, immunization with PPV has recently been found to be largely unsuccessful in the UK elderly population (Joint Committee on Vaccination and Immunisation, 2009). In addition, PPV is known to elicit a T-cell-independent immune response, which is underdeveloped in those under 2 years old (Stein, 1992). Conjugate vaccines were designed to improve efficacy in those under 2 years old.

PCV7 Prevenar (Pfizer, previously Wyeth) was first recommended for use in the US in the year 2000 (Committee on Infectious Diseases, 2000). The vaccine contains the capsular polysaccharide of seven serotypes, conjugated to CRM_{197}, a non-toxic diphtheria variant carrier protein (Eskola *et al.*, 2001). This immunogenic protein increases the vaccine efficacy in the young by inducing a T-cell-dependent response (Black *et al.*, 2000; Rennels *et al.*, 1998). The seven serotypes included in the vaccine – 4, 6B, 9V, 14, 18C, 19F and 23F – were selected as they caused the majority of invasive disease in the US (Hausdorff *et al.*, 2000). These serotypes are also associated with high antibiotic resistance (Hicks *et al.*, 2007; Tyrrell *et al.*, 2009).

PCV7 impact in the UK

PCV7 vaccination in the UK was predicted to result in a decrease in pneumococcal disease incidence (Clarke *et al.*, 2006), as described in the US (CDC, 2005), where PCV7 also resulted in a reduction in pneumococcal antibiotic non-susceptibility (Richter *et al.*, 2009). In September 2006, PCV7 was added to the UK routine childhood immunization programme to help reduce the burden of pneumococcal disease (Department of Health, 2006).

Serotype surveillance data for IPD in England, Wales and Scotland since PCV7 vaccine introduction are being continuously collected (www.hpa.org.uk, www.hps.scot.nhs.uk). Current data for England and Wales (Kaye *et al.*, 2009) show a 41 % decrease in the number of IPD cases in those aged 5 years and under between 2005–2006 (797 cases) and 2007–2008 (470 cases). This is primarily a result of a dramatic decrease in the number of IPD cases caused by vaccine types (VTs) in children $\leqslant$5 years old. This decrease can also be clearly seen when comparing the cumulative total of cases reported in the under 5s by week 20 of 2006, 2007 and 2010 (Table 1). VT disease, previously accounting for 70 % of cases in this age group during 2005–2006, reduced to only 24 % in 2007–2008. These overall trends have also been observed in Scotland (Shakir *et al.*, 2009).

In addition to the decrease in IPD in the vaccinated population, herd immunity to VT pneumococci has been induced in the UK population as an indirect effect of infant PCV7 immunization. A decrease in VT IPD incidence has been seen in children over the age of 5 and adults, who are largely unvaccinated (Kaye *et al.*, 2009). The level of herd immunity has been suggested to increase with the number of doses given (Haber *et al.*, 2007). Prior to this observation, surveillance carried out in the US demonstrated a 42 % fall in the incidence of IPD in infants $<$90 days old (Carter, 2006), indicating that herd immunity can also extend to those not yet old enough to be vaccinated or to have completed the vaccination course. A US model also predicted that even incomplete coverage and/or limited dose schedules would still confer herd immunity (Haber *et al.*, 2007). Herd immunity, primarily due to reduced exposure through decreased carriage and transmission from the vaccinated population, contributes extensively to the overall impact and cost-effectiveness of vaccination (CDC, 2005; Melegaro & Edmunds, 2004). Without such effects, the PCV7 introduction may not have been considered economically viable in the UK (Melegaro & Edmunds, 2004).

Serotype replacement

Following PCV7 introduction in the US, a shift in the prevalent serotypes circulating in the population and causing disease was observed, termed 'serotype replacement' (McEllistrem *et al.*, 2003). This was predicted to be mirrored in the UK (Spratt & Greenwood, 2000).

Table 1. Approximate number of IPD reports in those $<$5 years old by week 20 of 2006, 2007 or 2010

Data adapted from Current Epidemiology of Invasive Pneumococcal Disease (IPD) graphs, HPA website (www.hpa.org.uk).

Year	In PCV7	Not in PCV7	Total cases
2006	400	150	550
2007	275	175	450
2010	25	375	400
Increase ↑/decrease ↓	↓	↑	↓

Although there has been a dramatic reduction in VT IPD, the phenomenon of 'replacement disease' has occurred in the UK. Replacement disease is due to the increase in non-vaccine serotype (NVT) IPD cases, which has greatly offset the decrease in VT IPD (Table 1). The total number of IPD cases in those over 5 years of age did not change significantly between 2005–2006 (5514 cases) and 2007–2008 (5496 cases). Importantly, a recent increased incidence of IPD caused by PCV7 NVTs has been detected in all age groups, particularly involving serotypes 7F, 19A and 22F (Kaye *et al.*, 2009). These NVTs have also been observed to cause an increased incidence of IPD in countries outside the UK using PCV7, such 7F in Portugal (Sá-Leão *et al.*, 2009) and 19A and 22F in the US (Hicks *et al.*, 2007). The post-PCV7 19A increase in the US was particularly associated with one multilocus sequence type, ST320, a clone which had high antibiotic resistance (Hanage *et al.*, 2007; Pillai *et al.*, 2009). This may have been the driving force in its increase (Dagan *et al.*, 2009). However, the 19A clone increasing in the UK is predominantly ST199 (Pichon *et al.*, 2008), not ST320, suggesting factors other than antibiotic resistance were involved in causing this increase.

Vaccine inclusion of related serotypes within a serogroup was previously assumed to confer some level of cross-protection (Hausdorff *et al.*, 2000). However, serotype 19A IPD incidence has increased in the UK despite the fact that the related serotype 19F was included in the vaccine. The 19F polysaccharide is known to be the least immunogenic of the PCV7 VTs (Pletz *et al.*, 2008), and in addition cross-reaction of antibodies for 19F to 19A has also been shown to be weak *in vitro* (Lee *et al.*, 2009). In carriage, a significant increase in the prevalence of serotype 6C has been observed since PCV7 introduction in the UK (Nahm *et al.*, 2009; Tocheva *et al.*, 2010). This is despite the presence of the 6B polysaccharide in PCV7, which does provide protection against 6A (Väkeväinen *et al.*, 2001). 6B cross-reactivity does not extend to 6C, therefore the PCV7 elicits negligible or no immune protection against this serotype (Park *et al.*, 2007a). Serotypes related to the VTs have contributed to serotype replacement more than was perhaps first expected, potentially because they are in a prime position to fill the specific niche vacated by their counterpart in an environment under vaccine pressure. Replacement disease has dramatically reduced the effectiveness of PCV7 vaccination and is likely to be a major factor in the decision to replace PCV7 with PCV13 in April 2010.

The most common serotypes causing IPD can change rapidly. In Scotland, IPD-causing serotypes changed considerably from 2005/2006 to 2009 (Table 2). In 2009, 7F was reported to be the most common IPD-causing serotype in the under 5s, accounting for 12 % of cases (Kaye *et al.*, 2009). Before the introduction of PCV7, this serotype caused little disease; in fact, 7F was not isolated from a single reported IPD case for children under 5 years old in Scotland during 2006 (unpublished data). Dramatic changes in serotype prevalence can occur over time. This demonstrates the importance of long-term epidemiological surveillance in allowing appropriate response and action to changes.

Table 2. Serotypes (no. of cases) involved in IPD in Scotland: rank order of incidence in those $\leqslant$5 years old

Rank	2006*	2009†
1	14 (26)	7F (50)
2	1 (8)	1 (35)
3	19F (5)	8 (32)
4	6A (5)	3 (30)
5	6B (5)	19A (24)
6	9V (5)	22F (24)

*Our unpublished data.
†Data from Shakir *et al.* (2009).

It must also be noted that serotype distribution can fluctuate substantially in the absence of vaccination. A highly significant increase in serotype 1 was observed within the UK prior to routine PCV7 immunization, highlighting that other factors are also involved in pneumococcal serotype dynamics (Jefferies *et al.*, 2010; Kirkham *et al.*, 2006).

Although not the only cause, PCV7 is likely to have been playing an important part in serotype fluctuations in the UK by reducing VTs and creating a niche, which is being filled by NVTs or other species of bacteria. Ongoing surveillance and research will help uncover the reasons why certain organisms appear better at filling this niche than others, and why some cause invasive disease while others cause little or none at all.

PCV7 coverage

Historically, PCV7 covered 90 % of the serotypes causing IPD in the US. Crucially, even before PCV7 introduction, the disease serotype coverage in the UK was much lower. In Scotland, only 76.5 % of all IPD cases in those aged under 5 years old were calculated to be covered by PCV7 (Clarke *et al.*, 2006), whilst in the developing world coverage was suggested to drop as low as 45 % for IPD (Hausdorff *et al.*, 2000).

Pneumococci are the primary causal agent for pneumonia. The impact of PCV7 on pneumonia is hard to quantify as confirmed pneumococcal pneumonia cases are difficult to define. Some result in IPD, while the majority probably do not. Despite this variation, studies have indicated that PCV7 efficacy for clinical and X-ray-defined pneumonia is between 5 % and 25 % (Black *et al.*, 2002; Lucero *et al.*, 2009). In addition, the incidence of pneumonia-related hospitalization has decreased by up to 39 % since PCV7 introduction in the US in children under 2 years old (Grijalva *et al.*, 2007; CDC, 2009). All-cause pneumonia

cases remained stable during US PCV7 introduction, suggestive of other species contributing to disease replacement (CDC, 2009). Direct UK data regarding the impact of PCV7 on lower respiratory tract infection are not available, although it is likely that serotype replacement is occurring for pneumococcal pneumonia, as has been reported for IPD.

S. pneumoniae is also a leading cause of AOM. This non-invasive disease is a particular issue in young children and has a high economic burden worldwide (Melegaro *et al.*, 2006). Pneumococcal conjugate vaccines have been shown to offer some protection against AOM (Eskola *et al.*, 2001; Prymula *et al.*, 2006); additionally the incidence has been shown to decrease in the US post-PCV7 (Black *et al.*, 2004; Block *et al.*, 2004; Eskola *et al.*, 2001; Prymula *et al.*, 2006). US cases of otitis media due to NVTs were also seen to increase in incidence in the post-PCV7 era, a 10 % rise in NVTs was reported by one study (Block *et al.*, 2004), along with observations of capsular switch events (McEllistrem *et al.*, 2003). Serotype data for AOM in the UK and Europe are scarce (Rodgers *et al.*, 2009), although data from the US suggest that serotype replacement, capsular switch and species replacement may limit the effectiveness of pneumococcal vaccination against otitis media in the UK. *Haemophilus influenzae* is also an important cause of AOM, specifically, non-typable *H. influenzae* (NTHi) AOM, which was seen to increase by 15 % following widespread PCV7 vaccination in the US (Casey *et al.*, 2010). Increases in NTHi may also be filling part of the niche left by pneumococcal VTs in the UK.

Over time, PCV immunization will continue to impact on serotype prevalence and affect the incidence of pneumococcal disease. This impact will reduce current vaccine efficacy, confirming the need for ongoing vaccine development to ensure control of pneumococcal disease.

Vaccine progression

PHiD-CV and PCV13 are now both licensed in the UK, with Prevenar 13 having replaced Prevenar in the UK immunization programme. These vaccines target additional serotypes that are important to current disease incidence and are not well targeted by PCV7 (Table 3).

The PHiD-CV developed by GlaxoSmithKline includes two capsular polysaccharide types conjugated to either diphtheria (serotype 19F) or tetanus (serotype 18C) toxoid, and eight others (1, 4, 5, 6B, 7F, 9V, 14, 23F) conjugated to NTHi protein D (Wysocki *et al.*, 2009). This is said to give PHiD-CV the extra ability of providing some protection against AOM caused by NTHi and therefore may influence the impact and cost-effectiveness of this vaccine (Wysocki *et al.*, 2009). Originally, the GlaxoSmithKline experimental vaccine included 11 serotypes, yet the inclusion of serotype 3 was rejected due to a lack of inducible immunogenicity during clinical trials (Prymula *et al.*, 2006). The PCV13 developed by Wyeth (now Pfizer) targets the same pneumococcal serotypes as the PHiD-CV plus three additional serotypes, all conjugated to the immunogenic diphtheria toxoid (Scott *et al.*, 2007). Notably, 22F is not targeted by the PHiD-CV or PCV13, and this serotype has dramatically increased in IPD prevalence in children under 2 years of age in England and Wales (Kaye *et al.*, 2009). 22F is also now ranked sixth in Scotland for IPD in children under 5 (Table 2).

PHiD-CV and PCV13 coverage

Based on national (England and Wales) surveillance data, the percentage of serotypes causing cases of IPD covered by PCV7, PHiD-CV and PCV13 was calculated (Kaye *et al.*, 2009). In 2007–2008, only 24 % of IPD cases in those under 5 years old were caused by serotypes covered by PCV7, in stark contrast to the 76.5 % UK estimate based on IPD coverage in this age group prior to vaccine implementation (Clarke *et al.*, 2006). The serotype coverage of IPD in children under 5 years of age for PHiD-CV and PCV13 was 53 % and 74 %, respectively, for 2007/2008, a dramatic decrease from the 81 % and 92 % 2005/2006 coverage. A fundamental observation is that the potential coverage of PHiD-CV and PCV13 had already decreased prior to implementation due to the routine use of PCV7 and the associated serotype replacement, as well as shifts in pneumococcal epidemiology caused by other non-vaccine factors.

Clinical trials and mathematical models offer a basis for prediction of vaccine impact. One study used an algorithm that suggested that the PHiD-CV will be at least as effective as PCV7 in protecting against pneumococcal invasive disease worldwide (Hausdorff *et al.*, 2009), although the design of such an algorithm is complex and based on assumptions which may affect the model output. In clinical trials, the study population will naturally be exposed to multifaceted epidemiological factors that will affect vaccine

Table 3. Serotypes included in the 7-, 10- and 13-valent PCVs

Data from Black *et al.* (2000), Scott *et al.* (2007) and Wysocki *et al.* (2009). Bold and underlined text indicates serotypes not included in PCV7.

Vaccine	Manufacturer	Serotypes	UK status
PCV7 (Prevenar)	Pfizer	4, 6B, 9V, 14, 18C, 19F, 23F	Licensed
PHiD-CV (Synflorix)	GlaxoSmithKline	**1**, 4, **5**, 6B, **7F**, 9V, 14, 18C, 19F, 23F	Licensed
PCV13 (Prevenar 13)	Pfizer	**1**, **3**, 4, **5**, **6A**, 6B, **7F**, 9V, 14, 18C, **19A**, 19F, 23F	Licensed in use

impact in the target population. A German clinical study, powered to show immunological non-inferiority, showed that PCV13 should be just as effective as PCV7 at protecting against the seven serotypes included within the PCV7, as well as inducing sufficient immunity for the further VTs (Kieninger *et al.*, 2008). Importantly, PCV13 was shown to induce an opsonophagocytic activity to serotype 19A, which indicates that it will be efficient in preventing cases of serotype 19A invasive disease (Kieninger *et al.*, 2008).

Both PHiD-CV and PCV13 are likely to be effective in reducing IPD and non-invasive disease. However, the relative effect of PHiD-CV compared to PCV13 immunization on the prevention of combined pneumococcal and NTHi invasive or all-cause disease has not yet been established. As well as the effect on invasive diseases, pneumococcal carriage in individuals and the population is affected by routine immunization of the population, and the serotype effects appear to differ from those seen in invasive disease (unpublished data).

Invasive potential

Pneumococcal serotypes are known to differ in their invasiveness (Smith *et al.*, 1993). Traditionally the serotypes chosen for vaccine inclusion have been based on the rank order incidence of disease. These serotypes are often the most prevalent in carriage but they are not necessarily those with the highest potential for invasiveness (Brueggemann *et al.*, 2003, 2004). By targeting the serotypes in rank order of disease incidence, any serotype replacement that occurs may result in increased prevalence of a particularly invasive serotype. Several studies have calculated the potential of an individual serotype causing a disease case, taking into account factors such as the prevalence in carriage (Bättig *et al.*, 2006; Brueggemann *et al.*, 2004; Hanage *et al.*, 2005). One study highlighted that moderately prevalent NVT serotypes 3, 8, 33 and 38 all had similar potential to cause invasive disease as the VT 6B, 19F and 23F, previously responsible for a considerable proportion of disease cases in the pre-vaccine era (Brueggemann *et al.*, 2004). If a shift in prevalence occurred, for example, reduced VT 23F with increased NVT 8, this could then potentially result in serotype 8 having a similar disease incidence as VT 23F previously. NVT 19A and 7F, which have recently increased as a cause of IPD in the UK (Kaye *et al.*, 2009), have also previously been reported to be associated with invasive disease (Brueggemann *et al.*, 2003; Sjöström *et al.*, 2006). This may indicate that greater virulence of a serotype has been central to the rise in case numbers rather than an expansion of clones with these serotypes.

The presence of serotypes in PHiD-CV and PCV13 additional to those contained within PCV7 will, to some degree, help to protect against the emergence of some previously under-represented serotypes with significant invasive potential.

Future work

Data are only just becoming available on post-PCV7 pneumococcal carriage in the UK due to the time vaccine implementation has taken to translate into altered carriage and for the collection of comparable data. Further serotype data are also required for pneumonia and AOM, although these are not collected routinely by UK surveillance systems. A more detailed understanding of the dynamics of serotype prevalence is central to sustaining the control of pneumococcal disease.

Due to the global variation in serotype prevalence, vaccine design and use would preferably be specific to a geographical area, although this is unrealistic due to the time and cost involved in vaccine development. Future vaccines may have broader global application if design can employ more complex epidemiological models to simulate serotype replacement. One alternative is the use of vaccines based on highly conserved, immunogenic pneumococcal surface proteins that are involved in bacterial virulence. Current candidate proteins include pneumococcal surface protein A (PspA), pneumococcal surface protein C (PspC), pneumolysin (Ply) and caseinolytic protease (ClpP) (Cao *et al.*, 2007; Hamel *et al.*, 2004). Proteins could potentially be used in combination to maximize synergistic effects potentially providing protection from most, if not all, isolates of pneumococci (Cao *et al.*, 2007; Morsczeck *et al.*, 2008). Protein-based vaccines could potentially give broad protection from pneumococcal infection independent of serotype and invasiveness. However, such vaccines would also be predicted to have major effects on overall pneumococcal carriage with as yet unknown clinical significance of non-pneumococcal bacterial replacement. Establishing protection against otitis media and carriage would also ideally require stimulation of a mucosal antibody response (Zhang *et al.*, 2002). If proven to be safe in both direct and indirect effects, additional advantages of protein-only vaccines could be the induction of T-cell responses and ease of vaccine formulation resulting in reduced production costs when compared to conjugate vaccines. At present, candidate proteins are being evaluated in murine models, including the demonstration of passive immunity from polyclonal antibodies against specific pneumococcal proteins (Cao *et al.*, 2007, 2009; Morsczeck *et al.*, 2008; Ogunniyi *et al.*, 2007).

Summary

Four winters after PCV7 introduction to the UK routine infant immunization programme, the positive impact is clear. There has been a significant reduction in the incidence of PCV7 serotypes causing pneumococcal IPD in those under 5 years old, together with the induction of herd immunity. However, serotype replacement has occurred, such that IPD incidence in those under 2 years of age is now similar to that prior to PCV7 introduction.

Serotype replacement has been observed for invasive and non-invasive pneumococcal disease worldwide, and it is

evident that the overall effectiveness of PCV7 on the total pneumococcal disease burden has reduced. In April 2010, PCV13 replaced PCV7 in the UK infant immunization programme, and it is important to note that no older child catch-up campaign has been implemented (Department of Health, 2010). The presence of additional serotypes within PCV13 will help to combat the serotype replacement observed. Many additional factors will influence the serotype shifts in carriage and disease, including capsular switch events and presence of antimicrobial resistance. Nevertheless, pneumococcal vaccines based on a limited number of serotypes will continue to have a limited lifespan due to the selection pressure vaccines exert and the diversity of the bacteria. Increased vaccine coverage and control of pneumococcal disease is required worldwide, not only in the UK and other Western countries.

Acknowledgements

Part of the work described in this review was funded by an investigator-led unrestricted research grant from Wyeth Vaccines.

References

Bättig, P., Hathaway, L. J., Hofer, S. & Mühlemann, K. (2006). Serotype-specific invasiveness and colonization prevalence in *Streptococcus pneumoniae* correlate with the lag phase during in vitro growth. *Microbes Infect* **8**, 2612–2617.

Black, S., Shinefield, H., Fireman, B., Lewis, E., Ray, P., Hansen, J. R., Elvin, L., Ensor, K. M., Hackell, J. & other authors (2000). Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. *Pediatr Infect Dis J* **19**, 187–195.

Black, S. B. M., Shinefield, H. R. M., Ling, S. M. M., Hansen, J., Fireman, B. M., Spring, D. M., Noyes, J. M., Lewis, E. M., Ray, P. M. & other authors (2002). Effectiveness of heptavalent pneumococcal conjugate vaccine in children younger than five years of age for prevention of pneumonia. *Pediatr Infect Dis J* **21**, 810–815.

Black, S., Shinefield, H., Baxter, R., Austrian, R., Bracken, L., Hansen, J., Lewis, E. & Fireman, B. (2004). Postlicensure surveillance for pneumococcal invasive disease after use of heptavalent pneumococcal conjugate vaccine in Northern California Kaiser Permanente. *Pediatr Infect Dis J* **23**, 485–489.

Block, S. L., Hedrick, J., Harrison, C. J., Tyler, R., Smith, A., Findlay, R. & Keegan, E. (2004). Community-wide vaccination with the heptavalent pneumococcal conjugate significantly alters the microbiology of acute otitis media. *Pediatr Infect Dis J* **23**, 829–833.

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

Brueggemann, A. B., Peto, T. E., Crook, D. W., Butler, J. C., Kristinsson, K. G. & Spratt, B. G. (2004). Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. *J Infect Dis* **190**, 1203–1211.

Burman, L. A., Norrby, R. & Trollfors, B. (1985). Invasive pneumococcal infections: incidence, predisposing factors, and prognosis. *Rev Infect Dis* **7**, 133–142.

Calix, J. J. & Nahm, M. H. (2010). A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. *J Infect Dis* **202**, 29–38.

Cao, J., Chen, D., Xu, W., Chen, T., Xu, S., Luo, J., Zhao, Q., Liu, B., Wang, D. & other authors (2007). Enhanced protection against pneumococcal infection elicited by immunization with the combination of PspA, PspC, and ClpP. *Vaccine* **25**, 4996–5005.

Cao, J., Li, D., Gong, Y., Yin, N., Chen, T., Wong, C. K., Xu, W., Luo, J., Zhang, X. & other authors (2009). Caseinolytic protease: a protein vaccine which could elicit serotype-independent protection against invasive pneumococcal infection. *Clin Exp Immunol* **156**, 52–60.

Carter, R. J. F. (2006). Infants too young to receive pneumococcal conjugate vaccine benefit from herd immunity. *Thorax* **61**, 610.

Casey, J. R., Adlowitz, D. G. & Pichichero, M. E. (2010). New patterns in the otopathogens causing acute otitis media six to eight years after introduction of pneumococcal conjugate vaccine. *Pediatr Infect Dis J* **29**, 304–309.

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

CDC (2009). Pneumonia hospitalizations among young children before and after introduction of pneumococcal conjugate vaccine – United States, 1997–2006. *MMWR Morb Mortal Wkly Rep* **58**, 1–4.

Clarke, S. C., Jefferies, J. M., Smith, A. J., McMenamin, J., Mitchell, T. J. & Edwards, G. F. S. (2006). Potential impact of conjugate vaccine on the incidence of invasive pneumococcal disease among children in Scotland. *J Clin Microbiol* **44**, 1224–1228.

Coffey, T. J., Enright, M. C., Daniels, M., Morona, J. K., Morona, R., Hryniewicz, W., Paton, J. C. & Spratt, B. G. (1998). Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Mol Microbiol* **27**, 73–83.

Committee on Infectious Diseases (2000). Policy statement: recommendations for the prevention of pneumococcal infections, including the use of pneumococcal conjugate vaccine (Prevnar), pneumococcal polysaccharide vaccine, and antibiotic prophylaxis. *Pediatrics* **106**, 362–366.

Dagan, R., Givonâ-Lavi, N., Leibovitz, E., Greenberg, D. & Porat, N. (2009). Introduction and proliferation of multidrug-resistant *Streptococcus pneumoniae* serotype 19A clones that cause acute otitis media in an unvaccinated population. *J Infect Dis* **199**, 776–785.

Department of Health (2006). Important changes to the childhood immunisation programme, PL/CMO/2006/1.

Department of Health (2010). Introduction of Prevenar 13® into the Childhood Immunisation Programme. Department of Health, Gateway reference: 13581.

Eskola, J., Kilpi, T., Palmu, A., Jokinen, J., Haapakoski, J., Herva, E., Takala, A., Kayhty, H., Karma, P. & other authors (2001). Efficacy of a pneumococcal conjugate vaccine against acute otitis media. *N Engl J Med* **344**, 403–409.

Grijalva, C. G., Nuorti, J. P., Arbogast, P. G., Martin, S. W., Edwards, K. M. & Griffin, M. R. (2007). Decline in pneumonia admissions after routine childhood immunisation with pneumococcal conjugate vaccine in the USA: a time-series analysis. *Lancet* **369**, 1179–1186.

Haber, M., Barskey, A., Baughman, W., Barker, L., Whitney, C. G., Shaw, K. M., Orenstein, W. & Stephens, D. S. (2007). Herd immunity and pneumococcal conjugate vaccine: a quantitative model. *Vaccine* **25**, 5390–5398.

Hamel, J., Charland, N., Pineau, I., Ouellet, C., Rioux, S., Martin, D. & Brodeur, B. R. (2004). Prevention of pneumococcal disease in mice immunized with conserved surface-accessible proteins. *Infect Immun* **72**, 2659–2670.

Hanage, W. P., Kaijalainen, T. H., Syrjanen, R. K., Auranen, K., Leinonen, M., Makela, P. H. & Spratt, B. G. (2005). Invasiveness of

serotypes and clones of *Streptococcus pneumoniae* among children in Finland. *Infect Immun* **73**, 431–435.

Hanage, W. P., Huang, S. S., Lipsitch, M., Bishop, C. J., Godoy, D., Pelton, S. I., Goldstein, R., Huot, H. & Finkelstein, J. A. (2007). Diversity and antibiotic resistance among nonvaccine serotypes of *Streptococcus pneumoniae* carriage isolates in the post-heptavalent conjugate vaccine era. *J Infect Dis* **195**, 347–352.

Hausdorff, W. P., Bryant, J., Paradiso, P. R. & Siber, G. R. (2000). Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* **30**, 100–121.

Hausdorff, W. P., Dagan, R., Beckers, F. & Schuerman, L. (2009). Estimating the direct impact of new conjugate vaccines against invasive pneumococcal disease. *Vaccine* **27**, 7257–7269.

Hicks, L. A., Harrison, L. H., Flannery, B., Hadler, J. L., Schaffner, W., Craig, A. S., Jackson, D., Thomas, A., Beall, B. & other authors (2007). Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998–2004. *J Infect Dis* **196**, 1346–1354.

Hussain, M., Melegaro, A., Pebody, R. G., George, R., Edmunds, W. J., Talukdar, R., Martin, S. A., Efstratiou, A. & Miller, E. (2005). A longitudinal household study of *Streptococcus pneumoniae* nasopharyngeal carriage in a UK setting. *Epidemiol Infect* **133**, 891–898.

Jefferies, J. M., Smith, A. J., Edwards, G. F. S., McMenamin, J., Mitchell, T. J. & Clarke, S. C. (2010). Temporal analysis of invasive pneumococcal clones from Scotland illustrates fluctuations in diversity of serotype and genotype in the absence of pneumococcal conjugate vaccine. *J Clin Microbiol* **48**, 87–96.

Jin, P., Kong, F., Xiao, M., Oftadeh, S., Zhou, F., Liu, C., Russell, F. & Gilbert, G. L. (2009). First report of putative *Streptococcus pneumoniae* serotype 6D among nasopharyngeal isolates from Fijian children. *J Infect Dis* **200**, 1375–1380.

Joint Committee on Vaccination and Immunisation (2009). *Minutes of the Pneumococcal Subgroup on Tuesday 15th January*. Department of Health.

Kaye, P., Malkani, R., Martin, S., Slack, M., Trotter, C., Jit, M., George, R. & Miller, E. (2009). Invasive pneumococcal disease (IPD) in England & Wales after 7-valent conjugate vaccine (PCV7); potential impact of 10 and 13-valent vaccines. *Presented at The 27th Annual Meeting of the European Society for Paediatric Infectious Diseases*, 9–13 June 2009, Brussels.

Kieninger, D. M., Kueper, K., Steul, K., Juergens, C., Ahlers, N., Baker, S., Giardina, P., Gruber, W. & Scott, D. (2008). Safety and immunologic non-inferiority of 13-valent pneumococcal conjugate vaccine compared to 7-valent pneumococcal conjugate vaccine given as a 4-dose series with routine vaccines in healthy infants and toddlers. *Presented at The 48th Annual ICAAC/IDSA 46th Annual Meeting*, 25–28 October 2008, Washington DC.

Kirkham, L.-A. S., Jefferies, J. M. C., Kerr, A. R., Jing, Y., Clarke, S. C., Smith, A. & Mitchell, T. J. (2006). Identification of invasive serotype 1 pneumococcal isolates that express nonhemolytic pneumolysin. *J Clin Microbiol* **44**, 151–159.

Kyaw, M. H., Christie, P., Clarke, S. C., Mooney, J. D., Ahmed, S., Jones, I. G. & Campbell, H. (2003). Invasive pneumococcal disease in Scotland, 1999–2001: use of record linkage to explore associations between patients and disease in relation to future vaccination policy. *Clin Infect Dis* **37**, 1283–1291.

Lee, H., Nahm, M. H., Burton, R. & Kim, K. (2009). Immune response in infants to the heptavalent pneumococcal conjugate vaccine against vaccine-related serotypes 6A and 19A. *Clin Vaccine Immunol* **16**, 376–381.

Lucero, M. G., Dulalia, V. E., Nillos, L. T., Williams, G., Parreno, R. A., Nohynek, H., Riley, I. D. & Makela, H. (2009). Pneumococcal conjugate vaccines for preventing vaccine-type invasive pneumococcal disease and X-ray defined pneumonia in children less than two years of age. *Cochrane Database Syst Rev* CD004977.

Magee, A. D. & Yother, J. (2001). Requirement for capsule in colonization by *Streptococcus pneumoniae*. *Infect Immun* **69**, 3755–3761.

McEllistrem, M. C., Adams, J., Mason, E. O. & Wald, E. R. (2003). Epidemiology of acute otitis media caused by *Streptococcus pneumoniae* before and after licensure of the 7-valent pneumococcal protein conjugate vaccine. *J Infect Dis* **188**, 1679–1684.

Melegaro, A. & Edmunds, W. J. (2004). Cost-effectiveness analysis of pneumococcal conjugate vaccination in England and Wales. *Vaccine* **22**, 4203–4214.

Melegaro, A., Edmunds, W. J., Pebody, R., Miller, E. & George, R. (2006). The current burden of pneumococcal disease in England and Wales. *J Infect* **52**, 37–48.

Morsczeck, C., Prokhorova, T., Sigh, J., Pfeiffer, M., Bille-Nielsen, M., Petersen, J., Boysen, A., Kofoed, T., Frimodt-Møller, N. & other authors (2008). *Streptococcus pneumoniae*: proteomics of surface proteins for vaccine development. *Clin Microbiol Infect* **14**, 74–81.

Mulholland, K. (2007). Perspectives on the burden of pneumonia in children. *Vaccine* **25**, 2394–2397.

Nahm, M. H., Lin, J., Finkelstein, J. A. & Pelton, S. I. (2009). Increase in the prevalence of the newly discovered pneumococcal serotype 6C in the nasopharynx after introduction of pneumococcal conjugate vaccine. *J Infect Dis* **199**, 320–325.

Ogunniyi, A. D., Grabowicz, M., Briles, D. E., Cook, J. & Paton, J. C. (2007). Development of a vaccine against invasive pneumococcal disease based on combinations of virulence proteins of *Streptococcus pneumoniae*. *Infect Immun* **75**, 350–357.

Park, I. H., Park, S., Hollingshead, S. K. & Nahm, M. H. (2007a). Genetic basis for the new pneumococcal serotype, 6C. *Infect Immun* **75**, 4482–4489.

Park, I. H., Pritchard, D. G., Cartee, R., Brandao, A., Brandileone, M. C. C. & Nahm, M. H. (2007b). Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *J Clin Microbiol* **45**, 1225–1233.

Pichon, B., Beasley, L., Slack, M., Efstratiou, A., Miller, E. & George, R. (2008). Effect of the introduction of the pneumococcal conjugate vaccine in the UK childhood immunisation scheme on the genetic structure of paediatric invasive pneumococci. In *The 6th International Symposium on Pneumococci and Pneumococcal Diseases*, 8–12 June 2008, Reykjavik, Iceland.

Pillai, D. R., Shahinas, D., Buzina, A., Pollock, R. A., Lau, R., Khairnar, K., Wong, A., Farrell, D. J., Green, K. & other authors (2009). Genome-wide dissection of globally emergent multi-drug resistant serotype 19A *Streptococcus pneumoniae*. *BMC Genomics* **10**, 642.

Pletz, M. W., Maus, U., Krug, N., Welte, T. & Lode, H. (2008). Pneumococcal vaccines: mechanism of action, impact on epidemiology and adaption of the species. *Int J Antimicrob Agents* **32**, 199–206.

Prymula, R., Peeters, P., Chrobok, V., Kriz, P., Novakova, E., Kaliskova, E., Kohl, I., Lommel, P., Poolman, J. & other authors (2006). Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both *Streptococcus pneumoniae* and non-typable *Haemophilus influenzae*: a randomised double-blind efficacy study. *Lancet* **367**, 740–748.

Rennels, M. B., Edwards, K. M., Keyserling, H. L., Reisinger, K. S., Hogerman, D. A., Madore, D. V., Chang, I., Paradiso, P. R., Malinoski, F. J. & other authors (1998). Safety and immunogenicity of heptavalent pneumococcal vaccine conjugated to CRM197 in United States infants. *Pediatrics* **101**, 604–611.

Richter, S. S., Heilmann, K. P., Dohrn, C. L., Riahi, F., Beekmann, S. E. & Doern, G. V. (2009). Changing epidemiology of antimicrobial-resistant *Streptococcus pneumoniae* in the United States, 2004–2005. *Clin Infect Dis* **48**, e23–e33.

Rodgers, G. L., Arguedas, A., Cohen, R. & Dagan, R. (2009). Global serotype distribution among *Streptococcus pneumoniae* isolates causing otitis media in children: potential implications for pneumococcal conjugate vaccines. *Vaccine* **27**, 3802–3810.

Sá-Leão, R., Nunes, S., Brito-Avô, A., Frazão, N., Simões, A. S., Crisóstomo, M. I., Paulo, A. C. S., Saldanha, J., Santos-Sanches, I. & other authors (2009). 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* **15**, 1002–1007.

Scott, D. A., Komjathy, S. F., Hu, B. T., Baker, S., Supan, L. A., Monahan, C. A., Gruber, W., Siber, G. R. & Lockhart, S. P. (2007). Phase 1 trial of a 13-valent pneumococcal conjugate vaccine in healthy adults. *Vaccine* **25**, 6164–6166.

Shakir, E., Cameron, C., Denham, B. & McMenamin, J. (2009). Respiratory and immunisation quarterly report. *HPS Wkly Rep* **43**, 407–408.

Silva, N. A., McCluskey, J., Jefferies, J. M. C., Hinds, J., Smith, A., Clarke, S. C., Mitchell, T. J. & Paterson, G. K. (2006). Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. *Infect Immun* **74**, 3513–3518.

Sjöström, K., Spindler, C., Ortqvist, A., Kalin, M., Sandgren, A., Kühlmann-Berenzon, S. & Henriques-Normark, B. (2006). Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. *Clin Infect Dis* **42**, 451–459.

Smith, T., Lehmann, D., Montgomery, J., Gratten, M., Riley, I. D. & Alpers, M. P. (1993). Acquisition and invasiveness of different serotypes of Streptococcus pneumoniae in young children. *Epidemiol Infect* **111**, 27–39.

Spratt, B. G. & Greenwood, B. M. (2000). Prevention of pneumococcal disease by vaccination: does serotype replacement matter? *Lancet* **356**, 1210–1211.

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

Tocheva, A. S., Jefferies, J. M. C., Christodoulides, M., Faust, S. N. & Clarke, S. C. (2010). Increase in serotype 6C pneumococcal carriage, United Kingdom. *Emerg Infect Dis* **16**, 154–155.

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

Väkeväinen, M., Eklund, C., Eskola, J. & Käyhty, H. (2001). Cross-reactivity of antibodies to Type 6B and 6A polysaccharides of *Streptococcus pneumoniae*, evoked by pneumococcal conjugate vaccines, in infants. *J Infect Dis* **184**, 789–793.

Weinberger, D. M., Trzcinski, K., Lu, Y., Bogaert, D., Brandes, A., Galagan, J., Anderson, P. W., Malley, R. & Lipsitch, M. (2009). Pneumococcal capsular polysaccharide structure predicts serotype prevalence. *PLoS Pathog* **5**, e1000476.

Wysocki, J., Tejedor, J. C., Grunert, D., Konior, R., Garcia-Sicilia, J., Knuf, M., Bernard, L., Dieussaert, I. & Schuerman, L. (2009). Immunogenicity of the 10-valent pneumococcal non-typeable *Haemophilus influenzae* protein D conjugate vaccine (PHiD-CV) when coadministered with different *Neisseria meningitidis* serogroup C conjugate vaccines. *Pediatr Infect Dis J* **28**, S77–S88.

Zhang, Q., Choo, S. & Finn, A. (2002). Immune responses to novel pneumococcal proteins Pneumolysin, PspA, PsaA, and CbpA in adenoidal B cells from children. *Infect Immun* **70**, 5363–5369.

www.expert-reviews.com/toc/erv/11/8

EXPERT REVIEWS

Pneumococcal 13-valent conjugate vaccine for the prevention of invasive pneumococcal disease in children and adults

Expert Rev. Vaccines 11(8), 889–902 (2012)

Rebecca A Gladstone[1], Johanna M Jefferies[1,2] Saul N Faust[1,3,4] and Stuart C Clarke*[1,2,4]

[1]*Academic Unit of Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, UK*
[2]*Health Protection Agency, Public Health Laboratory, Southampton, UK*
[3]*NIHR Wellcome Trust Clinical Research Facility, University Hospital Southampton NHS Foundation Trust, Southampton, UK*
[4]*Southampton NIHR Respiratory Biomedical Research Unit, University Hospital Southampton NHS Foundation Trust, Southampton, UK*
**Author for correspondence:*
Tel.: +44 2 380 798 895
Fax: +44 2 380 796 992
s.c.clarke@southampton.ac.uk

Pneumococcal disease remains a global problem despite the availability of effective conjugate vaccines. The 13-valent pneumococcal conjugate vaccine (PCV13) extends the valency of PCV7 by including six additional serotypes highly associated with invasive pneumococcal disease (IPD). Comparisons between PCV13 and PCV7 or the pneumococcal polysaccharide vaccine have established noninferiority of PCV13 for both safety and immunogenicity profiles for use in children and adults, respectively. At the end of 2011, PCV13 had been approved and launched in 104 countries worldwide, with 54 including the vaccine in their pediatric national immunization program. Surveillance data from early adopters of PCV13 has indicated reductions are occurring in both overall IPD and IPD caused by the six non-PCV7 serotypes; early reports of serotype replacement in carriage are also emerging. While serotype replacement for PCV7 was observed to varying degrees for both carriage and disease, the extent to which this will occur for PCV13 is yet to be determined.

Keywords: conjugate • efficacy • herd protection • immunogenicity • pneumococcal vaccines • *Streptococcus pneumoniae*

Streptococcus pneumoniae, also known as pneumococcus, is responsible for a significant burden of infectious disease worldwide [1] and is estimated to have caused greater than 800,000 deaths per annum in children aged less than 5 years in the prevaccine era [2]. While the majority of this burden lies in the developing world [3], the burden in Europe and within the UK is still substantial with high mortality and morbidity associated with cases [4,5]. Manifestations include invasive pneumococcal disease (IPD; defined by isolation of pneumococci from normally sterile sites such as blood and cerebrospinal fluid), pneumonia and acute otitis media (AOM). Despite its pathogenic record, pneumococci are also carried asymptomatically, residing in the nasopharynx of approximately one-third of children under the age of 5 years in developed countries [6,7] and as many as 45–70% of children under 5 years in the developing world [8,9].

The high economic burden and fatality rates of IPD resulted in the production and licensing of a capsular polysaccharide-based 23-valent vaccine in the early 1980s (PPV23) (Table 1). Infants are a primary risk group for IPD and do not have fully developed T-cell-independent responses to polysaccharides, which prevents effective PPV23 use in this population [10]. The immunogenicity of saccharide–protein conjugations in infants has long been established by stimulating the production of memory cells through T-cell help [11]. This knowledge gave rise to the development of polyvalent pneumococcal conjugate vaccines (PCV) (Table 1).

In 2000 and 2006, the 7-valent PCV7 (Prevenar™, Pfizer) was added to the national immunization programs (NIPs) of the USA and UK, respectively. Surveillance of pneumococcal disease and carriage in countries that have added PCV7 to their NIPs has revealed that PCV7 is extremely effective at reducing both carriage

10.1586/ERV.12.68

ISSN 1476-0584

and disease cases of vaccine types (VT). A 98% decrease in pneumococcal VT IPD was observed in the UK for children under 2 years of age [12] and a 69% reduction in the carriage of VT in children of 4 years of age and under [6]. The USA reported similar success in reducing VT IPD with a >92% decrease in VT IPD for all age groups combined [13,14], near eradication of VT carriage was also described [7]. Despite the proven efficacy of PCV7 against VT, the replacement of VT with non-vaccine types (NVT) has gone some way to offset total reductions in IPD [12,14]. Although absolute rate increases of NVT disease in the USA were reported to be relatively small, NVT replacement disease was more pronounced in the UK. An 84% decrease in total VT IPD only resulted in a total IPD reduction of 34% for all age groups combined by 2009/10 when compared to pre-PCV7 UK surveillance data [12]. A move towards higher valency vaccines was made in response to observed replacement disease. Two US FDA and EMA approved PCVs with increased valency currently exist. These are GlaxoSmithKline's (GSK) 10-valent pneumococcal *Haemophilus influenzae* protein D conjugate vaccine (PCV10, Synflorix™) and Pfizer's 13-valent pneumococcal conjugate vaccine (PCV13, Prevenar 13™). While PCV13 clearly exceeds the valency of PCV10, the inclusion of the *H. influenzae* protein D in PCV10 offers the potential of additional protection against nontypeable *H. influenzae* (NTHi) disease including AOM. Characterizing the effectiveness of these two vaccines therefore involves complex comparisons owing to their different formulations [15–17]. In February 2010, the US Advisory Committee on Immunization Practises recommended PCV13 for all children aged 2–59 months [18], and in April 2010 PCV13 superseded PCV7 in the UK's routine childhood immunization schedule [19]. A summary of licensed pneumococcal vaccines and their trade names can be found in Table 1.

PCV13 formulation

Developed by Wyeth (since taken over by Pfizer) as an extension to their 7-valent vaccine, Prevenar 13 includes the purified capsular polysaccharide of the seven PCV7 serotypes plus six additional serotypes (Table 1). The additional serotypes were selected as they are the remaining major global causes of IPD [20]. Protection against vaccine-related serotypes, these being serotypes that belong to a serogroup which is represented in PCV7 but that are not contained within the vaccine themselves, was not as successful as first expected. For example, serotype 19A's IPD incidence has increased in the USA, UK and elsewhere despite the fact that the related serotype 19F was included in PCV7 [12,21,22]. This led to the inclusion of serotype 19A in PCV13. The inclusion of serotype 6A in PCV13, however, is expected to offer functional protection against 6C [23], a serotype that was observed to clonally expand after PCV7 introduction in both carriage and disease [6,12,24,25]. Early observations support this prediction with decreases in 6C IPD and carriage after PCV13 use [26,27]. Serotypes 1 and 5 are particularly associated with outbreaks [28,29] and significantly contribute to the burden of pneumococcal disease in the developing world [30,31]. Their inclusion in the vaccine is expected to further reduce the burden of global pneumococcal disease [28,29,31].

The 13 pneumococcal polysaccharides are all covalently bound to a carrier protein, CRM_{197}, the nontoxic diphtheria toxin variant, to form glycoconjugates [32,101]. One 0.5 ml prefilled syringe contains approximately 2.2 µg of each serotype, except for 6B of which there is 4.4 µg, resulting in the inclusion of approximately 32 µg of the CRM_{197} protein conjugate [32,101]. In addition to the saccharide–protein conjugates, each dose contains 100 µg of polysorbate 80 and 295 µg succinate buffer, which act as pharmacologically inactive stabilizers for the active ingredients. The formulation is completed with approximately 125 µg of the adjuvant aluminum in the form of aluminum phosphate ($AlPO_4$) [32,101].

Table 1. Serotypes included in pneumococcal vaccines.

Vaccine	Trade name	Manufacturer	Conjugation	Serotypes
PCV7	Prevenar™	Wyeth (now Pfizer)	Yes	4, 6B, 9V, 14, 18C, 19F, 23F
PCV10	Synflorix™	GlaxoSmithKline	Yes	**1**, 4, **5**, 6B, **7F**, 9V, 14, 18C, 19F, 23F
PCV13	Prevenar 13™	Pfizer	Yes	1, **3**, 4, 5, **6A**, 6B, 7F, 9V, 14, 18C, **19A**, 19F, 23F
PPV23	Pneumovax®	Merck	No	1, **2**, 3, 4, 5, 6B, 7F, **8**, **9N**, 9V, **10A**, **11A**, **12F**, 14, **15B**, **17F**, 18C, 19F, 19A, **20**, **22F**, 23F, **33F**

Underlined serotypes indicate those not included in PCV-7.
Bold serotypes are those not included in any vaccine of lower valency.
PCV: Pneumococcal conjugate vaccine; PPV: Pneumococcal polysaccharide vaccine.

Safety & immunogenicity

To enable licensing, noninferiority trials can be conducted to provide immunological evidence that a medical intervention is not inferior to an existing intervention. PCV13 has been compared with PCV7 for licensing in children and to PPV23 in adults. Analogous results to PCV13's predecessors were required for both its safety and immunogenicity profile.

Safety

PCV13 has been shown to have a safety profile largely equivalent to PCV7 or PPV23 for young children and adults respectively [32–41]. For children, any site or systemic reactions were mild or moderate, predominantly comprising tenderness or swelling at the injection site or mild fever [32–35,37–41]. A minority of these pediatric studies observed serious adverse events (SAEs) that were atypical of childhood maladies; one vaccine-related SAE occurred in each of the two studies by Yeh *et al.* and Gadzinowski *et al.*, with no subsequent deaths [32,33].

Common adverse reactions that occur in ≥10% of adults of 50 years and older included diarrhea, headache, rash and other generic symptoms and administrative site-specific reactions [101].

Immunogenicity

As PCV7 is established in the marketplace and post-licensure surveillance has been on-going, data were available to allow calculation of immunogenic thresholds, which correlate to protection from IPD in children, using the relationship between invasive disease vaccine efficacy and antibody concentrations. Cut-offs for serotype-specific antibody concentrations could only be determined for 19F, thus a serotype-independent antibody threshold was used [42]. Analysis of these relationships led Jódar *et al.* to support the 0.2 µg/ml control-adjusted pneumococcal antibody concentration as a minimum cut-off and the corresponding 1:8 opsonophagocytic assay (OPA) titer (a measure of antibody activity), considered to be predictive of protection against invasive disease in vaccinated children [42]. The WHO also recommends 0.2–0.35 µg/ml and 1:8 OPA titers as putative predictions of protection [43,102,103]. The extension of the therapeutic indications for PCV13 IPD prevention in adults aged 50 years and older was based on noninferiority to PPV23 and a pneumococcal conjugate has not been used in this population before. Therefore, the true correlates of protection will not be available until efficacy data is obtained from PCV13 use in adults. They may or may not be similar to the immunogenicity requirements of children.

Nunes and Madhi comprehensively reviewed the immunogenicity and safety of PCV13 in young children previously for *Expert Review of Vaccines* in 2011 and reported that after the primary vaccination series most PCV7 serotypes in PCV13 produced IgG concentrations and functional antibody responses comparable to those seen with PCV7 vaccination [40]. Administration of the booster dose, at 12–24 months depending on the study, resulted in further reductions of any disparity for this age group [40]. Additionally, their review of PCV13 clinical trials revealed that immunogenic analysis for those serotypes included in PCV13 but not PCV7 surpassed presumptive protection thresholds and were noninferior in all cases except for serotype 3, where there were disparities between studies. Five out of 14 studies examined by Nunes and Madhi reported a reduction of the geometric mean titer of anti-3 capsular IgG antibody from primary series to after the booster dose. This suggests a lack of immunological memory, which may impact on the long-term efficacy of this vaccine against serotype 3 disease [37,40]. Suboptimal immunogenicity results for serotype 3 in the experimental GSK vaccine resulted in the exclusion of this serotype from the final PCV10, suggesting that this is a property of the serotype which has a particularly thick capsule [17]. OPAs have subsequently provided evidence of functional anti-serotype 3 antibody production which is predicted to confer the protection needed based on correlates or immunogenicity [18,33]. PCV13 vaccine efficacy estimates from the UK of 62% and 66% for serotypes 1 and 3, respectively, suggest functional protection; however, the confidence intervals were reported to cross into negative values [26].

PCV13 is the first and currently the only pneumococcal conjugate vaccine licensed for use in adults. Adults at increased risk of pneumococcal disease may have been vaccinated with the polysaccharide-only vaccine PPV23. However, there has been debate surrounding the effectiveness of PPV23 and the quality of the studies that have reported data on its effectiveness in those aged >50 years, particularly in the UK [44,104–106]. Amendments to the therapeutic indications of PCV13 were extended to include protection against IPD only in adults aged 50 years or older. This was authorized on the basis of the noninferiority to PPV23 immunogenicity alone (Table 2) [107]. The clinical trial subjects accurately reflected this age group in terms of chronic diseases and immunogenicity was not seen to differ between healthy adults and those with chronic disease [107]. A number of studies in adults have shown that PCVs can induce immunogenicity comparable to or exceeding levels produced by PPV23 for those serotypes shared with the PCV, although these studies primarily compare PPV23 with PCV7 and it is still not established what the threshold of protection for pneumonia alone is [45–49]. Serotype 6A is unique to PCV13, therefore noninferiority could not be demonstrated; instead, the accepted immunogenicity threshold was a fourfold increase in capsular-specific antibody response when compared to the control population. For adults previously vaccinated with PPV23 there have been reports of blunting of the immune response to PCV when compared to those vaccinated with PCV without prior PPV23 vaccination. However, this does not result in noninferiority and therefore there may not be clinical repercussions to this observation [50–52]. The phenomenon of pneumococcal polysaccharide hyporesponsiveness is also observed

Table 2. Clinical immunogenicity trials in adults.

ClinicalTrials.gov identifier	Pfizer study number	Study function	Age (years)	Vaccination status
NCT00427895	004	PCV13 immunological noninferiority to PPV23	50–64	PPV not received
NCT00492557	3008	PCV13 + TIV immunological noninferiority to PCV13 alone	≥65	PPV not received
NCT00500266	3000	PCV13 safety in elderly that had previously received PPV	≥68	≥1 dose PPV received ≥3 years previous
NCT00521586	3001	PCV13 + TIV immunological noninferiority to PCV13 alone	50–59	PPV not received
NCT00546572	3005	PCV13 immunological noninferiority to PPV23	≥70	≥1 dose PPV received ≥5 years previous

Summary of pivotal clinical trials that supported the amendment of the therapeutic indications of PCV13, to include prevention of invasive pneumococcal disease in adults.
PCV: Pneumococcal conjugate vaccine; PPV: Pneumococcal polysaccharide vaccine; TIV: Trivalent influenza vaccine.
Data taken from [107] and Pfizer.

for subsequent doses of PPV23 itself and mimics the reduction in immunological response to vaccination after real infection, suggesting the mechanisms may be related [52,53].

Although conjugates are of considerable lower valency, PCV13 does contain a predominance of serotypes causing disease in the elderly and other at-risk adults, but its use may result in replacement disease by serotypes contained within the PPV23 [54]. An adult's mature immune system is capable of responding to polysaccharide alone without conjugation; however, this does not always translate into clinical efficacy.

Pneumonia is the key contributor to adult pneumococcal disease [55,56]. Huss *et al.* performed an extensive meta-analysis that revealed, when only trials of high methodological quality were included, that PPV offers no protection against pneumonia and that its effectiveness against IPD may only be 50% [44]. Adults who may benefit from PCV13 belong to a diverse set of risk groups that are immunocompromised to a varying extent and primarily include the elderly but also those with chronic conditions, including HIV-infected individuals and those with chronic obstructive pulmonary disease (COPD). PCV7 trials in adults with COPD have shown that the conjugated serotypes were more immunogenic than with PPV23 vaccination and, furthermore, a HIV-patient trial showed that PCV7 was protective in this population where PPV23 was not previously found to be protective [47,57].

Dosing

Publicly accessible information on PCV13 product characteristics are available on the electronic Medicines Compendium website [101]. The following section uses pertinent information contained within this resource. The document states that for active immunization against invasive disease, pneumonia and AOM caused by the 13 pneumococcal serotypes, PCV13 should be administered as a 0.5 ml intramuscular injection in the vastus lateralis thigh muscle in infants or in the upper arm deltoid muscle for other young children and toddlers. The manufacturer primarily recommend a three-dose primary series and a booster dose (3 + 1) for routine infant immunization of children under 6 months of age [101]. The manufacturer also endorsed for children <6 months, and later extended to those under 11 months, a reduced dose schedule of 2 + 1. The 2 + 1 series has been shown to be immunologically equivalent to the 3 + 1 in a study of the PCV13 experimental precursor PCV9 and is therefore considered to offer comparable protection [58]. The USA and UK adopted 3 + 1 and 2 + 1 schedules, respectively, for their NIPs [18,19,108]. However, the WHO has recommended an alternative expanded program on immunization (EPI) schedule for PCV of 6 weeks, 10 weeks and 14 weeks as a highly accelerated, booster-deficient dosing schedule for children, aiming to protect infants as early in life as possible. This has been adopted by the Global Alliance for Vaccines and Immunisation (GAVI) but is not officially endorsed by the manufacturer [59].

For those that have previously received PCV7 immunization, the product characteristics document advocates switching to PCV13 at any point in the schedule and provides guidelines on how to complete the immunization course, as summarized in Table 3. This is supported by a study by Grimprel *et al.*, which showed that the continuation of the vaccination schedule with PCV13 in children that have previously received PCV7 is both safe and immunogenic [35]. Additional recommendations are made by the manufacturers, the US CDC and the UK Department of Health for at-risk older children, adults and those over the age of 50 years, and are also summarized in Table 3. The PCV13 summary of product characteristics are based on the results of clinical immunogenicity trials (Table 2) and state that PCV13 can be administered to adults of 50 years and older regardless of their PPV vaccination status. PPV23 was shown to reduce IgG responses to PCV13 in pivotal clinical immunogenicity trials in adults (Table 2) compared to PCV13 alone, indicating that PPV23 not only blunts the immune response to itself but also to conjugates. Conversely, PCV13 followed by PPV23 augments the immunogenicity of PPV23 [52,53,60]. Therefore, in circumstances where PPV23 administration is considered appropriate, it is recommended that PCV13 be given first [101,107]. This is a single dose of PCV13; the need for a booster has not yet been determined [101]. Currently the USA and UK do not recommend routine immunization with PCV13 for adults over the age of 50 years.

Concomitant vaccination

Concomitant vaccination should always be given at different injection sites for PCV13 [101]. Published studies have gone on to show that concomitant vaccination of children with routine vaccines against diphtheria, tetanus, pertussis, *H. influenzae* type b, polio, hepatitis B, meningococcal serogroup C, measles, mumps, rubella and varicella are both safe and immunogenic and do not interfere with the immune response to coadministered vaccinations [39,61–63]. The trivalent influenza vaccine (TIV) is routinely administered in the elderly and other at-risk adult populations; therefore, it is important that concomitant vaccination of PCV13 and TIV is possible. Schwarz *et al.* reported that coadministration of TIV with PCV13 resulted in lower IgG geometric mean concentrations (GMCs), but was not inferior to PCV13 alone with the exception of 19F, which was just below the threshold [36]. The mechanism and clinical significance is not yet fully understood; however, the concomitant vaccination in adults has been deemed to be satisfactory for both its safety and immunogenicity [36].

Cost–effectiveness

Vaccine effectiveness can be estimated in advance using models in the absence of surveillance data. Analysis of vaccine cost–effectiveness is used as a guide for decision-making in healthcare by assessing the burden of disease. Measures such as life-years gained (LYG) and disability-adjusted life-years (DALY) or quality-adjusted life-years (QALY) are used to assess an intervention in terms of cost–benefit, taking into account the worth of lives saved or extended rather than just the costs associated with implementation and healthcare. DALY and QALY also include indicators of the quality of life. The effectiveness of PCV13 over PCV7 in children in the USA was modeled for a 10-year period using PCV7 data by Rubin *et al.* They estimate that PCV13 use will result in the prevention of 106,000 cases of IPD and savings of

Table 3. Recommendations for dosing by age group and source.

Group	Age group	PCV7 dose	Schedule	Primary series	Interval	Booster age
Pfizer						
Children	6 weeks–6 months	0	3 + 1	3	1 month	11–15 months
	6 weeks–6 months	0	2 + 1	2	2 months	11–15 months
	7–11 months	0	2 + 1	2	1 month	>2 years
	12–23 months	0	2	2	2 months	NA
	25 years	0	1	1	NA	NA
	0–6 weeks	0	0	0	NA	NA
Catch-up children	12–23 months	1	2	2	2 months	NA
	2–5 years	2	1	1	NA	NA
Adults	>50 years	0	1	1	NA	NA
Department of Health, UK						
Children eligible for routine immunization	2–12 months	0	2 + 1	2	2 months	12–13 months
	12 months–2 years	0	1	1	NA	NA
Catch-up children and at-risk children	2 months–2 years	1	1 + 1	1	NA	13 months
	2 months–2 years	2	Booster	0	NA	13 months
At-risk children with asplenia/splenic dysfunction or immunosuppressed	12 months–5 years	0	2	2	2 months	None
All at-risk children	2–12 months	0	2 + 1	2	2 months	12–13 months
At-risk children without asplenia/splenic dysfunction or immunosuppression	12 months–5 years	0	1	1	NA	NA
All at-risk children	>5	0	NR	NA	NA	NA
At-risk adults	All ages	0	NR	NA	NA	NA
Adults	>65 years	0	NR	NA	NA	NA
US CDC						
Children eligible for routine immunization	2–6 months	0	3 + 1	3	2 months	12–15 months
	7–11 months	0	2 + 1	2	2 months	12–15 months
Children	12–23 months	0	2	2	2 months	None
	24–59 months	0	1	1	NA	None
At-risk children	24–59 months	0	2	2	2 months	None
Catch-up children and at-risk children	2–11 months	1	2 + 1	2	2 months	≥12 months
	12–23 months	1	2	2	2 months	None
	2–23 months	2	+1	0	NA	≥12 months
	2–23 months	3	+1	0	NA	≥12 months
Children eligible for routine immunization	2–23 months	4	+1	0	NA	14–59 months
At-risk children	2–23 months	4	+1	0	NA	14–71 months
Adults	>50 months	NR	NR	NA	NA	NA
At-risk adults	All ages	NR	NR	NA	NA	NA
WHO						
Children eligible for routine immunization	6 weeks–12 months	0	3	3	4 weeks	None

Summary of the risk groups recommended for vaccination, dosing schedule and recommending body. Accurate as of 10 January 2012.
NA: Not applicable; NR: Not recommended.
Data taken from [75,101,108].

approximately US$ 11 billion over the course of a decade. Further cost savings were associated with a catch-up program that targets older children no longer eligible for routine vaccination; however, the negative effect of replacement disease cannot be accurately predicted in models [64]. On mainland Europe mathematical models have also been used to predict effectiveness and compare vaccines. One Italian mathematical model of PCV13 vaccination in Florence simulated 5 years of cost–benefit analysis in the pediatric population. The predictions made support the catch-up program for children aged less than 24 months, recommended by the Italian Ministry of Health, and suggests that expanding the catch-up program to children up to 5 years old would be beneficial for this region. This particular study did not model the impact of PCV10 [65]. Strutton *et al.* have produced data that predicted that a PCV13 NIP would prevent >30% of the current cases of IPD in Germany, Greece and The Netherlands, and would be more cost effective than either PCV7 or PCV10 [66]. Conversely, estimates from a Norwegian study that incorporated the burden of both pneumococcal and NTHi related diseases including IPD, community-acquired pneumonia (CAP) and AOM suggested that PCV10 would be capable of further cost savings over PCV13, in part owing to the efficacy of the vaccine against *H. influenzae* AOM [17,67].

To accurately estimate effectiveness and perform valid comparisons, models must try to take into account both the indirect effects of vaccination including herd protection and additional complexities with the accurate assessment of healthcare and social costs. In the case of PCV10, the impact on AOM caused by *H. influenzae* results in additional complexities and requires further assumptions. The validity of a model is dependent upon the assumptions and data on which it is built. The geographical differences in the epidemiology of disease and in structuring and management of healthcare systems mean that predictions are often specific to the countries from which the data is gathered, with limited extrapolation globally.

In Argentina, an upper middle income country, routine immunization with either PCV13 or PCV10 was similarly predicted to be cost effective. PCV13 would result in higher LYG and DALYs averted due to pneumococcal disease through its higher valency, but when the prevention of all-cause AOM is taken into account PCV10 was predicted to save further healthcare costs through its action against *H. influenzae* AOM. The model assumed PCV10 would have 33.6% efficacy against nontypeable *H. influenzae*, responsible for 40% of cases in Argentina, based on efficacy trials performed by Prymula *et al* [17,68]. Similar reports have been made elsewhere in Latin America and these are reviewed in a comprehensive systematic evaluation by Giglio *et al* [69].

The Gambia is considered one of the lowest income GAVI-eligible countries and has a significant burden of pneumococcal disease. Kim *et al.* used PCV9 data to predict the cost–effectiveness of PCV10 and PCV13 in The Gambia when compared to no vaccination using DALY and the GAVI-negotiated price of US$3.50 per dose [70]. PCV13 was consequently estimated to avert an extra 180 DALYs when compared to PCV10 with a $240 greater saving of $910 over PCV10 associated with each averted DALY for this country [70].

Decisions on the pneumococcal vaccine selection and schedule will vary, dependent on the needs of the country in which it is to be implemented. Estimates of cost–effectiveness play an important role in these decisions and have to be balanced by the resources available, burden of disease, highest risk groups, manifestations and alternative healthcare priorities.

Effectiveness in adults

Despite the proven efficacy of pneumococcal vaccination against IPD, complicating the issue of efficacy and effectiveness of any pneumococcal vaccination of adults is the protection offered against nonsystemic pneumonia. This is due to the difficulties in confirming the etiology to a species and serotype level without invasive procedures. Effectiveness against pneumonia without bacteremia has previously been reported to be low for PPV23 in the elderly, who are at high risk of developing this manifestation [44,71]. Effectiveness data for PCV13 in adults is lacking, as amendments to the therapeutic indications were recently granted on noninferiority immunogenicity studies for IPD in adults. The absence of substantial PCV13 effectiveness data for IPD or pneumonia in adults, difficulties in directly comparing a high-valency free polysaccharide vaccine and a 13-valent conjugate, PPV23-induced hyporesponsiveness, and serotype replacement all contribute complications to decision-making in adult pneumococcal vaccination. Cost–effectiveness of PCV13 versus PPV23 in adults using QALY has recently been modeled for populations in the USA. In the article, PCV13 was predicted to be more cost effective than PPV23, except in scenarios where the effectiveness of PCV13 against nonbacteremic pneumonia was assumed to be low or herd protection from the childhood NIP was assumed to be high [72]. Again, the lack of effectiveness data made it difficult for the authors to decide which assumptions are valid, resulting in wide parameters and alternative scenarios which had differing results.

Nonetheless, conjugate vaccines have been shown to be effective in children against IPD and nonbacteremic pneumonia [73], with good safety and comparable or superior immune stimulation for the included serotypes. Furthermore, highly powered and large-scale clinical trials of PCV13, including the Dutch CAPiTA randomized control trial which has enrolled in excess of 85,000 patients, started in 2008, will help to determine the best course of action for adult at-risk populations – results may be available as soon as 2013 [74].

Surveillance of effectiveness

Surveillance data from early adopters of routine PCV13 immunization are becoming available for analysis of the impact of pediatric PCV13 use. In the UK, Miller *et al.* reported that IPD in children aged <2 years old caused by serotypes unique to PCV13 were reduced by half during the study period [12,26]. Additionally, they conclude that a single dose of PCV13 in an older child catch-up program would further reduce the IPD burden. In the USA the CDC reported that between May 2010 and April 2011 there

were 135 cases of IPD caused by the additional PCV13 serotypes; of these a complete vaccination history was available for 81 children. In total, 78% were eligible for, but had not received PCV13 vaccination as part of the catch-up program, predominantly in children aged 24–59 months. The CDC therefore concluded that the transition to PCV13 was suboptimal in this age group in part due to reduced attendance at clinics [75]. In Germany both PCV13 and PCV10 are in use, and since 2010 early reports suggest that IPD, due to serotypes 1, 3 and 7F contained within the higher valency vaccines, has already reduced. Reductions in 19A IPD have not yet been observed in this population; this may be owing to suboptimal conjugate vaccine use and uptake as 19A is only targeted by PCV13 and not PCV10 [76]. The exact contribution of PCV13 to the decreases is impossible to determine owing to the shared additional serotypes in both PCV13 and PCV10 [76].

The efficacy of conjugate vaccines against IPD is center stage in clinical trials. Nonetheless, a considerable economic burden and sequela is associated with chronic or recurrent acute OM in young children and prolonged or periodic hospitalization from uncomplicated pneumonia in the elderly. This makes assessment of vaccine efficacy for the latter diseases important. Published efficacy data for pneumonia and AOM is sparse even for PCV7 owing to complications in case definition and determination of the definitive etiological agent without invasive procedures, in an environment where more than one species may be contributing. A randomized double-blind study performed by Prymula *et al.* has provided evidence that PCV10 offered protection against AOM not only for AOM caused by *H. influenzae* but additionally for AOM caused by the conjugated pneumococcal serotypes, which can be plausibly extrapolated to the action of PCV13 on VT AOM, though no head-to-head studies exist [17]. Studies are in progress, which should increase our understanding of efficacy for these diseases and the role of asymptomatic carriage in diseases such as AOM, increasing the evidence base with which conjugate vaccine comparisons can be made with further validity [27,74,77].

Serotype replacement after introduction of PCV7 has been well documented globally for both carriage and disease, albeit to varying extents. The extended valency of PCV13 is still limited when compared to the total pneumococcal serotypes, including a number known to be associated with disease. It is therefore biologically plausible to suggest that this is likely to arise for PCV13 in time. Changes to the prevalence of PCV13 NVT in carriage is expected in consequent years in the UK, as seen with PCV7 serotypes [6]. Attal *et al.* are one of the first groups to have reported a reduction in carriage of PCV13-specific serotypes among children with AOM in France, since PCV13 introduction in 2010 [27].

Serotype replacement in disease is yet to be reported for PCV13 but early reports from a French carriage study on AOM combined with knowledge of PCV7 serotype replacement suggests that this may follow. Any shift in the epidemiology of serotypes that results will be of particular clinical importance if replacement involves any NVTs that have a high invasive potential [78].

Although data from the USA and UK suggest that PCV13 is effective for the prevention of IPD, the extent to which NVTs will contribute to IPD and carriage is yet to be established. Limited data are available on PCV7 serotype replacement for developing countries. Only two GAVI-eligible countries, The Gambia and Rwanda, had been using PCV7 prior to PCV13, since 2009. Consequently it is difficult to predict the effect of this phenomenon for both disease and carriage in geographically distinct regions of the developing world [6,12,14,79–81]. High pneumococcal carriage rates may even augment increases in NVT disease in populations where there is high-level transmission and potentially an improved environment for horizontal gene transfer, including the exchange of the capsule locus that codes for the serotype [9]. As vaccination and surveillance continues, a clearer picture will emerge as to the additional efficacy of PCV13. These data will be vastly supplemented by those additional data from the increasing number of countries introducing PCV13.

Global conjugate vaccination

A significant proportion of pneumococcal disease still occurs in the developing world [3], populations that historically have not introduced PCVs into routine vaccination schedules until the intervention of not-for-profit organizations such as the WHO and GAVI. For example, GAVI secured an advance market commitment (AMC) for pneumococcal vaccination whereby GAVI committed money to the purchase of vaccines before development, acting as an incentive for companies to bring products to market. The pharmaceutical companies, in turn, agreed to a sustainable fixed price that will be maintained after GAVI funding runs out, allowing countries to budget for the future. As of November 2011, 37 of the 73 GAVI-eligible countries had been approved for PCV implementation [Accelerated Vaccine Introduction – progress report. Report to the GAVI Alliance Board. Pers. comm. (2011)]. Fifteen of the 19 that have selected a specific pneumococcal vaccine have so far opted for PCV13 while Kenya, Ethiopia, Pakistan and Madagascar have selected PCV10. These choices may be due to the fact that the cost–benefit analysis against IPD is simpler to interpret and inform decisions than those analyses that also take into account all-cause AOM, where PCV10 may have an advantage [17].

By the end of 2011, 14 of these 15 countries had introduced PCV13. For a further 18 developing countries, the introduction of a currently unspecified PCV will occur during 2012 and 2013. Such widespread conjugate vaccine implementation through the Accelerated Vaccine Introduction (AVI) initiative has raised concerns that supply shortages may occur which could influence conjugate vaccine choice; in response to this, AVI are actively ensuring an increase in availability is made to answer demand.

An estimated 74–82% of IPD cases in Africa, Asia and South America could be prevented by vaccination with PCV13 [3]. Vaccination with PCV13 over PCV7 in the developing world has the added benefit of active immunization against serotypes 1 and 5, which rank in the top three serotypes, behind the PCV7 serotype 14, for causing IPD in young children in GAVI alliance eligible countries [3]. However, African studies using a 9-valent PCV formulation reported that immune responses to serotype 1 did not always reach protective levels in a 3 + 0 schedule, with vaccine failures occurring in vaccinated children >1 years old [82–84]. These results are possibly a result of this alternative dosing schedule lacking a booster dose. The efficacy of PCV13 against IPD caused by serotype 1 in Africa

Table 4. Countries that include PCV13 in their national immunization program and their dosing schedules.

Country	Dosing schedule	Launched
Africa		
Morocco	2 + 1	2010
Cameroon†	3 + 0	2011
Democratic Republic of Congo†	3 + 0	2011
Benin†	3 + 0	2011
Central African Republic†	3 + 0	2011
The Gambia†	3 + 0	2011
Malawi†	3 + 0	2011
Mali†	3 + 0	2011
Rwanda†	3 + 0	2011
Sierra Leone†	3 + 0	2011
South Africa	3 + 0	2011
Burundi†	3 + 1	2011
Asia		
Israel	2 + 1	2010
Kazakhstan	2 + 1	2010
Oman	3 + 0	2010
Yemen†	3 + 0	2011
Hong Kong	3 + 1	2010
Macau	3 + 1	2011
Singapore	3 + 1	2010
Bahrain	3 + 1	2010
Kuwait	3 + 1	2010
Qatar	3 + 1	2010
Saudi Arabia	3 + 1	2010
United Arab Emirates	3 + 1	2010
Europe		
Belgium	2 + 1	2011
Denmark	2 + 1	2010
France	2 + 1	2010
Greenland	2 + 1	2010
Ireland	2 + 1	2010
Italy	2 + 1	2010
Lichtenstein	2 + 1	2011
Luxembourg	2 + 1	2010
Norway	2 + 1	2010
Sweden	2 + 1	2010

Summary of PCV13 national immunization program globally and the dosing schedule.
†Use supported by the Global Alliance for Vaccines and Immunisation (GAVI).
Data supplied by GAVI Alliance media & communications and Pfizer.

remains to be determined but serotype 1 in PCV9 and PCV13 are identical in formulation and bio-chemical preparation [Pfizer, Pers. Comm.]. Klugman *et al.* suggest that a booster dose is required and can be administered as early as 9 months and up to 2 years of age to give protective immunization against serotype 1. This has repercussions for nations that have their PCV13 vaccinations supported by GAVI and use this accelerated 3 + 0 schedule [85].

At the close of 2011, PCV13 had been approved for use in 120 countries globally [Pfizer Marketing Department, Pers. Comm.]. Fifty four of these countries have already added PCV13 to their NIPs (Table 4), whether or not the doses will be provided free of charge and the resultant uptake rate will be specific to individual countries and will be in-part dependent on the healthcare structure of these countries. PCV10 is also used globally and is available in some countries alongside PCV13 [76]; however, a definitive list of countries is not freely available at present. Encouragingly for the developing world, the implementation of PCV13 in 26% of these countries was supported by GAVI. Dosing schedules for the 54 countries largely relates to the continent in which they reside. Asia predominantly has a 3 + 1 schedule and Africa a 3 + 0 schedule, reflecting GAVI support in this region, and European countries primarily use a 2 + 1 schedule. Schedules across North America (including the USA, Canada and Central American countries) were typically more varied (Table 4). The dosing schedules of 2 + 1, 3 + 0 and 3 + 1 were evenly split between the total 54 NIPs. In addition, 50 countries have launched PCV13 but have not yet included it within their immunization program and a further 13 have approved PCV13 but are yet to decide on a dosing schedule and launch its use. Globally the most frequently chosen dosing schedule has marginally been the two dose primary series with a booster, with 21, 17 and 16 countries using 2 + 1, 3 + 1 and 3 + 0, respectively.

Future pneumococcal vaccines

An investigational 15-valent PCV is being developed by Merck [86]. This investigational vaccine is formulated in a similar manner to PCV13, except for the inclusion of two additional serotypes, 22F and 33F [86]. Expanding valency of vaccines is a response to serotype replacement observed globally for carriage [6,7] and disease caused by PCV7 types and, by extrapolation, expected for PCV13 types [12,87]. Serotypes 22F and 33F were selected for inclusion due to increases in their prevalence in IPD, resulting in these serotypes being primary causes of IPD after PCV7 vaccine types [54,86]. A Phase I trial has been completed for healthy adults and toddlers of the 15-valent experimental vaccine (NCT01215175). A Phase II, randomized, multicenter, double-blind clinical trial of the vaccine is now being carried out by Merck to assess its non-inferiority to Prevenar 13™ in healthy infants, with its primary outcome due to complete in November 2012 (NCT01215188 [107]).

At least one new expanded valency vaccine is nearing the market; however, the uptake of third-generation conjugate vaccines will in part depend on the extent to which disease replacement is observed after widespread routine use of PCV13, specifically changes in the epidemiology of the additional serotypes. The ability to further expand the valency of conjugates in response to replacement is limited by both manufacturing restraints,

hypothetical antigenic overload (especially when these vaccines are routinely given concomitantly with immunogens from a wide range of pathogens) and the potentially prohibitive costs per dose incurred through expanding the valency.

Conjugate vaccines do not prime broad-spectrum pneumococcal immunity as they are composed of purified, free or conjugated polysaccharides from limited serotypes which promote serotype replacement. Serotype-independent vaccines would circumvent this issue and a number are in development, including killed whole cell vaccines, with formulations being tested in mouse models to determine the optimal aluminum adjuvant stimulation of the desired cytokine profiles and the efficacy against both carriage and disease [88,89]. Single and multicomponent pneumococcal protein-based vaccines are also being investigated by analyzing the protective capacities of a number of different proteins, protein combinations, protein fusions and adjuvants in mice [90–92]. It is important to remember that alternative pathogens are documented to colonize vaccinated individuals [93], intricate interactions balance this ecological niche and the pneumococcus is a coevolved commensal organism. While the pneumococcal capsule is established to be the primary virulence factor and a minority of serotypes perpetuate disease [78], it may be detrimental to try to completely remove pneumococcus with the use of 'universal' pneumococcal vaccines. The focus should arguably be restricted to clinical need.

Expert commentary

Data from multiple studies have shown PCV13 to be safe for both children and adults aged 50 years and above, even when given in combination with other vaccines as part of routine vaccination. Adverse effects are predominantly mild. Importantly, PCV13 has been shown to be at as least as immunogenic as PCV7 in children and PPV23 in adults, for the serotypes that they share in noninferiority clinical trials. These studies have gone on to show that the six additional serotypes to PCV7 are immunogenic and largely generate responses that are considered to be protective. A subset of pediatric studies, however, has highlighted that long-term immunogenicity to serotype 3 may be suboptimal with reduced antibody levels after a booster dose. Although immunological studies have raised concerns as to the strength of immunological memory to serotype 3 without a serotype-3 specific correlate of protection, there is as yet little evidence to demonstrate that this will translate into clinical vaccine failures. Furthermore, the cause of serotype-1 vaccine failures in Africa is not yet fully established. Serotype 1 is known to have a high invasive potential [78]. To protect against serotypes for which progression from carriage to disease can be swift, it can be supposed that protective levels of functional antibodies in the circulation and mucosa are required to prevent disease. In the absence of protective antibody levels, even when memory can be demonstrated, the lag between stimulation of the memory population at colonization to produce protective levels of antibody may be time enough for disease to take hold for pneumococci that are strongly associated with disease rather than carriage. Administration of a booster dose after the primary series may contribute to maintaining systemic and mucosal antibody while children remain in a primary risk age group.

Table 4. Countries that include PCV13 in their national immunization program and their dosing schedules (cont.).

Country	Dosing schedule	Launched
Europe (cont.)		
Switzerland	2 + 1	2011
UK	2 + 1	2010
Hungary	2 + 1	2010
Slovakia	2 + 1	2010
Germany	3 + 1	2009
Greece	3 + 1	2010
Spain	3 + 1	2010
Cyprus	3 + 1	2010
Czech Republic	3 + 1	2010
Turkey	3 + 1	2010
North America		
El Salvador	2 + 1	2010
Canada	2 + 1	2010
Mexico	2 + 1	2011
Honduras†	3 + 0	2011
Nicaragua†	3 + 0	2010
USA	3 + 1	2010
Costa Rica	3 + 1	2010
South America		
Uruguay	2 + 1	2010
Guyana†	3 + 0	2011
Oceania		
Australia	3 + 0	2011

Summary of PCV13 national immunization program globally and the dosing schedule.
†Use supported by the Global Alliance for Vaccines and Immunisation (GAVI).
Data supplied by GAVI Alliance media & communications and Pfizer.

Post-licensure evaluation of safety and effectiveness and routine national surveillance is essential for assessment of conjugate vaccine efficacy and the impact of replacement disease. Surveillance of bacterial carriage can provide important epidemiological information and clues to replacement patterns, but does not provide the whole picture and does not necessarily lead to accurate predictions of the effect on invasive disease [6,12].

Widespread use of PCV13 will encourage the establishment of herd protection and this will contribute to its economic viability and impact on disease through these indirect effects. However, the extent to which this will occur depends on the vaccine uptake and effectiveness of catch up programs.

Five-year view

Data on new interventions requires time to be generated. More than half of the countries of the world, across all inhabitable

continents, had approved PCV13 usage by 2012. Publication of completed clinical trials combined with appropriate post-licensure and national disease surveillance will be required to fully evaluate PCV13. Future data gathered will firstly help appraise the benefit of routine PCV13 vaccination in at-risk adults. Second, it will assess the efficacy of vaccination for serotypes 1 and 3 and noninvasive disease and finally evaluate the importance of booster doses. Ultimately these findings, along with any further VT replacement, will contribute to the lifespan of PCV13 use.

The use of PCV13 in adults older than 50 years is likely to increase following the licensure of PCV13 in this age group, and further still on the release of additional data on efficacy against pneumonia from studies including the CAPITA trial [74]. For immunocompromised children of all ages that have received PCV7 or those not previously vaccinated, many experts now recommend the sole use of PCV13 in two doses at least 1 month apart, without additional PPV23 vaccination, owing to observations of hyporesponsiveness and questionable efficacy. Some centers now test for serotype-specific immunogenicity to allow identification of those immunosuppressed children who require additional doses, a strategy that could also be considered for immunocompromised adults.

The extent of serotype replacement in both carriage and disease will need to be closely observed, given the clonal diversity of the pneumococcus, as it is a logical albeit not definite step once higher valency PCVs are implemented [94]. However, global differences in serotype replacement have demonstrated that many elusive factors are involved in this phenomenon. Within 5 years, alternative conjugate vaccines are likely to have reached the market and serotype-independent vaccines may well have progressed to clinical trials.

Financial & competing interests disclosure

JM Jefferies has received consulting fees from GlaxoSmithKline. SN Faust acts as principal investigator for clinical trials conducted on behalf of the University Hospital Southampton NHS Foundation Trust/University of Southampton that are sponsored by vaccine manufacturers but receives no personal payments from them. They have participated in advisory boards for vaccine manufacturers but receive no personal payments for this work. SC Clarke currently receives unrestricted research funding from Pfizer Vaccines (previously Wyeth Vaccines) and has participated in advisory boards and expert panels for GlaxoSmithKline, Pfizer and Novartis. They are investigators on studies conducted on behalf of the University Hospital Southampton NHS Foundation Trust/the University of Southampton/HPA that are sponsored by vaccine manufacturers but receives no personal payments from them. SN Faust, SC Clarke and JM Jefferies have received financial assistance from vaccine manufacturers to attend conferences. All grants and honoraria are paid into accounts within the respective NHS Trusts or Universities, or to independent charities. SN Faust receives support from the National Institute for Health Research (NIHR) funding for the Southampton Wellcome Trust Clinical Research Facility and the Southampton NIHR Respiratory Biomedical Research Unit. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Key issues

- Reports from early adopters of pneumococcal conjugate vaccine (PCV13), including the UK and USA, suggest that PCV13 is effective at protecting against the six additional pneumococcal serotypes, as surveillance has revealed a reduction in pediatric invasive pneumococcal disease caused by these serotypes.
- PCV13 is safe and at least as immunogenic as PCV7 or PPV23 for the shared serotypes for children and the elderly, respectively. However, serotype 3 has not exhibited strong immunological memory in a number of pediatric studies.
- Evidence supports the use of both pediatric 2 + 1 and 3 + 1 dosing schedules. Conversely, studies using the 3 + 0 booster-deficient schedule have suggested a lack of immunological memory for serotype 1, which can be rectified with a booster dose.
- High valency conjugate vaccines should be considered for primary pneumococcal immunization in adults to prevent pneumococcal disease, based on PPV effectiveness and PCV13 immunogenicity; data from current trials will further inform this decision on a country-by-country basis.
- PCV13 is rapidly being employed globally in over 100 countries to help combat the burden of pneumococcal disease in pediatric populations with 26% of PCV13 national immunization programs implemented in developing countries in 2011, enabled by the support of the Global Alliance for Vaccines and Immunisation (GAVI) for PCV13 or PCV10.
- Trials are ongoing that will shed light on the effectiveness of PCV13 in adults and for noninvasive pneumococcal diseases, including nonbacteremic pneumonia and acute otitis media.

References

Papers of special note have been highlighted as:
• of interest
•• of considerable interest

1 Mulholland K. Global burden of acute respiratory infections in children: implications for interventions. *Pediatr. Pulmonol.* 36(6), 469–474 (2003).

2 O'Brien KL, Wolfson LJ, Watt JP *et al.*; Hib and Pneumococcal Global Burden of Disease Study Team. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. *Lancet* 374(9693), 893–902 (2009).

3 Johnson HL, Deloria-Knoll M, Levine OS *et al.* Systematic evaluation of serotypes causing invasive pneumococcal disease among children under five: the pneumococcal global serotype project. *PLoS Med.* 7(10), e1000348 (2010).

4 Melegaro A, Edmunds WJ, Pebody R, Miller E, George R. The current burden of pneumococcal disease in England and Wales. *J. Infect.* 52(1), 37–48 (2006).

5 Isaacman DJ, McIntosh ED, Reinert RR. Burden of invasive pneumococcal disease and serotype distribution among *Streptococcus pneumoniae* isolates in young children in Europe: impact of the 7-valent pneumococcal conjugate vaccine and considerations for future conjugate vaccines. *Int. J. Infect. Dis.* 14(3), e197–e209 (2010).

6 Tocheva AS, Jefferies JM, Rubery H *et al*. Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of *Streptococcus pneumoniae* in young children. *Vaccine* 29(26), 4400–4404 (2011).

• **Effect of PCV7 on carriage of vaccine types.**

7 Hanage WP, Bishop CJ, Huang SS *et al*. Carried pneumococci in Massachusetts children: the contribution of clonal expansion and serotype switching. *Pediatr. Infect. Dis. J.* 30(4), 302–308 (2010).

8 Granat SM, Mia Z, Ollgren J *et al*. Longitudinal study on pneumococcal carriage during the first year of life in Bangladesh. *Pediatr. Infect. Dis. J.* 26(4), 319–324 (2007).

9 Bouskraoui M, Soraa N, Zahlane K *et al*. [Study of nasopharyngeal colonization by *Streptococcus pneumoniae* and its antibiotics resistance in healthy children less than 2 years of age in the Marrakech region (Morocco)]. *Arch. Pediatr.* 18(12), 1265–1270 (2011).

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

11 Insel RA, Anderson PW. Oligosaccharide-protein conjugate vaccines induce and prime for oligoclonal IgG antibody responses to the *Haemophilus influenzae* b capsular polysaccharide in human infants. *J. Exp. Med.* 163(2), 262–269 (1986).

12 Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. *Lancet Infect. Dis.* 11(10), 760–768 (2011).

13 Whitney CG, Pilishvili T, Farley MM *et al*. Effectiveness of seven-valent pneumococcal conjugate vaccine against invasive pneumococcal disease: a matched case-control study. *Lancet* 368(9546), 1495–1502 (2006).

14 Rosen JB, Thomas AR, Lexau CA *et al.*; CDC Emerging Infections Program Network. Geographic variation in invasive pneumococcal disease following pneumococcal conjugate vaccine introduction in the United States. *Clin. Infect. Dis.* 53(2), 137–143 (2011).

15 Talbird SE, Taylor TN, Knoll S, Frostad CR, García Martí S. Outcomes and costs associated with PHiD-CV, a new protein D conjugate pneumococcal vaccine, in four countries. *Vaccine* 28(Suppl. 6), G23–G29 (2010).

16 De Wals P, Black S, Borrow R, Pearce D. Modeling the impact of a new vaccine on pneumococcal and nontypable *Haemophilus influenzae* diseases: a new simulation model. *Clin. Ther.* 31(10), 2152–2169 (2009).

17 Prymula R, Peeters P, Chrobok V *et al*. Pneumococcal capsular polysaccharides conjugated to protein D for prevention of acute otitis media caused by both *Streptococcus pneumoniae* and non-typable *Haemophilus influenzae*: a randomised double-blind efficacy study. *Lancet* 367(9512), 740–748 (2006).

• **PCV10 efficacy against acute otitis media.**

18 CDC. Licensure of a 13-valent pneumococcal conjugate vaccine (PCV13) and recommendations for use among children – Advisory Committee on Immunization Practices (ACIP). *MMWR Morb. Mortal. Wkly Rep.* 59(9), 258–261 (2010).

19 Jensch I, Gámez G, Rothe M *et al*. PavB is a surface-exposed adhesin of *Streptococcus pneumoniae* contributing to nasopharyngeal colonization and airways infections. *Mol. Microbiol.* 77(1), 22–43 (2010).

20 Hausdorff WP, Feikin DR, Klugman KP. Epidemiological differences among pneumococcal serotypes. *Lancet Infect. Dis.* 5(2), 83–93 (2005).

21 Jacobs MR, Good CE, Bajaksouzian S, Windau AR. Emergence of *Streptococcus pneumoniae* serotypes 19A, 6C, and 22F and serogroup 15 in Cleveland, Ohio, in relation to introduction of the protein-conjugated pneumococcal vaccine. *Clin. Infect. Dis.* 47(11), 1388–1395 (2008).

22 Pillai DR, Shahinas D, Buzina A *et al*. Genome-wide dissection of globally emergent multi-drug resistant serotype 19A *Streptococcus pneumoniae*. *BMC Genomics* 10, 642 (2009).

23 Cooper D, Yu X, Sidhu M, Nahm MH, Fernsten P, Jansen KU. The 13-valent pneumococcal conjugate vaccine (PCV13) elicits cross-functional opsonophagocytic killing responses in humans to *Streptococcus pneumoniae* serotypes 6C and 7A. *Vaccine* 29(41), 7207–7211 (2011).

24 Nahm MH, Lin J, Finkelstein JA, Pelton SI. Increase in the prevalence of the newly discovered pneumococcal serotype 6C in the nasopharynx after introduction of pneumococcal conjugate vaccine. *J. Infect. Dis.* 199(3), 320–325 (2009).

25 Tocheva AS, Jefferies JM, Christodoulides M, Faust SN, Clarke SC. Increase in serotype 6C pneumococcal carriage, United Kingdom. *Emerging Infect. Dis.* 16(1), 154–155 (2010).

26 Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Effectiveness of the new serotypes in the 13-valent pneumococcal conjugate vaccine. *Vaccine* 29(49), 9127–9131 (2011).

•• **Early demonstration of additional PCV13 types effectiveness against invasive pneumococcal disease.**

27 Attal S, Bingen E, Bonnet E *et al*. Impact of 13-valent pneumococcal conjugate vaccine (PCV13) on nasopharyngeal (NP) flora in children with acute otitis media (AOM). Presented at: *51st International conference on Antimicrobial Agents and Chemotherapy (ICAAC)*, Chicago, IL, USA, 17–20 September 2011.

28 Balicer RD, Zarka S, Levine H *et al*. Control of *Streptococcus pneumoniae* serotype 5 epidemic of severe pneumonia among young army recruits by mass antibiotic treatment and vaccination. *Vaccine* 28(34), 5591–5596 (2010).

29 Gupta A, Khaw FM, Stokle EL *et al*. Outbreak of *Streptococcus pneumoniae* serotype 1 pneumonia in a United Kingdom school. *BMJ* 337, a2964 (2008).

30 Hausdorff WP, Bryant J, Paradiso PR, Siber GR. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin. Infect. Dis.* 30(1), 100–121 (2000).

31 Hausdorff WP. The roles of pneumococcal serotypes 1 and 5 in paediatric invasive disease. *Vaccine* 25(13), 2406–2412 (2007).

32 Gadzinowski J, Albrecht P, Hasiec B *et al*. Phase 3 trial evaluating the immunogenicity, safety, and tolerability of manufacturing scale 13-valent pneumococcal conjugate vaccine. *Vaccine* 29(16), 2947–2955 (2011).

33 Yeh SH, Gurtman A, Hurley DC *et al.*; 004 Study Group. Immunogenicity and safety of 13-valent pneumococcal conjugate

vaccine in infants and toddlers. *Pediatrics* 126(3), e493–e505 (2010).

34 Frenck R Jr, Thompson A, Yeh SH *et al.*; 3011 Study Group. Immunogenicity and safety of 13-valent pneumococcal conjugate vaccine in children previously immunized with 7-valent pneumococcal conjugate vaccine. *Pediatr. Infect. Dis. J.* 30(12), 1086–1091 (2011).

35 Grimprel E, Laudat F, Patterson S *et al.* Immunogenicity and safety of a 13-valent pneumococcal conjugate vaccine (PCV13) when given as a toddler dose to children immunized with PCV7 as infants. *Vaccine* 29(52), 9675–9683 (2011).

• **Key evidence for progression from PCV7 to PCV13.**

36 Schwarz TF, Flamaing J, Rümke HC *et al.* A randomized, double-blind trial to evaluate immunogenicity and safety of 13-valent pneumococcal conjugate vaccine given concomitantly with trivalent influenza vaccine in adults aged ≥65 years. *Vaccine* 29(32), 5195–5202 (2011).

•• **Concomitant PCV13 and influenza vaccination in the elderly.**

37 Kieninger DM, Kueper K, Steul K *et al.*; 006 study group. Safety, tolerability, and immunologic noninferiority of a 13-valent pneumococcal conjugate vaccine compared to a 7-valent pneumococcal conjugate vaccine given with routine pediatric vaccinations in Germany. *Vaccine* 28(25), 4192–4203 (2010).

38 Huang LM, Lin TY, Juergens C. Immunogenicity and safety of a 13-valent pneumococcal conjugate vaccine given with routine pediatric vaccines in Taiwan. *Vaccine* 30(12), 2054–2059 (2012).

39 Snape MD, Klinger CL, Daniels ED *et al.* Immunogenicity and reactogenicity of a 13-valent-pneumococcal conjugate vaccine administered at 2, 4, and 12 months of age: a double-blind randomized active-controlled trial. *Pediatr. Infect. Dis. J.* 29(12), e80–e90 (2010).

40 Nunes MC, Madhi SA. Review on the immunogenicity and safety of PCV-13 in infants and toddlers. *Expert Rev. Vaccines* 10(7), 951–980 (2011).

41 Gurtman A, Tansey SP, Thompson A *et al.* Safety of 13-valent pneumococcal conjugate vaccine in infants and children: meta-analysis of 13 clinical trials in 9 countries. Presented at: *28th European Society for Pediatric Infectious Diseases (ESPID) Annual Meeting*, Nice, France, 4–8 May 2010.

42 Jódar L, Butler J, Carlone G *et al.* Serological criteria for evaluation and licensure of new pneumococcal conjugate vaccine formulations for use in infants. *Vaccine* 21(23), 3265–3272 (2003).

43 Feavers I, Knezevic I, Powell M, Griffiths E. Challenges in the evaluation and licensing of new pneumococcal vaccines, 7-8 July 2008, Ottawa, Canada. *Vaccine* 27(28), 3681–3688 (2009).

44 Huss A, Scott P, Stuck AE, Trotter C, Egger M. Efficacy of pneumococcal vaccination in adults: a meta-analysis. *CMAJ* 180(1), 48–58 (2009).

45 Lode H, Schmoele-Thoma B, Gruber W *et al.* Dose-ranging study of a single injection of pneumococcal conjugate vaccine (1 ×, 2 ×, or 4 ×) in healthy subjects aged 70 years or older. *Vaccine* 29(31), 4940–4946 (2011).

46 Goldblatt D, Southern J, Andrews N *et al.* The immunogenicity of 7-valent pneumococcal conjugate vaccine versus 23-valent polysaccharide vaccine in adults aged 50-80 years. *Clin. Infect. Dis.* 49(9), 1318–1325 (2009).

47 Dransfield MT, Nahm MH, Han MK *et al.*; COPD Clinical Research Network. Superior immune response to protein-conjugate versus free pneumococcal polysaccharide vaccine in chronic obstructive pulmonary disease. *Am. J. Respir. Crit. Care Med.* 180(6), 499–505 (2009).

48 Scott DA, Komjathy SF, Hu BT *et al.* Phase 1 trial of a 13-valent pneumococcal conjugate vaccine in healthy adults. *Vaccine* 25(33), 6164–6166 (2007).

49 Jackson LA, Neuzil KM, Nahm MH *et al.* Immunogenicity of varying dosages of 7-valent pneumococcal polysaccharide-protein conjugate vaccine in seniors previously vaccinated with 23-valent pneumococcal polysaccharide vaccine. *Vaccine* 25(20), 4029–4037 (2007).

50 Musher DM, Manof SB, Liss C *et al.* Safety and antibody response, including antibody persistence for 5 years, after primary vaccination or revaccination with pneumococcal polysaccharide vaccine in middle-aged and older adults. *J. Infect. Dis.* 201(4), 516–524 (2010).

51 Musher DM, Rueda AM, Nahm MH, Graviss EA, Rodriguez-Barradas MC. Initial and subsequent response to pneumococcal polysaccharide and protein-conjugate vaccines administered sequentially to adults who have recovered from pneumococcal pneumonia. *J. Infect. Dis.* 198(7), 1019–1027 (2008).

52 de Roux A, Schmöle-Thoma B, Schmöele-Thoma B *et al.* Comparison of pneumococcal conjugate polysaccharide and free polysaccharide vaccines in elderly adults: conjugate vaccine elicits improved antibacterial immune responses and immunological memory. *Clin. Infect. Dis.* 46(7), 1015–1023 (2008).

53 Jackson L, Gurtman A, van Cleeff M *et al.* 13-valent pneumococcal conjugate vaccine (PCV13) enhances the response to subsequent PCV13 and 23-valent pneumococcal polysaccharide (PPSV23) vaccinations in adults 50 years and older. Presented at: *Infectious Disease Society of America (IDSA) Annual Meeting*, Boston, MA, USA, 20–23 October 2011.

54 Pilishvili T, Lexau C, Farley MM *et al.*; Active Bacterial Core Surveillance/ Emerging Infections Program Network. Sustained reductions in invasive pneumococcal disease in the era of conjugate vaccine. *J. Infect. Dis.* 201(1), 32–41 (2010).

55 Jansen AG, Rodenburg GD, de Greeff SC *et al.* Invasive pneumococcal disease in the Netherlands: syndromes, outcome and potential vaccine benefits. *Vaccine* 27(17), 2394–2401 (2009).

56 Weycker D, Strutton D, Edelsberg J, Sato R, Jackson LA. Clinical and economic burden of pneumococcal disease in older US adults. *Vaccine* 28(31), 4955–4960 (2010).

57 French N, Gordon SB, Mwalukomo T *et al.* A trial of a 7-valent pneumococcal conjugate vaccine in HIV-infected adults. *N. Engl. J. Med.* 362(9), 812–822 (2010).

58 Goldblatt D, Southern J, Ashton L *et al.* Immunogenicity and boosting after a reduced number of doses of a pneumococcal conjugate vaccine in infants and toddlers. *Pediatr. Infect. Dis. J.* 25(4), 312–319 (2006).

59 WHO. Pneumococcal conjugate vaccine for childhood immunization – WHO position paper. *Wkly Epidemiol. Rec.*, 82(12), 93–104 (2007).

60 Jackson LA, Janoff EN. Pneumococcal vaccination of elderly adults: new paradigms for protection. *Clin. Infect. Dis.* 47(10), 1328–1338 (2008).

61 Gimenez-Sanchez F, Kieninger DM, Kueper K *et al.*; 501 and 006 study groups. Immunogenicity of a combination vaccine containing diphtheria toxoid, tetanus toxoid, three-component acellular pertussis, hepatitis B, inactivated polio

virus, and *Haemophilus influenzae* type b when given concomitantly with 13-valent pneumococcal conjugate vaccine. *Vaccine* 29(35), 6042–6048 (2011).

• **Safe concomitant immunization of PCV13 with routine childhood vaccinations.**

62 Miller E, Andrews N, Waight P *et al*. Safety and immunogenicity of coadministering a combined meningococcal serogroup C and *Haemophilus influenzae* type b conjugate vaccine with 7-valent pneumococcal conjugate vaccine and measles, mumps, and rubella vaccine at 12 months of age. *Clin. Vaccine Immunol.* 18(3), 367–372 (2011).

63 Esposito S, Tansey S, Thompson A *et al*. Safety and immunogenicity of a 13-valent pneumococcal conjugate vaccine compared to those of a 7-valent pneumococcal conjugate vaccine given as a three-dose series with routine vaccines in healthy infants and toddlers. *Clin. Vaccine Immunol.* 17(6), 1017–1026 (2010).

64 Rubin JL, McGarry LJ, Strutton DR *et al*. Public health and economic impact of the 13-valent pneumococcal conjugate vaccine (PCV13) in the United States. *Vaccine* 28(48), 7634–7643 (2011).

65 Boccalini S, Azzari C, Resti M *et al*. Economic and clinical evaluation of a catch-up dose of 13-valent pneumococcal conjugate vaccine in children already immunized with three doses of the 7-valent vaccine in Italy. *Vaccine* 29(51), 9521–9528 (2011).

66 Strutton DR, Farkouh RA, Earnshaw SR *et al*. Cost-effectiveness of 13-valent pneumococcal conjugate vaccine: Germany, Greece, and The Netherlands. *J. Infect.* 64(1), 54–67 (2012).

67 Robberstad B, Frostad CR, Akselsen PE, Kværner KJ, Berstad AK. Economic evaluation of second generation pneumococcal conjugate vaccines in Norway. *Vaccine* 29(47), 8564–8574 (2011).

68 Urueña A, Pippo T, Betelu MS *et al*. Cost–effectiveness analysis of the 10- and 13-valent pneumococcal conjugate vaccines in Argentina. *Vaccine* 29(31), 4963–4972 (2011).

69 Giglio N, Micone P, Gentile A. The pharmacoeconomics of pneumococcal conjugate vaccines in Latin America. *Vaccine* 29(Suppl. 3), C35–C42 (2011).

70 Kim SY, Lee G, Goldie SJ. Economic evaluation of pneumococcal conjugate vaccination in The Gambia. *BMC Infect. Dis.* 10, 260 (2010).

71 Jackson LA, Neuzil KM, Yu O *et al*.; Vaccine Safety Datalink. Effectiveness of pneumococcal polysaccharide vaccine in older adults. *N. Engl. J. Med.* 348(18), 1747–1755 (2003).

72 Smith KJ, Wateska AR, Nowalk MP, Raymund M, Nuorti JP, Zimmerman RK. Cost–effectiveness of adult vaccination strategies using pneumococcal conjugate vaccine compared with pneumococcal polysaccharide vaccine. *JAMA* 307(8), 804–812 (2012).

73 Black SBM, Shinefield HRM, Ling SMM *et al*. Effectiveness of heptavalent pneumococcal conjugate vaccine in children younger than five years of age for prevention of pneumonia. *Pediatr. Infect. Dis. J.* 21(9), 810–815 (2002).

74 Hak E, Grobbee DE, Sanders EA *et al*. Rationale and design of CAPITA: a RCT of 13-valent conjugated pneumococcal vaccine efficacy among older adults. *Neth. J. Med.* 66(9), 378–383 (2008).

75 CDC. Invasive pneumococcal disease and 13-valent pneumococcal conjugate vaccine (PCV13) coverage among children aged </=59 months – selected U.S. regions, 2010–2011. *MMWR Morb. Mortal. Wkly Rep.* 60, 1477–1481 (2011).

76 Imöhl M, Rückinger S, Linden MVD, Kries RV. First Effects of Immunization with Higher Valent Pneumococcal Conjugate Vaccines in German Children on Numbers of Reported Invasive Pneumococcal Disease (IPD) Cases. Presented at: *51st International Conference on Antimicrobial Agents and Chemotherapy (ICAAC)*, Chicago, IL, USA, 17–20 September 2011.

77 Shea KM, Weycker D, Stevenson AE, Strutton DR, Pelton SI. Modeling the decline in pneumococcal acute otitis media following the introduction of pneumococcal conjugate vaccines in the US. *Vaccine* 29(45), 8042–8048 (2011).

78 Brueggemann AB, Peto TE, Crook DW, Butler JC, Kristinsson KG, Spratt BG. Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. *J. Infect. Dis.* 190(7), 1203–1211 (2004).

79 Flasche S, Van Hoek AJ, Sheasby E *et al*. Effect of pneumococcal conjugate vaccination on serotype-specific carriage and invasive disease in England: a cross-sectional study. *PLoS Med.* 8(4), e1001017 (2011).

80 Kaplan SL, Mason EO Jr, Wald ER *et al*. Decrease of invasive pneumococcal infections in children among 8 children's hospitals in the United States after the introduction of the 7-valent pneumococcal conjugate vaccine. *Pediatrics* 113(3 Pt 1), 443–449 (2004).

81 Huang SS, Platt R, Rifas-Shiman SL, Pelton SI, Goldmann D, Finkelstein JA. Post-PCV7 changes in colonizing pneumococcal serotypes in 16 Massachusetts communities, 2001 and 2004. *Pediatrics* 116(3), e408–e413 (2005).

82 Saaka M, Okoko BJ, Kohberger RC *et al*. Immunogenicity and serotype-specific efficacy of a 9-valent pneumococcal conjugate vaccine (PCV-9) determined during an efficacy trial in The Gambia. *Vaccine* 26(29–30), 3719–3726 (2008).

83 Klugman KP, Madhi SA, Huebner RE, Kohberger R, Mbelle N, Pierce N; Vaccine Trialists Group. A trial of a 9-valent pneumococcal conjugate vaccine in children with and those without HIV infection. *N. Engl. J. Med.* 349(14), 1341–1348 (2003).

84 Madhi SA, Adrian P, Kuwanda L, Cutland C, Albrich WC, Klugman KP. Long-term effect of pneumococcal conjugate vaccine on nasopharyngeal colonization by *Streptococcus pneumoniae* – and associated interactions with *Staphylococcus aureus* and *Haemophilus influenzae* colonization – in HIV-infected and HIV-uninfected children. *J. Infect. Dis.* 196(11), 1662–1666 (2007).

85 Klugman KP, Madhi SA, Adegbola RA, Cutts F, Greenwood B, Hausdorff WP. Timing of serotype 1 pneumococcal disease suggests the need for evaluation of a booster dose. *Vaccine* 29(18), 3372–3373 (2011).

•• **Accelerated vaccination schedule may allow serotype-1 vaccine failures.**

86 Skinner JM, Indrawati L, Cannon J *et al*. Pre-clinical evaluation of a 15-valent pneumococcal conjugate vaccine (PCV15-CRM197) in an infant-rhesus monkey immunogenicity model. *Vaccine* 29(48), 8870–8876 (2011).

• **Investigational extended valency conjugate vaccine.**

87 Hicks L, Harrison L, Flannery B *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.* 196(9), 1346–1354 (2007).

88 Hogenesch H, Dunham A, Hansen B, Anderson K, Maisonneuve JF, Hem SL. Formulation of a killed whole cell pneumococcus vaccine – effect of aluminum adjuvants on the antibody and IL-17 response. *J. Immune Based Ther. Vaccines* 9, 5 (2011).

89 Lu Y-J, Leite L, Gonçalves VM *et al*. GMP-grade pneumococcal whole-cell vaccine injected subcutaneously protects mice from nasopharyngeal colonization and fatal aspiration-sepsis. *Vaccine* 28(47), 7468–7475 (2011).

90 Moschioni M, De Angelis G, Harfouche C *et al*. Immunization with the RrgB321 fusion protein protects mice against both high and low pilus-expressing *Streptococcus pneumoniae* populations. *Vaccine* 30(7), 1349–1356 (2012).

91 Cao J, Li D, Gong Y *et al*. Caseinolytic protease: a protein vaccine which could elicit serotype-independent protection against invasive pneumococcal infection. *Clin. Exp. Immunol.* 156(1), 52–60 (2009).

92 Olafsdottir TA, Lingnau K, Nagy E, Jonsdottir I. Novel protein-based pneumococcal vaccines administered with the Th1-promoting adjuvant IC31 induce protective immunity against pneumococcal disease in neonatal mice. *Infect. Immun.* 80(1), 461–468 (2012).

93 Murphy TF, Bakaletz LO, Smeesters PR. Microbial interactions in the respiratory tract. *Pediatr. Infect. Dis. J.* 28(Suppl. 10), S121–S126 (2009).

94 Jefferies JM, Clarke SC, Webb JS, Kraaijeveld AR. Risk of red queen dynamics in pneumococcal vaccine strategy. *Trends Microbiol.* 19(8), 377–381 (2011).

Websites

101 The Electronic Medicines Compendium. Prevenar 13 suspension for injection: summary of product characteristics. www.medicines.org.uk/EMC/medicine/22689/SPC/Prevenar+13+suspension+for+injection/

102 WHO. Recommendations for the production and control of pneumococcal conjugate vaccines. WHO Technical Report Series, No. 927 (2005). www.who.int/biologicals/publications/trs/areas/vaccines/pneumo/ANNEX%202%20PneumococcalP64-98.pdf

103 WHO. WHO/Health Canada consultation on serological criteria for evaluation and licensing of new pneumococcal vaccines. Executive summary. (2008). www.who.int/biologicals/publications/meetings/areas/vaccines/pneumococcal/Executive%20Summary%20FINAL%20version%205%20pneumo%2030_July_08%20_2_.pdf

104 Department of Health. JCVI advice on the pneumococcal vaccination programme for people aged 65 years and older. Department of Health, Gateway reference: 16380 (2011). www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_128537.pdf

105 Department of Health. JCVI advice on the pneumococcal vaccination programme for people aged 65 years and older. Department of Health, Gateway reference: 15777 (2011). www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_125244.pdf

106 Department of Health. JCVI statement on the routine pneumococcal vaccination programme for adults aged 65 years and older. 20 July 2011 JCVI minutes (2011). www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@ab/documents/digitalasset/dh_128704.pdf

107 ClinicalTrials.gov www.clinicaltrials.gov

108 Department of Health. Pneumococcal immunization against infectious disease. *The Green Book*. 295–313, Gateway reference: 7523 (2011). www.wp.dh.gov.uk/immunisation/files/2012/07/chap-25-dh_130544.pdf

Vaccine 30 (2012) 6738–6744

Contents lists available at SciVerse ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Review

Sampling methods for the study of pneumococcal carriage: A systematic review

R.A. Gladstone[a], J.M. Jefferies[a,b], S.N. Faust[a,b,c], S.C. Clarke[a,b,d,*]

[a] Academic Unit of Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, UK
[b] Southampton NIHR Respiratory Biomedical Research Unit, University Hospital Southampton NHS Foundation Trust, Southampton, UK
[c] NIHR Wellcome Trust Clinical Research Facility, University Hospital Southampton NHS Foundation Trust, Southampton, UK
[d] Health Protection Agency, Southampton, UK

ARTICLE INFO

Article history:
Received 20 July 2012
Received in revised form 20 August 2012
Accepted 31 August 2012
Available online 12 September 2012

Keywords:
Streptococcus pneumoniae
Carriage
Pneumococcal conjugate vaccines (PCV)

ABSTRACT

Streptococcus pneumoniae is an important pathogen worldwide. Accurate sampling of *S. pneumoniae* carriage is central to surveillance studies before and following conjugate vaccination programmes to combat pneumococcal disease. Any bias introduced during sampling will affect downstream recovery and typing. Many variables exist for the method of collection and initial processing, which can make inter-laboratory or international comparisons of data complex. In February 2003, a World Health Organisation working group published a standard method for the detection of pneumococcal carriage for vaccine trials to reduce or eliminate variability. We sought to describe the variables associated with the sampling of *S. pneumoniae* from collection to storage in the context of the methods recommended by the WHO and those used in pneumococcal carriage studies since its publication. A search of published literature in the online PubMed database was performed on the 1st June 2012, to identify published studies that collected pneumococcal carriage isolates, conducted after the publication of the WHO standard method. After undertaking a systematic analysis of the literature, we show that a number of differences in pneumococcal sampling protocol continue to exist between studies since the WHO publication. The majority of studies sample from the nasopharynx, but the choice of swab and swab transport media is more variable between studies. At present there is insufficient experimental data that supports the optimal sensitivity of any standard method. This may have contributed to incomplete adoption of the primary stages of the WHO detection protocol, alongside pragmatic or logistical issues associated with study design. Consequently studies may not provide a true estimate of pneumococcal carriage. Optimal sampling of carriage could lead to improvements in downstream analysis and the evaluation of pneumococcal vaccine impact and extrapolation to pneumococcal disease control therefore further in depth comparisons would be of value.

Contents

Abbreviations: STGG, skim milk-trypton-glucose-glycerin; SGG, skim milk-glycerol-glucose; ETA, egg-thioglycolate-antibiotic media; PCR, polymerase chain reaction; WHO, World Health Organisation; PCV-7, 7-valent pneumococcal conjugate vaccine; PCV-13, 13-valent pneumococcal conjugate vaccine; PHiD-CV, Pneumococcal, with Haemophilus influenzae protein D, conjugate vaccine.

* Corresponding author at: Infectious Disease Epidemiology Group, Mailpoint 814, Level C, Sir Henry Wellcome Laboratories, South Block, Southampton General Hospital, Southampton SO16 6YD, UK. Tel.: +44 23 8079 8895; fax: +44 23 8079 6992.
E-mail addresses: S.C.Clarke@southampton.ac.uk, s.c.clarke@soton.ac.uk (S.C. Clarke).

0264-410X/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.vaccine.2012.08.080

1. Introduction

Annually *Streptococcus pneumoniae* is estimated to cause 800,000 child deaths worldwide, through diseases including meningitis and pneumonia [2]. Pneumococci colonise the upper respiratory tract [3] therefore sampling of carriage, recovery and downstream typing allows surveillance of circulating pneumococci. Pneumococcal conjugate vaccines PCV-7, PCV-13 and PHiD-CV have been introduced since 2000 [4–6]. Any changes to the pneumococcal reservoir may impact on disease epidemiology, as colonisation precedes infection [7]. Serotype replacement has been observed for both pneumococcal carriage and disease since PCV implementation [8–10]. Carriage surveillance should adequately reflect the current pneumococcal population for accurate evaluation of vaccine policy, clarifying the link between carriage and disease. However, a wide range of methods exist for sampling *S. pneumoniae* carriage, varying in site, swab composition, transport medium and storage conditions. To allow valid comparisons between laboratories, countries or continents isolation of pneumococci needs to be as uniform as possible.

In 2003, a WHO Working Group sought to standardise the study of pneumococcal carriage in order to strengthen data collected [11], an extract of their recommendations, for collection and storage, are summarised in Table 1. This WHO publication is widely cited by pneumococcal carriage studies, despite being designed primarily for evaluating vaccine trials. Use of a standard method removes variability, but, unless it is clearly the optimal method, pneumococcal carriage could be underestimated. Analysis of pneumococcal populations traditionally requires isolation and accompanying typing methods – principally culture dependant. Culture can exert bias through competitive inhibition of growth and minority group masking, occurring on both a species and interspecies level. Culture independent techniques for positive detection of pneumococci exist [12] along with culture independent molecular methods, with or without viable recovery, screening for prevalent serotypes – particularly those associated with disease [13,14]. Phenotypic confirmation of capsular antigens requires culture, the gold standard method being Quellung [18]. Genotypic serotyping, including PCR methods, does not confirm the expression of a capsule, only the presence of the target gene which may not be functional and potentially be detected in a closely related species due to homologous recombination, but it can allow more focused phenotypic analysis [15]. Caution should also be used when dealing with serotyping based on sequence especially using small numbers of reference sequences as allelic variants of single serotypes such as 22F and 19F have been reported [16,17].

Table 1
Summary of WHO standard method for collection and storage of carried pneumococci.

Site	Swab	Swab transport media	Storage
Nasopharynx	Calcium alginate/Dacron®	Skim milk-tryptone-glucose-glycerin (STGG)	Swab shaft in STGG at −70°C

Description of the site, swab, swab transport media and storage conditions recommended by the WHO working group [10], for collection and storage of carried pneumococci.

Analysis of samples relies on efficiency of sampling and pre-analysis treatment which has been demonstrated to affect even culture independent analysis [12]. Any bias introduced, for or against detection or recovery, before analysis may influence final conclusions from surveillance. A large volume of literature details methods for pneumococci typing, while fundamental to pneumococcal epidemiology, this extensive subject warrants review and investigation outside the remit of this article. Therefore, we will discuss only preliminary stages of collecting carried pneumococci prior to capsular and genetic typing.

First it is important to consider that optimal sampling, detection and recovery will vary depending on the study aims and resources available. One universal method may not be feasible due to inherent inflexibility. However, an adaptable framework may allow studies to cater for circumstances such as international transport of specimens [19], and environmental temperate storage where temperature controlled environments are not available [20]. The best feasible assessment of *S. pneumoniae* prevalence in carriage and valid comparisons of carriage rate, when differing sensitivities of detection methods are unknown, requires detailed evaluation of the wide range of pneumococcal collection variables.

2. Sampling site

S. pneumoniae can be isolated from the nasal passage, nasopharynx and oropharynx [21–23]. Nasal washings [24] or nasopharyngeal aspirates permit collection of pneumococci, although aspirates are not usually available from healthy individuals [23,25]. Oropharyngeal swabbing has been demonstrated to be significantly poorer than nasopharyngeal swabbing for isolation of pneumococci in children [21,22] whilst the combined used of oropharyngeal and nasopharyngeal is suggested for sampling adults [22]. Nasal swabs are reported by Rapola et al. to give comparable results to nasopharyngeal swabs in Finland, although significant differences in isolation rates may be masked by small study size [23]. Abdullahi et al. reported that nasal washing is of increased sensitivity than nasopharyngeal swabbing but noted that nasal washings were less well tolerated by participants [24], demonstrating that ethical limitations may prevent use of methods with superior sensitivity.

No single sample can detect all pneumococcal colonisation, sampling a combination of sites or repeated sampling of a single site could increase detection rates [21,22,24]. The added benefit of detecting pneumococci or additional serotypes using multiple sites may not outweigh the discomfort/inconvenience to study subjects or change the clinical course of care for those with disease who may be sampled for carriage.

The most sensitive sample differs with age group, with implications for sampling carriage in the elderly to evaluate PCV use in this risk group [22,26,27]. No single study has compared all relevant sampling sites to determine the most sensitive and pragmatic sample source for any age group. Despite this, the current consensus is that the nasopharyngeal swab is most appropriate for sampling pneumococcal carriage.

3. Swab type

The swab material has not been examined extensively for its influence on detection sensitivity. Two studies from the 1930s

and 1940s suggested that fatty acids contained within cotton are inhibitory to pneumococci [28,29]. No further studies have confirmed or disputed the impact of cotton swabs. A study by Rubin et al. has however demonstrated some differences in both culture detection rate and culture independent PCR sensitivity that vary with swab composition for pneumococci [12]. Calcium alginate swabs were observed to be superior to rayon or Dacron® for culture or PCR detection, whilst Dacron®, was inferior to both calcium alginate and rayon for culture detection [12]. However during the 1990s there were suggestions that calcium alginate could be inhibiting PCR with conflicting results [30,31]. Nylon flocked swab have recently come to market whose multi-length fibres are proposed to improve sample absorbency and release of specimens, however there is little literature on the their use for nasopharyngeal bacterial sample collection, with one study on *Staphylococcus aureus* reporting no difference in sensitivity to rayon swabs [32]. Any unknown impact of swab composition could be greater in studies with longer transportation periods and those that store the original swab for retrospective analysis. As there appear to be differences between swabs it would be advantageous to determine the swab type which is least detrimental to identification of pneumococci in samples.

4. Swab transport medium

Swab transport medium is responsible for maintaining viability of bacteria before analysis whether it is transferred to a plate for culture or to storage medium. O'Brien et al. reported that, when compared to direct plating on solid media, use of the WHO-recommended STGG liquid medium resulted in subsequent equivalent rates of recovery [33]. Additionally greater numbers of pneumococci can be recovered from STGG than SGG over four days at 20–30 °C [34]. Gray described the superiority of ETA media to Amies or Stuarts media in situations where rapid processing or freezing (a requirement of numerous media types for maintaining pneumococci) is not possible. Relatively few studies have been performed to determine the optimal media for transportation and storage of pneumococci, comparing a limited number of media types [33–35].

5. Storage for recovery

The best possible recovery of pneumococcal isolates from the original samples is crucial for future analysis. Direct storage of the swab or swab contents lends itself to high sensitivity molecular techniques independent of culture which do not require abundant presence of viable bacteria; again this genotypic approach has its limitation in extrapolation to true phenotype and species specificity. Alternatively, storage of single colonies allows researchers to return to a specific cultured isolate. The WHO recommends the selection of two morphologically distinct presumptive pneumococcal colonies. Long term storage of isolates at −70 to −80 °C is considered the gold standard [11,36]. It is important to note that the act of long term storage itself or frequent freeze thaw cycles has been shown to alter the phenotype of pneumococci, for example sensitivity to optochin [37]. The WHO recommends STGG medium for long term storage, for the original swab and primary cultures. This medium has been shown to give enhanced ability for recovery of pneumococci on solid media than direct plating, after nine weeks storage, particularly for samples with low colony counts [33]. Trypton deficient SGG has however been shown to be as good as STGG for long term storage at −70 °C [34,38] which could reduce costs associated with carriage studies, important for developing countries where PCV evaluation may increase as GAVI supports PCV immunisation programs. Rubin et al. describe enhanced recovery of plated pneumococci when inoculated with a swab that has been stored in STGG, when compared to inoculation with STGG that contains a swab [12].

Any study using the supernatant for further analysis could benefit from sample vortexing, ensuring dispersal of microbial matter from the swab, endorsed by the WHO [11]. An alternative method for enhancement of pneumococcal recovery is use of a broth enrichment step, either prior to plating or prior to direct molecular assessment of serotype presence. Broth enrichment for culture and PCR has been reported to have increased sensitivity for detection of multiple serotype colonisation, results from these genotypic and phenotypic techniques are complementary in nature as no one technique detected all serotypes detected [14,39]. Broth enrichment detects more multiple serotypes and low density groups however it may still introduce some unknown culture bias with differential enrichment of some low density groups. Specimens are often shared and transported between laboratories for research, where the WHO recommend Dorset egg media for transportation of pneumococcal strains, Inverarity et al. point out that this product cannot be imported into certain countries due to avian influenzae related concerns about egg products [19].

Again there are not extensive evidence-based comparisons of long term storage options of recovery for pneumococci. Other media may have an equal or superior ability to STGG for maintaining pneumococci, including amendments such as enrichment, not yet investigated. Storage should endeavour to ensure that samples in long term storage are representative of the original sample.

A variety of methods exist to sample carried pneumococci population and studies could potentially employ a mix of techniques to collect pneumococci despite publication of a WHO standard method which includes collection guidelines. We conducted a systematic evaluation of published literature to determine the extent to which studies follow the WHO standard method for collection and storage and what the consensus variables were.

6. Methods

A search of published literature in the online PubMed database was performed on 1st June 2012, to identify published studies that collected pneumococcal carriage isolates, conducted after publication of the WHO standard method in February 2003 [11]. Quoted confidence intervals were calculated at 95% (CI). Quoted p values were generated using McNemars test on paired proportions to determine whether differences in proportions were statistically significant.

7. Search and exclusion strategy

Studies available on PubMed that contained 'carriage' and '*Streptococcus pneumoniae*' anywhere in the article, excluding reviews, were targeted by the search criteria. This search encompassed a broad range of articles to ensure all relevant articles were included. The search output was further filtered by excluding articles that did not report collection of carriage isolates of *S. pneumoniae* (Fig. 1), or articles that described samples already included for analysis.

8. Results

480 articles were returned by the PubMed database [1]. Of these articles, 375 were excluded, primarily as they commenced collection before the WHO publication (174) or did not collect carriage isolates (123) (Fig. 1). Selection returned 105 relevant articles (Supplementary reference list). These articles were examined in order to determine sample site, swab type, transport media and

Table 2
Collection and storage protocol – pooled data for sample site, swab type, swab transport media and storage, from studies post WHO publication of a standard method for detection of carriage pneumococci.

	Sample site	Swab type	Swab transport media	Storage
Choice not specified by article	**3/105** (3%)	**36/105** (34%)	**27/105** (26%)	**58/105** (55%)
Choice specified by article	**102/105** (97%)	**69/105** (66%)	**78/105** (74%)	**47/105** (45%)
Choice specified that conformed to WHO recommendation	**84/102** (82%)	**27/69** (39%)	**35/78** (45%)	**31/47** (66%)
Choice specified that was an alternative to WHO recommendation	**18/102** (18%)	**42/69** (61%)	**43/78** (55%)	**16/47** (34%)

Number of **articles** (%) for each criterion (swab, site, swab transport media) individually, which conformed to the pneumococcal carriage WHO standard method [10], for collection and storage, used an alternative to the WHO recommendation, or failed to specify the choice of site, swab, swab transport media or storage.

long-term storage conditions. Of these 105 studies, only 40 provided information for all four fields.

The information regarding site, swab, swab transportation media and storage conditions from all 105 included studies was pooled, to determine the consensus usage for each parameter individually (Table 2). Information that was not specified by an article was treated as missing data and excluded from proportions for the following analysis. The preferred sample site was the nasopharynx (82%, CI 75–90%). Alternative sites included the oropharynx, nasal passages and throat [40–42]. Fifteen percent of articles collected samples from two different sites while 1 article collected from a second site when collection from the first site was not possible [25].

For the majority of studies (61%, CI 50–73%), the swab chosen was not a WHO-recommended swab. Fourteen percent of studies (CI 6–22%) continued to use cotton swabs despite evidence that this could be detrimental to pneumococci.

Approximately 45% of studies, that specified their media choice, used the recommended STGG medium (CI 34–56%). Studies predominantly used Amies (24%, CI 15–33%), direct plating (15%, CI 7–23%) or Stuart's (10%, CI 3–17%) as an alternative. Greater than half of studies (55%, CI 46–65%) did not specify storage conditions and, of those that did (47 articles) more used <−70 °C STGG storage than an alternative (66%, CI 55–80%).

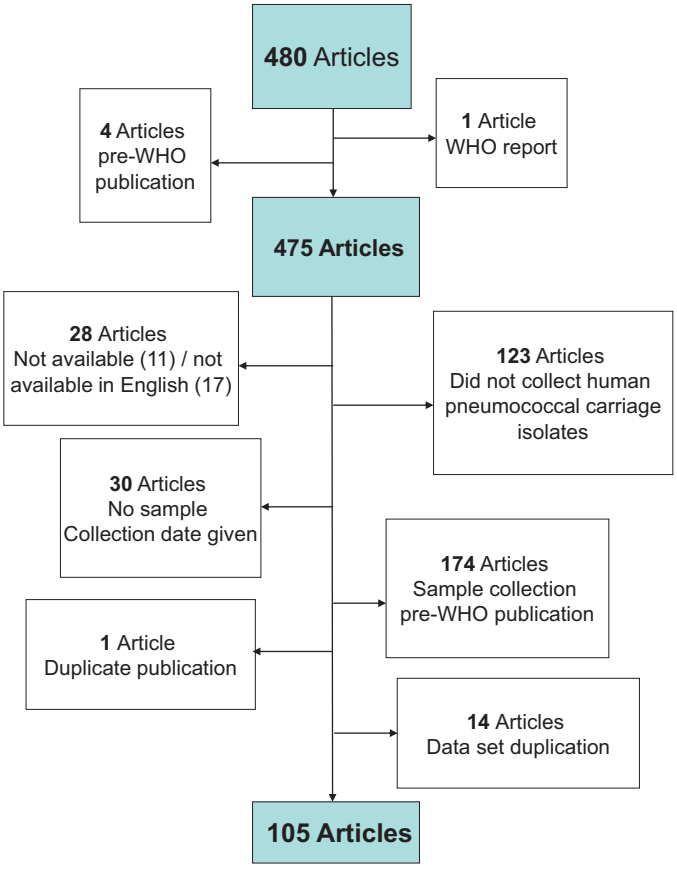


Fig. 1. Search results and exclusion of articles.

Table 3
Level of conformance to WHO standard method for collection and storage of carried pneumococci.

Conformance	0/3	1/3	2/3	3/3
Articles (%)	**4/66** (6%)	**26/66** (39%)	**17/66** (26%)	**19/66** (29%)

Number of **articles** (%) that conformed to the WHO standard method [10] for collection and storage of carried pneumococci, out of three, for swab, site and swab transport media – where specified (66 articles). Storage conditions were excluded from this analysis due to insufficient incidence of description, articles that did not specify their choice for any of swab, site or swab transport media, were also excluded.

66 studies provided full information for sample site, swab and media. Storage conditions were excluded from this analysis due to insufficient incidence of description. These 66 studies were analysed to determine the overall level of conformance to WHO guidelines by grouping them for conforming to none, one, two or three of the three WHO collection criteria (Table 3). Less than 29% of studies conformed exactly to the WHO recommendations for these three criteria (CI 18–40%), studies equally conformed to 2/3 criteria and 3/3 criteria. For the 26% (CI 15–37%) of studies that conformed to 2/3 WHO criteria the criterion that was not conformed to was primarily split between swab transport media type (6/17) and swab type (10/17). Those that did not use STGG for transport media and did not specify storage conditions are unlikely to have used STGG for storage. The McNemars test assessed the significance of differences in proportions for conforming to one but not the other of the following pairs; swab and site ($n = 69$), site and media ($n = 78$), or media and swab ($n = 66$). All articles were included where required information was specified. The greater proportion of studies using the nasopharynx but not calcium alginate/Dacron® (52%, CI 40–64%) as opposed to using calcium alginate or Dacron® but not the nasopharynx (3%, CI −1 to 7%) gave a p value of <0.001. The greater proportion of studies using the nasopharynx but not STGG (46%, CI 35–57%) as opposed to using STGG but not the nasopharynx (4%, CI 0–8%) also gave a highly significant p value of <0.001. The differing proportions of studies using STGG but not calcium alginate/Dacron® (18%, CI 9–28%), when compared to the proportion of studies using calcium alginate/Dacron® but not STGG (9%, CI 2–16%), was not statically significant $p = 0.238$. Use of the Nasopharynx and not calcium alginate/Dacron® or not STGG was observed significantly more often than expected by chance whereas there was no significant difference between using STGG and not calcium alginate/Dacron® and vice versa.

Temporal analysis of data, by year of the initiation of sample collection, revealed that the absolute number of articles adopting exactly the WHO standard method appeared to remain relatively stable (Fig. 2). However, the total number of studies that commenced collection and specified collection conditions each year, decreased over time, suggesting there has been a temporal increase in percentage uptake, which is difficult to interpret when such small numbers of studies represent latter years (Fig. 2).

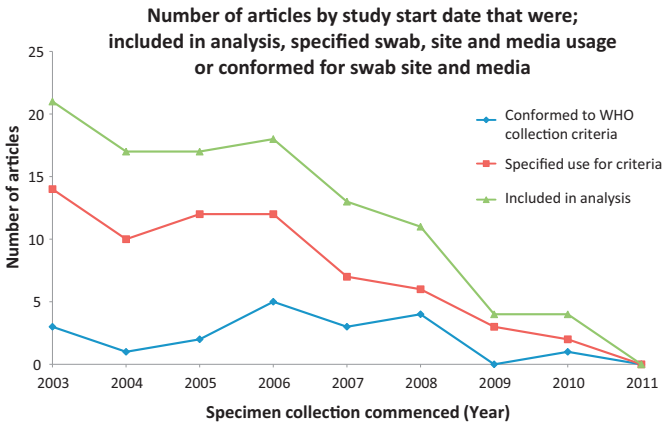


Fig. 2.

9. Discussion

Pneumococcal carriage studies are essential for evaluation of pneumococcal vaccines. The optimal method for pneumococcal sampling will allow the most accurate reporting of carriage. Detection of multiple pneumococcal serotypes [14,43–45], or co-colonising bacterial pathogens, is increasingly desirable to allow evaluation of vaccine impact on a specific biological niche. A number of studies (21/105) determined the carriage rate of *S. pneumoniae* alongside other respiratory pathogens, for example *H. influenzae* and *S. aureus*. This suggests that in an era when the importance of polymicrobial interactions are being realised, identification of multiple potential pathogen carriage is becoming commonplace [9,46–48]. PHiD-CV was designed to have a direct influence on paediatric *H. influenzae* colonisation without a conclusive evidence base of impact thus far [47,49], while PCV-7 and PCV-13 could have indirect effects by creating niche space for current pathogenic microbial respiratory tract inhabitants [50,51]. Optimal sampling of each organism within a panel is still necessary and description of the best method for simultaneous collection would be valuable to a broader range of carriage studies. These broad scope carriage studies may also have additional and specific requirements of any protocol. Carriage studies could be further broadened to include detection of viruses with the potential to use viral transport media if a study were to rely upon culture independent detection. However for further analysis it must be remembered that in a molecular era of increasing sensitivity caution should be used in assuming genotype is equivalent to phenotype. This is important in a species where extensive recombination is recognised [15–17] and particularly true for carriage isolates where the definition of the pneumococcus is less distinct than with invasive isolates due to their intrinsic diversity [52,53].

During the course of the analysis it was surprising to observe that numerous studies neglected to mention the geographical location which affects carriage rates, reported to range globally from 14.9 to 87.2% [54], or the years of collection: critical for deduction of whether results are pre or post PCV implementation. In this article, absence of collection dates resulted in exclusion of 30 studies from analysis.

Dependence on publication, availability on PubMed and journal access could introduce bias and therefore is a limitation of our approach. Temporal analysis of the use of the WHO standard method is complex due to reliance on publication. Studies that commenced collection in early years are more likely to have completed or published data than those commencing in recent years. The dramatic reduction in the number of studies published over time is in part due to publication bias but also due to an initial surge of carriage studies that began after licencing of PCV7, a second surge may well follow as PCV13 is introduced globally a critical time to re-evaluate study protocol. It is important to consider that a positive, or negative, bias may also be present in articles that specify detection methods in comparison to those that do not. The level of conformity to the WHO method of collection may be under or overestimated due to the absence of specific method details in a number of articles. As the number of articles is limited data may be more susceptible to any bias present. Despite these limitations, the data does provide a useful picture of the level of uniformity in pneumococcal carriage sampling. As research studies are carried out over the course of years a time lag exists between conception, implementation and publication, it could be that investigators were not aware of the WHO standard method during their study planning stages and cannot always validate a change in protocol part way through a study.

10. Conclusion

In an ideal situation, study methods for assessing pneumococcal carriage from sampling through to typing would be standardised so that directly comparable data are collected. Despite the existence of the WHO standard method, most carriage studies do not conform to preliminary stages of collection. When looking at each criteria individually (Table 1), the majority of studies sampled the nasopharynx but did not always conform to the WHO standard method for swab type or swab transport media. Alternative media choice and swab type were equally responsible for divergence. These trends may have resulted from the lack of data and evidence to support the WHO swab or swab transport media type and where there are, potentially cheaper, alternatives available. The optimal procedure for sampling is known to vary according to the study age group, environmental and geographical constraints and study objectives [19,20,26,27].

The nasopharynx is widely used and accepted for sampling pneumococci and this uniformity adds value to comparisons between studies. Although nasopharyngeal washes may increase sensitivity, swabbing the nasopharynx is clearly tolerated well enough to allow numerous studies to be completed. It must be noted though that collecting nasopharyngeal swabs requires trained healthcare professionals that is consequently expensive and can limit the size of studies and is prohibitive of large scale community based studies. The evidence that is available and discussed here is lacking but suggests that the calcium alginate, WHO recommended swab, but not perhaps Dacron™, is suitable for pneumococcal sampling. Further evidence in this area would be of benefit particular to determine the impact of this swab on other upper respiratory microbes. The storage of bacterial isolates at below −70 °C is gold standard there is no evidence that disputes the validity of this approach but there is more to be said of how it is stored in terms of cultured isolates and original specimens. There is argument for storing both cultured isolates and the original swab and for a broth enrichment step to enhance recovery of multiple serotypes. STGG has been reported to have superior sensitivity to other medias for transport and storage, more evidence is required to compare use of an alternative being used not just direct plating but Amies and Stuarts for transport. Storage of samples and or swabs in STGG is supported by current evidence; STGG is still not commercially available and this may discourage studies from using this media.

The increased ability of an evidence-based standardised method, to detect the presence of pneumococci, would reduce the number of variables between future carriage studies. This would serve to augment the value of collected data for extrapolation to pneumococcal disease control and evaluation of the impact of

pneumococcal vaccines on carriage. Additional evidence to support protocol adaptations for resource limited studies would be of benefit including the limits of transport media to maintain pneumococci over a range of temperatures. Overall evidence available does not contradict the WHO standard method; there is value in using a standard method for collection which should be adopted if available. Evidence that supports the WHO method or an alternative does however need to be more substantial.

Although this analysis primarily focused on the impact of steps prior to culture or downstream analysis it is also important to acknowledge that culturing or independent methods that allow typing to be performed can potentially have a huge impact on results. The data presented here for collection suggest that a large number of studies do not follow exactly the WHO standard method, for detection of upper respiratory tract carriage of pneumococci, at least for the primary steps in the protocol.

Conflicts of interest

J.M.J. has received consulting fees from GlaxoSmithKline. S.N.F. acts as principal investigator for clinical trials conducted on behalf of University Hospital Southampton NHS Foundation Trust/University of Southampton that are sponsored by vaccine manufacturers but receives no personal payments from them. S.N.F. has participated in advisory boards for vaccine manufacturers but receives no personal payments for this work. SCC currently receives unrestricted research funding from Pfizer Vaccines (previously Wyeth Vaccines) and has participated in advisory boards and expert panels for GSK, Pfizer and Novartis. SCC is an investigator on studies conducted on behalf of University Hospital Southampton NHS Foundation Trust/University of Southampton/HPA that are sponsored by vaccine manufacturers but receives no personal payments from them. S.N.F., S.C.C., J.M.J. and R.A.G. have received financial assistance from vaccine manufacturers to attend conferences. All grants and honoraria are paid into accounts within the respective NHS Trusts or Universities, or to independent charities.

Funding

SNF receives support from the National Institute for Health Research funding via the Southampton NIHR Wellcome Trust Clinical Research Facility and the Southampton NIHR Respiratory Biomedical Research Unit.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.vaccine.2012.08.080.

References

[1] US National Library of Medicine National Institutes of Health, PubMed. [cited 31.10.11]; available from: http://www.ncbi.nlm.nih.gov/pubmed/
[2] O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, McCall N, et al. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet 2009;374(September (9693)):893–902.
[3] Hussain M, Melegaro A, Pebody RG, George R, Edmunds WJ, Talukdar R, et al. A longitudinal household study of *Streptococcus pneumoniae* nasopharyngeal carriage in a UK setting. Epidemiol Infect 2005;133(05):891–8.
[4] Department of Health (2006). Important changes to the childhood immunisation programme, PL/CMO/2006/1.
[5] Department of Health. Introduction of Prevenar 13® into the Childhood Immunisation Programme. Department of Health, Gateway reference: 13581 (2010).
[6] Talbird SE, Taylor TN, Knoll S, Frostad CR, Garcia Marti S. Outcomes and costs associated with PHiD-CV, a new protein D conjugate pneumococcal vaccine, in four countries. Vaccine 2010;28(November (Suppl 6)):G23–9.
[7] Bogaert D, de Groot R, Hermans PWM. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. Lancet Infect Dis 2004;4(3):144–54.
[8] Miller E, Andrews NJ, Waight PA, Slack MP, George RC. Herd immunity and serotype replacement 4 years after seven-valent pneumococcal conjugate vaccination in England and Wales: an observational cohort study. Lancet Infect Dis 2011;11(October 10):760–8.
[9] Tocheva AS, Jefferies JMC, Rubery H, Bennett J, Afimeke G, Garland J, et al. Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of *Streptococcus pneumoniae* in young children. Vaccine 2011;29(26):4400–4.
[10] Hanage WP, Finkelstein JA, Huang F-L, Pelton SI, Stevenson AE, Kleinman K, et al. Evidence that pneumococcal serotype replacement in Massachusetts following conjugate vaccination is now complete. Epidemics 2010;2(2):80–4.
[11] O'Brien KLMM, Nohynek HM. The Who Pneumococcal Vaccine Trials Carriage Working Group. Report from a WHO Working Group: standard method for detecting upper respiratory carriage of *Streptococcus pneumoniae*. Pediatr Infect Dis J 2003;22(2):e1–11.
[12] Rubin LG, Rizvi A, Baer A. Effect of swab composition and use of swabs versus swab-containing skim milk-tryptone-glucose-glycerol (STGG) on culture- or PCR-based detection of *Streptococcus pneumoniae* in simulated and clinical respiratory specimens in STGG transport medium. J Clin Microbiol 2008;46(August (8)):2635–40.
[13] CDC. PCR deduction of pneumococcal serotypes. 2011; http://www.cdc.gov/ncidod/biotech/strep/pcr.htm [accessed 06.11].
[14] Carvalho MD, Pimenta FC, Jackson D, Roundtree A, Ahmad Y, Millar EV, et al. Revisiting pneumococcal carriage using broth-enrichment and PCR techniques for enhanced detection of carriage and serotypes. J Clin Microbiol 2010;48(5):1611–8.
[15] Carvalho MD, Jagero G, Bigogo GM, Junghae M, Pimenta FC, Moura I, et al. Potential non-pneumococcal confounding of PCR-based determination of serotype in carriage. J Clin Microbiol 2012;50(9):3146–7.
[16] 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;158(June (Pt. 6)):1560–9.
[17] Pimenta FC, Gertz Jr RE, Roundtree A, Yu J, Nahm MH, McDonald RR, et al. Rarely occurring 19A-like cps locus from a serotype 19F pneumococcal isolate indicates continued need of serology-based quality control for PCR-based serotype determinations. J Clin Microbiol 2009;47(July (7)):2353–4.
[18] Neufeld F. Ueber die Agglutination der Pneumokokken und uber die Theorien der Agglutination. Zeitschrift fur Hygiene Infektionskrankheiten 1902;34:pp. 54–72.
[19] Inverarity D., Diggle M., Edwards G., Mitchell T. An Evaluation of Media Suitable for the Transportation by Air of Streptococcus pneumoniae 2006.
[20] Joshi HH, Gertz Jr RE, da Gloria Carvalho M, Beall BW. Use of silica desiccant packets for specimen storage and transport to evaluate pneumococcal nasopharyngeal carriage among Nepalese children. J Clin Microbiol 2008;46(September (9)):3175–6.
[21] Capeding M, Nohynek H, Sombrero L, Pascual L, Sunico E, Esparar G, et al. Evaluation of sampling sites for detection of upper respiratory tract carriage of *Streptococcus pneumoniae* and *Haemophilus influenzae* among healthy Filipino infants. J Clin Microbiol 1995;33(November (11)):3077–9.
[22] Watt JP, O'Brien KL, Katz S, Bronsdon MA, Elliott J, Dallas J, et al. Nasopharyngeal versus oropharyngeal sampling for detection of pneumococcal carriage in adults. J Clin Microbiol 2004;42(November (11)):4974–6.
[23] Rapola S, Salo E, Kiiski P, Leinonen M, Takala AK. Comparison of four different sampling methods for detecting pharyngeal carriage of *Streptococcus pneumoniae* and *Haemophilus influenzae* in children. J Clin Microbiol 1997;35(May (5)):1077–9.
[24] Abdullahi O, Wanjiru E, Musyimi R, Glass N, Scott JAG. Validation of nasopharyngeal sampling and culture techniques for detection of *Streptococcus pneumoniae* in children in Kenya. J Clin Microbiol 2007;27(1):59–64.
[25] Leach AJ, Stubbs E, Hare K, Beissbarth J, Morris PS. Comparison of nasal swabs with nose blowing for community-based pneumococcal surveillance of healthy children. J Clin Microbiol 2008;46(June (6)):2081–2.
[26] Boersma WG, Lowenberg A, Holloway Y, Kuttschrutter H, Snijder JAM, Koeter GH. The role of antigen detection in pneumococcal carriers: a comparison between cultures and capsular antigen detection in upper respiratory tract secretions. Scand J Infect Dis 1993;25(1):51–6.
[27] Hendley JO, Sande MA, Stewart Jr PM. JMG. Spread of *Streptococcus pneumoniae* in Families. I. Carriage rates and distribution of types. J Infect Dis 1975;132(1):55–61.
[28] Pollock MR. Unsaturated fatty acids in cotton wool plugs. Nature 1948;161(May (4100)):853.
[29] Wright HD. A substance in cotton-wool inhibitory to the growth of the pneumococcus. J Pathol Bacteriol 1934;38:499–501.
[30] Poddar SK, Sawyer MH, Connor JD. Effect of inhibitors in clinical specimens on Taq and Tth DNA polymerase-based PCR amplification of influenza A virus. J Med Microbiol 1998;47(December (12)):1131–5.
[31] Wadowsky RM, Laus S, Libert T, States SJ, Ehrlich GD. Inhibition of PCR-based assay for *Bordetella pertussis* by using calcium alginate fiber and aluminum shaft components of a nasopharyngeal swab. J Clin Microbiol 1994;32(April (4)):1054–7.
[32] De Silva S, Wood G, Quek T, Parrott C, Bennett CM. Comparison of flocked and rayon swabs for detection of nasal carriage of *Staphylococcus aureus* among pathology staff members. J Clin Microbiol 2010;48(August (8)):2963–4.
[33] O'Brien KL, Bronsdon MA, Dagan R, Yagupsky P, Janco J, Elliott J, et al. Evaluation of a medium (STGG) for transport and optimal recovery of *Streptococcus*

pneumoniae from nasopharyngeal secretions collected during field studies. J Clin Microbiol 2001;39(March (3)):1021–4.
[34] Charalambous BM, Batt SL, Peek AC, Mwerinde H, Sam N, Gillespie SH. Quantitative Validation of Media for Transportation and Storage of *Streptococcus pneumoniae*. J Clin Microbiol 2003;41(December (12)):5551–6.
[35] Gray B. Egg-Based Media for Delayed Processing of Nasopharyngeal Swabs in Colonization Studies of *Streptococcus pneumoniae*. Eur J Clin Microbiol Infect Dis 2002;21(9):666–70.
[36] Gibson LF, Khoury JT. Storage and survival of bacteria by ultra-freeze. Lett Appl Microbiol 1986;3(6):127–9.
[37] Robson RL, Essengue S, Reed NA, Horvat RT. Optochin resistance in *Streptococcus pneumoniae* induced by frozen storage in glycerol. Diagn Microbiol Infect Dis 2007;58(June (2)):185–90.
[38] Siberry G, Brahmadathan KN, Pandian R, Lalitha MK, Steinhoff MC, John TJ. Comparison of different culture media and storage temperatures for the long-term preservation of *Streptococcus pneumoniae* in the tropics. Bull World Health Organ 2001;79(1):43–7.
[39] Kaltoft MS, Skov Sørensen UB, Slotved H-C, Konradsen HB. An easy method for detection of nasopharyngeal carriage of multiple *Streptococcus pneumoniae* serotypes. J Microbiol Methods 2008;75(3):540–4.
[40] Aslan G, Emekdas G, Bayer M, Serin MS, Kuyucu N, Kanik A. Serotype distribution of *Streptococcus pneumoniae* strains in the nasopharynx of healthy Turkish children. Indian J Med Res 2007;125(April (4)):582–7.
[41] 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;39(October (8)):1148–54.
[42] Hare KM, Stubbs E, Beissbarth J, Morris PS, Leach AJ. Swab transport in Amies gel followed by frozen storage in skim milk tryptone glucose glycerol broth (STGGB) for studies of respiratory bacterial pathogens. J Microbiol Methods 2010;81(3):253–5.
[43] Turner P, Hinds J, Turner C, Jankhot A, Gould K, Bentley SD, et al. Improved detection of nasopharyngeal cocolonization by multiple pneumococcal serotypes by use of latex agglutination or molecular serotyping by microarray. J Clin Microbiol 2011;49(May (5)):1784–9.
[44] Rivera-Olivero IA, Blommaart M, Bogaert D, Hermans PW, de Waard JH, Multiplex PCR. reveals a high rate of nasopharyngeal pneumococcal 7-valent conjugate vaccine serotypes co-colonizing indigenous Warao children in Venezuela. J Med Microbiol 2009;58(May (Pt. 5)):584–7.
[45] Leung MH, Oriyo NM, Gillespie SH, Charalambous BM. The adaptive potential during nasopharyngeal colonisation of *Streptococcus pneumoniae*. Infect Genet Evol 2010;11(8):1989–95.
[46] Janapatla R-P, Chang H-J, Hsu M-H, Hsieh Y-C, Lin T-Y, Chiu C-H. Nasopharyngeal carriage of Streptococcus pneumoniae, *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Alloiococcus otitidis* in young children in the era of pneumococcal immunization, Taiwan. Scand J Infect Dis 2011;43:1–6.
[47] Prymula R, Hanovcova I, Splino M, Kriz P, Motlova J, Lebedova V, et al. Impact of the 10-valent pneumococcal non-typeable *Haemophilus influenzae* Protein D conjugate vaccine (PHiD-CV) on bacterial nasopharyngeal carriage. Vaccine 2011;29(10):1959–67.
[48] Dahlblom V, Soderstrom M. Bacterial interactions in the nasopharynx – effects of host factors in children attending day-care centers. J Infect Public Health 2012;5(April (2)):133–9.
[49] Prymula R, Kriz P, Kaliskova E, Pascal T, Poolman J, Schuerman L. Effect of vaccination with pneumococcal capsular polysaccharides conjugated to *Haemophilus influenzae*-derived protein D on nasopharyngeal carriage of *Streptococcus pneumoniae* and *H. influenzae* in children under 2 years of age. Vaccine 2009;28(December (1)):71–8.
[50] Lee GM, Huang SS, Rifas-Shiman SL, Hinrichsen VL, Pelton SI, Kleinman K, et al. Epidemiology and risk factors for *Staphylococcus aureus* colonization in children in the post-PCV7 era. BMC Infect Dis 2009;9:110.
[51] van Gils EJM, Veenhoven RH, Rodenburg GD, Hak E, Sanders EAM. Effect of 7-valent pneumococcal conjugate vaccine on nasopharyngeal carriage with Haemophilus influenzae and Moraxella catarrhalis in a randomized controlled trial. Vaccine 2011;29(44):7595–8.
[52] Arbique JC, Poyart C, Trieu-Cuot P, Quesne G, Carvalho MdGS, Steigerwalt AG, et al. Accuracy of phenotypic and genotypic testing for identification of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. J Clin Microbiol 2004;42(October (10)):4686–96.
[53] Simoes AS, Valente C, de Lencastre H, Sa-Leao R. Rapid identification of non-capsulated *Streptococcus pneumoniae* in nasopharyngeal samples allowing detection of co-colonization and reevaluation of prevalence. Diagn Microbiol Infect Dis 2011;71(3):208–16.
[54] Cardozo DM, Nascimento-Carvalho CM, Souza FR, Silva NM. Nasopharyngeal colonization and penicillin resistance among pneumococcal strains: a worldwide 2004 update. Braz J Infect Dis 2006;10(August (4)):293–304.

OPEN ACCESS Freely available online

PLOS ONE

Clonal Expansion within Pneumococcal Serotype 6C after Use of Seven-Valent Vaccine

Nicholas J. Loman[1], Rebecca A. Gladstone[2], Chrystala Constantinidou[3], Anna S. Tocheva[2], Johanna M. C. Jefferies[2,6], Saul N. Faust[2,4,6], Leigh O'Connor[1], Jacqueline Chan[3], Mark J. Pallen[3]*[9], Stuart C. Clarke[2,5,6][9]

1 Centre for Systems Biology, School of Biosciences, University of Birmingham, Birmingham, United Kingdom, **2** Molecular Microbiology Group, Sir Henry Wellcome Laboratories, Academic Unit of Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton, United Kingdom, **3** Division of Microbiology and Infection, Warwick Medical School, University of Warwick, Coventry, United Kingdom, **4** NIHR, Wellcome Trust Clinical Research Facility, University Hospital Southampton Foundation NHS Trust, Southampton, United Kingdom, **5** Health Protection Agency, Southampton, United Kingdom, **6** Southampton NIHR, Respiratory Biomedical Research Unit, University Hospital Southampton Foundation NHS Trust, Southampton, United Kingdom

Abstract

Streptococcus pneumoniae causes invasive infections, primarily at the extremes of life. A seven-valent conjugate vaccine (PCV7) is used to protect against invasive pneumococcal disease in children. Within three years of PCV7 introduction, we observed a fourfold increase in serotype 6C carriage, predominantly due to a single clone. We determined the whole-genome sequences of nineteen *S. pneumoniae* serotype 6C isolates, from both carriage (n = 15) and disease (n = 4) states, to investigate the emergence of serotype 6C in our population, focusing on a single multi-locus sequence type (MLST) clonal complex 395 (CC395). A phylogenetic network was constructed to identify different lineages, followed by analysis of variability in gene sets and sequences. Serotype 6C isolates from this single geographical site fell into four broad phylogenetically distinct lineages. Variation was seen in the 6C capsular locus and in sequences of genes encoding surface proteins. The largest clonal complex was characterised by the presence of lantibiotic synthesis locus. In our population, the 6C capsular locus has been introduced into multiple lineages by independent capsular switching events. However, rapid clonal expansion has occurred within a single MLST clonal complex. Worryingly, plasticity exists within current and potential vaccine-associated loci, a consideration for future vaccine use, target selection and design.

Citation: Loman NJ, Gladstone RA, Constantinidou C, Tocheva AS, Jefferies JMC, et al. (2013) Clonal Expansion within Pneumococcal Serotype 6C after Use of Seven-Valent Vaccine. PLoS ONE 8(5): e64731. doi:10.1371/journal.pone.0064731

Editor: Eliane Namie Miyaji, Instituto Butantan, Brazil

Received January 17, 2013; **Accepted** April 17, 2013; **Published** May 28, 2013

Funding: This work was funded, in part, by an award to SCC by The Bupa Foundation. NJL is funded by a grant from the Biotechnology and Biological Sciences Research Council (BBSRC) BB/E011179/1. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: SCC currently receives unrestricted research funding from Pfizer Vaccines (previously Wyeth Vaccines). JMJ and SCC have received consulting fees from GlaxoSmithKline. SNF acts as principal investigator for clinical trials conducted on behalf of Southampton University Hospitals NHS Trust/University of Southampton that are sponsored by vaccine manufacturers but receives no personal payments from them. SNF has participated in advisory boards for vaccine manufacturers but receives no personal payments for this work. SNF, SCC and JMJ have received financial assistance from vaccine manufacturers to attend conferences. All grants and honoraria are paid into accounts within the respective NHS Trusts or Universities, or to independent charities. This does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

* E-mail: m.pallen@warwick.ac.uk

[9] These authors contributed equally to this work.

Introduction

Streptococcus pneumoniae (the pneumococcus) causes life-threatening invasive disease (pneumonia, septicaemia, and meningitis), primarily in children and the elderly, as well as other less severe infections (sinusitis, and acute otitis media). The global burden of pneumococcal invasive disease was estimated at almost 15 million cases in 2000.[1] Asymptomatic carriage is known to precede invasive disease and young children are the major reservoir of pneumococci in human populations, around one-third of children under five and over half of those under two carried *S. pneumoniae* asymptomatically in the nasopharynx.[2–4]

The pneumococcus is naturally transformable and shows considerable genotypic and phenotypic diversity, particularly in capsular serotype, of which over 90 are known. Capsular serotype is associated with the ability to cause invasive disease—around a quarter of the known serotypes account for the majority of invasive infections; some serotypes are at least ten-fold more likely to cause invasive disease than others.[5]

The seven-valent pneumococcal conjugate vaccine (PCV7) was licensed in the USA and Europe just over ten years ago. The vaccine consists of the polysaccharide capsule of seven serotypes associated with paediatric invasive disease in North America (4, 6B, 9V, 14, 18C, 19F and 23F), conjugated to diphtheria toxoid. It was added to the childhood immunisation schedule in the USA in 2001 but not added to the UK's schedule until 2006. The use of PCV7 resulted in a reduction of invasive pneumococcal disease both in the North America and Europe.[6–9] However, since the introduction of PCV7, "serotype replacement" has been observed, with an increase in the proportion of invasive and non-invasive disease caused by serotypes not represented in the vaccine.[10,11] In the UK, the increased incidence of invasive pneumococcal

disease was caused by serotypes not included in the seven-valent vaccine to some extent offset the impact of PCV7.[9] For this reason, PCV7 has recently been replaced in UK and US vaccination schedules with PCV13, which provides coverage for six additional serotypes.

Serotype 6C was first described in 2007 as a subtype of 6A that differed in reactivity with monoclonal antibodies from the majority of 6A strains.[12] PCV7 contains polysaccharide from the 6B serotype, which provides protection against serotypes 6A and 6B.[13] However, such protection does not extend to serotype 6C.[14] Serotype 6C appears to be rare in pre-vaccination populations.[15–21] However, a worrying increase in the incidence of serotype 6C disease and carriage has been observed in diverse populations around the world since the introduction of PCV7.[22–32] Locally, since the introduction of PCV7, we have seen a fourfold increase in the proportion of nasopharyngeal serotype 6C isolates among pneumococci isolated from our study population in Southampton, England—from 3.8% of all isolates in the winter of 2006/7 to 13.5% and 13.7% in the winters of 2007/8 and 2008/9 respectively.[2,33] Worryingly, we have also seen fatal cases of serotype 6C invasive disease.

Sustained surveillance and identification of factors influencing serotype distribution are essential for the continued control of pneumococcal disease and for rational vaccine design. Multi-locus sequence typing (MLST) has been used extensively to investigate the population structure of *S. pneumoniae*.[34] Although there is a correlation between MLST type and serotype, isolates from within a serotype can belong to a number of individual clonal complexes or sequence types and isolates of the same sequence type can express different capsular polysaccharides. Vaccine usage can result in capsular switching, where an existing sequence type from one capsular serotype remains prevalent by acquiring a different capsular locus.[35] Phylogenetic analyses of biosynthetic loci suggest that all 6C isolates belong to a single clade.[36] However, MLST studies on serotype 6C have shown it to encompass over two-dozen sequence types, several shared with the 6A serotype.[26,27,30,36]

It is clear that pneumococcal strains from the same sequence type and/or serotype can differ significantly in gene content and virulence-associated phenotypes.[37] However, as MLST samples neutral sequence variation within a handful of housekeeping genes, it cannot always provide information about differences in gene repertoire or sequence variation within loci associated with virulence. Furthermore, MLST cannot discriminate between very closely related isolates. We therefore turned to a more informative approach: high-throughput whole-genome sequencing. Several recent studies have shown that this approach is capable of the ultimate resolution of a single nucleotide base-pair change (SNP) between isolates while also providing valuable information on differences in virulence loci and gene content.[38–41]

Recombination is the major driving force for short-term genome evolution in the pneumococcus—an early MLST study suggested that a single nucleotide site is approximately 50 times more likely to change through recombination than mutation, while a more recent whole-genome-sequencing study estimated that 88% of SNPs in the Pneumococcal Molecular Epidemiology Network clone1 (PMEN1) lineage were the result of recombination events.[41,42] This high level of recombination distorts, but does not eliminate, phylogenetic signals of descent within pneumococcal lineages. However, given that recombination is so pervasive in pneumococci, evolutionary relationships between isolates are best represented by phylogenetic networks rather than by trees.[43]

Study goals

We undertook whole-genome sequencing of multiple local isolates from serotype 6C to investigate genetic diversity within a single serotype in a constrained geographical location and time period, focusing on the sequence type driving current 6C expansion in our study population, ST1692. We used whole-genome sequencing of serotype 6C pneumococcal isolates from Southampton to address the following questions:

1. Can genome sequencing provide information comparable or superior to MLST on the evolution and spread of serotype 6C lineages within a single geographical location?
2. Are there differences in gene content among the serotype 6C strains circulating in our community and might these differences account for variation in the prevalence of different lineages?
3. Is there heterogeneity in capsule regions or other virulence factors that might influence virulence and impact on the development of future vaccines?

Methods

Bacterial isolates

Samples were collected during the winters of 2006-07, 2007-08 and 2008-09. To obtain pneumococcal carriage isolates, nasopharyngeal swabs were collected from children aged four years and under with written parental/guardian informed consent; ethical approval for this proceedure was obtained from Southampton and South West Hampshire Research Ethics Committee 'B' [REC 06/Q1704/105]. Samples were collected regardless of health and vaccination status, gender or ethnicity and no demographic data were collected. Children were swabbed only once. All serotype 6C invasive disease isolates (n = 4) from blood or cerebrospinal fluid specimens received by the HPA South East regional microbiology laboratory between 2006 and 2010 were included; these were all from adult patients. Microbiological procedures were performed as previously described.[2] Pneumococcal capsular typing was performed on genomic DNA isolated from sub-cultured isolates by multiplex PCR.[2] Multi-locus sequence type (MLST) was determined by Qiagen Genomic Services, using the MLST website www.mlst.net to assign sequence type (ST).[2,34]

Thirty-two serotype 6C isolates were obtained from the carriage samples, falling into nine sequence types, of which ST1692 (belonging to clonal complex 395) was the most common (n = 18); three of the four invasive-disease isolates also belonged to ST 1692/CC395. CC395 was defined as all isolates within ST395 or within sequence types which shared 6/7 of MLST alleles with ST395. CC395 therefore encompassed nine isolates from ST1692, two representatives of ST1714 and a single representative of ST395. Fifteen of the 32 6C carriage isolates (47%) and four invasive disease isolates were selected for whole-genome sequencing (Table 1). These included at least one representative of each of the nine observed sequence types, with twelve isolates from CC395 (Table 1, Figure 1, 2).

Procedures

DNA was extracted for whole-genome sequencing from an approximately 10^{10} bacterial cell pellet of obtained from six hours liquid culture in 10 ml of liquid culture Brain-Heart-Infusion (BHI) for whole-genome sequencing. Extraction was performed using the Qiagen 100/G genomic tips by following the manufacturer's instruction for Gram-positive bacteria. All strains were shotgun-sequenced by 454 (Roche, Welwyn Garden City) using

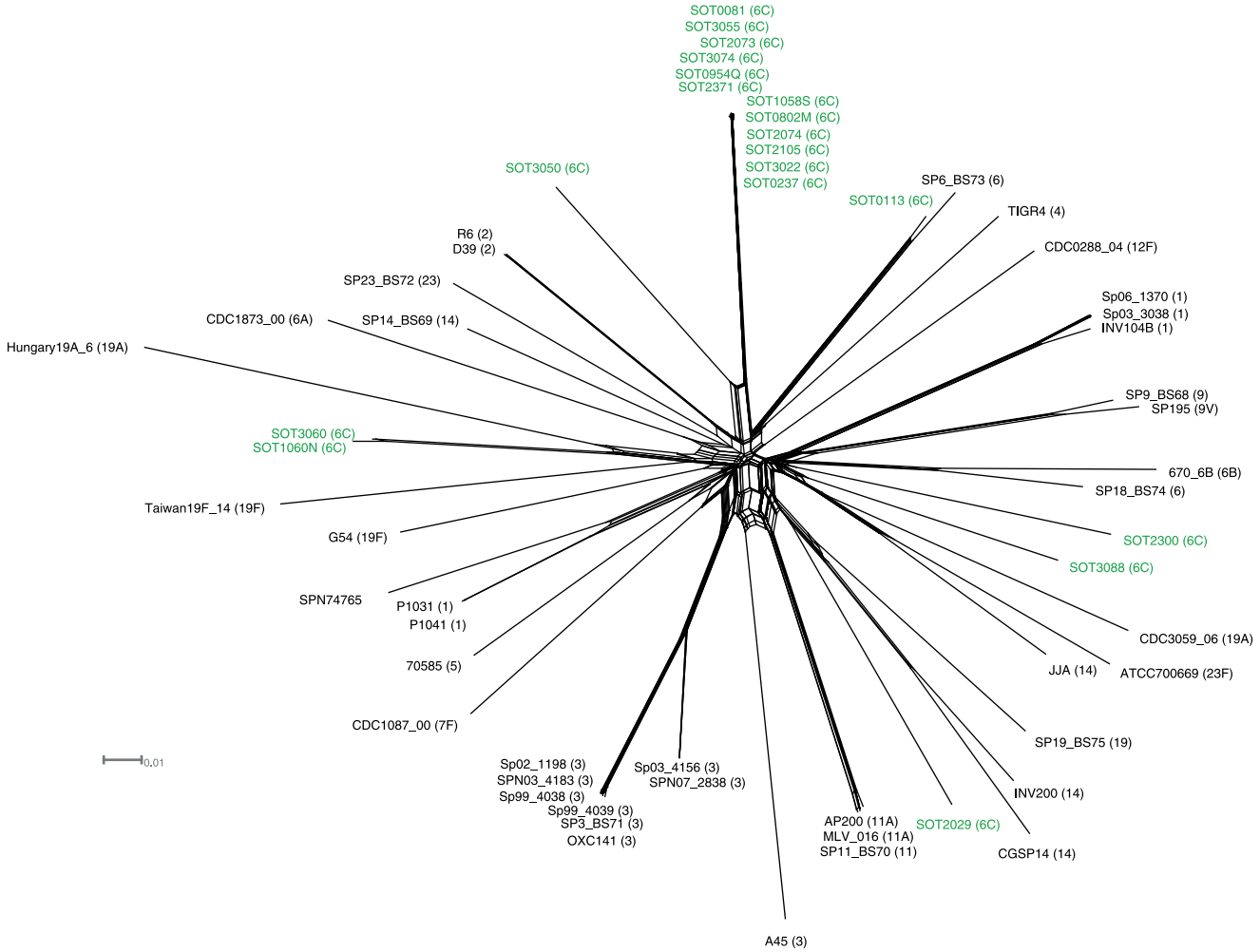


Figure 1. A phylogenetic split-network showing the relationship between Southampton 6C strains and other strains with whole-genome data from the public database. To generate the split-network, single nucleotide variants in concatenated multiple alignments of *S. pneumoniae* core genome coding sequences were input to a Neighbour-Net analysis in SplitsTree. Strains sequenced in this study are coloured in green. The serotype of each strain is shown in parenthesis.
doi:10.1371/journal.pone.0064731.g001

Titanium reagents at the University of Birmingham sequencing service. The reference strain SOT2073 was also sequenced with 454 mate-pair sequencing with an insert size of 8 kb to produce a single-scaffold reference sequence. All sequences were submitted to the short-read archive (http://www.ncbi.nlm.nih.gov/sra) and are available under the accession number SRP013270.

An average of 40 million bp were generated per isolate, yielding approximately 19-fold coverage per genome. Mean read lengths were 331 bp. A *de novo* assembly was performed for each strain, using Newbler version 2.5 (Roche) under default settings except that the '-rip' was invoked, which ensures that each read is placed in a single contig only. *De novo* assembly of each strain produced a mean of 167 contigs across all assemblies with a mean N50 (a statistical measure of the average length of a set of sequences) of 39753 base-pairs. Assembly of the strain SOT2073 data, including paired-end information, generated a single scaffold, 2,129,664bp in length, generally co-linear with the genome sequence of the well-characterised strain R6 (data not shown).

Assemblies were submitted to the xBASE annotation service (http://www.xbase.ac.uk/annotation) to provide an initial set of gene predictions. Orthologous genes were predicted from these annotated genomes in conjunction with complete *S. pneumoniae* genome sequences retrieved from GenBank using OrthoMCL (www.orthomcl.org) Whole-genome phylogenies were calculated from the core genome (defined as the set of single-copy orthologous genes present in all strains). Individual alignments of orthologues were produced by MUSCLE and trimmed for quality in Gblocks.[44,45] The resulting alignments were concatenated into a single super-alignment and phylogenetic networks created using SplitsTree drawing on genomes of representative strains in the public databases. Additionally, whole-genome alignments were produced in Mauve and analysed with ClonalFrame to remove signals of recombination to provide phylogenetic signals of vertical descent. The datasets generated by this study, including assemblies, annotations, orthologue predictions and alignments have been deposited in a Github repository (http://github.com/nickloman/pneumococcus-6C-study).

For accurate read mapping of the capsule locus reads were mapped against the SOT2073 reference sequence using the gsMapper component of Newbler. SNP calls were also produced in this way. SNPs were filtered using a variant frequency cut-off of 100% and the effect on protein sequence determined using xBASE.

Table 1. Pneumococcal isolates selected for whole genome sequencing.

Identifier	Year	Specimen type	Clinical Status and Outcome	ST
SOT0802M	2008	Blood	COPD exacerbation: Recovered	ST1692
SOT0954Q	2009	Blood	Pneumonia: Fatal	ST1692
SOT1058S	2010	CSF	Meningitis: Fatal	ST1692
SOT1060N	2010	Blood	Sepsis: Fatal	ST1150
SOT0081	2006/7	NP swab	Carriage	ST1692
SOT0113	2006/7	NP swab	Carriage	ST65
SOT0237	2006/7	NP swab	Carriage	ST1714
SOT2029	2007/8	NP swab	Carriage	ST3460
SOT2073	2007/8	NP swab	Carriage	ST1692
SOT2074	2007/8	NP swab	Carriage	ST1692
SOT2105	2007/8	NP swab	Carriage	ST1692
SOT2300	2007/8	NP swab	Carriage	ST1862
SOT2371	2007/8	NP swab	Carriage	ST395
SOT3022	2008/9	NP swab	Carriage	ST1692
SOT3050	2008/9	NP swab	Carriage	ST1600
SOT3055	2008/9	NP swab	Carriage	ST1692
SOT3060	2008/9	NP swab	Carriage	ST1150
SOT3074	2008/9	NP swab	Carriage	ST1714
SOT3088	2008/9	NP swab	Carriage	ST398

Abbreviations: CSF: cerebrospinal fluid; NP: nasopharyngeal.
doi:10.1371/journal.pone.0064731.t001

Results

Clonal diversity and clonal expansion

We identified SNPs in the genes conserved in all our isolates and other representative *S. pneumoniae* strains. From these, we generated a phylogenetic split-network (Figure 1), which largely reproduces the major pneumococcal lineages described by Donati *et al.*[43] Our serotype 6C isolates fall into four lineages. The largest 6C lineage consists of the twelve isolates belonging to clonal complex 395, together with SOT3050 from ST1600 and SOT0113 from ST65. Two serotype 6C isolates from ST1150—one associated with carriage (SOT3060), the other with fatal sepsis (SOT1060N) —form a distinct phylogenetic cluster. Two other seemingly unrelated serotype 6C carriage isolates, SOT2300 (ST1862) and SOT3088 (ST398) also form a distinct cluster. One carriage isolate from this serotype, SOT2029 from ST3460, sits separate from all other serotype 6C strains.

We next focused on diversity within this clonal complex, creating a phylogenetic network of CC395 strains (Figure 2). Tight clustering was found between four pairs of strains. The close relationship between SOT2073 and SOT2074 is unsurprising because these strains were collected from siblings in the same family during the same hospital visit. Nonetheless, there are seven SNPs that separate these genomes, three of them apparently acquired through recombination. Four of the seven SNPs occur in coding regions and three representing non-synonymous changes, emphasising the relatively fast pace of genome evolution in this species.

There are two pairs of carriage isolates (SOT0081/SOT3022 and SOT0237/SOT3074) where the isolates were obtained two years apart, yet cluster together tightly, suggesting that conserved genotypes can persist year on year. More worryingly, two of the four invasive isolates, SOT0802M and SOT1058S cluster

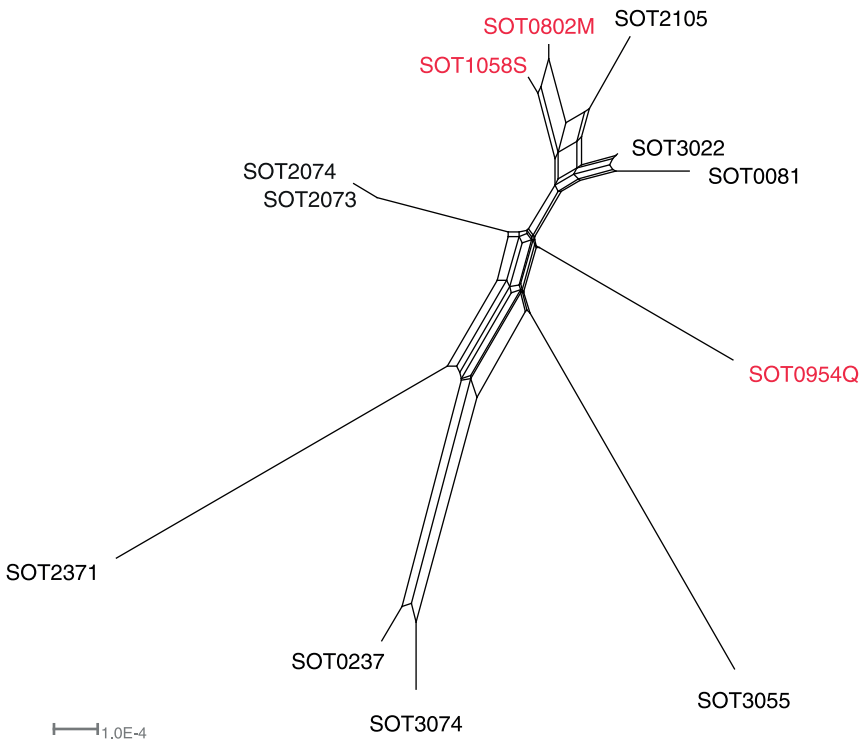


Figure 2. A phylogenetic split-network drawn from single nucleotide variants showing the relationship between CC395 isolates. Invasive isolates are coloured in red.
doi:10.1371/journal.pone.0064731.g002

together, yet were also obtained from samples two years apart, suggesting the persistence of a virulent clone carrying a serotype not covered by the PCV-7 vaccine.

Sequence variation in loci associated with surface structures

Mapping alignments revealed that the number of SNPs separating each isolate from SOT2073 ranged from 8 to 19,829 among our serotype 6C isolates. Up to 128 SNPs were seen in isolates from ST1692 and up to 172 SNPs within the CC395 clonal complex (Table 2). Surprisingly, the majority of SNPs were non-synonymous, probably reflecting selective pressure on key virulence determinants. Known or putative virulence factors are commonly mutated including the IgA1 protease, choline-binding proteins, surface proteins PspA and PspC and endo-beta-N-acetylglucosaminidase. All isolates from CC395, together with isolate SOT3050 (ST1600), encode a second PspA-domain protein, not seen in any of the other sequence types. This protein contains two domains: a glucan domain with homology to PspA and a peptidase/caspase domain. The pair of invasive isolates, SOT0802M and SOT1058S, also show variation in the choline-binding protein PspC, in particular in the length of a proline-alanine repeat motif, linking the C-terminal choline-binding domain and the active peptide domain.

The 6C capsular locus is thought to have originated at least three decades ago in a single recombination event that replaced the 6A *wciN* gene (encoding a glysoyltransferase) with the *wciN*-beta allele.[36,46] However, we found heterogeneity in capsular gene sequences—in particular, several stretches from the capsular locus of isolate SOT0113 showed <95% sequence identity to the reference sequence from SOT2073 and other 6C isolates. Also, a previous described insertion in the *wzy* gene was found in SOT0113, but not in the other isolates examined.[47]

Table 2. Single nucleotide polymorphisms.

Strain	ST	Total SNP	Filtered	CDS	Non-Synonymous
SOT2074	1692	81	8	4	3
SOT2105	1692	217	67	49	30
SOT3022	1692	233	73	51	29
SOT802M	1692	202	100	77	45
SOT1058S	1692	237	109	77	46
SOT0081	1692	189	111	80	48
SOT954Q	1692	251	111	70	45
SOT3055	1692	304	128	102	70
SOT0237	1714	334	156	122	83
SOT3074	1714	353	161	123	82
SOT2371	395	367	172	123	81
SOT0113	65	16772	2466	2042	636
SOT2029	3460	18835	12148	10360	2977
SOT3050	1600	16534	13777	11551	3494
SOT3060	1150	17824	13892	11935	3466
SOT3088	398	17735	14506	12275	3721
SOT1060N	1150	19298	15832	13403	3874
SOT2300	1862	24959	19829	17209	5054

The number of filtered SNPs separating isolates from SOT2073 ranges from 8 to 19,829. Within ST1692 the largest number of SNPs is 304;within CC395, the largest number is 367.

doi:10.1371/journal.pone.0064731.t002

Accessory genome of 6C serotype strains

Next, we examined the accessory genome of our serotype 6C isolates, that is the set of genes and gene clusters not present in all strains of *S. pneumoniae*. The largest differences in gene content between our 6C isolates and the SOT2073 reference genome were due to prophages, for example a near-identical ~13 kb prophage in the two ST1714 isolates and the ~32-kb *Streptococcus* phage 11865 in SOT3055. Prophages are known to carry virulence-related genes in many other bacterial species and, although a relatively underexplored topic, this is probably also true in the pneumococcus.[48,49]

Our survey revealed several putative resistance genes within the 6C serotype. As noted, SOT2029 clusters separately from all other serotype 6C strains. Interestingly, this isolate, shows resistance to erythromycin and tetracycline but not penicillin (minimum inhibitory concentrations of >256, 6 and 0.064 respectively) and carries the *ermB* and *tetM* genes flanked by genes associated with conjugative transposons. All of the other isolates were sensitive to penicillin, tetracycline and erythromycin, and none contained sequences for *ermB* and *tetM* genes. All of our isolates from CC395 carry a lantibiotic synthesis locus not seen in any of the other serotype 6C sequence types we have studied. A very similar locus has been described in two other genome-sequenced *S. pneumoniae* isolates, CGSP14 and INV200.[50] However, these belong to a different serotype and are not related to CC395 in our phylogenetic network, suggesting that this cluster has undergone horizontal gene transfer.

Discussion

This study illustrates the evolution of a single pneumococcal serotype, 6C, during a period of vaccine pressure. The scattered phylogenetic distribution of serotype 6C isolates provides convincing evidence for historical capsular serotype switching,[32] whereby the 6C capsular locus has been introduced into multiple pneumococcal lineages by independent recombination events. However, the phylogenetic network also confirms that, in our local population, the rise in prevalence of serotype 6C is largely due to clonal expansion within a single clonal complex.

The presence of a lantibiotic synthesis locus within and unique to all genome-sequenced members of this clonal complex provides a possible explanation for the success and clonal expansion of CC395 in our study group. Lantibiotics are bacteriocins that contain the modified amino acid lanthionine and which are produced by, and act on, Gram-positive bacteria.[51] Their ecological role is thought to impart colonisation resistance, preventing related strains from gaining a foothold in a specific environment [51,52]. Our genomic survey suggests that lantibiotic production may have enhanced the fitness of CC395 in the presence of other *S. pneumoniae* lineages in a competitive and shifting environment, potenitally explaining its increasing predominance in our sample set.

As expected from previous studies, we found little or no variation in the complement or sequences of metabolic genes. Instead, we found a serotype 6C accessory genome largely composed of prophage genes. However, we did find further evidence of sequence variation within loci encoding surface structures, both proteins and capsular polysccahride, presumably driven by selective forces imposed by the host immune system. This is worrying for two reasons. Firstly, although the new vaccine

PCV13 appears to provide protection against the 6C serotype,[53] our findings suggest that sufficient plasticity might exist within the 6C capsular locus for new vaccine-escape mutants to emerge; Elberse *et al.*[54] have made similar but distinct observations on plasticity within the 6C capsular locus. Secondly, given that surface proteins such as PspA and PspC are candidates for inclusion in next-generation protein–based pneumococcal vaccines, the sequence variation seen in these proteins raises the concern that the *S. pneumoniae* will respond to the introduction of these vaccines with the same kind of rapid evolution seen after use of polysaccharide vaccines. In other words, we may be facing an example of Red Queen evolutionary dynamics,[55] where we may need continual innovation in our vaccine repertoire to maintain the same level of control of infection. It is also worrying to see the emergence of antibiotic resistance in one of our 6C lineages.

This study confirms that genomic epidemiology surveys are no longer the sole preserve of the large sequencing centres and suggest that these techniques are poised to become cost-effective replacements for existing *S. pneumoniae* typing methods such as MLST, gene-specific PCR assays and PCR capsule typing. More generally, high-throughput sequencing provides a new tool in our armamentarium with the potential to keep pace with, or even outrun, the evolution and spread of microbial pathogens as the field developes further.

Author Contributions

Conceived and designed the experiments: SCC SF JJ MP NL. Performed the experiments: AT RG CC LO. Analyzed the data: RG NL JC. Contributed reagents/materials/analysis tools: SCC MP. Wrote the paper: RG JJ NL MP.

References

1. O'Brien KL, Wolfson LJ, Watt JP, Henkle E, Deloria-Knoll M, et al. (2009) Burden of disease caused by Streptococcus pneumoniae in children younger than 5 years: global estimates. Lancet 374: 893–902.
2. Tocheva AS, Jefferies JM, Rubery H, Bennett J, Afimeke G, et al. (2011) Declining serotype coverage of new pneumococcal conjugate vaccines relating to the carriage of Streptococcus pneumoniae in young children. Vaccine 29: 4400–4404.
3. Gray BM, Converse GM 3rd, Dillon HC Jr (1980) Epidemiologic studies of Streptococcus pneumoniae in infants: acquisition, carriage, and infection during the first 24 months of life. J Infect Dis 142: 923–933.
4. Hussain M, Melegaro A, Pebody RG, George R, Edmunds WJ, et al. (2005) A longitudinal household study of Streptococcus pneumoniae nasopharyngeal carriage in a UK setting. Epidemiol Infect 133: 891–898.
5. Brueggemann AB, Griffiths DT, Meats E, Peto T, Crook DW, et al. (2003) Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential. J Infect Dis 187: 1424–1432.
6. Harboe ZB, Valentiner-Branth P, Benfield TL, Christensen JJ, Andersen PH, et al. (2010) Early effectiveness of heptavalent conjugate pneumococcal vaccination on invasive pneumococcal disease after the introduction in the Danish Childhood Immunization Programme. Vaccine 28: 2642–2647.
7. Ruckinger S, van der Linden M, Reinert RR, von Kries R, Burckhardt F, et al. (2009) Reduction in the incidence of invasive pneumococcal disease after general vaccination with 7-valent pneumococcal conjugate vaccine in Germany. Vaccine 27: 4136–4141.
8. Whitney CG, Farley MM, Hadler J, Harrison LH, Bennett NM, et al. (2003) Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. N Engl J Med 348: 1737–1746.
9. Gladstone RA, Jefferies JM, Faust SN, Clarke SC (2011) Continued control of pneumococcal disease in the UK - the impact of vaccination. J Med Microbiol 60: 1–8.
10. McEllistrem MC, Adams J, Mason EO, Wald ER (2003) Epidemiology of acute otitis media caused by Streptococcus pneumoniae before and after licensure of the 7-valent pneumococcal protein conjugate vaccine. J Infect Dis 188: 1679–1684.
11. Hicks LA, Harrison LH, Flannery B, Hadler JL, Schaffner W, et al. (2007) Incidence of pneumococcal disease due to non-pneumococcal conjugate vaccine (PCV7) serotypes in the United States during the era of widespread PCV7 vaccination, 1998-2004. J Infect Dis 196: 1346–1354.
12. Park IH, Pritchard DG, Cartee R, Brandao A, Brandileone MC, et al. (2007) Discovery of a new capsular serotype (6C) within serogroup 6 of Streptococcus pneumoniae. J Clin Microbiol 45: 1225–1233.
13. Vakevainen M, Eklund C, Eskola J, Kayhty H (2001) Cross-reactivity of antibodies to type 6B and 6A polysaccharides of Streptococcus pneumoniae, evoked by pneumococcal conjugate vaccines, in infants. J Infect Dis 184: 789–793.
14. Park IH, Moore MR, Treanor JJ, Pelton SI, Pilishvili T, et al. (2008) Differential effects of pneumococcal vaccines against serotypes 6A and 6C. J Infect Dis 198: 1818–1822.
15. du Plessis M, von Gottberg A, Madhi SA, Hattingh O, de Gouveia L, et al. (2008) Serotype 6C is associated with penicillin-susceptible meningeal infections in human immunodeficiency virus (HIV)-infected adults among invasive pneumococcal isolates previously identified as serotype 6A in South Africa. Int J Antimicrob Agents 32 Suppl 1: S66–70.
16. Hermans PW, Blommaart M, Park IH, Nahm MH, Bogaert D (2008) Low prevalence of recently discovered pneumococcal serotype 6C isolates among healthy Dutch children in the pre-vaccination era. Vaccine 26: 449–450.
17. Campos LC, Carvalho Mda G, Beall BW, Cordeiro SM, Takahashi D, et al. (2009) Prevalence of Streptococcus pneumoniae serotype 6C among invasive and carriage isolates in metropolitan Salvador, Brazil, from 1996 to 2007. Diagn Microbiol Infect Dis 65: 112–115.
18. Porat N, Park IH, Nahm MH, Dagan R (2010) Differential circulation of Streptococcus pneumoniae serotype 6C clones in two Israeli pediatric populations. J Clin Microbiol 48: 4649–4651.
19. Ho PL, Ang I, Chow KH, Lai EL, Chiu SS (2010) The prevalence and characteristics of Streptococcus pneumoniae isolates expressing serotypes 6C and 6D in Hong Kong prior to the introduction of the 7-valent pneumococcal conjugate vaccine. Diagn Microbiol Infect Dis 68: 439–444.
20. Vickers I, Fitzgerald M, Murchan S, Cotter S, O'Flanagan D, et al. (2011) Serotype distribution of Streptococcus pneumoniae causing invasive disease in the Republic of Ireland. Epidemiol Infect 139: 783–790.
21. Yao KH, Liu ZJ, Yu JG, Yu SJ, Yuan L, et al. (2011) Type distribution of serogroup 6 Streptococcus pneumoniae and molecular epidemiology of newly identified serotypes 6C and 6D in China. Diagn Microbiol Infect Dis 70: 291–298.
22. Jacobs MR, Good CE, Bajaksouzian S, Windau AR (2008) Emergence of Streptococcus pneumoniae serotypes 19A, 6C, and 22F and serogroup 15 in Cleveland, Ohio, in relation to introduction of the protein-conjugated pneumococcal vaccine. Clin Infect Dis 47: 1388–1395.
23. Leach AJ, Morris PS, McCallum GB, Wilson CA, Stubbs L, et al. (2009) Emerging pneumococcal carriage serotypes in a high-risk population receiving universal 7-valent pneumococcal conjugate vaccine and 23-valent polysaccharide vaccine since 2001. BMC Infect Dis 9: 121.
24. Nahm MH, Lin J, Finkelstein JA, Pelton SI (2009) Increase in the prevalence of the newly discovered pneumococcal serotype 6C in the nasopharynx after introduction of pneumococcal conjugate vaccine. J Infect Dis 199: 320–325.
25. Sa-Leao R, Nunes S, Brito-Avo A, Frazao N, Simoes AS, et al. (2009) 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 15: 1002–1007.
26. Carvalho Mda G, Pimenta FC, Gertz RE Jr, Joshi HH, Trujillo AA, et al. (2009) PCR-based quantitation and clonal diversity of the current prevalent invasive serogroup 6 pneumococcal serotype, 6C, in the United States in 1999 and 2006 to 2007. J Clin Microbiol 47: 554–559.
27. Jacobs MR, Bajaksouzian S, Bonomo RA, Good CE, Windau AR, et al. (2009) Occurrence, distribution, and origins of Streptococcus pneumoniae Serotype 6C, a recently recognized serotype. J Clin Microbiol 47: 64–72.
28. Millar EV, Pimenta FC, Roundtree A, Jackson D, Carvalho Mda G, et al. (2010) Pre- and post-conjugate vaccine epidemiology of pneumococcal serotype 6C invasive disease and carriage within Navajo and White Mountain Apache communities. Clin Infect Dis 51: 1258–1265.
29. Hanna JN, Humphreys JL, Jennison A, Penny M, Smith HV (2010) Serotype 6C invasive pneumococcal disease in indigenous people in north Queensland. Commun Dis Intell Q Rep 34: 122–123.
30. Green MC, Mason EO, Kaplan SL, Lamberth LB, Stovall SH, et al. (2011) Increase in prevalence of Streptococcus pneumoniae serotype 6C at Eight Children's Hospitals in the United States from 1993 to 2009. J Clin Microbiol 49: 2097–2101.
31. Kim KH, Hong JY, Lee H, Kwak GY, Nam CH, et al. (2011) Nasopharyngeal pneumococcal carriage of children attending day care centers in Korea: comparison between children immunized with 7-valent pneumococcal conjugate vaccine and non-immunized. J Korean Med Sci 26: 184–190.
32. Hanage WP, Bishop CJ, Huang SS, Stevenson AE, Pelton SI, et al. (2011) Carried pneumococci in Massachusetts children: the contribution of clonal expansion and serotype switching. Pediatr Infect Dis J 30: 302–308.
33. Tocheva AS, Jefferies JM, Christodoulides M, Faust SN, Clarke SC (2010) Increase in serotype 6C pneumococcal carriage, United Kingdom. Emerg Infect Dis 16: 154–155.
34. Enright MC, Spratt BG (1998) A multilocus sequence typing scheme for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiology 144 (Pt 11): 3049–3060.

35. Jefferies JM, Smith A, Clarke SC, Dowson C, Mitchell TJ (2004) Genetic analysis of diverse disease-causing pneumococci indicates high levels of diversity within serotypes and capsule switching. J Clin Microbiol 42: 5681–5688.
36. Bratcher PE, Park IH, Oliver MB, Hortal M, Camilli R, et al. (2011) Evolution of the capsular gene locus of Streptococcus pneumoniae serogroup 6. Microbiology 157: 189–198.
37. Silva NA, McCluskey J, Jefferies JM, Hinds J, Smith A, et al. (2006) Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. Infect Immun 74: 3513–3518.
38. Lewis T, Loman NJ, Bingle L, Jumaa P, Weinstock GM, et al. (2010) High-throughput whole-genome sequencing to dissect the epidemiology of Acinetobacter baumannii isolates from a hospital outbreak. J Hosp Infect 75: 37–41.
39. Pallen MJ, Loman NJ, Penn CW (2010) High-throughput sequencing and clinical microbiology: progress, opportunities and challenges. Curr Opin Microbiol 13: 625–631.
40. Hornsey M, Loman N, Wareham DW, Ellington MJ, Pallen MJ, et al. (2011) Whole-genome comparison of two Acinetobacter baumannii isolates from a single patient, where resistance developed during tigecycline therapy. J Antimicrob Chemother 66: 1499–1503.
41. Feil EJ, Smith JM, Enright MC, Spratt BG (2000) Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data. Genetics 154: 1439–1450.
42. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, et al. (2011) Rapid pneumococcal evolution in response to clinical interventions. Science 331: 430–434.
43. Donati C, Hiller NL, Tettelin H, Muzzi A, Croucher NJ, et al. (2010) Structure and dynamics of the pan-genome of Streptococcus pneumoniae and closely related species. Genome Biol 11: R107.
44. Talavera G, Castresana J (2007) Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56: 564–577.
45. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.
46. Park IH, Park S, Hollingshead SK, Nahm MH (2007) Genetic basis for the new pneumococcal serotype, 6C. Infect Immun 75: 4482–4489.
47. Mavroidi A, Godoy D, Aanensen DM, Robinson DA, Hollingshead SK, et al. (2004) Evolutionary genetics of the capsular locus of serogroup 6 pneumococci. J Bacteriol 186: 8181–8192.
48. Romero P, Croucher NJ, Hiller NL, Hu FZ, Ehrlich GD, et al. (2009) Comparative genomic analysis of ten Streptococcus pneumoniae temperate bacteriophages. J Bacteriol 191: 4854–4862.
49. Tobe T, Beatson SA, Taniguchi H, Abe H, Bailey CM, et al. (2006) An extensive repertoire of type III secretion effectors in Escherichia coli O157 and the role of lambdoid phages in their dissemination. Proc Natl Acad Sci U S A 103: 14941–14946.
50. Ding F, Tang P, Hsu MH, Cui P, Hu S, et al. (2009) Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant Streptococcus pneumoniae serotype 14. BMC Genomics 10: 158.
51. Engelke G, Gutowski-Eckel Z, Hammelmann M, Entian KD (1992) Biosynthesis of the lantibiotic nisin: genomic organization and membrane localization of the NisB protein. Appl Environ Microbiol 58: 3730–3743.
52. Hillman JD, Novak J, Sagura E, Gutierrez JA, Brooks TA, et al. (1998) Genetic and biochemical analysis of mutacin 1140, a lantibiotic from Streptococcus mutans. Infect Immun 66: 2743–2749.
53. Cooper D, Yu X, Sidhu M, Nahm MH, Fernsten P, et al. (2011) The 13-valent pneumococcal conjugate vaccine (PCV13) elicits cross-functional opsonophagocytic killing responses in humans to Streptococcus pneumoniae serotypes 6C and 7A. Vaccine 29: 7207–7211.
54. Elberse K, Witteveen S, van der Heide H, van de Pol I, Schot C, et al. (2011) Sequence diversity within the capsular genes of Streptococcus pneumoniae serogroup 6 and 19. PLoS One 6: e25018.
55. Jefferies JM, Clarke SC, Webb JS, Kraaijeveld AR (2011) Risk of red queen dynamics in pneumococcal vaccine strategy. Trends Microbiol 19: 377–381.