Genome Evolution in *Streptococcus pneumoniae.*

Kelly L. Wyres

(Department of Zoology and St Catherine's College,
University of Oxford)

Thesis submitted for examination for the degree of
Doctor of Philosophy in Zoology
(Trinity Term 2012)

Supervisor: Dr. Angela B. Brueggemann

Kelly Wyres, St. Catherine's College
Thesis submitted for examination for the degree of DPhil in Zoology, Trinity Term 2012.

Genome Evolution in *Streptococcus pneumoniae*

ABSTRACT

Streptococcus pneumoniae (the pneumococcus) is a bacterial pathogen responsible for >1.6 million annual deaths globally. Pneumococcal penicillin-resistance is conferred by acquisition of 'altered' penicillin-binding protein (*pbp*) genes. The first penicillin-nonsusceptible pneumococci were identified in the late 1960s. Global pneumococcal penicillin-nonsusceptibility rates rapidly increased in the 1980s/90s.

Since 2000, protein-conjugate vaccines, targeting 7, 10 or 13 of the ≥94 different pneumococcal capsule types (serotypes), have been introduced in many countries. Following vaccine implementation there has been a decline in vaccine-type pneumococcal disease and an increase in non-vaccine-type disease. These epidemiological changes result from "serotype replacement" and/or "serotype switching". The former describes the expansion of non-vaccine-type clones in the absence of vaccine-type pneumococci. The latter describes serotype change following recombination at the capsule polysaccharide synthesis (*cps*) locus.

To fully understand how pneumococci respond to vaccine- and antibiotic-induced selective pressures, we must better understand the evolutionary history of this pathogen. This thesis describes the study of a global collection of 426 pneumococci, dated 1937 - 2007. Serotype, genotype and penicillin-susceptibility data were collected. Nucleotide sequences of three *pbp* genes (for 389 isolates) and whole-genome sequences (for 96 isolates) were also generated.

The data demonstrated the long-term persistence of certain clones within pneumococcal populations, and that *pbp* and large-fragment (>30 kb) *cps* ± *pbp* recombination was occurring prior to both widespread antibiotic use *and* vaccine implementation. The data highlighted the promiscuous nature of the globally-distributed PMEN1 clone and its contribution to the spread of pneumococcal penicillin-resistance. PMEN1 also donated multiple, large regions (1.7 - 32.3 kb) of its genome to at least two un-related clones. Finally, six "Tn*916*-like" genetic elements, conferring resistance to non-penicillin antibiotics, were newly identified. These included two of the oldest ever described.

These results provided a unique insight into the history of pneumococcal evolution and the importance of genetic recombination.

DECLARATION

This thesis of ~45,000 words, is submitted for the degree of Doctor of Philosophy in Zoology, in Trinity Term 2012. None of the contents of this thesis have previously been submitted for any other degree.

The data herein described were generated by myself under the supervision of Dr. Angela B. Brueggemann and/or by our collaborators, as described in Chapter 2, *General Methods* and summarised at the start of each Results chapter (Chapters 3 – 6). I completed all analyses under the supervision of Dr. Brueggemann.

The results described in Chapter 3 were presented in-part at the 51st Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago 2011 (see Appendix 3). Chapters 4 and 5 have been submitted for publication in peer-reviewed journals. These manuscripts were written by myself and Dr. Brueggemann, and were subsequently reviewed by each of the co-authors (as listed at the start of these chapters). All other sections of this thesis were written by me.

Signed,

Kelly L. Wyres

ACKNOWLEDGEMENTS

The work in this thesis was supported by the Wellcome Trust through a Career Development Fellowship awarded to Dr. Angela Brueggemann. I would like to thank Dr. Brueggemann for agreeing to supervise and fund my DPhil studies as part of this fellowship. Angela is an excellent teacher, and is always very generous with her time. I have learned a great deal during my DPhil course and had the opportunity to travel to several scientific conferences. I am extremely grateful for all of Angela's guidance and support.

Secondly I would like to thank my collaborators, Dr. Lotte Lambertsen, Prof. Karl Krisstinsson, Prof. Josefina Liñares, Prof. Regine Hakenbeck, Prof. Michael Jacobs, Dr. Anne von Gottberg, Prof. Keith Klugman, Dr. Lesley McGee and Dr. Bernard Beall for the donation of bacterial strains and/or the provision of data. I also thank the members of the CDC Active Bacterial Core Surveillance Team who were responsible for the original collection of some isolates. I thank Dr. Nicholas Croucher for providing guidance concerning the assembly and analysis of whole-genome sequence data; Dr. Stephen Bentley and Prof. Julian Parkhill for whole-genome sequencing provision; the staff at the Wellcome Trust Sanger Institute for completion of whole-genome sequencing; and the Wellcome Trust for providing funding for whole-genome sequencing. Furthermore, I thank all of the above named collaborators for their thoughtful comments concerning the manuscripts related to Chapters 4 and 5 of this thesis.

I thank Dr. Keith Jolley and Prof. Martin Maiden for the set-up, maintenance and continued development of the BIGS database infrastructure, which was used extensively throughout my research. The Brueggemann Group BIGS database is hosted online by the Department of Zoology, University of Oxford.

I would like to thank Prof. Martin Maiden and Dr. Angus Buckling for their comments and suggestions at the time of my transfer from PRS to DPhil status. I thank Andries van Tonder for retrieval of data from the international Genbank database, technical computer assistance and general discussion of ideas. I also thank Roberta Fisher, Helen King and Nathan Barker for help in proof-reading sections of this thesis.

On a personal note, I would like to thank my parents, David and Rosemary Wyres, for their love, support and encouragement, without which I could not have come this far. I would also like to thank my friends, in particular Tommy Duncan, Matthew Hodges, Helen Thompson, Jim Fouracre, Roberta Fisher and Dorothea Hill for their support and for never failing to brighten my day. Finally, I would very much like to thank Nathan Barker for always listening, for keeping me sane, and for always making me happy. Thank you.

CONTENTS

ACRONYMS AND ABBREVIATIONS

Note – Abstracts and the main texts of Chapters 3 – 6 were written so that they may be understood in isolation. As such, acronyms/abbreviations defined in Chapters 1 - 2 are redefined as appropriate.

ABB	Angela B. Brueggemann
AMC	Advanced Market Commitment
ANSORP	Asian Network for Surveillance of Resistant Pathogens
AVG	Anne von Gottberg
BIGSdb	Bacterial Isolate Genome Sequence database
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
BWB	Bernard W. Beall
CC	Clonal Complex
CDS	Coding Sequence(s)
CLSI	Clinical Laboratory Standards Institute
cps	capsular polysaccharide synthesis
CSF	Cerebrospinal Fluid
CSP	Competence Stimulating Peptide
DLV	Double Locus Variant
DNA	DeoxyriboNucleic Acid
DNAse	DeoxyriboNucleic Acid-ase
dNTPs	deoxyNucleoside TriPhosphates
dsDNA	double-stranded DeoxyriboNucleic Acid
EARS-Net	European Antimicrobial Surveillance-Network
eBURST	e-Based Upon Related Sequence Types
ETOH	ethanol
GAVI	Global Alliance for Vaccines and Immunizations
goeBURST	global optimal e-Based Upon Related Sequence Types
H_2O	water
HGT	Horizontal Gene Transfer
I	Intermediate
ICE	Integrative Conjugative Element
IPD	Invasive Pneumococcal Disease
JL	Josefina Liñares
JP	Julian Parkhill
kb	kilo base
KGK	Karl G. Kristinsson
KLW	Kelly L. Wyres
KPK	Keith P. Klugman
LM	Lesley McGee
LML	Lotte M. Lambertsen
M	Molar
MEGA	Macrolide Efflux Genetic Assembly
MEGA5	Molecular Evolutionary Genetics Analysis 5
MIC	Minimum Inhibitory Concentration
min(s)	minute(s)

ml	milli-litre(s)
MLEE	Multi-Locus Enzyme Electrophoresis
MLST	Multi-locus Sequence Typing
mM	milli-Molar
MRJ	Michael R. Jacobs
MUSCLE	MUltiple Sequence Comparison by Log Expectation
NaCl	sodium chloride
NaOAc	sodium acetate
ng	nano-gram(s)
NJC	Nicholas J. Croucher
NRDB	Non-Redundant DataBase
NT	NonTypeable
NVT	Non-Vaccine Type
PBP	Penicillin-Binding Protein
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PCV	Pneumococcal Conjugate Vaccine
PEG	PolyEthylene Glycolate
PFGE	Pulsed-Field Gel Electrophoresis
PMEN	Pneumococcal Molecular Epidemiology Network
PNSP	Penicillin-Nonsusceptible Pneumococci
PPV	Pneumococcal Polysaccharide Vaccine
PSP	Penicillin-Susceptible Pneumococci
R	Resistant
RH	Regine Hakenbeck
RNAse	RiboNucleic Acid-ase
rpm	revolutions per minute
rRNA	ribosomal RiboNucleic Acid
S	Susceptible
s	second(s)
SDB	Stephen D. Bentley
SLV	Single Locus Variant
ssDNA	single-stranded DeoxyriboNucleic Acid
SSI	Statens Serum Institut
ST	Sequence Type
STARS	Sequence Typing Analysis Retrieval System
TE Buffer	Tris-EthyleneDiamineTetraAcetic acid Buffer
TMP/SMX	TriMethoPrim-SulfaMethoXazole
UK	United Kingdom
UNICEF	United Nations International Children's Emergency Fund
USA	United States of America
VT	Vaccine Type
WHO	World Health Organization
µg	micro-gram(s)
µl	micro-litre(s)
σ^2	sigma factor X

1. INTRODUCTION

1.1 What is *Streptococcus pneumoniae*?

S. pneumoniae, also known as the "pneumococcus", is a gram positive, facultatively anaerobic bacterium. Like all streptococci, the pneumococcus is catalase-negative and exhibits chain-like growth under gram stain conditions [67]. Unlike other streptococci, the pneumococcus is both optochin-susceptible and bile-soluble [67, 206]. This bacterium metabolises by fermentation to produce lactic acid; growth on blood agar is characterised by glistening white/grey colonies with sunken centres and an α-haemolytic halo [206].

S. pneumoniae the pathogen

S. pneumoniae is a major cause of otitis media, sinusitis, pneumonia, meningitis and bacteraemia. Globally this bacterium is responsible for over 1.6 million deaths each year [241]. Children below the age of 5 years and adults over the age of 65 years are those most affected [111, 203]. Immunocompromised individuals, such as those with HIV/AIDS, also suffer an increased risk of disease [244].

In the year 2000 there were an estimated 13.8 million global cases of pneumococcal pneumonia and 103,000 cases of meningitis among children <5 years of age [183]. Incidence rates of invasive pneumococcal disease (IPD) and pneumonia are much higher in developing than industrialised countries, with an estimated 3,627 annual cases per 100,000 children <5 years of age in Africa compared to 920 per 100,000 in

the Americas and 504 per 100,000 in Europe (Figure 1.1.1, [183]). Developing countries also suffer the highest case fatality rates which can be up to 20% for bacteraemia cases and over 50% for meningitis cases [69].

Despite the introduction of effective vaccination programmes from the year 2000 (see section 1.6 *Pneumococcal Vaccines*), the burden of pneumococcal disease remains significant even in industrialised countries. Among persons of all ages in the USA in 2004 there were an estimated 866,000 cases of pneumococcal pneumonia, 3,000 cases of meningitis and 12,000 cases of bacteraemia [111]. The associated direct medical costs at this time were estimated at ~$3.5 billion US [111]. Much of this cost (~$2.5 billion) was attributed to the treatment of pneumococcal pneumonia. (Pneumococci are responsible for between 30% and 50% of adult community-acquired pneumonia infections resulting in hospitalisation [69].) Additionally, a significant proportion of the total pneumococcal-associated healthcare costs were attributed to the treatment of non-invasive diseases such as otitis media (~$270 million) and sinusitis (~$300 million). Whilst adults are those most likely to suffer pneumococcal sinusitis, young children are those most likely to suffer otitis media, of which 7 million cases are reported in the USA each year [242].

It is worth noting here that the accuracy of estimates of pneumococcal disease burden is subject to the limitations of the current detection methods used in microbiology laboratories. For example, since cerebrospinal fluid (CSF) is a normally sterile body fluid (as are other body fluids such as blood or pleural fluid), culture of a bacterial organism in the CSF is the 'gold standard' for the diagnosis of bacterial meningitis.

However, positive cultures cannot be obtained in all cases [17a], for example, only 136 of 154 (88%) of cases of bacterial meningitis studied in the USA were CSF culture-positive [181a]. Furthermore, when antibiotic treatment is started prior to specimen collection, bacterial culture-positive rates are reduced [128a, 181a]. Gram staining of CSF samples may also be used for the identification of bacteria, but detection rates vary widely: estimates of Gram stain positive cases range between 60% and 90% [231a]. Other diagnostic tests, such as latex agglutination or PCR-based assays, can detect bacteria in culture-negative samples, but such tests are not universally available and may be subject to sensitivity and/or specificity failures [17a, 231a].

Matters are more complicated for the diagnosis of pneumococcal pneumonia. Guidelines from The Infectious Diseases Society of America state that community-acquired pneumonia may be diagnosed by a combination of "suggestive clinical features" (e.g. productive cough) plus X-ray demonstration of lung infiltrate [155b]. However, demonstration of lung infiltrate may not be possible when X-rays are taken early in the progression of disease, or following antibiotic treatment [155a]. Etiological diagnosis requires further testing, such as blood or sputum culture, but such tests are only recommended/carried out in certain cases (e.g. severe pneumonia cases) [155b]. Additionally, bacterial growth from blood samples would only be expected for cases of bacteraemic pneumonia, and a review of 13,043 community-acquired pneumonia cases indicated that only 7% were bacteraemic [167a]. Importantly, while culture of sputum samples may be more useful in the diagnosis of non-bacteraemic pneumonia, such samples may be contaminated by bacteria colonising the oropharynx, potentially leading to misdiagnosis of the etiologic agent [155a]. Furthermore, a large proportion

of patients are unable to produce a sputum sample of sufficient quality for testing, and even among suitable samples a significant proportion do not yield a positive bacterial culture [81a]. Urinary antigen-based assays may allow for a diagnosis in culture-negative cases, but are also restricted by their sensitivity [2a, 81b]. Thus using currently available diagnostic procedures, the true scale of community-acquired pneumonia, and the proportion of this disease attributable to *S. pneumoniae,* is likely to be underestimated.

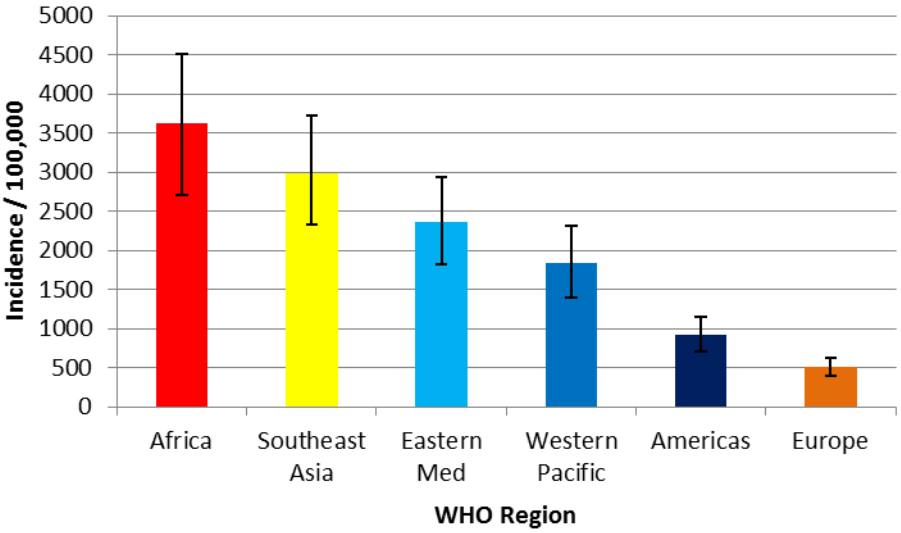


Figure 1.1.1 Burden of pneumococcal pneumonia and IPD among children <5 years old stratified by WHO (World Health Organisation) region.
Data from [183] are estimated incidences of pneumococcal pneumonia and IPD per 100,000 children <5 years of age in 2000. Data are stratified by WHO regions. Error bars depict uncertertainty ranges as calculated in [183]. Information regarding the geographic range of each region can be found at www.who.int. Eastern Med = Eastern Mediterranean.

S. pneumoniae the asymptomatic coloniser

Despite its immense capacity to cause disease, the natural niche of the pneumococcus is the human nasopharynx, where it is usually carried asymptomatically. Colonisation requires an interaction between pneumococcal surface exposed proteins, such as

Pneumococcal surface adhesin A, and carbohydrates on the surface of resting host epithelial cells [13]. Passive transfer of pneumococci (e.g. by aerosolisation) to sites not usually colonised (e.g. eustachian tubes, sinuses) may be the first step in the progression from asymptomatic carriage to disease [179a, 231b]. Under normal circumstances, however, pneumococci are successfully cleared by the immune system. When immune clearance fails, e.g. following edema or congestion resulting from allergy, co-existing viral infection or toxin exposure, pneumococci can replicate and cause local inflammation (the key pathogenesis of otitis media and sinusitis) [179a]. The development of pneumonia results when pneumococci are carried to the alveoli and subsequently multiply/spread through the lung. The required attachment to alveolar epithelial cells is enhanced by preceding viral infection [89a, 163a]. Alternatively, infection of the airways may result from direct invasion of damaged mucosa (for example due to smoking or co-infection of another pathogen) [231b].

Pneumococcal invasion of the blood, either at the site of colonisation or at a primary infection site can result in dissemination around the body. Subsequent invasion of the endothelium can result in the development of a secondary infection site. This so called, “hematogenous,” infection is implicated in the progression to meningitis. Disease of the meninges results when pneumococci cross the blood-brain barrier and infect the CSF, e.g. when pneumococcal adherence to brain microvascular endothelial cells is followed by transcytosis to the basal surface of the cell [200a]. This process is mediated by the eukaryotic platelet activating factor. Alternatively, damage to the tight junctions of the endothelium cells may allow intercellular passage from the blood to the CSF [170a].

Pneumococci are transmitted between asymptomatic carriers through aerosol droplets, mucus and saliva [13, 146]. Rates of carriage are highest among children, primarily those <5 years of age, and lower amongst adults [105, 156, 191]. Poor living conditions and overcrowding facilitate pneumococcal transmission, and are associated with an increased risk of carriage [45]. Other risk factors among children include day care attendance and having one or more sibling(s) [27, 169, 185, 191].

Carriage is a transient process whereby pneumococci are repeatedly acquired and lost. Colonisation with a single strain may last several weeks [106, 112, 217]. Presumably the duration of carriage is dependent upon both the extent/efficacy of immune clearance *and* a combination of inter-/intra- species competition effects. Indeed carriage duration is significantly greater for children than adults [106], likely as a result of age-related differences in immune response [13]. Additionally, carriage duration varies between strains expressing different polysaccharide capsule types ([112, 217], see section 1.5 *Pneumococcal Polysaccharide Capsules*), which are known to differ in their immunogenicity [128]. Competition between pneumococcal strains [49, 150] and with other colonising bacterial species [15, 48, 154] has also been demonstrated.

1.2 Phylogeny of the genus *Streptococcus*

S. pneumoniae is a member of the genus *Streptococcus*. This genus includes many species, both pathogenic and commensal. Broadly, the genus can be divided into two groups; the pyogenic and the viridans streptococci. Pyogenic streptococci include numerous pathogenic species such as *Streptococcus pyogenes,* the causal agent of

'strep throat' infections. The viridans streptococci form the major component of the human oral micro-flora and can be further classified into five groups based upon 16s rRNA sequence data [130]. These groups have been named Anginosus (including *Streptococcus anginosus, Streptococcus constellatus* and *Streptococcus intermedius*), Salivarius (including *Streptococcus salivarius* and *Streptococcus vestibularis*), Bovis (including *Streptococcus bovis, Streptococcus equinus* and *Streptococcus alactolyticus*), Mutans (including *Streptococcus mutans* and *Streptococcus sobrinus*) and finally, the Mitis group which includes *Streptococcus pneumoniae* and the predominantly commensal species *Streptococcus oralis, Streptococcus mitis, Streptococcus gordonii, Streptococcus infantis, Streptococcus pseudopneumoniae, Streptococcus sanguinis, Streptococcus parasanguinis, Streptococcus cristatus, Streptococcus peroris, Streptococcus australis* and *Streptococcus sinensis* [108, 130].

16s rRNA sequence data are a useful mechanism of discrimination between the main groups of streptococci, but are not sufficient for discrimination between *S. pneumoniae* and its closest relatives [130]. As an alternative, sequence data for a variety of genes shared between *S. pneumoniae, S. oralis, S. mitis* and *S. pseudopneumonaie* have been used as a basis for phylogeny construction. Analyses based upon single gene sequences do not show consistent results, suggesting that transfer of genetic material between members of different species (inter-species horizontal gene transfer (HGT)) is an important factor. However, analyses based upon concatenated sequences of between four and six housekeeping genes consistently suggested that *S. pneumoniae* and *S. mitis* were most closely related to each other and form a cluster separate to that of *S. oralis* [35, 56, 108, 133].

A point of disagreement is the placement of *S. pseudopneumoniae* within these phylogenies. An analysis by Do and colleagues supported a closer relationship of *S. pseudopneumoniae* with *S. mitis* than *S. oralis* (*S. pneumoniae* was not included in this analysis, [56]). Two other independent analyses supported this finding and suggested that *S. pseudopneumoniae* was most closely related to *S. pneumoniae,* even more so than *S. mitis* [108, 133]. However, a second analysis by Do and colleagues, using a different phylogenetic method, suggested that *S. pseudopneumoniae* was actually most closely related to *S. oralis.* The authors attributed these differences to the fact that some of the *S. pseudopneumoniae* genes clustered with those of *S. mitis* and some clustered with those of *S. oralis.* Consequently, the authors suggested that *S. pseudopneumoniae* was a hybrid species constructed by HGT.

The idea of such extensive HGT between these streptococci should not be surprising. In fact there is ample evidence of DNA exchange between these species [35, 52, 90] which will be discussed further below (see section 1.7 *Antimicrobials and Antimicrobial Resistance*).

1.3 Pneumococcal Population Biology

Bacteria reproduce asexually and historically it was believed that they were clonal organisms; however, for many bacterial species this is not the case [162]. Individual bacterial cells may be able to acquire DNA from other, potentially unrelated, cells through the processes of HGT (see section 1.4 *Horizontal Gene Transfer*). Variation

between related cells may thus be produced by either mutation *and/or* recombination of horizontally acquired DNA. The pneumococcus is a highly recombinogenic bacterial species [225] for which the likelihood that a single gene fragment will be altered by recombination is 10 times the likelihood that the fragment will be altered by mutation [70]. As a result, estimation of pneumococcal evolutionary histories and classification of isolates to clonal groups can be challenging.

Previously, pneumococci have been classified into genotypes by techniques such as pulsed-field gel electrophoresis (PFGE) [88], rRNA restriction analysis [57] and multilocus enzyme electrophoresis (MLEE) [39]. The current gold standard for genotyping pneumococci is a technique known as multilocus sequence typing (MLST) [61]. MLST is a genotyping method whereby the nucleotide sequences of ~500 bp regions of 7 housekeeping genes are determined, and unique sequences are assigned allele numbers. Each unique combination of alleles is assigned a unique sequence type (ST). The pneumococcal MLST scheme includes the following genes: *aroE* (shikimate kinase), *gdh* (glucose-6-phosphate 1-dehydrogenase), *gki* (glucose kinase), *recP* (transketolase), *spi* (signal peptidase 1), *xpt* (xanthine phosphoribosyltransferase) and *ddl* (D-alanine-D-alanine ligase). Allele sequences and ST allelic profiles are quality-checked and deposited into the international web database by a curator. To date (March 2012) >16,000 pneumococcal isolates and >8,000 unique pneumococcal STs have been submitted to the database, which is available at www.mlst.net and hosted by Imperial College, London. The number of assigned alleles at each locus ranged from 217 (*recP*) to 529 (*ddl*).

The rationale for choosing housekeeping genes as MLST loci was that they were considered to be under neutral selection. Consequently, variation in nucleotide sequence was assumed to accumulate slowly [61]. However, it is now clear that this assumption was not valid for all of the loci, in particular the *ddl* locus. *ddl* is the most variable MLST locus, located 783 bp downstream of the *pbp2b* locus in the pneumococcal genome. Acquisition of "altered" *pbp2b* nucleotide sequences via HGT is associated with the development of penicillin nonsusceptibility (see section 1.7 *Antimicrobials and Antimicrobial Resistance*). These "altered" *pbp2b* sequences differ from those of penicillin-susceptible pneumococci by up to ≥25% nucleotide divergence, and are thought to have originated from closely related streptococcal species [58]. It has been shown that the increased sequence variability of the *ddl* locus compared to that of the other MLST loci is the result of a genetic hitch-hiking effect whereby divergent *ddl* nucleotide sequences are acquired from closely related streptococci alongside *pbp2b.* Variation in the breakpoints of genetic recombination events results in an array of different mosaic patterns, further increasing *ddl* sequence diversity [61a].

Isolates genotyped by MLST can be assigned to clonal complexes (CCs) by the eBURST [224] or goeBURST algorithms [78]. In this approach isolates are assigned to CCs based on shared MLST alleles. Isolates assigned to the same CC are considered to have descended from a common ancestor (the group founder). STs are assigned to groups, within which each ST shares 6 alleles with at least one other ST represented within the group. Within each group the ST with the highest number of single locus variants (SLVs, defined as STs which differ by only a single allele), will be assigned as the group

founder. All possible SLVs of the group founder are assigned as such. Subsequently, SLVs of these STs (i.e. double locus variants (DLVs) of the group founder) are also assigned, and so on. The estimated relationships between STs can be represented by a (go)eBURST diagram, where STs are represented as nodes, and SLVs are connected by edges. The size of a node is proportional to the frequency of isolates assigned to that ST. A goeBURST diagram representing CC205 (CCs are named after the group founder ST) is shown in Figure 1.3.1.

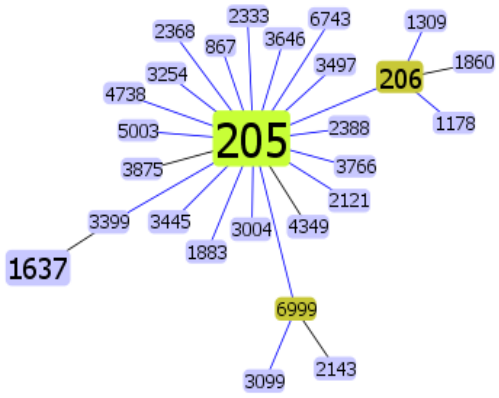


Figure 1.3.1 goeBURST representation of *S. pneumoniae* CC205.
goeBURST analysis was completed on isolates submitted to the international web database, www.mlst.net using the Phyloviz software (see Chapter 2). ST205, assigned as the group founder, is represented by the pale green node. Darker green nodes represent subgroup founders, i.e. SLVs of ST205 with >1 SLV of their own. Nodes representing SLVs are connected by edges.

Among recombinogenic species, allelic change results from both mutation and/or recombination. When rates of recombination are high, the true relationships between STs may be masked: STs which differ at two loci will only be grouped in the same CC if an appropriate connecting SLV is present within the data sample. Thus in some cases true descendants of the group founder ST may not be recognised, the likelihood of which is increased when the rate of allelic change is high (for example, when the rate of recombination is high). Furthermore, recombinogenic re-shuffling of alleles could result in the identification of false positive relationships between STs, particularly when sample sizes are large (e.g. when analysing the entire MLST database).

Nevertheless the use of MLST and (go)eBURST have played an important role in the study of pneumococcal molecular epidemiology [8, 20, 40, 78, 187]. Analyses of such data have highlighted the following characteristics of pneumococcal populations:

1) Many STs belong to just a few large CCs, members of which have been isolated from sites all around the world [61]. Forty-three such internationally disseminated clones have been designated the PMEN clones (see section 1.8 *Pneumococcal Molecular Epidemiology Network Clones*).
2) Isolates belonging to the same CC usually express the same serotype (capsule type), but isolates of the same serotype may belong to numerous CCs [18, 61, 187].
3) Some CCs are associated with antimicrobial resistance but others are not [92, 165].
4) There is some evidence of an association between certain genotypes and the potential to cause IPD [93]; such associations are largely, though not exclusively, a result of the association of genotype with serotype, and differences in serotype potential to cause disease (see section 1.5 *Pneumococcal Polysaccharide Capsules*) [18, 207].
5) Among isolates belonging to the same CC there is evidence of genetic variation, sometimes extensive variation, at non-MLST loci [46, 197, 230].

1.4 Horizontal Gene Transfer

There are three mechanisms through which bacteria are able to exchange DNA horizontally; transformation (the internalisation and assimilation of free DNA from the

surrounding media), conjugation (the transfer of DNA from one cell to another via a conjugative sex pilus/pore) and transduction (phage-mediated DNA transfer) [79]. Until recently it was believed that the predominant mechanism of pneumococcal HGT was transformation. However, as greater numbers of pneumococcal genomes become available it is increasingly apparent that conjugative and phage elements are common, and have likely played a crucial role in pneumococcal genome evolution.

Pneumococcal transformation

The scale of recombinational genetic change following transformation, and its importance for pneumococcal genome evolution has been highlighted by three recent studies: Among 240 pneumococci belonging to CC81 (defined by MLST data), 641 recombination events were identified and ranged from 3 bp to 72,038 bp in length [46]. In a second study, recombination of sequences between ~400 bp and ~235,000 bp in length were identified among three ST13 isolates and one ST2011 isolate from a single patient suffering recurrent otitis media [107]. The third study identified multiple recombination events affecting regions of the genomes of five ST695 pneumococci isolated in the USA. The recombinant genome regions ranged in size from 40 bp to ≥44,000 bp [84]. Interestingly, the latter two studies also highlighted the phenomenon of multi-fragment recombination, whereby a single genome showed evidence of recombination of multiple DNA fragments, each likely originating from the same donor cell. In these examples up to 11 and 27 such recombination fragments, respectively, were identified. The fragments represented a significant portion of the genome and were randomly dispersed around the recipient chromosomes.

The transformation mechanism

Bacterial transformation requires the induction of the so-called 'competent state'. In pneumococci, spontaneous competence development occurs in the early exponential growth phase [37] and requires a multi-step intercellular signalling cascade encoded by the *comABCDE* operon [123]. The signalling protein, competence stimulating peptide (CSP), is encoded by *comC* and is secreted from the cell via an ABC transporter. This ABC transporter and its accessory protein are encoded by *comA* and *comB* respectively. The CSP receptor, encoded by *comD,* is a histidine kinase activated by the conformational change induced upon CSP binding. ComD activation is presumed to lead to phosphorylation of the response regulator, ComE. This protein up-regulates transcription of the *comABCDE* operon and stimulates transcription of *comX*, encoding the sigma factor X (σ^X). The half-life of phosphorylated ComE is short, and so it can only exert its effects when the rate of phosphorylation is greater than the rate of degradation, a state corresponding to an extracellular CSP concentration of 1 - 10 ng/ml [121].

ComE regulated genes are the 'early competence genes'. The 'late competence genes' are those regulated by σ^X and are much greater in number [47]. These include several genes whose products are associated with DNA processing and repair, and several implicated in the fratricide pathway (see below). Overall, the expression of more than 180 genes is altered in response to raised extracellular CSP levels, the functions of many of which remain unknown [193].

DNA internalisation at the cell membrane, depicted in Figure 1.4.1, requires that the DNA molecule first cross the cell wall. It is suggested that a competence pseudopilus is responsible for this process. The pseudopilus is a composite protein structure comprising processed CglC and CglD pilin-like peptides. DNA bound to the pseudopilus is pulled across the cell wall as the pseudopilus retracts, a process dependent on the proton motive-force [33].

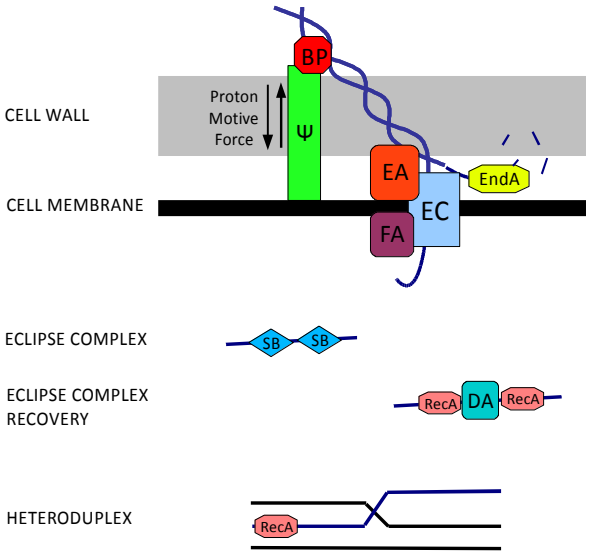


Figure 1.4.1 DNA internalisation and recombination by transformation. dsDNA, bound by an unknown binding protein (BP), is pulled across the cell wall by the pseudopilus (ψ), requiring the proton motive force. At the cell membrane dsDNA is bound by the ComEA receptor (EA) and a single strand is translocated across the membrane through the ComEC protein channel (EC). This process is assisted by the DNA translocase, ComFA (FA). The second strand is degraded by the nuclease, EndA. Following internalisation, association with SsbB (SB) results in the formation of the eclipse complex. Recovery from the eclipse complex is associated with binding of DprA (DA) and RecA, and is a pre-requisite for heteroduplex formation.

Uptake of DNA at the cell membrane is in the single-stranded form and requires a 'free end'. It is thought that a membrane-associated nuclease is able to non-specifically nick one strand of a dsDNA molecule, subsequently resulting in a double-stranded break [33]. The dsDNA molecule can bind to the receptor, ComEA and a single strand is transported across the membrane through the associated protein channel, ComEC. This process is assisted by the DNA translocase, ComFA [32]. The second DNA strand is degraded at the extracellular membrane surface by the nuclease, EndA.

Internalised ssDNA may be incorporated into the host chromosome, but ~75% of these molecules will be degraded and recycled within the cell [11]. Non-degraded ssDNA is protected from cellular nucleases by association with SsbB in the 'eclipse complex'. Up to 70,000 SsbB molecules may be present inside the cell, protecting up to ~1.15 Mb of internalised DNA [6]. It is suggested that such a volume of internalised nucleic acids may act as a reservoir for successive rounds of recombination. This hypothesis helps to explain the mechanisms allowing the large and multi-fragment recombination mentioned above.

DNA in the eclipse complex is associated with a loss of transformation ability. Recovery from the eclipse complex prior to recombination heteroduplex formation may require the DprA and RecA proteins. The CoiA and RadA competence-induced proteins have also been implicated in ssDNA processing, but their roles have not been fully elucidated [38].

In addition to its role in protection of ssDNA, RecA is thought to mediate the search for homologous regions in the host chromosome [32] and initiate heteroduplex formation [175]. Similarly, SsbB may also play a role in the stimulation of ssDNA processing just prior to recombination [6]. Extension of the heteroduplex structure involves the RecG-like MmsA protein. Resolution of the heteroduplex requires the, as yet unidentified, resolvase and ligase enzymes [38, 175].

Origins of transforming DNA

Previously it was believed that transforming DNA was obtained from the surrounding growth media as a result of leakage from dying cells. However, more recently it has been shown that competent pneumococci and other members of the Mitis phylogenetic group [37, 122] are able to actively induce lysis of non-competent neighbouring cells, resulting in the release of their DNA [176].

This process, termed "fratricide" (depicted in Figure 1.4.2), is regulated by σ^x [89] by stimulation of expression of *cibABC* and *cpbD*. CibAB is necessary but not sufficient to induce non-competent cell lysis, a process which also requires the cell wall hydrolases LytA and LytC. *lytA* and *lytC* are expressed by both competent and non-competent cells and are independent from the competence regulon.

CibAB represents a two component bacteriocin likely contributing to fratricide by increasing the permeability of target cell membranes and providing a trigger for cell wall hydrolase activity. The co-expressed *cibC* provides immunity for the competent cell against the effects of CibAB [37]. CbpD is an amidase/peptidase believed to bind target cells and cleave cell wall murein stem peptides, subsequently triggering LytA amidase and LytC lysozyme activity. Digestion of the cell wall leads to DNA release [176].

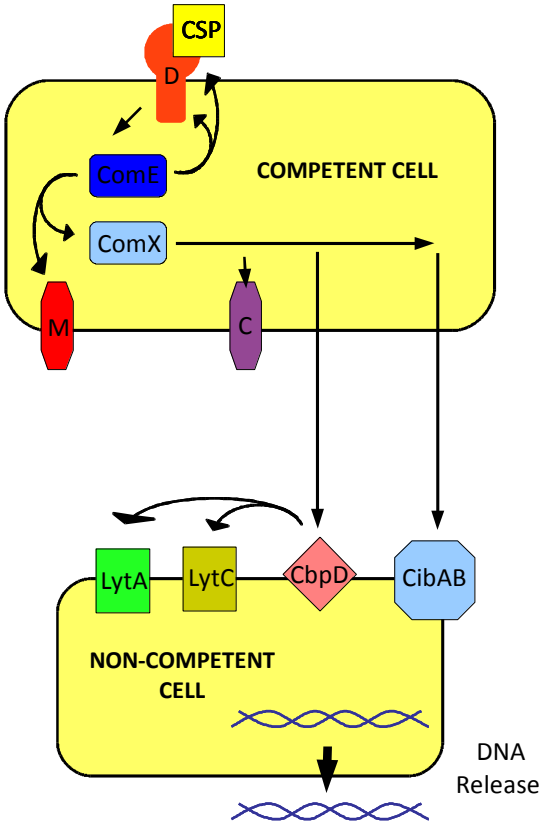


Figure 1.4.2 The pneumococcal fratricide pathway.
Cells are induced to competence by detection of extracellular CSP by the ComD (D) receptor. Upon CSP-ComD binding the transcription regulator ComE is activated, leading to increased production of the sigma factor ComX and the immunity protein, ComM (M). Production of the secreted proteins, CibAB and CpbD, and the immunity protein, CibC (C) is elicited by ComX. CibAB and CpbD are able to bind to the walls of cells not expressing CibC and ComM (non-competent cells) leading to recruitment of LytA and LytC, followed by cell lysis and DNA release.

Competent cells are protected from the effects of CbpD by the ComM immunity protein, encoded by the early competence gene, *comM* [122]. As a consequence of immunity, fratricide can only take place among mixed populations containing both competent and non-competent cells. Such a circumstance may arise due to the phenomenon of bistability (variation of gene expression by isogenic cell lines in identical growth conditions) [37] or by inter-strain variation of CSP and its ComD receptor.

Variation of CSP types (pherotypes) between pneumococcal isolates has been described, with the majority of pneumococci expressing one of two types [26, 121, 233]. One study identified an association between pherotype and genotype as defined by MLST/PFGE [26]. Subsequently, these authors suggested that HGT occurred more

commonly between isolates of the same pherotype than between isolates of different pherotypes, and that fratricide was primarily a result of bistability. If this were the case, cells expressing the same pherotype would be required to show heterogeneity in the development of competence, with those remaining non-competent also being susceptible to fratricide; such attributes have been demonstrated for the D39 laboratory strain and its derivatives [89].

In contrast, a second study found no association between pherotype and genotype, and indicated that inter-pherotype HGT was common [43a]. The latter finding is consistent with the hypothesis that fratricide occurs when a population of cells contains one or more subsets producing different CSP variants. In this situation, cells of the most common variant will likely be the first to reach competence, subsequently inducing lysis of those cells unable to detect the variant. Indeed, *in vitro* experiments by Johnsborg and colleagues showed that pneumococci were only induced to competence when exposed to extracellular CSP types complimentary to their ComD receptor, and that such competent cells were able to acquire novobiocin resistance markers from co-cultured, non-competent cells (that expressed an alternative ComD receptor) [122].

HGT is thought to occur during nasopharyngeal carriage and not during IPD. The extent to which different strains of pneumococci may co-colonise a single host is not clear, but recent evidence suggests that co-colonisation may be more common than previously thought [201, 233]. Co-colonisation of pneumococci expressing different

pherotypes has also been reported [233] and presumably permits HGT via the latter of the above described pathways.

The reasons for the differences in the findings of the two conflicting studies described above are not clear at this time. The former study included 483 invasive pneumococcal isolates, collected within a three year period in Portugal. The latter study included 88 clinical isolates collected within a two year period and from geographically diverse locations. It is therefore possible that differences in the study outcomes arose as a consequence of differences in sample size or the range of geographic locales from which the isolates were collected. Most often a larger sample size would be considered to be more reliable, particularly when size differences are substantial (e.g. the 5-6 fold difference represented here). However, differences in the findings could also have resulted from differences in the authors' methodologies, as suggested by Cornejo and colleagues (the authors of the latter study) [43a]. These authors suggested that the assumption of equal subpopulation sizes, which was an underlying assumption of the analyses by Corrolo and colleagues [26], was not valid. Whatever the causes of the differences in these findings, further study is clearly required to reach a resolution. An ideal study sample would be a large ($n \geq 200$) collection of pneumococci isolated from carriage (the environment within which HGT is thought to occur and which includes representatives of genotypes/serotypes that frequently cause disease as well as those which do not frequently cause disease).

Conjugation

Although it has never been directly observed, evidence from recent genetic data suggests that, at least at some point in its history, the pneumococcus has been able to exchange DNA via the process of conjugation. Furthermore, a transposable element known as *Tn5253,* has been successfully transferred between pneumococcal strains in the laboratory [196]. This composite element was formed from integration of a *Tn916*-like element into a *Tn5252*-like element. 29 of 55 surveyed antimicrobial-resistant pneumococcal isolates, representing a range of serotypes and STs, possessed ≥1 *Tn5253*-associated gene(s) [101]. Genomic analysis of an ST81 isolate, known as Spain^{23F}-1, also identified a ~81 kb *Tn5253*-like element, 'ICE*Sp*23FST81' [46].

The *Tn916* family of elements comprise three functional modules encoding the necessary proteins for conjugation, regulation and recombination. A fourth module includes accessory genes [202], such as the *tet(M)* gene conferring tetracycline resistance in Spain^{23F}-1. These elements have a broad host range stretching across numerous genera including *Mycobacterium, Bacillus, Staphylococcus, Neisseria, Haemophilus, Pseudomonas, Mycoplasma* and *Streptococcus* [202]. The *Tn5252* element carries a chloramphenicol acetyl transferase gene, conferring pneumococcal chloramphenicol resistance, and the *umuC/umuD* genes associated with the SOS response [101].

A more recent survey of 18 erythromycin-resistant pneumococcal isolates from China found that the *Tn916*-like, *Tn2010* and *Tn6002* elements were present in 13 and 5 isolates respectively. Both of these elements contained the *erm(B)* gene conferring

resistance to macrolides, lincosamides and streptogramin B. The former element also contained the macrolide efflux genetic assembly (MEGA) element [148], also associated with macrolide resistance. This element itself is not transposable but is incorporated into a number of other elements, including *Tn2010* and *Tn2009* [51].

Sequence comparison of the elements from the Chinese isolates showed that there was genetic variation, even among elements of the same type. Similarly, the study of 240 CC81 pneumococci found extensive variation within the ICE*Sp*23FST81 element [46]. These data suggest that conjugative elements are highly variable and may be common amongst pneumococci, which has important implications for pneumococcal genomic evolution.

Transduction

Bacterial viruses, known as bacterioophage, display two distinct life-cycles: Upon infection of the host, lytic phage can immediately hijack host cell machinery to produce new phage particles which are released through host cell lysis. Lysogenic (or temperate) phage can integrate their DNA into the host chromosome and remain in a dormant state until such time that they are reactivated, possibly by an environmental stimulus. Integration/excision from the host genome and subsequent phage genome replication require site-specific tyrosine or serine recombinases. Following reactivation and DNA replication, phage DNA is packaged into new phage particles and released through host cell lysis [79]. Accidental inclusion of host DNA within phage particles can result in transfer of this DNA to a new host cell. It is this phage-mediated movement of bacterial-associated DNA which is known as transduction [21, 79].

When phage DNA is integrated in the host genome (in the dormant state) it is known as a prophage. Over time prophage may lose the ability to reactivate if DNA degradation results in the loss or irreversible mutation of essential genes. However, the remnants of the prophage genome can remain in the host chromosome and be passed to its descendants or other bacterial cells via other HGT mechanisms. As such, a single bacterial genome may be comprised of up to 20% prophage DNA [28].

Pneumococcal phages or 'pneumophages' have so far been identified as members of three phage families; Syphoviridae, Myoviridae and Podoviridae. Pneumococcal prophage include MM1 and EJ-1 [153]. A comparative study of 10 pneumococcal prophage genomes identified three groups based on gene homology, gene synteny and site of integration in the host genome [205]. All of these prophage genomes were between 31 kb and 42 kb length and were organised into five functional gene clusters, associated with lysogeny, transcriptional regulation, packaging, morphology or host cell lysis. An independent survey of 791 pneumococci indicated a high prophage incidence, with the presence of at least one phage in 76% of isolates [198]. These data indicate that phage may also play an important role in the process of pneumococcal genome evolution.

1.5 Pneumococcal Polysaccharide Capsules

Like many bacterial species, pneumococci possess an extracellular polysaccharide capsule which is covalently attached to the peptidoglycan cell wall [245]. Encapsulated

pneumococci show increased bacterial survival during IPD compared to non-encapsulated strains. Protection is achieved through prevention of neutrophil-mediated killing [234] and a reduction of complement binding to prevent phagocytosis [81, 245]. The protective role of the capsule during nasopharyngeal colonisation is less certain, but may include prevention of mucus-mediated clearance [181].

To date 94 different pneumococcal capsule types have been identified [24, 24a, 102, 120, 160, 190] and separated into 48 serogroups based on serum antibody agglutination patterns [102]. The biochemical structures of each capsule type are distinct and play an important role in determining epidemiology. The differentiation of serotype (now serogroup) 20 into two independent sub-types was only very recently described [24a]. Future studies will likely also lead to reclassifications or descriptions of new serotypes, particularly as greater numbers of pneumococci are genetically characterised and immunological differentiation techniques are further developed. Pneumococcal conjugate vaccines (see section 1.6 *Pneumococcal Vaccines*) also influence the serotype distribution within a population and may lead to the discovery of serotypes not previously recognised. Indeed, serotype 6C was first revealed recently in the post-conjugate vaccine era, whien it was differentiated from serotype 6A during the development and validation of a novel serotyping system based on mouse monoclonal antibodies and a multiplexed immunoassay [190].

Capsule epidemiology

A recent review of 169 studies of serotype prevalence among IPD isolates from children <5 years of age, prior to the introduction of pneumococcal conjugate vaccines,

found that serotypes 14, 6B, 1, 23F, 5, 19F and 6A were the most common globally (in rank order) [124]. (For information regarding pneumococcal conjugate vaccines and the changes in pneumococcal serotype distributions in countries which have implemented vaccination programs, see section 1.6 *Pneumococcal Vaccines*.) Differences by geographic region were also indicated and had been noted previously [99]. For example, serotypes 5 and 1 were ranked second and third most common, respectively among 20 developing countries included in the review. In contrast these serotypes were ranked eighth and fifth, respectively among industrialised countries. An earlier global meta-analysis of serotypes causing IPD identified serotypes/groups 1, 19, 6 and 5 as the most common in Asia, and serotypes/groups 6, 14, 1 and 19 as the most common in Africa [100]. A separate meta-analysis of data collected across western Europe indicated that serotype 14 and serogroups 19 and 23 were the most common causes of IPD in children <18 years of age [118].

To complicate matters further, invasive pneumococcal serotypes also vary with patient age [16, 100, 203] and through time [7, 22, 36, 75, 77, 97, 139, 140, 177, 213]. In general, adult IPD is caused by a broader range of serotypes. Nevertheless, among adults the most common serotypes include the 'paediatric' serotypes/groups 6, 9 and 14 (so called because they also account for a substantial amount of disease in children), and the additional serotypes 1, 3, 8 and 12F [99].

Pneumococcal nasopharyngeal carriage may be an essential pre-requisite for invasive disease, but carriage serotype distributions are not necessarily the same as those of invasive isolates. A study of 2,799 children <7 years old in Italy identified serotypes 3,

19F, 23F, 19A, 6B and 14 as the most commonly carried [156], and similarly in Switzerland serotypes 3, 19F and 23F were the most common [138]. In England serotypes 6B, 9V, 19F and 23F were those most commonly identified from carriage [18], whilst in Turkey serotypes 11, nontypeable and 19F were identified most commonly [185]. In the Gambia serotypes 3, 6A and 23F were isolated most commonly [105] and in Bangladesh serotypes/groups 6, 19, 15, 23 and 10 were most common [85].

The differences in serotype distributions among carriage and IPD isolates reflect serotype-specific differences in the potential to cause disease, defined as 'invasiveness' or 'invasive disease potential'. Differences in invasive disease potential can be indicated by comparison of serotypes among carriage and invasive disease isolates from the same host population. One such study by Brueggemann and colleagues in Oxford showed that serotypes 4, 1, 14 and 18C were associated with a significantly increased invasive disease potential, but serotype 23F was associated with a significantly decreased invasive disease potential [18]. Similarly, a study by Hanage and colleagues found that serotypes 14, 18C and 19A were significantly associated with IPD in Finland, whereas serotypes 6A, 35F and 11A were significantly associated with carriage. Following the former study, serotype-specific invasiveness was positively correlated with differences in serotype-specific attack rates, i.e. the number of invasive disease episodes resulting from a given number of carriage episodes of a given serotype [217].

The reasons for serotype-specific differences in invasiveness are not completely understood, but may result from differences in capsule structure. It has been noted that the degree of capsule production is serotype-dependent and likely reflects metabolic constraints acting on differences in capsule structure complexity. Pneumococci expressing complex capsules tend to produce them in lesser volume and are less well adapted for growth in nutrient-restricted conditions [98]. Conversely, pneumococci expressing simpler capsules produce a greater volume and show greater resistance to neutrophil-mediated killing [237]. Rate of complement deposition on pneumococcal cells has also been shown to vary by serotype [166] and it is likely that those serotypes associated with a lower rate will have a greater invasive disease potential.

Wzy-dependent capsule synthesis

Capsule types may be biochemically distinct but almost all (92/94) are synthesised through the 'Wzy-dependent' pathway [10, 137, 245]. Briefly, the component monosaccharides are synthesised and successively linked to a lipid carrier at the cytosolic surface of the cell membrane. The resulting lipid-linked repeat unit is transferred across the membrane by a 'flippase'. Finally, repeat units are polymerised at the outer membrane before attachment to the peptidoglycan cell wall.

The genes encoding the enzymes required for Wzy-dependent capsular synthesis have been mapped to a region of the pneumococcal chromosome termed the *cps* (capsular polysaccharide synthesis) locus, situated between the *dexB* and *aliA* loci [10, 160, 245]. Throughout the 1990s and early 2000s the loci of strains of several different serotypes

were sequenced and characterised [137, 173, 174, 179, 199], and more recently the sequences for all but three (6C, 6D and 11E) of the Wzy-dependent capsule *cps* loci were made available in a single study [10].

This sequence data shows that the *cps* locus varies in size from ~10 kb to ~30 kb [10] and contains a variable number of genes e.g. 11 genes for serotype 1 [179] and 18 genes for serotype 18C [119]. Generally the *cps* genes have the same orientation as that of *dexB* and *aliA,* and appear to be expressed as a single transcriptional unit. A promoter sequence is located upstream of the first gene (*wzg*) and a stem-loop termination structure is located downstream of the last gene [10, 113, 137, 173].

In many cases the regions between the *cps* genes and the flanking genes, *dexB* and *aliA*, contain components of complete or partial mobile genetic elements [245]. For example, the serotype 1 *cps* locus is flanked by IS1167 [179] and the serotype 23F locus is preceded by IS1202 [199]. The *cps* genes themselves follow a common pattern: The first and second genes (*wzg* and *wzh*) are highly conserved amongst all serotypes [173, 245]. Whilst the third and fourth genes (*wzd* and *wze*) can be divided into two conserved groups [172, 245]. Together these four genes, also known as *cpsA, cpsB, cpsC* and *cpsD* respectively, are the type non-specific regulatory genes [10].

The type-specific genes are located 3' of *wzg-wzd* and encode transferases, polymerases and synthases specific to capsular type. Like *wzd* and *wze, wchA* (or *cpsE*), can also be divided into two conserved groups [245]. This gene encodes the initial glucose phosphate transferase, linking the activated glucose monosaccharide to

the membrane bound lipid carrier [10]. The *wzy* and *wzx* genes encode a polysaccharide polymerase and a flippase, respectively, and are present in the type-specific *cps* regions for all of the Wzy-dependent capsule types [10].

Shared homology of the type-specific region may be found between strains of different serotypes, most commonly between members of the same serogroup. For example, the *cps* loci of serotype 19B strains (*cps19B*) contain two regions of homology to those of serotype 19F strains (*cps19F*) [174]. Both the 19F and 19B capsule share similar structures, with 19B possessing an extra sugar in its polysaccharide backbone and an extra disaccharide side chain. The 10.5 kb region of *cps19B* which is not shared with *cps19F* includes five genes. Homology analyses of the predicted protein products of these genes suggests that they encode the necessary transferases and transport proteins for the addition of the 19B-specific sugars to the capsule [174].

In some cases shared homology of type-specific regions can be detected between strains of different serogroups. For example, part of the *cps23F* locus shows homology to that of *cps19F* [199]. The genes in this region have been shown to encode proteins associated with the synthesis and processing of dTDP-rhamnose, which is present in both the 23F and 19F capsules [173].

Synthase-dependent capsule synthesis

Two of the currently known pneumococcal capsule types do not appear to be synthesised through the Wzy-dependent pathway. These types, 3 and 37, are thought

to be produced by the synthase-dependent synthesis pathway, requiring only a single glycosyltransferase [245].

The type 3 capsule contains the repeated unit, D-glucuronic acid β (1➔ 4) linked to D-glucose [80], and thus its production requires the synthesis and polymerisation of the sugar precursors, UDP-glucose and UDP-glucuronic acid [54]. The necessary proteins are transcribed from genes located at the *cps* locus [245]; however, in serotype 3 strains the organisation of this locus does not follow the patterns observed for the Wzy-dependent capsules. In this case five complete and one interrupted genes are located at the locus [4] and are likely transcribed as two separate units [4, 54].

The *cps* loci of serotype 37 pneumococci are defective, and the type 37 capsule is produced by the synthase gene, *tts,* located in a different region of the genome. This capsule is composed of β-D-Glucose-(1➔2)-β-D-Glucose linked by β-1,3 bonds. The *tts* gene encodes a protein with sequence motifs characteristic of cellulose synthases and β-glucosyltransferases. Interestingly, in several type 37 strains the possession of the *tts* gene is associated with a genome re-arrangement event. In all cases *tts* is flanked at its 5' end by *gpmA* and, in several cases, at its 3' end by *pyrDA*. However, in other strains the *gpmA* and *pyrDA* genes are located >380 kb apart. Llull and colleagues showed that this genome rearrangement followed transformation of the M24 pneumococcal laboratory strain with type 37 chromosomal DNA, but did not result in every case. These authors suggest that the IS1167 sequence, associated with the *tts* gene, may play a role in its integration and the possible rearrangement of the chromosome [152].

Evolution of the capsule genes

Analysis of the G+C content of the *cps* genes shows that the type-specific genes generally have a higher G+C content than the type non-specific genes, and this together with the evidence suggesting that capsule genes of different species cluster closely to some of those of the pneumococcus suggests that, at least parts of the *cps* loci have been acquired by HGT from other species [10, 232]. Indeed, it has been shown that capsular switching events, whereby a strain can exchange its capsular genes through HGT, can and do occur in natural pneumococcal populations [19, 41, 42].

In some cases features of specific *cps* loci have allowed speculation about their particular origins. The analyses of Bentley *et al*. and Mavroidi *et al*. [10, 160] allow further suggestions to be made about the evolution of different *cps* loci. These authors suggest that differences between loci in the same subcluster are a result of divergence from a common ancestor. This could be the case for types 9A and 9V which both have highly similar capsule loci, differing by only two nucleotides [10]. In some cases loci in the same subcluster show almost complete synteny with the exception of a single gene or region, for example, the five genes present in the *cps19B* locus that are not homologous to those of the *cps19F* locus. These genes have a lower G+C content than the homologous regions (27.2 - 29.7 mol% compared to 30.3 - 42.3 mol%) [174] suggesting that acquisition of these genes resulted from HGT. In addition, analyses have shown that differences between some *cps* loci are due solely to the presence/absence of different mobile genetic elements. This is true of types 44 and 46

which differ from type 12F only by differences in insertion sequence transposase genes [10]. Thus it seems that the variation of capsule types and capsular loci observed in pneumococcal populations is a result of a combination of HGT events, movement of mobile genetic elements and mutation.

Variation among the *cps* loci of strains expressing the same serotype has also been noted for a few types. Nucleotide sequences of three genes within the *cps* loci of >100 serogroup 6 isolates revealed within-serotype allelic diversity, and that *cps6B* loci cluster into two groups differing by >5% nucleotide divergence [17, 161]. Additionally, the *cps* loci of two isolates indicated a recombinogenic hybridisation of sequences representing each of the two clusters. The *wciP, wzy* and the first half of the *wzx* gene were identical to those of isolates representing *cps6B* cluster 2. However, the remaining *wzx* region was identical to that of isolates representing *cps6B* cluster 1 [161].

An independent analysis of the entire length of 11 *cps6A* loci and 6 *cps6B* loci indicated 6 and 3 subtypes respectively [60]. Major differences between subtypes were primarily located within the *wzg* and *rmlA-D* genes (required for rhamnose synthesis). In fact, a high concentration of nucleotide substitutions within a 166 bp region of the *wzg* gene indicated that this region may represent a mutation hotspot. One *cps6B* subtype was highly divergent from the other three subtypes at every *cps* gene, and the authors suggested that this sequence may represent *cps6B* cluster 2 as defined in the previous studies.

A similar analysis of the complete sequences of 10 *cps19A* and 10 *cps19F* loci revealed three and four subtypes, respectively [60]. Among the *cps19A* subtypes, the principle differences were again located in the *wzg* and *rmlA-D* genes. Interestingly, the *rmlD* genes of two subtypes were reverse-oriented. It is not clear to what extent the reversion affected the expression of this gene, but it did not seem to prevent production of the type 19A capsule. Finally, closer comparison of serogroup 6 and serogroup 19 *rmlA-D* genes provided additional evidence of *cps* recombination between representatives of the same serotype [60].

The analyses above highlight the immense variability of the pneumococcal *cps* locus. Such variation was further evidenced by a study of nontypeable and atypical encapsulated pneumococci by Salter and colleagues [210]. An atypical *cps33C* locus appeared to represent a mixture of reference *cps33C* and *cps33B* genes. A *cps22F* locus contained two novel genes which were highly divergent from those of the reference *cps22F* locus, but were predicted to encode peptides with the same function. Further, the comparison of *cps* loci of nontypeable isolates indicated several categories of genetic change leading to nontypeability: Firstly, the *cps* loci of three isolates were completely deleted, with only a transposase remaining between the *dexB* and *aliA* loci. Secondly, the *cps* loci of five isolates were partially deleted. Finally, the *cps* loci of 10 isolates from Thailand and two isolates from The Netherlands were replaced by insertion sequences and a newly identified gene named "*nspA"*. This novel gene was highly variable and was predicted to encode a surface exposed, potentially immunogenic protein. Park and colleagues also identified the same gene among 9 pneumococci from South Korea [189]. These authors named the gene, "*pspK",* and

showed that it was associated with enhanced nasopharyngeal colonisation ability when compared to other nontypeable, or *cps* knockout mutant isolates. The prevalence of the *nspA*/*pspK* gene among pneumococcal populations, and its importance during pneumococcal disease are yet to be determined.

Given the potential variability and plasticity of the *cps* locus, and the impact of serotype on pneumococcal epidemiology, it is essential that we better understand the processes and history of genetic change in this region of the chromosome.

1.6 Pneumococcal Vaccines

There are currently two types of licensed pneumococcal vaccines. These are the 23-valent polysaccharide vaccine and the protein-conjugate vaccines (PCV, 7-, 10- or 13-valent). Both types of vaccine are designed to elicit an immune response to the pneumococcal capsular polysaccharides in the vaccine formulas (vaccine types, (VT)), and thus protection against only a subset of pneumococci is provided. The capsule types contained within each of the vaccines are listed in Table 1.6.1. The types contained within PCV7 were chosen as those most commonly isolated from IPD in the USA. In fact, serotype 14 was considered to be the most common cause of IPD, prior to conjugate vaccine introduction, among all regions in a global meta-analysis [124]. Serotype 6B was ranked second in all regions except Africa, where it was the fifth most common cause of IPD. Serotypes 19F and 23F were shown to be important causes of disease in all regions. Serotype 18C was a highly ranking cause of IPD in regions largely comprising economically developed countries e.g. North America and Europe.

Serotypes 4 and 9V were also implicated as important causes of disease in these regions [18, 93, 132]. In contrast, serotypes 1 and 5 were more common causes of IPD in developing countries such as those in Africa, Asia and South America [100, 124].

The additional capsule types included in PCV10 and/or PCV13 were chosen because they were important causes of pre-vaccine implementation disease in countries other than the USA (e.g. serotypes 1 and 5) and/or represented serotypes which increased in prevalence among IPD isolates after PCV7 introduction (e.g. 1, 3, 7F and 19A – see below for further details).

Polysaccharide vaccine

The 23-valent pneumococcal polysaccharide vaccine (produced by Merck & Co., INC), was licensed in the USA in 1977 and recommended for individuals at risk of serious pneumococcal disease such as the immunocompromised, the elderly and persons suffering chronic pulmonary, cardiac or renal disease [214]. The immune response elicited by this vaccine is a T-cell independent response (i.e. anti-pneumococcal antibody production by B-lymphocytes is independent of T-helper lymphocyte action). Consequently, the vaccine is not effective in children <2 years of age [76]. For persons over the age of 65, the vaccine is effective at reducing the risk of IPD, but may not be effective in the reduction of community-acquired pneumonia [114, 214]. Nevertheless, the range of serotypes for which protection is elicited and its efficacy in protection against IPD has led to the continued use of this vaccine.

Table 1.6.1 Capsular polysaccharide vaccine types (serotypes) included within the currently licensed pneumococcal vaccines.

Serotype	PPV	PCV7	PCV10	PCV13
1	x		x	x
2	x			
3	x			x
4	x	x	x	x
5	x		x	x
6A				x
6B	x	x	x	x
7F	x		x	x
8	x			
9N	x			
9V	x	x	x	x
10A	x			
11A	x			
12F	x			
14	x	x	x	x
15B	x			
17F	x			
18C	x	x	x	x
19A	x			x
19F	x	x	x	x
20	x			
22F	x			
23F	x	x	x	x
33F	x			

PPV = pneumococcal polysaccharide vaccine.
PCV = pneumococcal conjugate vaccine.

Pneumococcal conjugate vaccines

PCVs differ from the polysaccharide vaccine in that each of the component capsular polysaccharides is conjugated to a carrier protein (diphtheria toxoid for PCV7 and PCV13, a conserved *Haemophilus* protein for PCV10), making the vaccine more complex and costly to produce, but capable of eliciting a more robust immune response [50, 76]. PCVs have been shown to elicit T-cell dependent antibody responses in vaccinees previously non-responsive to the polysaccharide vaccine [247], and in children <2 years old [62]. PCV7 (produced by Pfizer Vaccines) was shown to be

highly effective in the protection of young children against IPD [12], and was shown to significantly reduce carriage of PCV7-type (VT^7) pneumococci [182]. Additionally, this vaccine was shown to be moderately effective in protecting against acute otitis media [63]. PCV7 was licensed for use in young children and infants in the USA in 2000 [1], was introduced into the childhood vaccination programme in Canada in 2002 [131] and in the UK in 2006. Subsequently, many other countries licensed and introduced PCV7 into their national childhood immunisation programmes.

Following PCV7 introduction in the USA and Canada a significant decline in IPD in young children was observed [132, 240]. In addition, declines in IPD amongst adults and children above the age of vaccination were also seen, suggesting that vaccination of infants and toddlers had produced a level of herd immunity protecting the unvaccinated population. Herd immunity was most likely the result of the reduction in VT^7 carriage among vaccinees, and subsequent disruption of pneumococcal transmission from these vaccinees to older children and adults [147, 194].

Despite reports of an overall reduction of pneumococcal disease following PCV7 introduction, it has been shown that IPD [104, 132, 178, 236], acute otitis media [63] and carriage [110, 131, 168] of NVT^7 (non-PCV7-type) pneumococci has subsequently increased. In the USA this increase has been attributed to the increased prevalence of pneumococci expressing serotypes 1, 3, 6C, 7F, 15, 19A, 22F and 33F [104, 116, 129]. In fact, serotype 19A became the most common cause of IPD in some regions [116, 129, 192]. These changes in pneumococcal serotype distributions are a result of two

phenomena: serotype replacement, the expansion of pre-existing NVT^7 pneumococcal clones [151, 187]; and capsular switching, the acquisition and expression of NVT^7 *cps* genes by previously VT^7-associated clones [19].

The relative frequency with which these phenomena occur is not fully understood. It has been shown in the USA that at least five independent capsular switching events have occurred to produce serotype 19A variants of the ST695 clone. This clone was previously only ever associated with the VT serotype 4. It is believed that the primary variant, detected in 2003, originated in the north-eastern states and has subsequently spread westwards across the country to become the third most common cause of serotype 19A IPD [9, 84].

PCV10 (also known as PHiD-CV, produced by GlaxoSmithKline) and PCV13 (produced by Pfizer Vaccines) are now available, and are used as a replacement for PCV7. PCV13 has been used in the USA [30] and UK childhood immunisation programmes since 2010. As part of an innovative programme run through the GAVI (Global Alliance for Vaccines and Immunizations) Alliance and supported by the WHO and UNICEF, PCV10 and PCV13 vaccination programmes have also begun in developing countries. This scheme, which is part funded by a $1.5 billion US donation from the Bill and Melinda Gates foundation and the governments of Canada, Italy, Norway, Russia and the UK, is a pilot Advance Market Commitment (AMC) programme [31]. Between 2013 and 2023, GlaxoSmithKline and Pfizer Vaccines have each agreed to supply 30 million yearly doses of their vaccines for distribution in GAVI eligible countries [221]. A further

combined 7 million doses were pledged for 2010, with 24 million and 20 million pledged for 2011 and 2012, respectively.

GAVI eligible countries are those for which the annual per capita income is <$1,500 US. The AMC price per PCV dose is set at $7 for the first two years and $3.50 thereafter, which is considerably lower than the $96 per dose price in the USA. To date, PCV vaccination programmes have begun in 15 GAVI eligible countries, including Cameroon, Ethiopia, Honduras, Malawi, Nicaragua, Rwanda, and Yemen. In total, 55 countries have been approved for the scheme [GAVI Alliance online, www.gavialliance.org].

PCV13 has been shown to elicit anti-pneumococcal immune responses comparable to those of PCV7 for the 7 shared serotypes. Functional responses to the 6 additional serotypes are also elicited [64]. Children vaccinated with at least one dose of PCV13 had a lower rate of pneumococcal carriage than those vaccinated with PCV7, and showed a significant reduction in carriage of serotype 19A and 7F pneumococci [43]. Concordantly, early reports of post-PCV13 IPD among children <2 years old in the USA indicated that there had been a reduction of IPD due to VT^{13} pneumococci, primarily due to reductions in serotype 19A and 7F disease [171]. Continued surveillance of pneumococcal serotype epidemiology is clearly necessary to assess the long-term impact of second-generation PCVs, and to detect any prevalence changes of $NVT^{10/13}$ pneumococci.

1.7 Antimicrobials and Antimicrobial Resistance

Over the past three decades pneumococcal antimicrobial resistance has become of increasing concern, with *in vivo* resistance detected for each of the major drug classes used to treat pneumococcal infections. Table 1.7.1 describes antimicrobials which exhibit pneumococcal bactericidal/static activity, and the corresponding mechanisms of pneumococcal resistance. Rates of resistance are highly variable by drug, geographic region, host age and even by study. Pneumococcal β-lactam-resistance rates and mechanisms will be discussed in detail below, followed by a brief description of resistance rates for other antimicrobials.

β-lactams

The β-lactams are a broad spectrum class of antimicrobial drugs active against a variety of bacteria including *Escherichia coli, Staphylococcus aureus, Neisseria meningitidis* and *S. pneumoniae.* This class of drugs includes the penicillins, the penicillin-like drugs (e.g. meticillin, amoxicillin and ampicillin) and the cephalosporins (e.g. cefotaxime, cefuroxime and ceftriaxone). The penicillins and penicillin-like drugs are bacteriocidal, causing eventual lysis of the susceptible bacterial cell [243]. The cephalosporins however, are bacteriostatic, preventing growth of susceptible cells without leading to lysis. The mechanism through which these effects are achieved is by binding and acetylation of penicillin-binding proteins (PBPs), preventing their normal functionality.

Table 1.7.1 Anti-pneumococcal antimicrobial drugs, modes of action and pneumococcal resistance mechanisms.

Antimicrobial Class	Examples	Mode of Action	Bacteriostatic or Bacteriocidal	Resistance Mechanism	Primary Resistance Determinants
β-lactams (Penicillins)	Penicillin G, Amoxicillin, Ampicillin, Meticillin	Binding and acetylation of penicillin binding proteins to prevent cell wall cross-linking	Bacteriocidal	Target site alteration (acquisition of altered genes through HGT)	*pbps*
β-lactams (Cephalosporins)	Cefotaxime, Cefuroxime, Ceftriaxone	Binding and acetylation of penicillin binding proteins to prevent cell wall cross-linking	Bacteriostatic	Target site alteration (acquisition of altered genes through HGT)	*pbps*
Macrolides	Erythromycin, Clarithromycin, Azithromycin	50s ribosomal subunit binding, preventing peptide elongation during protein synthesis.	Bacteriostatic but can be bacteriocidal at high doses.	Efflux pump or target site alteration (methylation)	*erm, mef*
Chloramphenicol		50s ribosomal subunit binding, preventing peptide elongation during protein synthesis.	Bacteriostatic but can be bacteriocidal at high doses.	Drug degradation	*cat*
Tetracyclines	Chloretetracycline, Oxytetracycline, Doxycycline	30s ribosomal subunit binding disrupting protein synthesis.	Bacteriostatic	Production of ribosomal protection proteins	*tetM, tetO*
Trimethoprim-Sulfamoxazole (Co-trimoxazole)		Inhibits dihydrofolate reductase and dihydropteroate synthase.		Target site alteration (point mutation)	*folA*
Fluoroquinolones	Ciprofloxacin, Levofloxacin, Ofloxacin	Bind to DNA gyrase, Type II and IV topoisomerases	Bacteriocidal	Efflux pump or target site alteration (point mutation)	*pmrA, parC, gyrA*

References [2, 83, 158, 227, 228, 243]

The pneumococcus has six types of PBP, five of which are high molecular weight molecules (PBP1a, PBP1b, PBP2a, PBP2b and PBP2x) and one of which is a low molecular weight molecule (PBP3). These proteins catalyse the latter steps of murein biosynthesis, cross-linking peptidoglycan molecules in the bacterial cell wall [91]. The high molecular weight PBPs can be further divided into two classes; class A, including PBP1a, PBP1b and PBP2a exhibit both glycan chain polymerisation and peptidoglycan transpeptidase activity. Class B, comprising PBP2b and PBP2x, have only been associated with transpeptidase activity [186]. Each PBP has three conserved active site domains, Ser-X-X-Lys, Ser-X-Asn and Lys-Thr (Ser)-Gly which vary in position, but show conserved synteny within the protein structures.

The primary mechanism of pneumococcal β-lactam-resistance appears to be altered PBPs, encoded by divergent *pbp* genes. These protein alterations reduce the affinity of PBPs for β-lactams and thus prevent β-lactam binding, and in some cases, PBP acetylation. However, mutations in other genes encoding a variety of proteins have also been implicated. These include the *murMN/fibAB* operon [74, 235], the *ciaRH* operon [157] and the *clpL* locus [231].

Studies of pneumococcal *pbps* have shown that nucleotide sequences of β-lactam-susceptible strains are highly conserved [59, 141] but sequences of resistant strains contain blocks of DNA diverging from that of susceptible strains by up to 25% [91, 142]. In general, there is a correlation between the β-lactam minimum inhibitory concentration (MIC) of a strain and the level of divergence of its *pbp* genes [86]. It is widely assumed that divergent *pbp* gene blocks originated from a bacterial species

other than *S. pneumoniae*: Three penicillin-resistant isolates of major clones in eastern Europe have been shown to possess mosaic *pbp2b* blocks with similarity to that of the *S. oralis* M3 strain [216], and the divergent regions of many *pbp* sequences also show high similarity to sequences from *S. mitis* [35, 58]. Furthermore, it has been shown that penicillin resistance determinants can be transferred between *S. mitis* and *S. pneumoniae,* and subsequently between *S. pneumoniae* and *S. oralis* or *S. sanguis* through the process of HGT [216].

High-level β-lactam resistance requires acquisition of at least two altered high molecular weight PBPs, most commonly altered PBP2x, PBP1a or PBP2b. PBP2x has one of the highest affinities for β-lactams [159], and is thus often the first PBP altered in laboratory experiments selecting for resistance [25, 87]. Unlike PBP2x, alterations of PBP2b are associated with penicillin-resistance only, and are not required for cephalosporin-resistance [164] since cephalosporins do not bind PBP2b [87]. Expression of low affinity PBP2x and PBP2b may be a necessary pre-requisite for acquisition of low affinity PBP1a, which is associated with high level resistance to both penicillin and cephalosporins [218].

Resistance-conferring *pbps* contain many nucleotide substitutions compared to those of susceptible strains. Many of these nucleotide substitutions are synonymous [219], but a reasonable proportion do encode protein amino acid substitutions. *In* vivo, it seems that no given mutation is completely necessary or sufficient to confer resistance, but rather that certain combinations of mutations within specific *pbp* genetic backgrounds are definitive [220]. Notwithstanding, several mutations of

increased importance for resistance have been identified, either because they are found commonly amongst resistant strains or have been shown to affect resistance in laboratory manipulation experiments. Most of these mutations are located within the PBP transpeptidase domains, which are also thought to be the penicillin-binding domains. Furthermore, a large proportion of these mutations affect amino acids in or near the conserved active site motifs [23, 25, 34, 53, 73, 142, 184, 212, 226, 246].

Pneumococcal penicillin-resistance is generally defined in accordance with the Clinical and Laboratory Standards Institute (CLSI) guidelines. Prior to 2008 the CLSI breakpoint for penicillin-nonsusceptibility was defined as MIC >0.06 μg/ml (Table 1.7.2). It was noted that non-meningitis IPD and respiratory tract infections due to penicillin-nonsusceptible isolates, as defined by these guidelines, could be treated successfully with an appropriate dosage of penicillin. However, penicillin treatment of penicillin-nonsusceptible pneumococcal meningitis was subject to treatment failure, reflecting difficulties in maintaining sufficient drug concentrations in the cerebral spinal fluid [238]. In 2008 the CLSI guidelines were changed to reflect treatment outcomes by disease syndrome and treatment route (Table 1.7.2). For reasons of continuity and simplicity, penicillin-nonsusceptibility will be defined according to the old (pre-2008) CLSI guidelines throughout this thesis, unless otherwise stated.

Table 1.7.2 CLSI pneumococcal penicillin-nonsusceptibility breakpoints.

CLSI Guidelines*:	Penicillin-susceptibility determined by MIC (μg/ml)		
	Susceptible	**Intermediate**	**Resistant**
Pre-2008			
All infections (all routes)	≤ 0.06	0.12 - 1.0	≥ 2.0
Post-2008			
Meningitis (intravenous route)	≤ 0.06	-	≥ 0.12
Non-meningitis (intravenous route)	≤ 2.0	4.0	≥ 8.0
Non-meningitis (oral route)	≤ 0.06	0.12 - 1.0	≥ 2.0

* Type of infection (treatment route). Reference [238].

Pneumococcal penicillin-resistance was first reported in the 1960s in the USA and Australia [94, 135]. Throughout the next ~10 years, penicillin-nonsusceptible pneumococci (PNSP) were sporadically reported in Australia/Oceania [95, 96], Europe [109, 170] and North America [55, 143, 155, 180, 188, 229], and by the late 1970s the first high-level PNSP were reported from South Africa [3, 117]. Over the following two decades penicillin-resistance spread rapidly around the globe [71, 149]: PNSP represented 17.0% of invasive pneumococcal isolates in Spain between 1979 and 1985 but between 1993 and 1997 this figure had risen to 38.8% [72]. Among community-acquired respiratory tract infection isolates from France and the USA, PNSP levels increased from 7.7% and 5.6% in 1992, to 35.8% and 20.4% in 2001 respectively [71]. Conversely, resistance rates in some countries e.g. the UK and Germany remained low throughout this time [71]. (See Figure 1.9.1 for a timeline of the development and spread of pneumococcal penicillin-nonsusceptibility.) Antimicrobial consumption increased greatly in the 1980s and 1990s [72, 163] and it is likely that the selective pressures imposed by antimicrobial use contributed to the spread of penicillin-resistance. Indeed, countries or regions with a greater level of β-lactam consumption generally exhibit higher rates of penicillin-nonsusceptibility [103, 200]. Additionally,

among children <6 years of age, cephalosporin treatment within 30 days prior to sampling was associated with increased risk of β-lactam-resistant pneumococcal carriage [211].

Recently estimated penicillin-nonsusceptibility rates from large multi-centre studies are shown in Table 1.7.3. Rates are highly variable by geographic region, with the highest nonsusceptibility levels in Asia and the lowest in Northern Europe. Few multi-centre studies of pneumococcal antimicrobial susceptibilities in Africa have been reported; a single study from Malawi indicated that PNSP levels in this region were ~15%, whilst a meta-analysis of smaller studies in Sub-Saharan Africa indicated that PNSP represented <3% of pneumococcal isolates [5].

Pneumococci isolated from children are more likely to be penicillin-nonsusceptible than those isolated from adults [126], and pneumococci expressing serotypes commonly carried or causing disease among children are also more likely to be penicillin-nonsusceptible [223]. Consequently, following the introduction of PCV7, (which targeted these serotypes), there was a reduction of penicillin-nonsusceptibility levels in the USA [240]. More recently however, rates of penicillin-resistance among NVT^7 pneumococci have increased [82, 195], and the increase of pneumococcal disease and carriage due to NVT^7 pneumococci has subsequently been accompanied by a secondary increase of PNSP levels (Figure 1.7.1) [167].

Table 1.7.3 Recently estimated pneumococcal penicillin-nonsusceptibility rates.

Region	Country	Year(s)	Reference	PNSP (%)
Asia	Hong Kong	2008-2009	[134]	98.2
	Vietnam	2008-2009	[134]	79.0
	South Korea	2008-2009	[134]	77.1
	Taiwan	2008-2009	[134]	73.6
	China	2008-2009	[134]	71.6
	Thailand	2008-2009	[134]	57.1
	Malaysia	2008-2009	[134]	26.6
	Philippines	2008-2009	[134]	7.7
North America	USA	2009	[127]	41.0
Africa	Malawi	2009	[66]	~15.0
Australasia	Australia	2006	[204]	10.6
Europe	Spain	2010	[65]	29.8
	France	2010	[65]	27.6
	Ireland	2010	[65]	18.0
	Portugal	2010	[65]	14.7
	Finland	2010	[65]	14.2
	Italy	2010	[65]	9.2
	Czech Republic	2010	[65]	4.9
	Sweden	2010	[65]	3.8
	Germany	2010	[65]	3.7
	Denmark	2010	[65]	3.6
	Norway	2010	[65]	3.6
	UK	2010	[65]	3.1
	Netherlands	2010	[65]	2.0
	Belgium	2010	[65]	0.4

Post-PCV7 introduction, serotype 19A IPD isolates were more likely to be penicillin-nonsusceptible than non-19A isolates [129]. Thus introduction of PCV13 in the USA, which includes serotype 19A polysaccharides, may be associated with a secondary reduction of PNSP levels. Continued surveillance of pneumococci in the USA and elsewhere in the era of second generation vaccines will be crucial to monitor these trends and any subsequent increase of penicillin-nonsusceptibility amongst other NVT pneumococci.

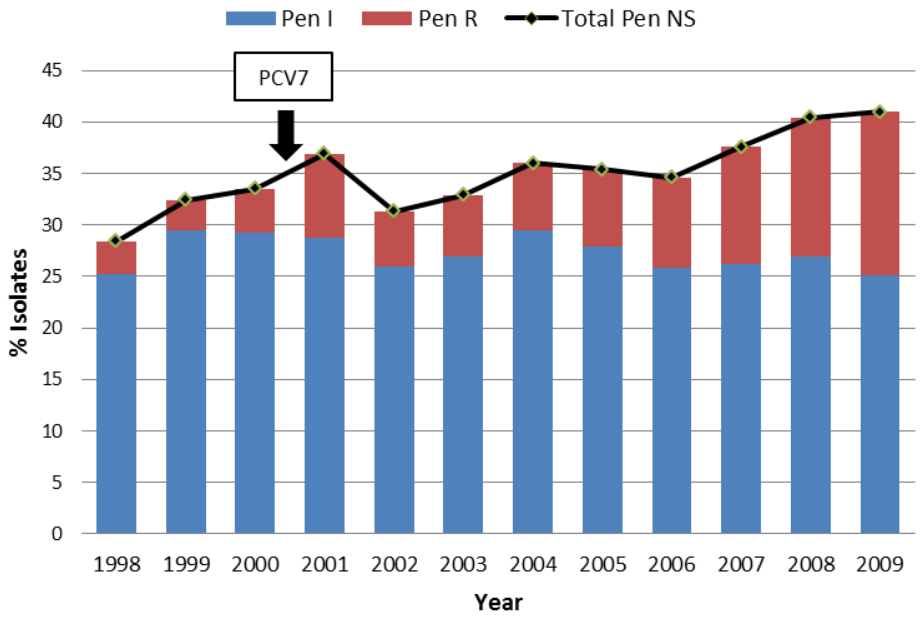


Figure 1.7.1 Rates of peumococcal penicillin-nonsusceptibility in the USA before and after pneumcooccal conjugate vaccine introduction.
Data from the SENTRY Antimicrobial Surveillance Program [127] show the percentage of penicillin-intermediate (Pen I: 0.12 µg/ml ≤ MIC ≤1.0 µg/ml), penicillin-resistant (Pen R: 2.0 µg/ml ≤ MIC) and the combined total penicillin-nonsusceptible (Total Pen NS: 0.064 µg/ml < MIC) isolates among pneumococci causing community-acquired respiratory tract infections in the USA between 1998 and 2009. Introduction of the first generation pneumococcal conjugate vaccine (PCV7) is marked.

Non-β-lactams

As for the β-lactams, rates of pneumococcal resistance to other antimicrobial drugs vary by geographic region and host age. Respiratory tract isolates from children <5 years of age were significantly more likely to be resistant to macrolides, chloramphenicol, tetracycline and trimethoprim-sulfamethoxazole (TMP/SMX) than those from older patients [126]. In contrast, there was a trend for increasing levels of fluoroquinolone-resistance with increasing host age, likely reflecting differences in prescribing practices among different patient groups.

Data from the European Antimicrobial Resistance Surveillance Network (EARS-Net) in 2010 suggested that pneumococcal macrolide-nonsusceptibility in Europe ranged from 3.8% in Norway to 30% in France ([65], in countries for which >100 isolates were tested). In 2009 macrolide-nonsusceptibility in the USA was estimated at 39.2% by the SENTRY study [127], whilst rates in Sub-Saharan Africa were estimated at only 7% [5]. Worryingly, the Asian Network for Surveillance of Resistant Pathogens (ANSORP) study indicated much higher levels of macrolide-resistance in Asia, with 96.4% of isolates from China resistant to erythromycin and >80% of isolates from South Korea, Taiwan and Vietnam nonsusceptible to this drug [134]. Another concern are the increasing rates of penicillin-erythromycin co-resistance which, in a 10 year period beginning in 1992, rose from 1.8% to 32.7%, 3.7% to 17.0% and 3.2% to 15.3% in France, Spain and the USA respectively [71].

Surveillance studies estimated that 3.9% - 6.9% of pneumococcal isolates from the UK [68, 125] were tetracycline-resistant, with nonsusceptibility rates of 1% - 84.8% in Asia [144], 24.5% in the USA [127], 13.8% in Germany, 35.1% in Spain and 42.3% in France [125]. TMP/SMX-nonsusceptibility rates rapidly increased in the 1990s and have more recently been reported at 72.2% in Thailand and 86.1% in China [134], 34.3% in the USA [127], 34.1% in Latin America [29], 28.2% in the UK, 48.5% in Spain, 49.7% in the Middle East and 56.2% in France [115].

Pneumococcal chloramphenicol-resistance levels are generally lower than those for β-lactams and the other drugs mentioned above; rates have been reported at 5.0% in

Sub-Saharan Africa [5], 3.8% in the UK, 8.2% in the USA, 20.3% in France, 26.1% in Spain, 27.1% in Singapore and 66.8% in Hong Kong [115]. Use of fluoroquinolones is often reserved for disease cases where pneumococcal resistance to other antibiotics is suspected. As such, rates of pneumococcal fluoroquinolone-resistance have remained low in most regions. Levofloxacin-nonsusceptibility rates of 0.2% for 10 Latin American countries [29] and 0.8% for the USA [127] have been reported. Ofloxacin-resistance rates of 0.5%, 0.7%, 0.7%, 0.9% and 1.2% were reported for Belgium, the USA, Germany, Spain and France respectively [71]. Low rates (<1%) have also been reported in Asia, with the exception of estimates of 3.3% and 9.1% levofloxacin-nonsusceptibility in Thailand and Taiwan, respectively.

1.8 Pneumococcal Molecular Epidemiology Network Clones

Much of the spread of pneumococcal antimicrobial resistance is attributed to the international dissemination of a small number of clones [14, 44, 88, 136, 145, 208, 209, 215, 222]. Consequently, in 1997 the Pneumococcal Molecular Epidemiology Network (PMEN) was established to monitor global antimicrobial-resistant pneumococci and standardise the associated nomenclature of internationally disseminated resistant clones. The first 16 PMEN clones were described by McGee and colleagues in 2001 [165]. These clones were characterised by serotype, PFGE, MLST and *pbp* fingerprinting. They comprised 6 unique serotypes and 16 unique STs. The original criteria for inclusion were as follows:

1) Wide geographic distribution of the clone within a country or internationally.
2) Establishment of the clone over a number of years.

3) Resistance to one or more antibiotics commonly used in clinical settings.

4) Data on the clone should be published or *in press*.

5) A representative of the clone must be available for submission to the American Type Culture Collection and for typing to confirm it is unique from the pre-existing clones.

The criteria were subsequently modified to allow inclusion of antimicrobial susceptible clones known to be important in disease, and to stipulate that clones must be disseminated over at least two continents. To date, there are 43 PMEN clones (Table 1.8.1). The typing, antibiogram and primary citation information for these clones can be found at http://www.sph.emory.edu/PMEN/index.html.

Table 1.8.1 PMEN clones and attributes of the type strains.

			Antibiotic Susceptibility‡					
Clone	**Serotype**	**ST***	**Penicillin**	**Cefotaxime**	**Erythromycin**	**Tetracycline**	**Chloramphenicol**	**Trimethoprim-Sulfamoxazole**
Spain^{23F}-1	23F	81	I	I	S	R	R	I
Spain^{6B}-2	6B	90	I	I	S	R	R	I
Spain^{9V}-3	9V	156	I	I	S	S	S	I
Tennessee^{23F}-4	23F	37	I	R	R	S	S	I
Spain^{14}-5	14	18	I	I	S	R	R	S
Hungary^{19A}-6	19A	268	I	I	R	R	R	I
S.Africa^{19A}-7	19A	75	I	S	S	S	S	R
S.Africa^{6B}-8	6B	185	I	S	S	S	S	I
England^{14}-9	14	9	S	S	R	S	S	S
CSR^{14}-10	14	20	R	I	R	R	R	S
CSR^{19A}-11	19A	175	R	S	R	R	R	I
Finland^{6B}-12	6B	270	R	I	R	R	R	R
S.Africa^{19A}-13	19A	41	I	S	R	R	R	R
Taiwan^{19F}-14	19F	236	R	I	R	R	R	I

Taiwan^{23F}-15	23F	242	I	I	R	R	S	S
Poland^{23F}-16	23F	173	R	R	R	R	R	I
Maryland^{6B}-17	6B	384	I	I	R	S	S	I
Tennessee^{14}-18	14	67	R	R	R	R	S	I
Columbia^{5}-19	5	289	S	S	S	R	R	S
Poland^{6B}-20	6B	315	I	S	R	R	R	S
Portugal^{19F}-21	19F	177	S	S	R	R	S	S
Greece^{6B}-22	6B	273	S	S	R	R	R	S
N.Carolina^{6A}-23	6A	376	I	I	R	S	S	I
Utah^{35B}-24	35B	377	I	I	S	S	S	S
Sweden^{15A}-25	15A	63	I	I	R	S	S	S
Columbia^{23F}-26	23F	338	I	I	S	S	S	S
Sweden^{1}-27	1	217	S	S	S	S	S	S
Sweden^{1}-28	1	306	S	S	S	S	S	S
USA^{1}-29	1	615	S	S	S	S	S	S
Greece^{21}-30	21	193	S	S	S	S	S	S
Netherlands^{3}-31	3	180	S	S	S	S	S	S
Denmark^{14}-32	14	230	I	I	S	R	S	I
Netherlands^{8}-33	8	53	S	S	S	S	S	S
Denmark^{12F}-34	12F	218	S	S	S	S	S	S
Netherlands^{14}-35	14	124	S	S	S	S	S	S
Netherlands^{18C}-36	18C	113	S	S	S	S	S	S
Netherlands^{15B}-37	15B	199	S	S	S	S	S	S
Sweden^{4}-38	4	205	S	S	S	S	S	S
Netherlands^{7F}-39	7F	191	S	S	S	S	S	S
Sweden^{1}-40	1	304	S	S	S	S	S	S
Portugal^{6A}-41	6A	327	S	S	S	S	S	I
Norway^{NT}-42	NT†	344	I	S	R	R	S	S
USA^{NT}-43	NT†	448	I	S	R	R	S	S

* Sequence type as defined by MLST.

‡ Susceptibilities described as S (susceptible), I (intermediate-resistant) or R (resistant) based on CLSI MIC interpretative standards 2000.

†Nontypeable

1.9 Aims of this Thesis

This thesis explores aspects of pneumococcal genomic evolution relevant to antimicrobial resistance and vaccine escape. A unique historical collection of 426 pneumococci (dated between 1937 and 2007, Figure 1.9.1) were used to study the

evolution of pneumococcal antimicrobial resistance and capsular switching. Chapter 3 compares the genotypes of historical isolates, including 104 early-PNSP (dated prior to 1990), to those of modern pneumococci, with particular reference to the PMEN clones. Nucleotide sequences of regions of three *pbp* genes were also generated for 389 selected isolates. *pbp* gene evolution amongst PSP and the statistical association between *pbp* nucleotide divergence and penicillin MIC are also described in Chapter 3. Chapter 4 describes variation among the *pbp* genes of PNSP and the contribution of the PMEN1 clone to the spread of pneumococcal penicillin resistance. Following whole-genome sequencing of 96 selected isolates, the ancestor of the PMEN1 clone was identified, as was the unusual ability of this clone to share multiple regions of its genome with unrelated pneumococci. Chapter 5 describes evolution at the *cps* locus by comparison of the genome sequences of ancestrally related isolates. Eighteen independent historical and modern capsular switching events are described. Chapter 6 describes mobile genetic elements, containing non-β-lactam antimicrobial resistance determinants, which were identified among the genomes of 38 pneumococci dated prior to 1980 and an additional 61 pneumococci representing internationally-disseminated, temporally-persistent clones.

The inclusion of historical pneumococci pre-dating both the widespread use of antibiotics and the introduction of pneumococcal vaccines (Figure 1.9.1) has allowed a unique opportunity to characterise genomic changes in the absence of these selective pressures. Consequently, the findings detailed here will contribute to a better understanding of the way in which pneumococci have and may continue to respond to vaccine and treatment interventions.

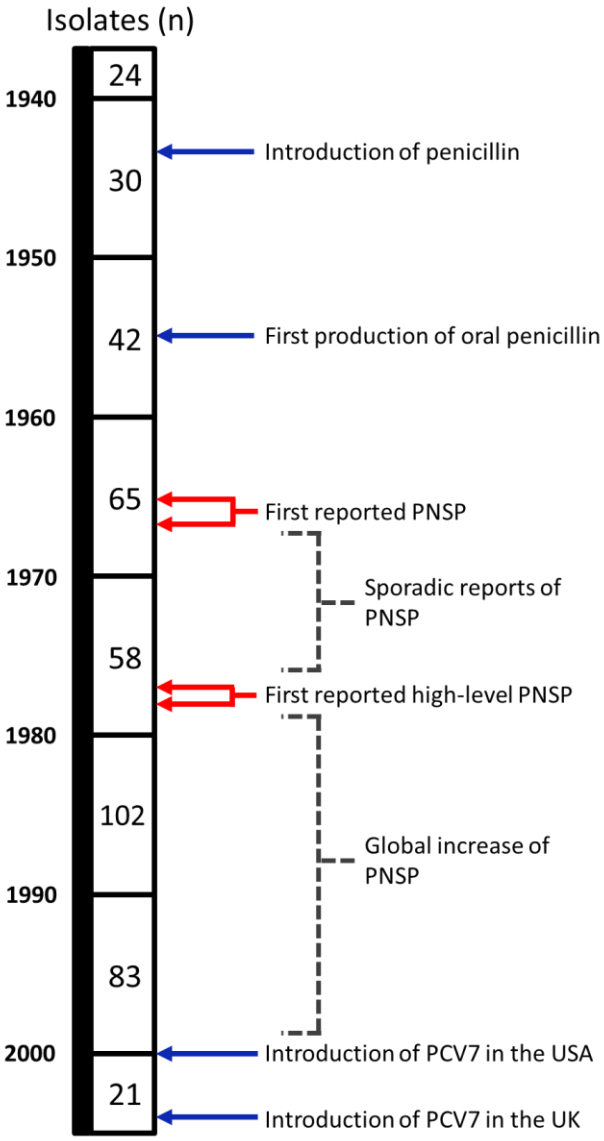


Figure 1.9.1 Timeline of the development and spread of pneumococcal penicillin-nonsusceptibility relative to the introduction of clinical interventions and the dates of isolates in the historical collection.
Isolates are grouped by decade. The date of one isolate is unknown, but presumed later than 1990.

1.10 References

1. **Abramson, J. S., C. J. Baker, M. C. Fisher, M. A. Gerber, H. C. Meissner, D. L. Murray, G. D. Overturf, C. G. Prober, M. B. Rennels, T. N. Saari, L. B. Weiner, R. J. Whitley, G. Peter, L. K. Pickering, N. E. MacDonald, L. Chilton, G. Delage, S. F. Dowell, R. F. Jacobs, M. G. Myers, W. A. Orenstein, P. A. Patriarca, E. O. Ledbetter, J. Kim, and Commitee 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.

2. **Ambrose, K., and D. S. Stephens.** 2004. Macrolide, quinolone, and other non-b-lactam antibiotic resistance in *Streptococcus pneumoniae*., *In* E. I. Tuomanen (ed.), The Pneumococcus. ASM Press, Washington D. C.

2a. **Andreo, F., J. Dominguez, J. Ruiz, S. Blanco, E. Arellano, C. Prat, J. Morera, and V. Ausina.** 2006. Impact of rapid urine antigen tests to determine the etiology of community-acquired pneumonia in adults. Respir Med. **100:**884-891.

3. **Appelbaum, P. C., J. N. Scragg, A. J. Bowen, A. Bhamjee, A. F. Hallett, and R. C. Cooper.** 1977. *Streptococcus pneumoniae* resistant to penicillin and chloramphenicol. Lancet **2:**995-997.

4. **Arrecubieta, C., E. García, and R. López.** 1995. Sequence and transcriptional analysis of a DNA region involved in the production of capsular polysaccharide in *Streptococcus pneumoniae* type 3. Gene **167:**1-7.

5. **Ashley, E. A., Y. Lubell, N. J. White, and P. Turner.** 2011. Antimicrobial susceptibility of bacterial isolates from community acquired infections in Sub-Saharan Africa and Asian low and middle income countries. Trop Med Int Health. **16:**1167-1179.

6. **Attaiech, L., A. Olivier, I. Mortier-Barrière, A. L. Soulet, C. Granadel, B. Martin, P. Polard, and C. J. P.** 2011. Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. PLoS Genet. **7:**e1002156.

7. **Barry, M. A., D. E. Craven, and M. Finland.** 1984. Serotypes of *Streptococcus pneumoniae* isolated from blood cultures at Boston City Hospital between 1979 and 1982. J Infect Dis. **149:**449-452.

8. **Beall, B., M. C. McEllistrem, R. E. Gertz, S. Wedel, D. J. Boxrud, A. L. Gonzalez, M.-J. Medina, R. Pai, T. A. Thompson, L. H. Harrison, L. McGee, C. G. Whitney, and Active Bacterial Core Surveillance Team.** 2006. Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. J Clin Microbiol. **44:**999-1017.

9. **Beall, B. W., R. E. Gertz, R. L. Hulkower, C. G. Whitney, M. R. Moore, and A. B. Brueggemann.** 2011. Shifting genetic structure of invasive serotype 19A pneumococci in the United States. J Infect Dis. **203:**1360-1368.

10. **Bentley, S. D., D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabbinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail, G. Samuel, I. C. Skovsted, M. S. Kaltoft, B. Barrell, P. R. Reeves, J. Parkhill, and B. G. Spratt.** 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. **2:**262-269.

11. **Bergé, M., I. Mortier-Barrière, B. Martin, and J.-P. Claverys.** 2003. Transformation of *Streptococcus pneumoniae* relies on DprA- and RecA-dependent protection of incoming DNA single strands. Mol Microbiol. **50:**527-536.

12. **Black, S., H. Shinefield, B. Fireman, E. Lewis, P. Ray, J. R. Hansen, L. Elvin, K. M. Ensor, J. Hackell, G. Siber, F. Malinoski, D. Madore, I. Chang, R. Kohberger, W. Watson, R. Austrian, and K. Edwards.** 2000. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Pediatr Infect Dis J. **19:**187-195.

13. **Bogaert, D., R. de Groot, and P. W. Hermans.** 2004. *Streptococcus pneumoniae* colonisation: the key to pneumococcal disease. Lancet Infect Dis. **4:**144-154.

14. **Bogaert, D., G. A. Syrogiannopoulos, I. N. Grivea, R. de Groot, N. G. Beratis, and P. W. M. Hermans.** 2000. Molecular epidemiology of penicillin-nonsusceptible *Streptococcus pneumoniae* among children in Greece. J Clin Microbiol. **38:**4361-4366.

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

16. **Brandileone, M. C., V. S. Vieira, R. C. Zanella, I. M. Landgraf, C. Melles, A. D. Taunay, J. C. de Moraes, and R. Austrian.** 1995. Distribution of serotypes of *Streptococcus pneumoniae* isolated from invasive infections over a 16-year period in the greater Sao-Paulo area, Brazil. J Clin Microbiol. **33:**2789-2791.

17. **Bratcher, P. E., I. H. Park, M. B. Oliver, M. Hortal, R. Camilli, S. K. Hollingshead, T. Camou, and M. H. Nahm.** 2011. Evolution of the capsular gene locus of *Streptococcus pneumoniae* serogroup 6. Microbiol. **157:**189-198.

17a. **Brouwer, M. C., A. R. Tunkel, and D. van de Beek.** 2010. Epidemiology, diagnosis, and antimicrobial treatment of acute bacterial meningitis. Clin. Microbiol. Rev. **23:**467-492.

18. **Brueggemann, A. B., D. T. Griffiths, E. Meats, T. E. A. Peto, D. W. Crook, and B. G. Spratt.** 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.
19. **Brueggemann, A. B., R. Pai, D. Crook, and B. Beall.** 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog. **3:**e168.
20. **Brueggemann, A. B., and B. G. Spratt.** 2003. Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. J Clin Microbiol. **41:**4966-4970.
21. **Brüssow, H., C. Canchaya, and W.-D. Hardt.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev. **68:**560-602.
22. **Butler, J. C., R. F. Breiman, H. B. Lipman, J. Hofmann, and R. R. Facklam.** 1995. Serotype distribution of *Streptococcus pneumoniae* infections among preschool children in the United States, 1978-1994: Implications for development of a conjugate vaccine. J Infect Dis. **171:**885-889.
23. **Cafini, F., R. del Campo, L. Alou, D. Sevillano, M. I. Morosini, F. Baquero, J. Prieto, and the Spanish Pneumococcal Network.** 2006. Alterations of the penicillin-binding proteins and *murM* alleles of clinical *Streptococcus pneumoniae* isolates with high-level resistance to amoxicillin in Spain. J Antimicrob Chemother. **57:**224-229.
24. **Calix, J. J., and M. H. Nahm.** 2010. A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. J Infect Dis. **202:**29-38.
24a. **Calix, J. J., R. J. Porambo, A. M. Brady, T. R. Larson, J. Yother, C. Abeygunwardana, and M. H. Nahm.** 2012. Biochemical, genetic and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pnuemococcal serotype. J Biol Chem. ***In press***.
25. **Carapito, R., L. Chesnel, T. Vernet, and A. Zapun.** 2006. Pneumococcal b-Lactam resistance due to a conformational change in penicillin-binding protein 2x. J Biol Chem. **281:**1771-1777.
26. **Carrolo, M., F. R. Pinto, J. Melo-Cristino, and M. Ramirez.** 2009. Pherotypes are driving genetic differentiation within *Streptococcus pneumoniae.* BMC Microbiol. **9:**191-200.
27. **Cartwright, K.** 2002. Pneumococcal disease in Western Europe: burden of disease, antibiotic resistance and management. Eur J Pediatr **161:**188-195.
28. **Casjens, S.** 2003. Prophages and bacterial genomics: what have we learned so far? Mol Microbiol. **49:**277-300.
29. **Castanheira, M., A. C. Gales, R. E. Mendes, R. N. Jones, and H. S. Sader.** 2004. Antimicrobial susceptibility of *Streptococcus pneumoniae* in Latin America: results from five years of the SENTRY antimicrobial surveillance program. Clin Microbiol Infect. **10:**645-651.
30. **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.
31. **Cernuschi, T., E. Furrer, N. Schwalbe, A. Jones, E. R. Berndt, and S. McAdams.** 2011. Advance market commitment for pneumococcal vaccines: putting theory into practice. Bull World Health Organ. **89:**913-918.
32. **Chen, I., P. J. Christie, and D. Dubnau.** 2005. The ins and outs of DNA transfer in bacteria. Science **310:**1456-1460.
33. **Chen, I., and D. Dubnau.** 2004. DNA uptake during bacterial transformation. Nat Rev Microbiol. **2:**241-249.
34. **Chesnel, L., R. Carapito, J. Croizé, O. Dideberg, T. Vernet, and A. Zapun.** 2005. Identical penicillin-binding domains in penicillin-binding proteins of *Streptococcus pneumoniae* clinical isolates with different levels of b-lactam resistance. Antimicrob Agents Chemother. **49:**2895-2902.
35. **Chi, F., O. Nolte, C. Bergmann, M. Ip, and R. Hakenbeck.** 2007. Crossing the barrier: evolution and spread of a major class of mosaic *pbp2x* in *Streptococcus pneumoniae*, *S. mitis* and *S. oralis*. Int J Med Microbiol. **297:**503-512.
36. **Choi, E. H., S. H. Kim, B. W. Eun, S. J. Kim, N. H. Kim, J. Lee, and H. J. Lee.** 2008. *Streptococcus pneumoniae* serotype 19A in children, South Korea. Emerg Infect Dis. **14:**275-281.
37. **Claverys, J.-P., B. Martin, and L. S. Havarstein.** 2007. Competence-induced fratricide in streptococci. Mol Microbiol. **64:**1423-1433.

38. **Claverys, J. P., B. Martin, and P. Polard.** 2009. The genetic transformation machinery: composition, localization, and mechanism. Fems Microbiol. Rev. **33:**643-656.
39. **Coffey, T. J., S. Berron, M. Daniels, M. E. GarciaLeoni, E. Cercenado, E. Bouza, A. Fenoll, and B. G. Spratt.** 1996. Multiply antibiotic-resistant *Streptococcus pneumoniae* recovered from Spanish hospitals (1988-1994): Novel major clones of serotypes 14, 19F and 15F. Microbiology. **142:**2747-2757.
40. **Coffey, T. J., M. Daniels, M. C. Enright, and B. G. Spratt.** 1999. Serotype 14 variants of the Spanish penicillin-resistant serotype 9V clone of *Streptococcus pneumoniae* arose by large recombinational replacements of the *cpsA-pbp1a* region. Microbiology. **145:**2023-2031.
41. **Coffey, T. J., C. G. Dowson, M. Daniels, J. Zhou, C. Martin, B. G. Spratt, and J. M. Musser.** 1991. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthesis genes, in natural populations of *Streptococcus pneumoniae*. Mol Microbiol. **5:**2255-2260.
42. **Coffey, T. J., M. C. Enright, M. Daniels, J. K. Morona, R. Morona, W. Hryniewicz, P. J. C, and B. G. Spratt.** 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.
43. **Cohen, R., C. Levy, E. Bingen, E. Bonnet, M. Koskas, S. Attal, I. Nave, B. Fritzell, and E. Varon.** 2011. Impact of 13-valent pneumococcal conjugate vaccine (PCV13) on nasopharyngeal (NP) flora in children with acute otitis media (AOM). 51st Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago.
43a. **Cornejo, O. E., L. McGee, and D. E. Rozen.** 2010. Polymorphic competence peptides do not restrict recombination in *Streptococcus pneumoniae.* Mol Biol Evol. **27:**694-702.
44. **Corso, A., E. P. Severina, V. F. Petruk, Y. R. Mauriz, and A. Tomasz.** 1998. Molecular characterisation of penicillin-resistant *Streptococcus pneumoniae* isolates causing respiratory disease in the United States. Microb Drug Resist. **4:**325-337.
45. **Crook, D. W., A. B. Brueggemann, K. L. Sleeman, and T. E. A. Peto.** 2004. Pneumococcal carriage., *In* E. I. Tuomanen (ed.), The Pneumococcus. ASM Press, Washington.
46. **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, B. Pichon, S. Baker, C. M. Parry, L. M. Lambertsen, D. Shahinas, D. R. Pillai, T. J. Mitchell, G. Dougan, A. Tomasz, K. P. Klugman, J. Parkhill, W. P. Hanage, and S. D. Bentley.** 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331:**430-434.
47. **Dagkessamanskaia, A., M. Moscoso, V. Hénard, S. Guiral, K. Overweg, M. Reuter, B. Martin, J. Wells, and J. Claverys.** 2004. Interconnection of competence, stress and CiaR regulons in *Streptococcus pneumoniae*: competence triggers stationary phase autolysis of *ciaR* mutant cells. Mol Microbiol. **51:**1071-1086.
48. **Dahlblom, V., and M. Söderström.** 2012. Bacterial interactions in the nasopharynx - Effects of host factors in children attending day-care centers. J Infect Public Health. **5:**133-139.
49. **Dawid, S., A. M. Roche, and J. N. Weiser.** 2007. The *blp* bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both *in vitro* and *in vivo.* Infect Immun. **75:**443-451.
50. **de Roux, A., B. Schmöele-Thoma, G. R. Siber, J. G. Hackell, A. Kuhnke, N. Ahlers, S. A. Baker, A. Razmpour, E. A. Emini, P. D. Fernsten, W. C. Gruber, S. Lockhart, O. Burkhardt, T. Welte, and H. M. Lode.** 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:**1015-1023.
51. **Del Grosso, M., R. Camilli, F. Iannelli, G. Pozzi, and A. Pantosti.** 2006. The *mef*(E)-carrying genetic element (mega) of *Streptococcus pneumoniae*: Insertion sites and association with other genetic elements. Antimicrob Agents Chemother. **50:**3361-3366.
52. **Denapaite, D., R. Brückner, M. Nuhn, P. Reichmann, B. Henrich, P. Maurer, Y. Schähle, P. Selbmann, W. Zimmermann, R. Wambutt, and R. Hakenbeck.** 2010. The genome of *Streptococcus mitis* B6 - what Is a commensal? PLoS One. **5:**e9426.
53. **Dias, R., D. Felix, and M. Canica.** 2008. Diversity of penicillin binding proteins among clinical *Streptococcus pneumoniae* strains from Portugal. Antimicrob Agents Chemother. **52:**2693-2695.
54. **Dillard, J. P., M. W. Vandersea, and J. Yother.** 1995. Characterization of the cassette containing genes for type 3 capsular polysaccharide biosynthesis in *Streptococcus pneumoniae.* J Exp Med. **181:**973-983.

55. **Dixon, J. M. S., A. E. Lipinski, and M. E. Graham.** 1977. Detection and prevalence of pneumococci with increased resistance to penicillin. CMAJ. **117:**1159-1161.
56. **Do, T., K. A. Jolley, M. C. J. Maiden, S. C. Gilbert, D. Clark, W. G. Wade, and D. Beighton.** 2009. Population structure of *Streptococcus oralis.* Microbiology. **155:**2593-2602.
57. **Doit, C., C. Loukil, F. Fitoussi, P. Geslin, and E. Bingen.** 1999. Emergence in France of multiple clones of clinical *Streptococcus pneumoniae* isolates with high-level resistance to amoxicillin. Antimicrob Agents Chemother. **43:**1480-1483.
58. **Dowson, C. G., T. J. Coffey, C. M. Kell, and R. A. Whiley.** 1993. Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. Mol Microbiol. **9:**635-643.
59. **Dowson, C. G., A. Hutchison, J. A. Brannigan, R. C. George, H. David, J. Liñares, A. Tomasz, J. M. Smith, and B. G. Spratt.** 1989. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. Proc Natl Acad Sci U S A. **86:**8842-8846.
60. **Elberse, K., S. Witteveen, H. van der Heide, I. van de Pol, C. Schot, A. van der Ende, G. Berbers, and L. Schouls.** 2011. Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. PLoS One. **6:**e25018.
61. **Enright, M. C., and B. G. Spratt.** 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology. **144:**3049-3060.
61a. **Enright, M. C., and B. G. Spratt.** 1999. Extensive variation in the *ddl* gene of penicillin-resistant *Streptococcus pneumoniae* results from a hitch-hiking effect driven by the penicillin-binding protein 2b gene. Mol Biol Evol. **16:**1687-1695.
62. **Eskola, J., and M. Antilla.** 1999. Pneumococcal conjugate vaccines. Pediatr Infect Dis J. **18:**543-551.
63. **Eskola, J., T. Kilpi, A. Palmu, J. Jokinen, J. Haapakoski, E. Herva, A. K. Takala, H. Käyhty, P. Karma, R. Kohberger, G. Siber, and P. H. Mäkelä.** 2001. Efficacy of a pneumococcal conjugate vaccine against acute otitis media. N Engl J Med. **344:**403-409.
64. **Esposito, S., S. Tansey, A. Thompson, A. Razmpour, J. Liang, T. R. Jones, G. Ferrera, A. Maida, G. Bona, C. Sabatini, L. Pugni, E. A. Emini, W. C. Gruber, D. A. Scott, and N. Principi.** 2010. 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:**1017-1026.
65. **European Centre for Disease Prevention and Control (ECDC)** 2012. European Antimicrobial Resistance Surveillance Network (EARS-Net) database. European Centre for Disease Prevention and Control (ECDC). [Online.]
66. **Everett, D. B., M. Mukaka, B. Denis, S. B. Gordon, E. D. Carrol, J. J. van Oosterhout, E. M. Molyneux, M. Molyneux, N. French, and R. S. Heyderman.** 2011. Ten years of surveillance for invasive *Streptococcus pneumoniae* during the era of antiretroviral scale-up and cotrimoxazole prophylaxis in Malawi. PLoS One. **6:**e17765.
67. **Facklam, R.** 2002. What happened to the streptococci: overview of taxonomic and nomenclature changes. Clin. Microbiol. Rev. **15:**613-630.
68. **Farrell, D. J., D. Felmingham, J. Shackcloth, L. Williams, K. Maher, R. Hope, D. M. Livermore, R. C. George, G. Brick, S. Martin, R. Reynolds, and the BSAC Working Parties on Resistance Surveillance.** 2008. Non-susceptibility trends and serotype distributions among *Streptococcus pneumoniae* from community-acquired respiratory tract infections and from bacteraemias in the UK and Ireland, 1999 to 2007. J Antimicrob Chemother. **62:**ii87-ii95.
69. **Fedson, D. S., J. Anthony, and G. Scott.** 1999. The burden of pneumococcal disease among adults in developed and developing countries: what is and is not known. Vaccine **17:**S11-S18.
70. **Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt.** 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics. **154:**1439-1450.
71. **Felmingham, D., A. R. White, M. R. Jacobs, P. C. Appelbaum, J. Poupard, L. A. Miller, and R. N. Gruneberg.** 2005. The Alexander Project: the benefits from a decade of surveillance. J Antimicrob Chemother. **56:**ii3-ii21.

72. **Fenoll, A., J. J. Granizo, L. Aguilar, M. J. Giménez, L. Aragoneses-Fenoll, G. Hanquet, J. Casal, and D. Tarragó.** 2007. Temporal trends of invasive *Streptococcus pneumoniae* serotypes and antimicrobial resistance patterns in Spain from 1979 to 2007. J Clin Microbiol. **47:**1012-1020.
73. **Ferroni, A., and P. Berche.** 2001. Alterations to penicillin-binding proteins 1A, 2B and 2X amongst penicillin-resistant clinical isolates of *Streptococcus pneumoniae* serotype 23F from the nasopharyngeal flora of children. J Med Microbiol. **50:**828-832.
74. **Filipe, S. R., E. Severina, and A. Tomasz.** 2001. The role of the *murMN* operon in penicillin resistance and antibiotic tolerance of *Streptococcus pneumoniae*. Microb Drug Resist. **7:**303-316.
75. **Finland, M., and M. W. Barnes.** 1977. Changes in occurence of capsular serotypes of *Streptococcus pneumoniae* at Boston City Hospital during selected years betwen 1935 and 1974. J Clin Microbiol. **5:**154-166.
76. **Finn, A.** 2004. Bacterial polysaccharide-protein conjugate vaccines. Br Med Bull. **70:**1-14.
77. **Foster, D., K. Knox, A. S. Walker, D. T. Griffiths, H. Moore, E. Haworth, T. Peto, A. B. Brueggemann, and D. W. Crook on behalf of the Oxford Invasive Pneumococcal Surveillance Group.** 2008. Invasive pneumococcal disease: epidemiology in children and adults prior to implementation of the conjugate vaccine in the Oxfordshire region, England. J Med Microbiol. **57:**480-487.
78. **Francisco, A. P., M. Bugalho, M. Ramirez, and J. A. Carriço.** 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics. **10:**152.
79. **Frost, L. S., R. Leplae, A. O. Summers, and A. Toussaint.** 2005. Mobile genetic elements: The agents of open source evolution. Nat Microbiol Rev. **3:**722-732.
80. **García, E., P. García, and R. López.** 1993. Cloning and sequencing of a gene involved in synthesis of the capsular polysaccharide of *Streptococcus pneumoniae* type 3. Mol Gen Genet. **239:**188-195.
81. **García, E., and R. López.** 1997. Molecular biology of the capsular genes of *Streptococcus pneumoniae*. FEMS Microbiol Lett. **149:**1-10.
81a. **García-Vázquez, E., M. A. Marcos, J. Mensa, A. de Roux, J. Puig, C. Font, G. Francisco, and A. Torres.** 2004. Assessment of the usefulness of sputum culture for diagnosis of community-acquired pneumonia using the PORT predictive scoring system. Arch Intern Med. **164:**1807-1811.
81b. **Genné, D., H. H. Siegrist, and R. Lienhard.** 2006. Enhancing the etiologic diagnosis of community-acquired pneumonia in adults using the urinary antigen assay (Binax NOW). Int J Infect Dis. **10:**124-128.
82. **Gertz, R. E., Z. Y. Li, F. C. Pimenta, D. Jackson, B. A. Juni, R. Lynfield, J. H. Jorgensen, M. D. Carvalho, and B. W. Beall for the Active Bacterial Core Surveillance Team.** 2010. Increased penicillin nonsusceptibility of nonvaccine-serotype invasive pneumococci other than serotypes 19A and 6A in post-7-valent Conjugate vaccine era. J Infect Dis. **201:**770-775.
83. **Gill, M. J., N. P. Brenwald, and R. Wise.** 1999. Identification of an efflux pump gene, *pmrA*, associated with fluoroquinolone resistance in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **43:**187-189.
84. **Golubchik, T., A. B. Brueggemann, T. Street, R. E. Gertz Jr, C. C. A. Spencer, T. Ho, E. Giannoulatou, R. Link-Gelles, R. M. Harding, B. Beall, T. E. A. Peto, M. R. Moore, P. Donnelly, D. W. Crook, and R. Bowden.** 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. Nat Genet. **44:**352-355.
85. **Granat, S. M., Z. Mia, J. Ollgren, E. Herva, M. Das, L. Piirainen, K. Auranen, and P. H. Mäkelä.** 2007. Longitudinal study on pneumococcal carriage during the first year of life in Bangladesh. Pediatr Infect Dis J. **26:**319-324.
86. **Granger, D., G. Boily-Larouche, P. Turgeon, K. Weiss, and M. Roger.** 2006. Molecular characteristics of *pbp1a* and *pbp2b* in clinical *Streptococcus pneumoniae* isolates in Quebec, Canada. J Antimicrob Chemother. **57:**61-70.
87. **Grebe, T., and R. Hakenbeck.** 1996. Penicillin-binding proteins 2b and 2x of *Streptococcus pneumoniae* are primary resistance determinants for different classes of β-lactam antibiotics. Antimicrob Agents Chemother. **40:**829-834.

88. **Greenberg, D., D. P. Speert, E. Mahenthiralingam, D. A. Henry, M. E. Campbell, D. W. Scheifele, and the CPS / LCDC Impact Monitoring Network.** 2002. Emergence of penicillin-nonsusceptible *Streptococcus pneumoniae* invasive clones in Canada. J Clin Microbiol. **40:**68-74.
89. **Guiral, S., T. J. Mitchell, B. Martin, and J. Claverys.** 2005. Competence-programmed predation of noncompetent cells in the human pathogen *Streptococcus pneumoniae*: Genetic requirements. Proc Natl Acad Sci U S A. **102:**8710-8715.
89a. **Håkansson, A., A. Kidd, G. Wadell, H. Sabharwal, and C. Svanborg.** 1994. Adenovirus infection enhances in vitro adherence of *Streptococcus pneumoniae.* Infect Immun **62:**2707-2714.
90. **Hakenbeck, R., N. Balmelle, B. Weber, C. Gardes, W. Keck, and A. de Saizieu.** 2001. Mosaic genes and mosaic chromosomes: Intra- and interspecies genomic variation of *Streptococcus pneumoniae*. Infect Immun. **69:**2477-2486.
91. **Hakenbeck, R., T. Grebe, D. Zahner, and J. B. Stock.** 1999. β-lactam resistance in *Streptococcus pneumoniae*: penicillin-binding proteins and non-penicillin-binding proteins. Mol Microbiol. **33:**673-678.
92. **Hanage, W. P., C. Fraser, J. Tang, T. R. Connor, and J. Corander.** 2009. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. Science **324:**1454-1457.
93. **Hanage, W. P., T. H. Kaijalainen, R. K. Syrjänen, K. Auranen, M. Leinonen, P. H. Mäkelä, and B. G. Spratt.** 2005. Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. Infect Immun. **73:**431-435.
94. **Hansman, D., and M. M. Bullen.** 1967. A resistant pneumococcus. Lancet **2:**264-265.
95. **Hansman, D., L. Devitt, H. Miles, and I. Riley.** 1974. Pneumococci relatively insensitive to penicillin in Australia and New Guinea. Med J Aust. **2:**353-356.
96. **Hansman, D., L. Devitt, and I. Riley.** 1973. Pneumococci with increased resistance to penicillin. BMJ. **3:**405-405.
97. **Harboe, Z. B., T. L. Benfield, P. Valentiner-Branth, T. Hjuler, L. Lambertsen, M. Kaltoft, K. Krogfelt, H. C. Slotved, J. J. Christensen, and H. B. Konradsen.** 2010. Temporal trends in invasive pneumococcal disease and pneumococcal serotypes over 7 decades. Clin Infect Dis. **50:**329-337.
98. **Hathaway, L., B. Morand, S. Aebi, and K. Mühlemann.** 2010. Serotype 7F capsule genes delay growth in nutrient-limited conditions., 7th International Symposium of Pnuemococci and Pneumococcal Diseases Tel Aviv, Israel.
99. **Hausdorff, W. P., A. B. Brueggemann, J. G. Hackell, and J. A. G. Scott.** 2008. Pneumococcal serotype epidemiology., *In* G. Siber, K. P. Klugman, and P. H. Mäkelä (ed.), Pneumococcal vaccines: the impact of conjugate vaccine. ASM Press, Washington D.C.
100. **Hausdorff, W. P., J. Bryant, P. R. Paradiso, and G. R. Siber.** 2000. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. Clin Infect Dis. **30:**100-121.
101. **Henderson-Begg, S. K., A. P. Roberts, and L. M. C. Hall.** 2009. Diversity of putative Tn*5253*-like elements in *Streptococcus pneumoniae.* Int J Antimicrob Agents. **33:**364-367.
102. **Henrichsen, J.** 1995. Six newly recognised types of *Streptococcus pneumoniae*. J Clin Microbiol. **33:**2759-2762.
103. **Hicks, L. A., Y. Chien, T. H. Taylor Jr, M. Haber, K. P. Klugman, and the Active Bacterial Core Surveillance (ABCs) Team.** 2011. Outpatient antibiotic prescribing and nonsusceptible *Streptococcus pneumoniae* in the United States, 1996-2003. Clin Infect Dis. **53:**631-639.
104. **Hicks, L. A., L. H. Harrison, B. Flannery, J. L. Hadler, W. Schaffner, A. S. Craig, D. Jackson, A. Thomas, B. Beall, R. Lynfield, A. Reingold, M. M. Farley, and C. G. Whitney.** 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.
105. **Hill, P. C., A. Akisanya, K. Sankareh, Y. B. Cheung, M. Saaka, G. Lahai, B. M. Greenwood, and R. A. Adegbola.** 2006. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian villagers. Clin Infect Dis. **43:**673-679.
106. **Hill, P. C., J. Townend, M. Antonio, B. Akisanya, C. Ebruke, G. Lahai, B. M. Greenwood, and R. A. Adegbola.** 2010. Transmission of *Streptococcus pneumoniae* in rural Gambian villages: a longitudinal study. Clin Infect Dis. **50:**1468-1476.
107. **Hiller, N. L., A. Ahmed, E. Powell, D. P. Martin, R. Eutsey, J. Earl, B. Janto, R. J. Boissy, J. Hogg, K. Barbadora, R. Sampath, S. Lonergan, J. C. Post, F. Z. Hu, and G. D. Ehrlich.** 2010. Generation

of genetic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal paediatric infection. PLoS Pathog. **6:**e1001108.

108. **Hoshino, T., T. Fujiwara, and M. Kilian.** 2005. Use of phylogenetic and phenotypic analyses to identify nonhemolytic streptococci isolated from bacteremic patients. J Clin Microbiol. **43:**6073-6085.
109. **Howes, V. J., and R. G. Mitchell.** 1976. Meningitis due to relatively penicillin-resistant pneumococcus. BMJ. **1:**996-996.
110. **Huang, S. S., V. L. Hinrichsen, A. E. Stevenson, S. L. Rifas-Shiman, K. Kleinman, S. I. Pelton, M. Lipsitch, W. P. Hanage, G. M. Lee, and J. A. Finkelstein.** 2009. Continued impact of pneumococcal conjugate vaccine on carriage in young children. Pediatrics. **124:**E1-E11.
111. **Huang, S. S., K. M. Johnson, G. T. Ray, P. Wroe, T. A. Lieu, M. R. Moore, E. R. Zell, J. A. Linder, C. G. Grijalva, J. P. Metlay, and J. A. Finkelstein.** 2011. Healthcare utilization and cost of pneumococcal disease in the United States. Vaccine **29:**3398-3412.
112. **Hussain, M., A. Melegaro, R. G. Pebody, R. George, W. J. Edmunds, R. Talukdar, S. A. Martin, A. Efstratiou, and E. Miller.** 2005. A longitudinal household study of *Streptococcus pneumoniae* nasopharyngeal carriage in a UK setting. Epidemiol Infect. **133:**891-898.
113. **Iannelli, F., B. J. Pearce, and G. Pozzi.** 1999. The type 2 capsule locus of *Streptococcus pneumoniae.* J Bacteriol. **181:**2652-2654.
114. **Jackson, L. A., K. M. Neuzil, O. C. Yu, P. Benson, W. E. Barlow, A. L. Adams, C. A. Hanson, L. D. Mahoney, D. K. Shay, and W. W. Thompson for the Vaccine Safety Datalink.** 2003. Effectiveness of pneumococcal polysaccharide vaccine in older adults. N Engl J Med. **348:**1747-1755.
115. **Jacobs, M. R., D. Felmingham, P. C. Appelbaum, R. N. Gruneberg, and The Alexander Project Group.** 2003. The Alexander Project 1998-2000: susceptibility of pathogens isolated from community-acquired respiratory tract infection to commonly used antimicrobial agents. J Antimicrob Chemother. **52:**229-246.
116. **Jacobs, M. R., C. E. Good, S. Bajaksouzian, and 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:**1388-1395.
117. **Jacobs, M. R., H. J. Koornhof, R. M. Robins-Browne, C. M. Stevenson, Z. A. Vermaak, I. Freiman, B. Miller, M. A. Witcomb, M. Isaäcson, J. I. Ward, and R. Austrian.** 1978. Emergence of multiply resistant pneumococci. N Engl J Med. **299:**735-740.
118. **Jefferson, T., E. Ferroni, F. Curtale, P. G. Rossi, and P. Borgia.** 2006. *Streptococcus pneumoniae* in western Europe: serotype distribution and incidence in children less than 2 years old. Lancet Infect Dis. **6:**405-410.
119. **Jiang, S.-M., L. Wang, and P. R. Reeves.** 2001. Molecular characterisation of *Streptococcus pneumoniae* type 4, 6B, 8 and 18C capsular polysaccharide gene clusters. Infect Immun. **69:**1244-1255.
120. **Jin, P., F. Kong, M. Xiao, S. Oftadeh, F. Zhou, C. Liu, F. Russell, and G. L. Glibert.** 2009. First report of putative *Streptococcus pneumoniae* serotype 6D among nasopharyngeal isolates from Fijian children. J Infect Dis. **200:**1375-1380.
121. **Johnsborg, O., T. Blomqvist, M. Kilian, and H. L. Sigve.** 2007. Biologically active peptides in Streptococci., *In* R. Hakenbeck and S. Chhatwal (ed.), Molecular Biology of Streptococci. Horizon Scientific Press, Norwich, UK.
122. **Johnsborg, O., V. Eldholm, M. L. Bjørnstad, and L. S. Håvarstein.** 2008. A predatory mechanism dramatically increases the efficiency of lateral gene transfer in *Streptococcus pneumoniae* and related commensal species. Mol Microbiol. **69:**245-253.
123. **Johnsborg, O., and L. S. Håvarstein.** 2009. Regulation of natural genetic transformation and acquisition of transforming DNA in *Streptococcus pneumoniae.* Fems Microbiol. Rev. **33:**627-642.
124. **Johnson, H. L., M. Deloria-Knoll, O. S. Levine, S. K. Stoszek, L. F. Hance, R. Reithinger, L. R. Muenz, and 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:**e10000348.
125. **Jones, M. E., R. S. Bloser-Middleton, I. A. Critchley, J. A. Karlowsky, C. Thornsberry, and D. F. Sahm.** 2003. *In vitro* susceptibility of *Streptococcus pneumoniae*, *Haemophilus influenzae* and

Moraxella catarrhalis: a European multicentre study during 2000-2001. Clin Microbiol Infect. **9:**590-599.

126. **Jones, R. N., D. J. Biedenbach, and M. L. Beach.** 2003. Influence of patient age on the susceptibility patterns of *Streptococcus pneumoniae* isolates in North America (2000-2001): report from the SENTRY antimicrobial surveillance program. Diagn Microbiol Infect Dis. **46:**77-80.

127. **Jones, R. N., and H. S. Sader.** 2010. Declining antimicrobial susceptibility of *Streptococcus pneumoniae* in the United States: report from the SENTRY antimicrobial surveillance program (1998-2009). Diagn Microbiol Infect Dis. **68:**334-336.

128. **Kamboj, K. K., H. L. Kirchner, R. Kimmel, N. S. Greenspan, and J. R. Schreiber.** 2003. Significant variation in serotype-specific immunogenicity of the seven-valent *Streptococcus pneumoniae* capsular polysaccharide-CRM197 conjugate vaccine occurs despite vigorous T cell help induced by the carrier protein. J Infect Dis. **187:**1629-1638.

128a. **Kanegaye, J. T., P. Soliemanzadeh, and J. S. Bradley.** 2001. Lumbar puncture in pediatric bacterial meningitis: defining the time interval for recovery of cerebrospinal fluid pathogens after parenteral antibiotic pretreatment. Pediatrics **108:**1169-1174.

129. **Kaplan, S. L., W. J. Barson, P. L. Lin, S. H. Stovall, J. S. Bradley, T. Q. Tan, J. A. Hoffman, L. B. Givner, and E. O. Mason.** 2010. Serotype 19A is the most common serotype causing invasive pneumococcal infections in children. Pediatrics. **125:**429-436.

130. **Kawamura, Y., X.-G. Hou, F. Sultana, H. Miura, and T. Ezaki.** 1995. Determination of 16s rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptcoccus*. Int J Syst Bacteriol. **45:**406-408.

131. **Kellner, J. D., D. Scheifele, O. G. Vanderkooi, J. MacDonald, D. L. Church, and G. J. Tyrrell.** 2008. Effects of routine infant vaccination with the 7-valent pneumococcal conjugate vaccine on nasopharyngeal colonization with *Streptococcus pneumoniae* in children in Calgary, Canada. Pediatr Infect Dis J. **27:**526-532.

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

133. **Kilian, M., K. Poulsen, T. Blomqvist, L. S. Håvarstein, M. Bek-Thomsen, H. Tettelin, and U. B. S. Sorensen.** 2008. Evolution of *Streptococcus pneumoniae* and its close commensal relatives. PLoS One. **3:**e2683.

134. **Kim, S. H., J. Song, D. R. Chung, V. Thamlikitkul, Y. Yang, H. Wang, M. Lu, T. M. So, P. R. Hsueh, R. M. Yasin, C. C. Carlos, H. V. Pham, M. K. Lalitha, N. Shimono, J. Perera, A. M. Shibl, J. Y. Baek, C. I. Kang, K. S. Ko, and K. R. Peck.** 2012. Changing trends in antimicrobial resistance and serotypes of *Streptococcus pneumoniae* isolates in Asian countries: an Asian Network for Surveillance of Resistant Pathogens (ANSORP) study. Antimicrob Agents Chemother. **56:**1418-1426.

135. **Kislak, J. W., L. M. B. Razavi, A. K. Daly, and M. Finland.** 1965. Susceptibility of pneumococci to nine antibiotics. Am J Med Sci. **250:**261-268.

136. **Klugman, K. P., T. J. Coffey, A. Smith, A. Wasas, M. Meyers, and B. G. Spratt.** 1994. Cluster of an erythromycin-resistant variant of the Spanish multiply resistant 23F clone of *Streptococcus pneumoniae* in South Africa. Eur J Clin Microbiol Infect Dis. **13:**171-174.

137. **Kolkman, M. A. B., W. Wakarchuk, P. J. M. Nuijten, and B. A. M. van der Zeijst.** 1997. Capsular polysaccharide synthesis in *Streptococcus pneumoniae* serotype 14: molecular analysis of the complete *cps* locus and identification of genes encoding glycosyltransferases required for the biosynthesis of the tetrasaccharide subunit. Mol Microbiol. **26:**197-208.

138. **Krononberg, A., P. Zucs, S. Droz, and K. Mühlemann.** 2006. Distribution and invasiveness of *Streptococcus pneumoniae* serotypes in Switzerland, a country with low antibiotic selection pressure. J Clin Microbiol. **44:**2032-2038.

139. **Kyaw, M. H., S. Clarke, G. F. S. Edwards, I. G. Jones, and H. Campbell.** 2000. Serotypes/groups distribution and antimicrobial resistance of invasive pneumococcal isolates: implications for vaccine strategies. Epidemiol Infect. **125:**561-572.

140. **Lagos, R., A. Muñoz, O. San Martin, A. Maldonado, J. C. Hormazabal, W. C. Blackwelder, and M. M. Levine.** 2008. Age- and serotype-specific pediatric invasive pneumococcal disease: insights from systematic surveillance in Santiago, Chile, 1994-2007. J Infect Dis. **198:**1809-1817.

141. **Laible, G., B. G. Spratt, and R. Hakenbeck.** 1991. Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. Mol Microbiol. **5:**1993-2002.

142. **Lambertsen, L., M. Brendstrup, H. Friis, and J. J. Christensen.** 2010. Molecular characterisation of invasive penicillin non-susceptible *Streptococcus pneumoniae* from Denmark, 2001 to 2005. Scand J Infect Dis. **42:**333-340.

143. **Lauer, B. A., and L. B. Reller.** 1980. Serotypes and penicillin susceptibility of pneumococci isolated from blood. J Clin Microbiol. **11:**242-244.

144. **Lee, N. Y., J. Song, S. Kim, K. R. Peck, K. Ahn, S. Lee, Y. Yang, J. Li, A. Chongthaleong, S. Tiengrim, N. Aswapokee, T. Lin, J. Wu, C. Chiu, M. K. Lalitha, K. Thomas, T. Cherian, J. Perera, T. T. Yee, F. Jamal, U. C. Warsa, P. H. Van, C. C. Carlos, A. M. Shibl, M. R. Jacobs, and P. C. Appelbaum.** 2001. Carriage of antibiotic-resistant pneumococci among Asian children: a multinational surveillance by the Asian Network for Surveillance of Resistant Pathogens (ANSORP). Clin Infect Dis. **32:**1463-1469.

145. **Lefèvre, J. C., M. A. Bertrand, and G. Faucon.** 1995. Molecular analysis by pulsed field gel electrophoresis of penicillin-resistant *Streptococcus pneumoniae* from Toulouse, France. Eur J Clin Microbiol Infect Dis. **14:**491-497.

146. **Levine, H., S. Zarka, R. Dagan, T. Sela, V. Rozhavski, D. I. Cohen, and R. D. Balicer.** 2012. Transmission of *Streptococcus pneumoniae* in adults may occur through saliva. Epidemiol Infect. **140:**561-565.

147. **Lexau, C. A., R. Lynfield, R. Danila, T. Pilishvili, R. Facklam, M. M. Farley, L. H. Harrison, W. Schaffner, A. Reingold, N. M. Bennett, J. Hadler, P. R. Cieslak, C. G. Whitney, and the Active Bacterial Core Surveillance Team.** 2005. Changing epidemiology of invasive pneumococcal disease among older adults in the era of pediatric pneumococcal conjugate vaccine. JAMA. **294:**2043-2051.

148. **Li, Y., H. Tomita, Y. Lv, J. Liu, F. Xue, B. Zheng, and Y. Ike.** 2011. Molecular characterization of *erm*(B)- and *mef*(E)-mediated erythromycin-resistant *Streptococcus pneumoniae* in China and complete DNA sequence of Tn*2010*. J Appl Microbiol. **110:**254-265.

149. **Liñares, J., C. Ardanuy, R. Pallarés, and A. Fenoll.** 2010. Changes in antimicrobial resistance, serotypes and genotypes in *Streptococcus pneumoniae* over a 30-year period. Clin Microbiol Infect. **16:**402-410.

150. **Lipsitch, M., J. K. Dykes, S. E. Johnson, E. W. Ades, J. King, D. E. Briles, and G. M. Carlone.** 2000. Competition among *Streptococcus pneumoniae* for intranasal colonization in a mouse model. Vaccine **18:**2895-2901.

151. **Lipsitch, M., K. O'Neill, D. Cordy, B. Bugalter, K. Trzcinski, C. M. Thompson, R. Goldstein, S. Pelton, H. Huot, V. Bouchet, R. Reid, M. Santosham, and K. L. O'Brien.** 2007. Strain characteristics of *Streptococcus pneumoniae* carriage and invasive disease isolates during a cluster-randomized clinical trial of the 7-valent pneumococcal conjugate vaccine. J Infect Dis. **196:**1221-1227.

152. **Llull, D., R. Muñoz, R. López, and E. García.** 1999. A single gene (*tts*) located outside the cap locus directs the formation of *Streptococcus pneumoniae* type 37 capsular polysaccharide: Type 37 pneumococci are natural, genetically binary strains. J Exp Med. **190:**241-251.

153. **López, R., and E. García.** 2004. Recent trends on the molecular biology of pneumococcal capsules, lytic enzymes, and bacteriophage. Fems Microbiol. Rev. **28:**553-580.

154. **Lysenko, E. S., A. J. Ratner, A. L. Nelson, and J. N. Weiser.** 2005. The role of innate immune responses in the outcome of interspecies competition for colonization of mucosal surfaces. PLoS Pathog. **1:**e1.

155. **Mace, J. W., D. S. Janik, R. L. Sauer, and J. J. Quilligan.** 1977. Penicillin-resistant pneumococcal meningitis in an immunocompromised infant. J Pediatr. **91:**506-507.

155a. **Madhi, S. A., and K. P. Klugman.** 2007. World Health Organisation definition of "radiologically-confirmed pneumonia" may under-estimate the true public health value of conjugate pneumococcal vaccines. Vaccine **25:**2413-2419.

155b. **Mandell, L. A., R. G. Wunderink, A. Anzueto, J. G. Bartlett, G. D. Campbell, N. C. Dean, S. F. Dowell, T. M. File Jr., D. M. Musher, M. S. Niederman, A. Torres, and C. G. Whitney.** 2007. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. **44 Suppl 2:**S27-S72.

156. **Marchisio, P., S. Esposito, G. C. Schito, A. Marchese, R. Cavagna, N. Principi, and the Hercules Project Collaborative Group.** 2002. Nasopharyngeal carriage of *Streptococcus pneumoniae* in healthy children: Implications for the use of heptavalent pneumococcal conjugate vaccine. Emerging Infectious Diseases **8:**479-484.

157. **Mascher, T., M. Heintz, D. Zahner, M. Merai, and R. Hakenbeck.** 2006. The CiaRH system of *Streptococcus pneumoniae* prevents lysis during stress induced by treatment with cell wall inhibitors and by mutations in *pbp2x* involved in β-lactam resistance. J Bacteriol. **188:**1959-1968.

158. **Mason Jr, E. O., E. R. Wald, J. S. Bradley, W. J. Barson, S. L. Kaplan, and the United States Pediatric Multicentre Pneumococcal Surveillance Study Group.** 2003. Macrolide resistance among middle ear isolates of *Streptococcus pneumoniae* observed at eight United States pediatric centres: prevalence of M and MLS_B phenotypes. Pediatr Infect Dis J. **22:**623-627.

159. **Maurer, P., B. Koch, I. Zerfaß, J. Krauß, M. van der Linden, J. Frère, C. Contreras-Martel, and R. Hakenbeck.** 2008. Penicillin-binding protein 2x of *Streptococcus pneumoniae*: Three new mutational pathways for remodelling an essential enzyme into a resistance determinant. J Mol Biol. **376:**1403-1416.

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

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

162. **Maynard-Smith, J. M., N. H. Smith, M. Orourke, and B. G. Spratt.** 1993. How clonal are bacteria. Proc Natl Acad Sci U S A. **90:**4384-4388.

163. **McCraig, L. F., and J. M. Hughes.** 1995. Trends in antimicrobial drug prescribing among office-based physicians in the United States. JAMA. **273:**214-219.

163a. **McCullers, J. A., and J. E. Rehg.** 2002. Lethal synergism between influenza virus and *Streptococcus pneumoniae*: characterization of a mouse model and the role of platelet-activating factor receptor. J Infect Dis. **186:**341-350.

164. **McDougal, L. K., J. K. Rasheed, J. W. Biddle, and F. C. Tenover.** 1995. Identification of multiple clones of extended-spectrum cephalosporin-resistant *Streptococcus pneumoniae* isolates in the United States. Antimicrob Agents Chemother. **39:**2282-2288.

165. **McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefévre, A. Tomasz, and K. P. Klugman.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. J Clin Microbiol. **39:**2565-2571.

166. **Melin, M., K. Trzcinski, S. Meri, H. Kayhty, and M. Vakevainen.** 2010. The capsular serotype of *Streptococcus pneumoniae* is more important than the genetic background for resistance to complement. Infect Immun. **78:**5262-5270.

167. **Mera, R. M., L. A. Miller, H. Amrine-Madsen, and D. F. Sahm.** 2009. The impact of the pneumococcal conjugate vaccine on antimicrobial resistance in the United States since 1996: Evidence for a significant rebound by 2007 in many classes of antibiotics. Microb Drug Resist. **15:**261-268.

167a. **Metersky, M. L., A. Ma, D. W. Bratzler, and P. M. Houck.** 2004. Predicting bacteremia in patients with community-acquired pneumonia. Am J Respir Crit Care Med. **169:**342-347.

168. **Millar, E. V., K. L. O'Brien, J. P. Watt, M. A. Bronsdon, J. Dallas, C. G. Whitney, R. Reid, and M. Santosham.** 2006. Effect of community-wide conjugate pneumococcal vaccine use in infancy on nasopharyngeal carriage through 3 years of age: A cross-sectional study in a high-risk population. Clin Infect Dis. **43:**8-15.

169. **Millar, E. V., K. L. O'Brien, E. R. Zell, M. A. Bronsdon, R. Reid, and M. Santosham.** 2009. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Navajo and White Mountain Apache children before the introduction of pneumococcal conjugate vaccine. Pediatr Infect Dis J. **28:**711-716.

170. **Modde, H. K.** 1978. *Streptococcus pneumoniae* isolates relatively insensitive to penicillin G recovered from patients in Switzerland. Chemotherapy. **24:**227-230.

170a. **Mook-Kanamori, B. B., M. Geldhoff, T. van der Poll, and D. van de Beek.** 2011. Pathogenesis and pathophysiology of pneumococcal meningitis. Clin Microbiol Rev. **24:**557-591.

171. **Moore, M., R. Link-Gelles, M. M. Farley, A. Thomas, A. Reingold, L. Harrison, C. Lexau, S. Zansky, W. Schaffner, S. Petit, K. Gershman, K. Scherzinger, B. Beall, and C. Whitney.** 2011. Presented at the 51st Interscience Conference on Antimicrobial Agents and Chemotherapy, Chicago.

172. **Morona, J. K., R. Morona, and P. J. C.** 1999. Analysis of the 5' portion of the type 19A capsule locus identifies two classes of *cpsC, cpsD* and *cpsE* genes in *Streptococcus pneumoniae.* J Bacteriol. **181:**3599-3605.

173. **Morona, J. K., R. Morona, and J. C. Paton.** 1997. Characterization of the locus encoding the *Streptococcus pneumoniae* type 19F capsular polysaccharide biosynthetic pathway. Mol Microbiol. **23:**751-763.

174. **Morona, J. K., R. Morona, and J. C. Paton.** 1997. Molecular and genetic characterization of the capsule biosynthesis locus of *Streptococcus pneumoniae* type 19B. J Bacteriol. **179:**4953-4958.

175. **Mortier-Barriere, I., O. Humbert, B. Martin, M. Prudhomme, and J. Claverys.** 2000. Control of recombination rate during transformation of *Streptococcus pneumoniae*: an overview. Microb Drug Resist. **3:**233-242.

176. **Moscoso, M., and J. Claverys.** 2004. Release of DNA into the medium by competent *Streptococcus pneumoniae*: kinetics, mechanism and stability of the liberated DNA. Mol Microbiol. **54:**783-794.

177. **Mufson, M. A., and R. J. Stanek.** 1999. Bacteremic pneumococcal pneumonia in one American city: a 20-year longitudinal study, 1978-1997. Am J Med. **107:**34S-43S.

178. **Munoz-Almagro, C., I. Jordan, A. Gene, C. Latorre, J. J. Garcia-Garcia, and R. Pallarés.** 2008. Emergence of invasive pneumococcal disease caused by nonvaccine serotypes in the era of 7-valent conjugate vaccine. Clin Infect Dis. **46:**174-182.

179. **Muñoz, R., M. Mollerach, R. López, and E. García.** 1997. Molecular organization of the genes required for the synthesis of type 1 capsular polysaccharide of *Streptococcus pneumoniae*: Formation of binary encapsulated pneumococci and identification of cryptic dTDP-rhamnose biosynthesis genes. Mol Microbiol. **25:**79-92.

179a. **Musher, D. M.** 2004. A pathogenic categorization of clinical syndromes caused by *Streptococcus pnuemoniae.*, *In* E. I. Tuomanen, T. J. Mitchell, D. A. Morrison, and B. G. Spratt (ed.), The Pneumococcus. ASM Press, Washington D. C..

180. **Naraqi, S., K. G. P, and S. Kabins.** 1974. Relapsing pneumococcal meningitis: isolation of an organism with decreased susceptibility to penicillin G. J Pediatr. **85:**671-673.

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

181a. **Nigrovic, L. E., R. Malley, C. G. Macias, J. T. Kanegaye, D. M. Moro-Sutherland, R. D. Schremmer, S. H. Schwab, D. Agrawal, K. M. Mansour, J. E. Bennett, Y. L. Katsogridakis, M. M. Mohseni, B. Bulloch, D. W. Steele, R. L. Kaplan, M. I. Herman, S. Bandyopadhyay, P. Dayan, U. T. Truong, V. J. Wang, B. K. Bonsu, J. L. Chapman, and N. Kuppermann.** 2008. Effect of antibiotic pretreatment on cerebrospinal fluid profiles of children with bacterial meningitis. Pediatrics **122:**726-730.

182. **O'Brien, K. L., E. V. Millar, E. R. Zell, M. Bronsdon, R. Weatherholtz, R. Reid, J. Becenti, S. Kvamme, C. G. Whitney, and M. Santosham.** 2007. Effect of pneumococcal conjugate vaccine on nasopharyngeal colonization among immunized and unimmunized children in a community-randomized trial. J Infect Dis. **196:**1211-1220.

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

184. **Overweg, K., D. Bogaert, M. Sluijter, R. de Groot, and P. W. M. Hermans.** 2001. Molecular characteristics of penicillin-binding protein genes of penicillin-nonsusceptible *Streptococcus pneunmoniae* isolated in The Netherlands. Microb Drug Resist. **7:**323-334.

185. **Ozdemir, B., U. Beyazova, A. D. Camurdan, N. Sultan, S. Ozkan, and F. Sahin.** 2008. Nasopharyngeal carriage of *Streptococcus pneumoniae* in healthy Turkish infants. J Infect. **56:**332-339.

186. **Pagliero, E., L. Chesnel, J. Hopkins, J. Croizé, O. Dideberg, T. Vernet, and A. M. Di Guilmi.** 2004. Biochemical characterization of *Streptococcus pneumoniae* penicillin-binding protein 2b and its implication in beta-lactam resistance. Antimicrob Agents Chemother. **48:**1848-1855.

187. **Pai, R., M. R. Moore, T. Pilishvili, R. E. Gertz, W. C. G, B. Beall, and the Active Bacterial Core Surveillance Team.** 2005. Postvaccine genetic structure of *Streptococcus pneumoniae* serotype 19A from children in the United States. J Infect Dis. **192:**1988-1995.

188. **Paredes, A., L. H. Taber, M. D. Yow, D. Clark, and W. Nathan.** 1976. Prolonged pneumococcal meningitis due to an organism with increased resistance to penicillin. Pediatrics. **58:**378-381.

189. **Park, I. H., K. Kim, A. L. Andrade, D. E. Briles, L. S. McDaniel, and M. H. Nahm.** 2012. Nontypeable pneumococci can be divided into multiple *cps* types, including one type expressing the novel gene *pspK.* Mbio. **3:**e00035-00012.

190. **Park, I. H., D. G. Pritchard, R. Cartee, A. Brandao, M. C. C. Brandileone, and M. H. Nahm.** 2007. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. J Clin Microbiol. **45:**1225-1233.

191. **Pebody, R. G., O. Morgan, Y. Choi, R. George, M. Hussain, and 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:**555-561.

192. **Pelton, S. I., H. Huot, J. A. Finkelstein, C. J. Bishop, K. K. Hsu, J. Kellenberg, S. S. Huang, R. Goldstein, and W. P. Hanage.** 2007. Emergence of 19A as virulent and multidrug resistant pneumococcus in Massachusetts following universal immunization of infants with pneumococcal conjugate vaccine. Pediatr Infect Dis J. **26:**468-472.

193. **Peterson, S. N., C. K. Sung, R. Cline, B. V. Desai, E. C. Snesrud, P. Luo, J. Walling, H. Li, M. Mintz, G. Tsegaye, P. C. Burr, Y. Do, S. Ahn, J. Gilbert, R. D. Fleischmann, and D. A. Morrison.** 2004. Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. Mol Microbiol. **51:**1051-1070.

194. **Poehling, K. A., T. R. Talbot, M. R. Griffin, A. S. Craig, C. G. Whitney, E. Zell, C. A. Lexau, A. R. Thomas, L. H. Harrison, A. L. Reingold, J. L. Hadler, M. M. Farley, B. J. Anderson, and S. W.** 2006. Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine. JAMA. **295:**1668-1674.

195. **Porat, N., A. Arguedas, B. G. Spratt, R. Trefler, E. Brilla, C. Loaiza, D. Godoy, N. Bilek, and R. Dagan.** 2004. Emergence of penicillin-nonsusceptible *Streptococcus pneumoniae* clones expressing serotypes not present in the antipneumococcal conjugate vaccine. J Infect Dis. **190:**2154-2161.

196. **Provvedi, R., R. Manganelli, and G. Pozzi.** 1996. Characterization of conjugative transposon Tn*5251* of *Streptococcus pneumoniae.* FEMS Microbiol Lett. **135:**231-236.

197. **Rakov, A. V., K. Ubukata, and D. A. Robinson.** 2011. Population structure of hyperinvasive serotype 12F, clonal complex 218 *Streptococcus pneumoniae* revealed by multilocus boxB sequence typing. Infect Genet Evol. **11:**1929-1939.

198. **Ramirez, M., E. Severina, and A. Tomasz.** 1999. A high incidence of prophage carriage among natural isolates of *Streptococcus pneumoniae.* J Bacteriol. **181:**3618-3625.

199. **Ramirez, M., and A. Tomasz.** 1998. Molecular characterization of the complete 23F capsular polysaccharide locus of *Streptococcus pneumoniae.* J Bacteriol. **180:**5273-5278.

200. **Riedel, S., S. E. Beckmann, K. P. Helimann, S. S. Richter, J. Garcia-de-Lomas, M. Ferech, H. Goosens, and G. V. Doern.** 2007. Antimicrobial use in Europe and antimicrobial resistance in *Streptococcus pnuemoniae.* Eur J Clin Microbiol Infect Dis. **26:**485-490.

200a. **Ring, A., J. N. Weiser, and E. I. Tuomanen.** 1998. Pneumococcal trafficking across the blood-brain barrier. Molecular analysis of a novel bidirectional pathway. J Clin Invest. **102:**347-360.

201. **Rivera-Olivero, I. A., M. Blommaart, D. Bogaert, P. W. M. Hermans, and J. H. de Waard.** 2009. Multiplex PCR reveals a high rate of nasopharyngeal pneumococcal 7-valent conjugate vaccine serotypes co-colonizing indigenous Warao children in Venezuela. J Med Microbiol. **58:**584-587.

202. **Roberts, A. P., and P. Mullany.** 2009. A modular master on the move: the Tn*916* family of mobile genetic elements. Trends Microbiol. **17:**251-258.

203. **Robinson, K. A., W. Baughman, G. Rothrock, N. L. Barrett, M. Pass, C. Lexau, B. Damaske, K. Stefonek, B. Barnes, J. Patterson, E. R. Zell, A. Schuchat, and C. G. Whitney.** 2001. Epidemiology of invasive *Streptococcus pneumoniae* infections in the United States, 1995-1998: Oppurtunities for prevention in the conjugate vaccine era. JAMA. **285:**1729-1735.

204. **Roche, P. W., V. Krause, H. Cook, J. Barralet, D. Coleman, A. Sweeny, J. Fielding, C. Giele, R. Gilmour, R. Holland, R. Kampen, M. Brown, L. Gilbert, G. Hogg, and D. Murphy.** 2008. Invasive pneumococcal disease in Australia, 2006. Commun Dis Intell. **32:**18-30.

205. **Romero, P., N. J. Croucher, N. L. Hiller, F. Z. Hu, G. D. Ehrlich, S. D. Bentley, E. García, and T. J. Mitchell.** 2009. Comparative genomic analysis of ten *Streptococcus pneumoniae* temperate bacteriophages. J Bacteriol. **191:**4854-4862.
206. **Ruoff, K. L.** 1995. Streptococcus, *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), Manual of Clinical Microbiology. ASM Press, Washington D.C.
207. **Sá-Leão, R., F. Pinto, S. Aguiar, S. Nunes, J. A. Carriço, N. Frazão, N. Gonçalves-Sousa, J. Melo-Cristino, H. de Lencastre, and M. Ramirez.** 2011. Analysis of invasiveness of pneumococcal serotypes and clones circulating in Portugal before widespread use of conjugate vaccines reveals heterogeneous behavior of clones expressing the same serotype. J Clin Microbiol. **49:**1369-1375.
208. **Sá-Leão, R., S. E. Vilhelmsson, H. de Lencastre, K. G. Kristinsson, and A. Tomasz.** 2002. Diversity of penicillin-nonsusceptible *Streptococcus pneumoniae* circulating in Iceland after the introduction of penicillin-resistant clone Spain^{6B}-2. J Infect Dis. **186:**966-975.
209. **Sadowy, E., R. Izdebski, A. Skoczynska, P. Grzesiowski, M. Gniadkowski, and W. Hryniewicz.** 2007. Phenotypic and molecular analysis of penicillin-nonsusceptible *Streptococcus pneumoniae* isolates in Poland. Antimicrob Agents Chemother. **51:**40-47.
210. **Salter, S. J., K. A. Gould, L. M. Lambertsen, W. P. Hanage, M. Antonio, P. Turner, P. W. M. Hermans, H. J. Bootsma, K. L. O'Brien, and S. D. Bentley.** 2012. Variation at the capsule locus, *cps,* of mistyped and non-typeable *Streptococcus pneumoniae* isolates. Microbiology. **Advance online publication.**
211. **Samore, M. H., M. Lipsitch, S. C. Alder, B. Haddadin, G. Stoddard, J. Williamson, K. Sebastian, K. Carroll, O. Ergonul, Y. Carmeli, and M. A. Sande.** 2006. Mechanisms by which antibiotics promote dissemination of resistant pneumococci in human populations. Am J Epidemiol. **163:**160-170.
212. **Sanbongi, Y., T. Ida, M. Ishikawa, Y. Osaki, H. Kataoka, T. Suzuki, K. Kondo, F. Ohsawa, and M. Yonezawa.** 2004. Complete sequences of six penicillin-binding protein genes from 40 *Streptococcus pneumoniae* clinical isolates collected in Japan. Antimicrob Agents Chemother. **48:**2244-2250.
213. **Scheifele, D., S. Halperin, L. Pelletier, J. Talbot, and Members of the Canadian Paediatric Society/Laboratory Centre for Disease Control Immunization Monitoring Program.** 2000. Invasive pneumococcal infections in Canadian children, 1991-1998: Implications for new vaccination strategies. Clin Infect Dis. **31:**58-64.
214. **Shapiro, E. D., A. T. Berg, R. Austrian, D. Schroeder, V. Parcells, A. Margolis, R. K. Adair, and J. D. Clemens.** 1991. The protective efficacy of polyvalent pneumococcal polysaccharide vaccine. N Engl J Med. **325:**1453-1460.
215. **Shi, Z.-Y., M. C. Enright, P. Wilkinson, D. Griffiths, and B. G. Spratt.** 1998. Identification of three major clones of multiply antibiotic-resistant *Streptococcus pneumoniae* in Taiwanese hospitals by multilocus sequence typing. J Clin Microbiol. **36:**3514-3519.
216. **Sibold, C., J. Henrichsen, A. Konig, C. Martin, L. Chalkley, and R. Hakenbeck.** 1994. Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. Mol Microbiol. **12:**1013-1023.
217. **Sleeman, K. L., D. Griffiths, F. Shackley, L. Diggle, S. Gupta, M. C. Maiden, E. R. Moxon, D. W. Crook, and T. E. A. Peto.** 2006. Capsular serotype-specific attack rates and duration of carriage of *Streptococcus pneumoniae* in a population of children. J Infect Dis. **194:**682-688.
218. **Smith, A. M., and K. P. Klugman.** 1998. Alterations in PBP 1A essential for high-level penicillin resistance in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **42:**1329-1333.
219. **Smith, A. M., and K. P. Klugman.** 1995. Alterations in penicillin-binding protein 2B from penicillin-resistant wild-type *Streptococcus pneumoniae*. Antimicrob Agents Chemother. **39:**859-867.
220. **Smith, A. M., and K. P. Klugman.** 2005. Amino acid mutations essential to production of an altered PBP 2X conferring high-level β-lactam resistance in a clinical isolate of *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **49:**4622-4627.
221. **Snyder, C. M., W. Begor, and E. R. Berndt.** 2011. Economic perspectives on the advance market commitment for pneumococcal vaccines. Health Aff. **30:**1508-1517.
222. **Sogstad, M. K. R., E. A. Høiby, and D. A. Caugant.** 2006. Molecular characterization of non-penicillin-susceptible *Streptococcus pneumoniae* in Norway. J Clin Microbiol. **44:**3225-3230.

223. **Song, J. H., R. Dagan, K. P. Klugman, and B. Fritzell.** 2012. The relationship between pneumococcal serotypes and antibiotic resistance. Vaccine. **Advance online publication.**
224. **Spratt, B. G., W. P. Hanage, B. Li, D. M. Aanensen, and E. J. Feil.** 2004. Displaying the relatedness among isolates of bacterial species - the eBURST approach. FEMS Microbiol Lett. **15:**129-134.
225. **Spratt, B. G., and M. C. J. Maiden.** 1999. Bacterial population genetics, evolution and epidemiology. Philos Trans R Soc Lond Ser B Biol Sci. **354:**701-710.
226. **Stanhope, M. J., T. Lefébure, S. L. Walsh, J. A. Becker, P. Lang, P. D. Pavinski Bitar, L. A. Miller, M. J. Italia, and H. Amrine-Madsen.** 2008. Positive selection in penicillin-binding proteins 1a, 2b, and 2x from *Streptococcus pneumoniae* and its correlation with amoxicillin resistance development. Infect Genet Evol. **8:**331-339.
227. **Syrogiannopoulos, G. A., I. N. Grivea, A. Tait-Kamradt, G. D. Katopodis, N. G. Beratis, J. Sutcliffe, P. C. Appelbaum, and T. A. Davies.** 2001. Identification of an *erm*(A) erythromycin resistance methylase gene in *Streptococcus pneumoniae* isolated in Greece. Am J Epidemiol. **45:**342-344.
228. **Taylor, D. E., and A. Chau.** 1996. Tetracycline resistance mediated by ribosomal protection. Antimicrob Agents Chemother. **40:**1-5.
229. **Tempest, B., J. Carney, and B. Eberle.** 1974. Distribution of the sensitivities to penicillin of types of *Diplococcus pneumoniae* in an American Indian population. J Infect Dis. **130:**67-69.
230. **Thomas, J. C., M. Figueira, K. P. Fennie, A. S. Laufer, Y. Kong, M. E. Pichichero, S. I. Pelton, and M. M. Pettigrew.** 2011. *Streptococcus pneumoniae* clonal complex 199: genetic diversity and tissue-specific virulence. PLoS One. **6:**e18649.
231. **Tran, T. D., H. Kwon, E. Kim, K. Kim, D. E. Briles, S. Pyo, and R. Dong-Kwon.** 2011. Decrease in penicillin susceptibility due to heat shock protein ClpL in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **55:**2714-2728.
231a. **Tunkel, A. R., B. J. Hartman, S. L. Kaplan, B. A. Kaufman, K. L. Roos, W. M. Scheld, and R. J. Whitley.** 2004. Practice guidelines for the management of bacterial meningitis. Clin Infect Dis. **39:**1267-1284.
231b. **Tuomanen, E. I.** 2004. Attachment and invasion of the respiratory tract., *In* E. I. Tuomanen, T. J. Mitchell, D. A. Morrison, and B. G. Spratt (ed.), The Pneumococcus. ASM Press, Washington D. C..
232. **Varvio, S., K. Auranen, E. Arjas, and P. H. Mãkela.** 2009. Evolution of the capsular regulatory genes in *Streptococcus pneumoniae.* J Infect Dis. **200:**1144-1151.
233. **Vestrheim, D. F., P. Gaustad, I. S. Aaberge, and D. A. Caugant.** 2011. Pherotypes of pneumococcal strains co-existing in healthy children. Infect Genet Evol. **11:**1703-1708.
234. **Wartha, F., K. Beiter, B. Albiger, J. Fernebro, A. Zychlinsky, S. Normark, and B. Henriques-Normark.** 2007. Capsule and D-alanylated lipoteichoic acids protect *Streptococcus pneumoniae* against neutrophil extracellular traps. Cell Microbiol. **9:**1162-1171.
235. **Weber, B., K. Ehlert, A. Diehl, P. Reichmann, H. Labischinski, and R. Hakenbeck.** 2000. The *fib* locus in *Streptococcus pneumoniae* is required for peptidoglycan crosslinking and the PBP-mediated β-lactam resistance. FEMS Microbiol Lett. **188:**81-85.
236. **Weinberger, D. M., R. Malley, and M. Lipsitch.** 2011. Serotype replacement in disease after pneumococcal vaccination. Lancet **378:**1962-1973.
237. **Weinberger, D. M., K. Trzcinski, Y. Lu, D. Bogaert, A. Brandes, J. Galagan, P. W. Anderson, R. Malley, and M. Lipsitch.** 2009. Pneumococcal capsular polysaccharide structure predicts serotype prevalence. PLoS Pathog. **5:**e1000476.
238. **Weinstein, M. P., K. P. Klugman, and R. N. Jones.** 2009. Rationale for revised penicillin susceptibility breakpoints versus *Streptococcus pneumoniae*: Coping with antimicrobial susceptibility in an era of resistance. Clin Infect Dis. **48:**1596-1600.
239. **Weiser, J. N.** 2010. The pneumococcus: why a commensal misbehaves. J Mol Med. **88:**97-102.
240. **Whitney, C. G., M. M. Farley, J. Hadler, L. H. Harrison, N. M. Bennett, R. Lynfield, A. Reingold, P. R. Cieslak, T. Pilishvili, D. Jackson, R. R. Facklam, J. H. Jorgensen, and A. Schuchat.** 2003. Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. N Engl J Med. **348:**1737-1746.
241. **World Health Organisation.** 2007. Pneumococcal conjugate vaccine for childhood immunisation - WHO position paper. Wkly Epidemiol Rec. **82:**93-104.

242. **World Health Organisation Initiative for Vaccine Research** 2009. Acute Respiratory Infections (Update September 2009). World Health Organisation. [Online.]
243. **Yao, J. D. C., and R. C. Moellering Jr.** 1995. Antimicrobial Agents, *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), Manual of Clinical MIcrobiology, 6th ed. ASM Press, Washington D. C.
244. **Yin, Z., B. D. Rice, P. Waight, E. Miller, R. George, A. E. Brown, R. D. Smith, M. Slack, and V. C. Delpech.** 2012. Invasive pneumococcal disease among HIV-positive individuals, 2000-2009. AIDS **26:**87-94.
245. **Yother, J.** 2004. Capsules, *In* E. I. Tuomanen (ed.), The Pneumococcus. ASM Press, Washington D. C.
246. **Zhanel, G. G., X. Wang, K. Nichol, A. Nikulin, A. K. Wierzbowski, M. Mulvey, and D. J. Hoban.** 2006. Molecular characterisation of Canadian paediatric multidrug-resistant *Streptococcus pneumoniae* from 1998-2004. International Journal of Antimicrobial Agents **28:**465-471.
247. **Zielen, S., E. Bühring, N. Strand, J. Reichenbach, and D. Hofmann.** 2000. Immunogenicity and tolerance of a 7-valent pneumococcal conjugate vaccine non-responders to the 23-valent pneumococal vaccine. Infect Immun. **68:**1435-1440.

2. GENERAL METHODS

This chapter describes recurrent materials and methods relevant to this thesis. Methods specific to any given results chapter are described within the chapter itself.

2.1 Pneumococcal Historical Isolate Collection

A unique collection of 426 global PSP and PNSP dated 1937 - 2007 was studied, as summarised in Table 2.1. Isolates included one of the first PNSP reported from Australia [10], the first high-level PNSP reported from South Africa [1, 11] and the 43 PMEN reference strains [14]: 211 pneumococci dated 1937 - 1996 were provided by L. M. Lambertsen (LML) at the Statens Serum Insititut (SSI), Denmark. These isolates included both PSP (n = 203), PNSP (n = 7) and an isolate of unknown penicillin susceptibility (n = 1). Following the analyses of MLST and *pbp* sequence data from these isolates, the 43 PMEN clones (including both PSP (n = 18) and PNSP (n = 25)) and additional modern (1990s - 2000s) PSP (n = 38) and PNSP (n = 9) that had been recently genotyped by A. B. Brueggemann (ABB) / L. McGee (LM) and represented CCs of interest, were selected for inclusion. In addition, a literature search was performed to identify studies that reported early PNSP. The authors of those papers were contacted and isolates were requested; the majority of isolates were still available and were added to our collection. The available isolates comprised PNSP dated 1969 and 1972 from Papua New Guinea (n = 2), PSP (n = 8) and PNSP (n = 71) dated 1977 - 1989 from South Africa, PSP (n = 1) and PNSP (n = 26) dated 1981 - 1991 from the USA, PNSP dated 1987 - 1989 (n = 9) from Spain and PNSP dated 1986 - 1994 from Germany (n = 8). Finally, genomic sequence data for 16 isolates was retrieved from the Genbank

database. These isolates included both PSP (n = 5), PNSP (n = 4) and seven pneumococci of unknown penicillin-susceptibility. Sources and references for all isolates are detailed in Table 2.1. Demographic, serotype and penicillin-susceptibility information for all isolates are summarised by decade in Chapter 4, Supplementary Table 4.1. (The complete data set can be found as Appendices 1.1 and 1.2.)

This isolate collection represents the most comprehensive sample of historical pneumococci, including both PSP and PNSP, known to exist at this time. As such, the collection provides a unique opportunity to study pneumococcal evolution. However, the nature of the sample (i.e. not a systematic sample) does impose certain limitations. For example, isolates were not evenly distributed through time or by geographic location, and in fact those from Denmark and the USA were over-represented among the oldest isolates. Therefore it was not appropriate to address questions about the rate of pneumococcal evolution or to statistically test differences between isolates representing different locations or time periods.

The genetic diversity represented by the isolate collection, although extensive, was biased towards that represented by isolates from Europe (specifically Denmark for older isolates) and the USA (note however that these did include isolates representing internationally disseminated clones). Thus, it is likely that pneumococcal genotypes, which may be/were common in other geographical regions such as Asia or South America, but are/were rare in Europe and the USA, were not represented in the collection. It is assumed that the general trends described here would also be true for

less-well represented pneumococcal populations, but without sufficient isolate representatives it is not possible to know if this is the case.

Inclusion of isolates representing the earliest PNSP was largely limited to those which had been reported in the literature. Prior to the 1980s relatively few studies reported PNSP, and prior to the late-1970s it was assumed that such pneumococci were rare (see section 1.7 *Antimicrobials and Antimicrobial Resistance*). However, given that pneumococcal penicillin-susceptibility testing was not common place at this time, it is conceivable that greater numbers of PNSP existed but were not detected (particularly given that intermediate penicillin resistance is not associated with adverse clinical outcomes – also see section 1.7). Such pneumococci could have represented alternative genotypes / possess alternative *pbp* gene sequences which were not represented in this collection.

Another potential source of bias within the collection was that the majority of isolates were invasive isolates, rather than those isolated from asymptomatic carriage. Given that nasopharyngeal carriage is a pre-requisite for disease (see section 1.1 *What is Streptococcus pneumoniae?*), all invasive pneumococci must have colonised the nasopharynx at some point. In contrast, pneumococci which are carried in the nasopharynx are not all equally likely to cause disease (see section 1.5 *Pneumococcal Polysaccharide Capsules*). As such, this collection is biased towards those pneumococci which are more likely to cause invasive disease. Consequently, it is possible that the conclusions presented in this thesis are relevant mainly to invasive pneumococcal isolates and do not apply to the wider pneumococcal population.

An idealised sample would include randomly selected carriage isolates (thus also including those which case IPD) stratified by date of collection (e.g. by decade) and by geographic location (e.g. by continent). Unfortunately it is not possible to retrospectively assemble such a sample for numerous reasons including; 1) prior to the last two/three decades pneumococci were rarely collected from carriage; 2) older pneumococci are less likely to be available and/or viable because they were not saved/stored correctly, or records were lost/incomplete; 3) studies in geographic regions predominated by developing countries are less common than those in regions predominated by economically developed countries.

Table 2.1 Historical pneumococcal isolates.

Number of Isolates (†)	Years (n, Unknown)	Country (n)	Source	Reference(s)
211 (174)	1937 - 1996	Denmark (139), USA (48), Germany (6), France (2), India (2), Malaysia (2), Singapore (2), Australia (1), Belgium (1), The Gambia (1), The Netherlands (1), Norway (1), Scotland (1), Spain (1), Sweden (1), Unknown (2)	LML	[2]
43 (43)	1948 - 1998	The Netherlands (6), USA (6), Portugal (5), South Africa (3), Spain (3), CSR (2), Greece (2), Poland (2), Taiwan (2), Canada (1), Columbia (1), Denmark (1), England (1), Finland (1), France (1), Hungary (1), Sweden (1), Unknown (4)	LM	[14]
36 (36)	1979 - 1989	South Africa (36)	KPK, AVG	--
27 (27)	1969 - 1994	South Africa (17), Germany (8), Papua New Guinea (2)	RH	[8, 9]
27 (27)	1981 - 1991	USA (26), South Africa (1)	LM	[13]
26 (26)	1977 - 1978	South Africa (26)	MRJ	[1, 11, 17]
22 (22)	1999 - 2007 (1)	USA (22)	LM, BWB	--
17 (17)	1993 - 2005	Iceland (17)	KGK	--
9 (9)	1987 - 1989	Spain (9)	JL	[6, 15, 16, 20, 21]
8 (8)	1998 - 2002	USA (8)	ABB	--

† Number of isolates selected for *pbp* nucleotide sequencing.

LML = L. M. Lambertsen, LM = L. McGee, KPK = K. P. Klugman, AVG = A. von Gottberg, RH = R. Hakenbeck, MRJ = M. R. Jacobs, BWB = B. W. Beall, KGK = K. G. Kristinsson, JL = J. Liñares, ABB = A. B. Brueggemann.

2.2 Serotype and Antimicrobial Susceptibilities

Where not previously published, serotype data were provided by the appropriate isolate contributors. Penicillin and cefotaxime susceptibility data were provided by isolate contributors, with the exception of isolates from K. Kristinsson (KGK) and those from ABB. Penicillin and cefotaxime MICs for these isolates (n = 25) were determined by Etest as follows: Isolates were cultured onto Columbia agar with sheep blood (Oxoid) and incubated overnight at 37°C + 5% CO_2. A sterile swab was used to suspend bacterial growth in 1 ml PBS (phosphate buffered saline, pH 7.4, Fisher Bioreagents) to a turbidity matching 0.5 McFarland standard. Bacterial cell suspension was spread onto Mueller Hinton agar with sheep blood (Oxoid) using a new sterile swab. Single Etest strips (bioMérieux) were applied to each culture plate before overnight incubation at 37°C + 5% CO_2. MICs were determined as that indicated on the Etest strip at the edge of the zone of growth inhibition. The PMEN1 type strain, for which penicillin and cefotaxime MICs have been previously published [14], was used as a control.

2.3 DNA Extraction

Purified bacterial DNA was provided for isolates from LML, R. Hakenbeck (RH), KGK and ABB. DNA extracts for isolates provided by LM/ B. W. Beall (BWB), M. R. Jacobs (MRJ) and J. Liñares (JL) were prepared by the boilate method (below). This is a crude method whereby cells are suspended in a buffer and are lysed at high temperature. The resultant solution contains the DNA and is used as the extract. This method works well for polymerase chain reaction (PCR) and sequencing of MLST genes, but

unfortunately not as well for *pbp* gene sequencing. Consequently, it was decided to use the DNeasy kit DNA extraction method (below) for the remaining isolates in the collection (those provided by K. P. Klugman (KPK) /A. von Gottberg (AVG)).

Boilate DNA extraction

Pneumococcal cells were lysed to release chromosomal DNA as follows; 5 - 10 pneumococcal colonies grown overnight on Columbia agar with sheep blood (Oxoid) at 37°C + 5% CO_2 or ~4 - 9mm^3 frozen isolate stock were suspended in 60 µl TE buffer (pH 8, Ambion) and incubated at 95°C for 10 mins.

DNeasy kit DNA extraction

Single pneumococcal colonies were cultured on Columbia agar with sheep blood (Oxoid) and incubated overnight at 37°C + 5% CO_2. 3 - 4 sweeps of growth were subcultured onto tryptic soy agar (Oxoid) + 1000 U/ml catalase (from bovine liver, Oxoid) and spread with a swab before further overnight incubation at 37°C + 5% CO_2. (Addition of catalase was required for growth on tryptic soy agar in order to neutralise the hydrogen peroxide produced by the pneumococci. In the absence of catalase, hydrogen peroxide release would lead to premature cell lysis and subsequent DNA denaturation.) Pneumococcal growth was suspended in 1 ml PBS (pH 7.4, Fisher Bioreagents) and centrifuged at 7,600 rpm for 10 mins. Supernatants were removed and pellets resuspended in 180 µl lysis buffer + 2% lysozyme (from chicken egg white, Sigma), and incubated at 37°C for 30 mins. Addition of 25 µl proteinase K solution (Qiagen) and 200 µl AL buffer (Qiagen) was followed by incubation at 70°C for 30 mins. 200 µl 100% ETOH was added before thorough mixing by inversion. The mix was

transferred by pipette to a DNeasy mini column sitting in a 2 ml collection tube (Qiagen) and centrifuged at 10,000 rpm for 1 min. Flow-through plus the collection tube were discarded and the DNeasy mini column was placed in a new 2 ml collection tube. Addition of 500 µl Buffer AW1 (Qiagen) was followed by centrifugation at 10,000 rpm for 1 min. Flow-through plus the collection tube were discarded and the DNeasy mini column was placed in a new 2 ml collection tube. Addition of 500 µl Buffer AW2 (Qiagen) was followed by centrifugation at 13,000 rpm for 3 mins. DNeasy column was placed into a 1.5 ml microcentrifuge tube (Axygen) before addition of 115 µl TE Buffer and incubation at room temperature for 1 min. Final centrifugation at 10,000 rpm for 1 min produced ~110 µl elute, which contained the purified DNA.

2.4 Multilocus Sequence Typing (MLST)

The pneumococcal MLST scheme, first described by Enright and Spratt [5], requires PCR and sequencing of regions of 7 housekeeping genes as described in Chapter 1. MLST data for the 43 PMEN clones and the isolates provided by MRJ were previously published. MLST data for all isolates from RH were provided by RH. MLST data for all isolates from KGK and ABB were provided by ABB. MLST data for 5 of 9 and 22 of 49 isolates from JL and LM/BWB respectively were provided by JL/LM. MLST data for all remaining isolates (n = 278) were generated by PCR and conventional Sanger sequencing as detailed below.

PCR amplification

PCR reactions containing 1 µl DNA extract, 2.5 µl 10x buffer (Qiagen), 0.5 µl each of the forward and reverse primers (100 µM, Table 2.2), 0.5 µl dNTPs (10 mM, Invitrogen) and 0.15 µl Taq DNA polymerase (5 U/µl, Qiagen) were prepared to a final volume of 25 µl. Cycling parameters were as follows: Initial denaturation of 95°C for 3 mins, followed by 35 cycles of 95°C for 30 s, 55°C for 30 s and 72°C for 1 min, followed by a final extension period of 72°C for 10 mins.

Gel electrophoresis of PCR products

2 µl PCR product was combined with 5 µl 5x loading dye (Sigma) and loaded into 0.8% agarose, 2.5% ethidium bromide gel. Samples were electrophoresed at 140 V for 25 mins and viewed with a BioDoc-it UV transilluminator.

Precipitation of PCR products

60 µl of 20% PEG (polyethylene glycolate) / 2.5 M NaCl was added to each PCR product before incubation at room temperature for 30 mins. Incubated products were centrifuged at 3,200 rpm, 4°C for 1 hour. Supernatants were removed by inversion onto laboratory tissue paper and centrifugation at 500 rpm for 5 s, followed by further inversion onto laboratory tissue paper and centrifugation at 500 rpm for 1 min. Pellets were washed by addition of 150 µl 70% ETOH and centrifugation at 3,200 rpm, 4°C for 30 mins, followed by removal of supernatant by inversion onto laboratory tissue paper and centrifugation at 500 rpm for 1 min. The ETOH wash steps were repeated and final pellets were dried by incubation at 37°C for 2 mins. Purified PCR products were

re-hydrated with 10 - 50 µl DNAse/RNAse free H_2O (Sigma, dependent upon DNA gel band intensity).

Sequencing reaction

Sequencing reactions containing 2 µl rehydrated PCR product, 1.75 µl 5x sequencing buffer, 4 µl primer (1 µM, Table 2.2) and 0.5 µl Taq FS (Big Dye) were prepared in a final volume of 10 µl. Cycling parameters were as follows: 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 2 mins.

Precipitation of sequencing products

Sequencing products were precipitated by addition of 20 µl 4:11 [3 M NaOAc pH 5.2:H_2O] and 60 µl 95% ETOH followed by incubation at -20°C for 1 hour and centrifugation at 3,500 rpm, 4°C for 1 hour. Supernatants were removed by inversion onto laboratory tissue paper and centrifugation at 500 rpm for 1 min. Pellets were washed by addition of 150 µl 70% ETOH and centrifugation at 3,200 rpm, 4°C for 30 mins, followed by removal of supernatant by inversion onto laboratory tissue paper and centrifugation at 500 rpm for 1 min. The ETOH wash steps were repeated and final pellets were dried by incubation at 37°C for 2 mins.

Final sequencing products were separated and detected with an ABI Prism 3730 automated DNA sequencer (Applied Biosystems) at the Sequencing Laboratory, Dept. of Zoology, Oxford.

Table 2.2 Primers for pneumococcal MLST.

Locus	Forward Primer 5' - 3'	Reverse Primer 5' - 3'	Allele Length (bp)
aroE	CGT TTA GCT GCA GTT GTT GC*	CCC ACA CTG GTG GCA TTA AC*	405
gdh	ATG GAC AAA CCA GCN AGY TT‡	GCT TGA GGT CCC ATR CTN CC‡	460
gki	GGC ATT GGA ATG GGA TCA CC‡	TCT CCC GCA GCT GAC AC‡	483
recP	GCC AAC TCA GGT CAT CCA GG‡	AGA TGG CTT GCC TGA AGC$^{\Psi}$	450
spi	CGC TTA GAA AGG TAA GTT ATG A$^{\Psi}$	GTG ATT GGC CAG AAG CGG AA‡	474
xpt	CCA CTA CAA CGG GAA ATA TTT GA*	CTT GAG TTT GCA TTA GAG ATC TGC$^{\Psi}$	486
ddl	AGC GTG TTC TGG AAT CTG C$^{\Psi}$	AGG TCA ACC AAA CGC TCG$^{\Psi}$	441

* Reference [4a]; ‡ reference [5]; $^{\Psi}$ designed by A. B. Brueggemann in 2005.

Sequence analysis

MLST nucleotide sequences were assembled and edited using the STARS (Sequence Typing Analysis and Retrieval System, Staden Package [18]) software which enables the user to assemble forward and reverse sequence reads, trim the sequence to the appropriate allele length and edit any mis-called bases. Allele sequences/profiles were queried at the pneumococcal MLST database (www.spneumoniae.mlst.net) and new alleles/profiles were submitted to the curator for assignment.

2.5 Clonal Complex (CC) Designation

As described previously, when a large sample of isolates representing a recombinogenic bacterial species are analysed by the (go)eBURST method, it is possible that false positive relationships between STs may be inferred. In this situation isolates assigned to the same CC will not necessarily have descended from a (recent) common ancestor, an attribute which must be upheld under the original CC definition [4b]. In particular, when analysing the entire pneumococcal MLST isolate database (>16,000 isolates and >8,000 STs), goeBURST defines one very large "CC".

The majority of goeBURST defined CCs comprise between 2 and 279 STs (excluding the very large "CC" and STs designated as singletons, which differ from all other STs by at least two loci). In contrast, the very large "CC" comprises 1105 STs. This "CC" includes representatives of, what pneumococcal investigators have recognised as, several major epidemiologically different CCs, including multiple PMEN clones that express different serotypes / antimicrobial susceptibility profiles. As such it is unlikely that all the isolates within this "CC" share a recent common ancestor (although of course all pneumococci are considered to share a common ancestor at some point in their more distant evolutionary history).

In order that the original definition of a CC was upheld for the purposes of this thesis, isolates were assigned to CCs using a modified goeBURST method. This method, described below, resulted in a more conservative assignment of isolates to CCs, and separated the PMEN clone representatives mentioned above.

MLST data for all isolates and all those in the pneumococcal database were analysed by the goeBURST method implemented in the Phyloviz package [7] to predict group and sub-group founder STs. Within each goeBURST designated group, isolates were manually assigned to the "primary" CC only if they represented SLVs/DLVs of the predicted group founder, or represented additional SLVs of large ($n \geq 5$ STs) sub-group founders. Isolates representing STs that did not meet these criteria were not assigned to the primary CC; those that were not assigned to the primary CC were subsequently considered as new group founders of "secondary" CCs. Isolates were assigned to secondary CCs using the same criteria as described for primary CCs. Secondary CCs

were assigned sequentially, following the descending rank order of putative founder ST SLV frequency.

CCs were named according to the predicted founder(s) ST(s). When there was no clear group founder (generally among small CCs comprising only a few STs) the group was designated NoneX, where X was the ST of lowest numerical value in the group. When isolates could not be assigned to a CC due to a lack of SLVs within the sample, they were designated SingletonX, where X was the isolate ST.

2.6 *pbp* PCR Amplification and Sequencing

pbp2x (1,673 bp), *pbp1a* (901 bp) and *pbp2b* (1,272 bp) nucleotide sequences for 389 selected isolates, were retrieved from Genbank (n = 36 sequences) or determined by conventional PCR and Sanger sequencing (n = 1131 sequences, 23 of which were provided by RH). 174 of 211 isolates from the SSI were chosen for *pbp* nucleotide sequencing because they represented a wide range of serotypes, STs and isolation dates. All isolates from each of the other sources were also included. PCR reactions containing 1 µl DNA extract, 13.5 µl Hotstar Taq DNA Polymerase master mix (Qiagen) and 0.5 µl each of the forward and reverse primers (100 µM, Table 2.3) were prepared to a final volume of 25 µl. Cycling parameters were as follows: Initial denaturation of 95°C for 3 mins, followed by 35 (40 for *pbp2x*) cycles of 95°C for 30 s, 55°C (58°C for *pbp1a*) for 30 s and 72°C for 1 min followed by a final extension period of 72°C for 10 min. PCR products were detected by gel electrophoresis as described above. PCR products were precipitated by 20% PEG/2.5M NaCl, washed in 70% ETOH and

rehydrated in 10 - 20 µl nuclease-free H_2O (Sigma, dependent on gel band intensity). Sequencing reactions containing 2 µl rehydrated PCR product, 1.75 µl 5x sequencing buffer, 4 µl primer (1 µM, Table 2.3) and 0.5 µl Taq FS (Big Dye) were prepared in a final volume of 10 µl. Cycling parameters were as for MLST sequencing. Sequencing products were precipitated with 3 M NaOAc (pH 5.2, Sigma) solution, washed with 70% ETOH and subsequently separated and detected with an ABI Prism 3730 automated DNA sequencer (Applied Biosystems) as described for MLST.

2.7 *pbp* Sequence Analysis

pbp1a sequences were composed of a single forward and a single reverse sequence, and were assembled/edited using the STARS software as described above. *pbp2x* and *pbp2b* sequences, each composed of two forward plus two reverse sequences, were assembled using the preGap4 software and edited using the Gap4 software (Staden Package [18]), which function similarly to STARS. Unique *pbp* sequences were identified using the non-redundant databases (NRDB) tool available at www.pubmlst.org. Sequences were submitted to the tool in raw FASTA format and were output as a list of unique allele groups. Unique sequences were assigned unique allele numbers. Sequences were further analysed by alignment with both Multalin [3] and MUSCLE [4]. Following the latter method, alignments were imported to MEGA5 [19] for visual comparison. MEGA5 was also used to calculate pairwise nucleotide differences and produce variable site maps.

Table 2.3 Primers for PCR and conventional sequencing of regions of *pbp2x, pbp1a* and *pbp2b*.

Gene	Primer Name	Primer Sequence (Forward Primers (F) 5'-3', Reverse Primers (R) 3'-5')	Application	Allele Length (bp)
pbp2x	Sp36F	GAA GAA TCC TAT GTA AGA GAG C	PCR / sequencing	1,673
	Sp35F	GAG GAC TTT GTT TGG CGT GA	Sequencing only	
	Sp34R	GCT GAA GCC CGC TCC AAG AT	Sequencing only	
	Sp33R	GCT GGC CTG TAA TTT GCG CCT	PCR / sequencing	
	Sp37F*	ACA GGT CAA ATA CTA CAG ATG C	PCR / sequencing	
	Sp52R*	CCT TGC ACA TAG GCA ATA ACT CC	PCR / sequencing	
	*pbp2x*F1*	GCT GAG GAC GCA ACY TCC	PCR / sequencing	
	*pbp2x*F2*	AAG GAC TTT GTT TGG CGT GA	PCR / sequencing	
	*pbp2x*F4*	ACT TCT GGG ATG GAG AGT TCC	PCR / sequencing	
	*pbp2x*R1*	GTG ATT CTT TCA TAG CTG AGG C	PCR / sequencing	
	*pbp2x*F5*	ACC GAA ATG GAG TCC CGA TTG C	PCR / sequencing	
	*pbp2x*R4*	ATC TCG AGT CGT CGC ATC CGC	PCR / sequencing	
pbp1a	Sp47F	ACT ATT ATT TGT GCT TGG AGT GGT TGA GC	PCR / sequencing	901
	Sp43R	CCA AGA AGC TCA AAA ACA TCT GTG GG	PCR / sequencing	
	*pbp1a*F*	TTG ATG GGG GTT GTT GTG GAG C	PCR / sequencing	
	*pbp1aR2**	AGC CTT GCT GGC TGG AAT GC	PCR / sequencing	
	*pbp1aR3**	GGC TAT AAC CTT CTA ACT ACT GGG	PCR / sequencing	
pbp2b	*pbp2b*_2F	GCG TAT AGT TAT CAA ACT GCC CA	PCR / sequencing	1,272
	*pbp2b*_3F	ACC TGT TTT TGG GTT AAG GGC	Sequencing only	
	*pbp2b*_3R	CAA ATT TGA CAG AAC GCC AGC	Sequencing only	
	*pbp2b*_4R	ACA GCC ATT CGA TTC CGA TTC	PCR / sequencing	

* Alternative primers used following failure of initial primers and/or when sequences were suspected to be divergent in the primer binding regions.

2.8 DNA Extraction and Quantification for Whole-genome Sequencing

Ninety-six isolates from the historical collection were selected for whole-genome sequencing by the Illumina method (see Appendix 1.1 and Chapter 4). DNA extraction for 48 isolates (those originally provided by LML) was performed by LML at the SSI in Denmark. DNA was extracted from the remaining 48 isolates in Oxford as follows: Single pneumococcal colonies were cultured on Columbia agar with sheep blood (½ plate, Oxoid) and incubated overnight at 37°C + 5% CO_2. ⅓ of the growth from each ½ plate was subcultured onto each of 3 tryptic soy agar (Oxoid) + 1000 U/ml catalase (from Bovine liver, Oxoid) plates and spread with a swab. Cultures were incubated overnight at 37°C + 5% CO_2. Growth from each 1½ plates corresponding to the same strain was suspended in 1 ml PBS (pH 7.4, Fisher Bioreagents) and centrifuged at 7,600 rpm for 10 mins. Supernatants were removed and pellets resuspended in 180 µl lysis buffer + 2% lysozyme (from chicken egg white, Sigma), and incubated at 37°C for 30 mins. Addition of 25 µl proteinase K solution (Qiagen) and 200 µl AL buffer (Qiagen) was followed by incubation at 70°C for 30 mins. Extractions were completed using the DNeasy kit (Qiagen) as described above. Extracts representing the same strain were combined in a single microcentrifuge tube.

DNA was quantified using the Qubit Fluorometer and Quant-iT BR dsDNA assay kit (Invitrogen) as follows: Quant-iT working solution was prepared by addition of n µl Quant-iT reagent to 199 x n µl Quant-iT buffer (where n = number of samples to be quantified). The solution was mixed by inversion. 10 µl DNA extract was added to 190 µl working solution in a Quant assay tube and vortexed for 2 - 3 s. Samples were

incubated at room temperature for 2 – 3 mins and concentrations were determined by the Qubit reader after calibration with Qubit standards. Target DNA concentration was 20 - 99 ng/µl.

2.9 Whole-genome Sequencing

Library construction and paired-end Illumina sequencing was performed by staff at the Wellcome Trust Sanger Institute, Hinxton, UK. This process is briefly described below:

DNA was fragmented before PCR amplification, adaptor ligation and fragment size selection. Adaptor-ligated PCR products were hybridised to sequencing flow cells, each consisting of 8 sequencing lanes. PCR products representing a total of 12 independent pneumococcal genomes were mixed together in a single lane. The sequences were later separated *in silico* by comparison of index tags (unique for each isolate, incorporated in the adaptor sequences). Hybridised PCR products provided templates for synthesis of covalently linked ssDNA on the flow cell surface. Subsequent removal of the original PCR products was followed by bridge amplification, producing clusters of identical ssDNA molecules. Clusters were sequenced through successive rounds of; 1) addition of fluorescently labelled, reversibly terminated nucleotides; 2) laser excitation and; 3) removal of the fluorescence label/termination group. 54 rounds of nucleotide addition were undertaken to produce sequence reads of 54 nucleotide length.

Production of paired reads began by removal of sequencing products though denaturation. ssDNA fragments, which were covalently linked on the flow cell surface,

were used as templates for synthesis of covalently linked complementary strands. Removal of the original ssDNA, by cleavage of a specific sequence motif in its adaptor region, was followed by sequencing of the complimentary strand.

2.10 Whole-genome Sequence Analysis

Paired 54 nucleotide sequence reads were assembled to contigs using Velvet [22]. This software uses a de Bruijn graph algorithm to produce *de novo* assemblies from short read sequence data. The Velvet optimiser script was used to optimise k-mer lengths to the longest length achieving an average (arithmetic mean) of ≥20x coverage across the assembled sequences when sufficient sequence data were available. Resulting average coverage values were isolate-dependent and ranged from 12x to 352x (note that all except one isolate, ICE570, had an average read coverage of ≥23x). As expected, read coverage was not constant across the entire length of each genome, as some sequence regions were more difficult to sequence/assemble and thus had a lower than average read coverage. Conversely some regions, such as those containing repeats of greater length than the sequence reads or those which were present in more than one copy within the genome, had a higher than average coverage. However, as part of its error removal processes, Velvet will remove assembly contigs which fall below a minimum coverage cut-off value (minimum = 4), thus maintaining a constant level of accuracy across genomes and between isolates.

Whilst isolate-dependent coverage variation is not likely to significantly affect consensus sequence accuracy, it is known to affect assembly contig size [22]. Average

coverage below ~20-30x was shown to result in reduced contig N50 (the contig length for which the sum of lengths of all contigs of equal or greater size is equivalent to half of the total genome length). Only a single isolate (ICE570) included in the analyses described in this thesis had an average coverage of <20x, resulting in an N50 value of 1,771 bp. The mean N50 for all isolates was ~43,000 bp. It is not likely that this variation would have significantly affected the findings discussed in this thesis since 1,771 bp is larger than the length of most pneumococcal genes (relevant to Genome Comparator analyses – see below and chapter 5) and visual inspection of sequence alignments allowed for the detection and description of sequence gaps in regions of interest.

Velvet contigs were deposited in a customised BIGSdb (Bacterial Isolate Genome Sequence database) designed by K. Jolley at the University of Oxford [12]. BIGSdb allows the user to link unlimited sequence data with user-specified isolate demographic information. The user is able to search or query the database for isolate or sequence information, and run BLAST or other tools using sequences deposited in the database. One such tool named, "Genome Comparator", utilises the BLAST algorithm to search for sequence matches to a pre-defined sequence list or a user-specified reference sequence. In the latter case, Genome Comparator will individually BLAST sequences marked as "CDS" (coding sequences) within the user-specified Genbank file (specified by Genbank accession no.). Unique alleles at each coding sequence are assigned unique allele numbers, and these are output in table format. The user can specify the minimum % identity and minimum % alignment length for BLAST matches, and the BLASTN word size.

2.11 References

1. **Appelbaum, P. C., J. N. Scragg, A. J. Bowen, A. Bhamjee, A. F. Hallett, and R. C. Cooper.** 1977. *Streptococcus pneumoniae* resistant to penicillin and chloramphenicol. Lancet **2:**995-997.
2. **Bentley, S. D., D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabbinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail, G. Samuel, I. C. Skovsted, M. S. Kaltoft, B. Barrell, P. R. Reeves, J. Parkhill, and B. G. Spratt.** 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. **2:**262-269.
3. **Corpet, F.** 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. **16:**10881-10890.
4. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792-1797.

4a. **Enright, M. C., K. Knox, D. Griffiths, D. W. Crook, and B. G. Spratt.** 2000. Molecular typing of bacteria directly from cerebrospinal fluid. Eur J Clin Microbiol Infect Dis. **19:**627-630.

4b. **Enright, M. C., D. A. Robinson, G. Randle, E. J. Feil, H. Grundmann, and B. G. Spratt.** 2002. The evolutionary history of methicillin-resistant *Staphylococcus aureus* (MRSA). Proc Natl Acad Sci U S A. **99:**7687-7692.

5. **Enright, M. C., and B. G. Spratt.** 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology. **144:**3049-3060.
6. **Fernández, A., C. Cabellos, F. Tubau, J. Liñares, P. F. Viladrich, and F. Gudiol.** 2001. Relationship between penicillin and cephalosporin resistance of *Streptococcus pneumoniae* strains and its inflammatory activity in the experimental model of meningitis. Med Microbiol Immunol. **190:**135-138.
7. **Francisco, A. P., M. Bugalho, M. Ramirez, and J. A. Carriço.** 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics. **10:**152.
8. **Hakenbeck, R., T. Briese, L. Chalkley, H. Ellerbrok, R. Kalliokoski, C. Latorre, M. Leinonen, and C. Martin.** 1991. Antigenic variation of penicillin-binding proteins from penicillin-resistant clinical strains of *Streptococcus pneumoniae.* J Infect Dis. **164:**313-319.
9. **Hakenbeck, R., T. Briese, L. Chalkley, H. Ellerbrok, R. Kalliokoski, C. Latorre, M. Leinonen, and C. Martin.** 1991. Variability of penicillin-binding proteins from penicillin-sensitive *Streptococcus pneumoniae*. J Infect Dis. **164:**307-312.
10. **Hansman, D., and M. M. Bullen.** 1967. A resistant pneumococcus. Lancet **2:**264-265.
11. **Jacobs, M. R., H. J. Koornhof, R. M. Robins-Browne, C. M. Stevenson, Z. A. Vermaak, I. Freiman, B. Miller, M. A. Witcomb, M. Isaäcson, J. I. Ward, and R. Austrian.** 1978. Emergence of multiply resistant pneumococci. N Engl J Med. **299:**735-740.
12. **Jolley, K. A., and M. C. Maiden.** 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. **10:**595.
13. **McDougal, L. K., J. K. Rasheed, J. W. Biddle, and F. C. Tenover.** 1995. Identification of multiple clones of extended-spectrum cephalosporin-resistant *Streptococcus pneumoniae* isolates in the United States. Antimicrob Agents Chemother. **39:**2282-2288.
14. **McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefévre, A. Tomasz, and K. P. Klugman.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. J Clin Microbiol. **39:**2565-2571.
15. **Muñoz, R., C. G. Dowson, M. Daniels, T. J. Coffey, C. Martin, R. Hakenbeck, and B. G. Spratt.** 1992. Genetics of resistance to third-generation cephalosporins in clinical isolates of *Streptococcus pneumoniae*. Mol Microbiol. **6:**2461-2465.
16. **Pallarés, R., J. Liñares, M. Vadillo, C. Cabellos, F. Manresa, P. F. Viladrich, R. Martin, and F. Gudiol.** 1995. Resistance to penicillin and cephalosporin and mortality from severe pneumococcal pneumonia in Barcelona, Spain. N Engl J Med. **333:**474-480.

17. **Reinert, R. R., J. M. R, P. C. Appelbaum, S. Bajaksouzian, S. Cordeiro, M. van der Linden, and A. Al-Lahham.** 2005. Relationship between the original multiply resistant South African isolates of *Streptococcus pneumoniae* from 1977 to 1978 and contemporary international resistant clones. J Clin Microbiol. **43:**6035-6041.
18. **Staden, R., K. F. Beal, and J. K. Bonfield.** 2000. The Staden package, 1998. Methods Mol Biol. **132:**115-130.
19. **Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar.** 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. **28:**2731-2739.
20. **Viladrich, P. F., C. Cabellos, R. Pallarés, F. Tubau, J. Martinez-Lacasa, J. Liñares, and F. Gudiol.** 1996. High doses of cefotaxime in treatment of adult meningitis due to *Streptococcus pneumoniae* with decreased susceptibilities to broad spectrum cephalosporins. Antimicrob Agents Chemother. **40:**218-220.
21. **Viladrich, P. F., F. Gudiol, J. Liñares, R. Pallarés, I. Sabaté, G. Rufí, and J. Ariza.** 1991. Evaluation of vancomycin for therapy of adult pneumococcal meningitis. Antimicrob Agents Chemother. **35:**2467-2472.
22. **Zerbino, D. R., and E. Birney.** 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821-829.

3. PNEUMOCOCCAL HISTORY SINCE 1937: MULTILOCUS SEQUENCE TYPES AND PENICILLIN-BINDING PROTEIN GENES

This chapter describes serotype, penicillin susceptibility and MLST data for a collection of 426 pneumococcal isolates.

The isolates were provided by my supervisor and collaborators (listed in Chapter 2), as were the serotype data. I generated penicillin susceptibility data for 25 isolates and extracted DNA from 190 isolates. I also generated MLST data for 278 isolates. The remaining susceptibility data (n = 400 isolates – note that one isolate was of unknown penicillin susceptibility), DNA extractions (n = 236 isolates) and MLST data (n = 148 isolates) were provided by my supervisor and collaborators.

389 isolates were selected for PCR and sequencing of regions of three *pbp* genes. I generated a total of 1108 nucleotide sequences and a further 36 sequences were retrieved from the international Genbank database (www.ncbi.nlm.nih.gov/genbank/). RH provided 23 *pbp* sequences. (Note that some *pbp* sequences could not be obtained for technical reasons – see Appendix 2). I retrieved MLST and *pbp* sequence data for a further 16 isolates, for which genomic data was available in Genbank.

3.1 Abstract

Pneumococcal penicillin-nonsusceptibility was first detected in the 1960s, approximately 20 years after the introduction of penicillin. Resistance is mediated by acquisition of altered penicillin-binding protein genes (*pbps*) from closely-related species. In some regions of the world penicillin-nonsusceptible pneumococci (PNSP) now account for >50% of isolates recovered from patients or study subjects, the majority of which represent a small number of internationally disseminated clones. Here a unique set of pneumococcal isolates, dated 1937 – 2007, was studied with the aim to better understand pneumococcal molecular epidemiology and the development of penicillin-resistance.

Multilocus sequence typing and penicillin susceptibility data were collected for 426 isolates including the globally distributed Pneumococcal Molecular Epidemiology Network (PMEN) strains. Isolates were assigned to clonal complexes (CCs) by goeBURST. Regions of the *pbp2x, pbp1a and pbp2b* genes were sequenced. Data from 16 publicly available genomes were also used in analyses. The relationship between *pbp* nucleotide sequence divergence (as compared to the R6 penicillin-susceptible reference strain) and penicillin minimum inhibitory concentration (MIC) was tested statistically.

232 unique sequence types and 163 CCs were represented. The earliest relatives of 9 of the penicillin-susceptible PMEN clones were recovered prior to 1960. The four oldest PNSP (1967-1976) represented genotypes which were not associated with any PMEN clones, but just over half (55/100) of PNSP isolated between 1977 and 1989 did represent PMEN clones. Amongst penicillin-susceptible isolates there was

evidence of *pbp* evolution by both mutation and intra-species recombination, even prior to the 1960s. Amongst PNSP, *pbp* nucleotide sequence divergence was significantly correlated to penicillin MIC ($p < 2.2e^{-16}$) but only accounted for ~62% of the variation in the data (R^2 adj = 0.6219).

It was concluded that representatives of several PMEN clones have been circulating amongst pneumococcal populations for several decades and that pneumococcal *pbp* recombination has long been an inherent part of pneumococcal biology. Additionally, *pbp* nucleotide divergence is an important factor determining penicillin-resistance but not the sole factor.

3.2 Introduction

Penicillin was introduced in the 1940s and remains the first line therapy against pneumococcal disease. Recently however, pneumococcal penicillin resistance has become a concern. Resistance is primarily conferred by alterations of three of the six pneumococcal penicillin-binding proteins (PBPs) [19]. *pbp* nucleotide sequences of penicillin-nonsusceptible pneumococci (PNSP) differ from those of penicillin-susceptible pneumococci (PSP) by up to 25% divergence [19, 29]. These nucleotide sequences are thought to have been acquired from the closely related viridans streptococci through the process of genetic transformation [6, 12, 45]. In general, increasing *pbp* sequence divergence is associated with increasing resistance levels [17] as defined by penicillin minimum inhibitory concentrations (MICs).

The first reports of penicillin-nonsusceptible viridans streptococci were in the late 1940s [28] but PNSP were not reported until the late 1960s [20, 27]. The following 10

years saw sporadic reports of penicillin-intermediate resistant pneumococci (0.12 μg/ml ≤ MIC ≤ 1 μg/ml) from Australia/Oceania [21, 22], Europe [24, 38] and North America [11, 30, 34, 39, 40, 48]. In the late 1970s the first reports of high-level penicillin-resistant pneumococci (MICs ≥2 μg/ml) came from South Africa [2, 26]. The 1980s and 90s saw a rapid increase of PNSP levels to greater than 50% in some regions [14, 32]. By this time oral penicillin, which was first introduced in the mid-1950s, was readily available in many countries, and presumably widespread use of this drug imposed a selective pressure favouring penicillin-resistant pneumococci. Indeed, the force of antibiotic-induced selective pressure is evidenced by reports of significant positive associations between penicillin consumption and pneumococcal resistance rates [23, 43].

By the late 1990s it was clear that the majority of invasive disease-causing PNSP represented a small number of internationally disseminated clones [8, 25, 31, 41, 44]. These and other globally important disease causing clones have now been defined by the Pneumococcal Molecular Epidemiology Network (PMEN) [37]. In total 43 clones (detailed at http://www.sph.emory.edu/PMEN/) were characterised on the basis of genotype (determined by pulsed-field gel electrophoresis and multilocus sequence typing (MLST)), antibiotic susceptibility and *pbp*-fingerprinting. A recent study of genomic evolution between members of the first PMEN clone, 'Spain^{23F}-1' or 'PMEN1', indicated that this clone first emerged in Europe, circa 1970 [9]. However, the age of other major clones circulating within pneumococcal populations is unclear. Similarly, the reasons for the delay in development of pneumococcal penicillin-nonsusceptibility

compared to that of viridans streptococci, and the delay in the wide-scale spread of resistance are uncertain.

Within this study a unique collection of historical and modern pneumococci were characterised with the aim to better understand pneumococcal molecular epidemiology and the evolution of pneumococcal penicillin resistance. MLST and nucleotide sequence data for the three primary resistance-conferring *pbp* genes (*pbp2x, pbp1a* and *pbp2b*) were collected. These data provide the first insight into the genotypes of historical pneumococci compared to those of modern isolates, and of the diversity of their *pbp* genes. For the first time a study of *pbp* gene evolution prior to the development of pneumococcal resistance and temporally within penicillin-susceptible lineages has been possible. Finally, *pbp* sequence data from PNSP were used to statistically test the association between *pbp* nucleotide divergence and penicillin MIC.

3.3 Methods

Pneumococcal isolates included in this study are described in Chapter 2, as are protocols for penicillin susceptibility testing, MLST/clonal complex (CC) designation, *pbp* nucleotide sequencing and *pbp* sequence analysis.

Statistical analyses

Regression analyses were performed using the R software package (downloaded at http://www.r-project.org/). Data for pneumococcal isolates for which all *pbp* nucleotide sequences were ≤1% divergent from the R6 reference strain were excluded

(n = 239) to reduce the left-skew of the data distribution and because it was not possible to know which of these sequences most closely represented the true ancestral sequence (all of these isolates were PSP). Of those remaining, only isolates for which a complete data set (*pbp2x, pbp1a* and *pbp2b* nucleotide sequences plus penicillin MIC) was available were included in the analyses (n = 144 of 166). Log (penicillin MIC) was fitted as the response variable. *pbp2x, pbp1a* and *pbp2b* nucleotide divergence were fitted as explanatory variables. The optimum model was determined following step-wise addition of explanatory variables and their interaction terms. The threshold for statistical significance was set at $p < 0.05$.

3.4 Results

Genotypes among the historical isolate collection

232 unique multilocus sequence types (STs) were represented by isolates in this study. (Note that after repeated attempts using multiple primer sets the *ddl* sequence for isolate SA82 (1989, CC124, serotype 14) could not be determined, presumably due to sequence divergence within the primer-binding region.) 100 STs, representing 115 isolates, were newly recognised in this study. 3 *aroE*, 4 *gdh,* 4 *gki,* 6 *spi,* 6 *xpt* and 5 *ddl* alleles were newly identified.

Overall, 163 unique CCs were represented, 17 of which were represented by >5 isolates (Table 3.1). Among these 17 CCs, 13 were associated with PMEN clones. Five CCs (CCs 62, 113, 124, 191 and 218) were associated only with penicillin-susceptible PMEN clones and six CCs (CCs 66, 81, 156/162, 176, 439 and None41) were associated

only with penicillin-nonsusceptible clones. Two CCs (CCs 15 and 90) were associated with both penicillin-susceptible and penicillin-nonsusceptible PMEN clones.

Penicillin-susceptible representatives of three of the six penicillin-nonsusceptible PMEN CCs mentioned above were identified within this isolate collection (CCs 66, 156/162 and 439). In all three cases penicillin-susceptible members of these CCs had also been recognised previously (www.mlst.net). CC124 is associated with a penicillin-susceptible PMEN clone and representatives of this CC are predominantly (though not exclusively) penicillin-susceptible (www.mlst.net). Among this collection there were three CC124 representatives which were penicillin-nonsusceptible. These pneumococci were isolated in South Africa in 1978 (ST published previously [42]), 1985 and 1989, and had penicillin MICS of 0.25 µg/ml, 0.125 µg/ml and 0.25 µg/ml respectively.

The earliest representatives of 7 out of 17 PMEN-associated CCs mentioned above were dated prior to 1960, with two of these dated as early as 1937 (CC191) and 1939 (CC113). Additionally, CC306 (which is associated with the penicillin-susceptible PMEN28 clone), was represented by three isolates dated 1943, 1943 and 1944. The PMEN29 reference strain was dated 1948 and another representative of the same CC (CC615) was dated 1956. The earliest representatives of 3 out of 17 PMEN-associated CCs listed in Table 3.1, and one other PMEN-associated CC (CC199, represented by PMEN37) were dated between 1960 and 1980. These data indicate that many of the PMEN-lineages have been established within the pneumococcal population for several decades. The population structure of 12 out of 17 PMEN-associated lineages (Figure

3.1) indicated a high level of clonal expansion consistent with the global dissemination and subsequent evolution of these clones. Numerous single-locus variants of the group founder STs were identified, and one or more double-locus variants were identified in most cases. In contrast, CC None41 comprised just two STs.

Four CCs, represented by >5 isolates, were not associated with PMEN clones, and accordingly the structures of three of these (Figure 3.1) indicated a lower level of clonal expansion. CCs 447 and 3122 were represented only by penicillin-susceptible isolates. The six CC447 representatives were dated across six decades and originated from Denmark (n = 4) and the USA (n = 2), but the CC3122 representatives were all isolated from Denmark within a five year time period. In fact, all of the 12 CC3122, serotype 12A representatives were isolated in 1966 and likely represent a sampling bias within this collection. CC1094 was represented only by penicillin-nonsusceptible isolates, all of which were isolated in South Africa. The international MLST database included an additional 15 CC1094 representatives dated from 1988 to 2007, all of which were isolated from either South Africa or the neighbouring Mozambique.

Table 3.1 CCs represented by >5 isolates in this collection.

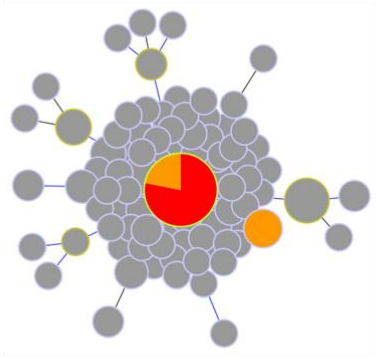
CC	Associated PMEN clone(s)†	Isolate Frequency	Years (n = Unknown)	Serotype(s) (n)	Penicillin Susceptibility†			
					n = S	n = I	n = R	n = Unknown$^{\Psi}$
81	PMEN1^{I}	28	1984 - 1999	23F (27), 19F (1)	-	10	18	-
1094	None	23	1978 - 1998	6A (19), 6B (3), 19A (1)	-	5	18	-
124	PMEN35^{S}	21	1952 - 2005	14 (19), 9L (1), 11C (1)	18	3	-	-
15	PMEN9^{S}, PMEN10^{R}	15	1967 - 2004/5	14 (15)	8	3	4	-
113	PMEN36^{S}	15	1939 - 2005	18C (10), 35C (2), 9V (1), 17F (1), 18B (1)	15	-	-	-
None41	PMEN13^{I}	15	1978 - 1998	19A (15)	-	3	12	-
66	PMEN18^{R}	14	1952 - 2005 (1)	9N (5), 14 (5), 19F (2), 7B (1), 23F (1)	10	3	1	-
3122	None	13	1962 - 1966	12A (12), 20 (1)	13	-	-	-
156/162	PMEN3^{I}	12	1962 - 1994	9V (9), 14 (2), 9A (1)	3	5	4	-
191	PMEN39^{S}	12	1937 - 2005	7F (11), 7A (1)	12	-	-	-
218	PMEN34^{S}	11	1952 - 2003/4	12F (6), 7F (5)	11	-	-	1*
90	PMEN2^{I}, PMEN12^{R}, PMEN22^{S}	10	1981 - 1995	6B (10)	1	6	2	1**
176	PMEN6^{I}, PMEN16^{R}, PMEN26^{I}	8	1978 - 1998	19A (6), 23F (2)	-	7	1	-
439	PMEN4^{I}	7	1945 - 1996	23F (6), 23A (1)	4	2	1	-
62	PMEN33^{S}	6	1952 - 1989 (1)	8 (3), 6B (1), 11A (1), 11D (1)	5	-	-	1*
447	None	6	1943 - 1996	37 (6)	6	-	-	-
490	None	6	1939 - 2007	6A (1), 6B (1), 6C (1), 10F (1), 18F (1), 22A (1)	5	1	-	-

† Penicillin susceptibilities defined as: S = Susceptible, MIC ≤ 0.06 µg/ml, I = Intermediate-resistant, MIC 0.12 - 1 µg/ml, R = Resistant, MIC ≥ 2.0 µg/ml

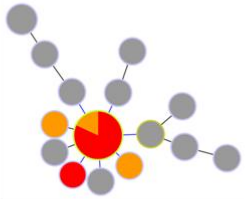
$^{\Psi}$ Penicillin susceptibility information was unavailable for three isolates for which genomic data were retrieved from Genbank.

* *pbp* nucleotide sequences indicate that this is a PSP. ** *pbp* nucleotide sequences indicate that this is a PNSP.

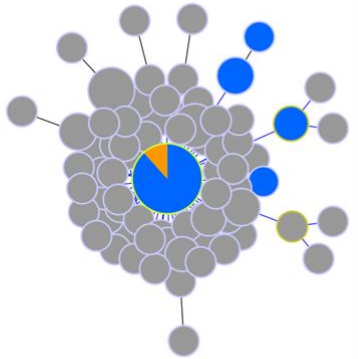
3.1A



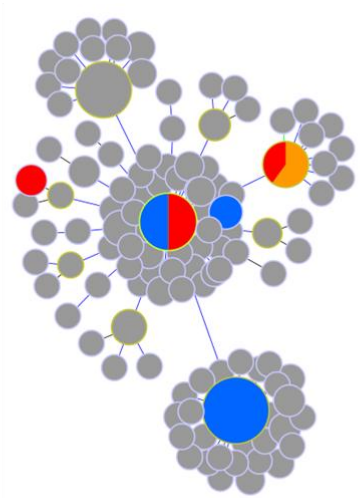
CC81
PMEN1



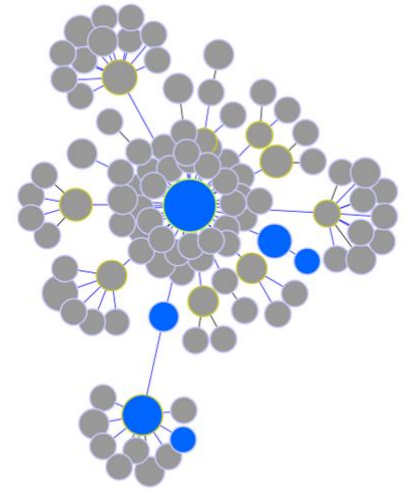
CC1094



CC124
PMEN35

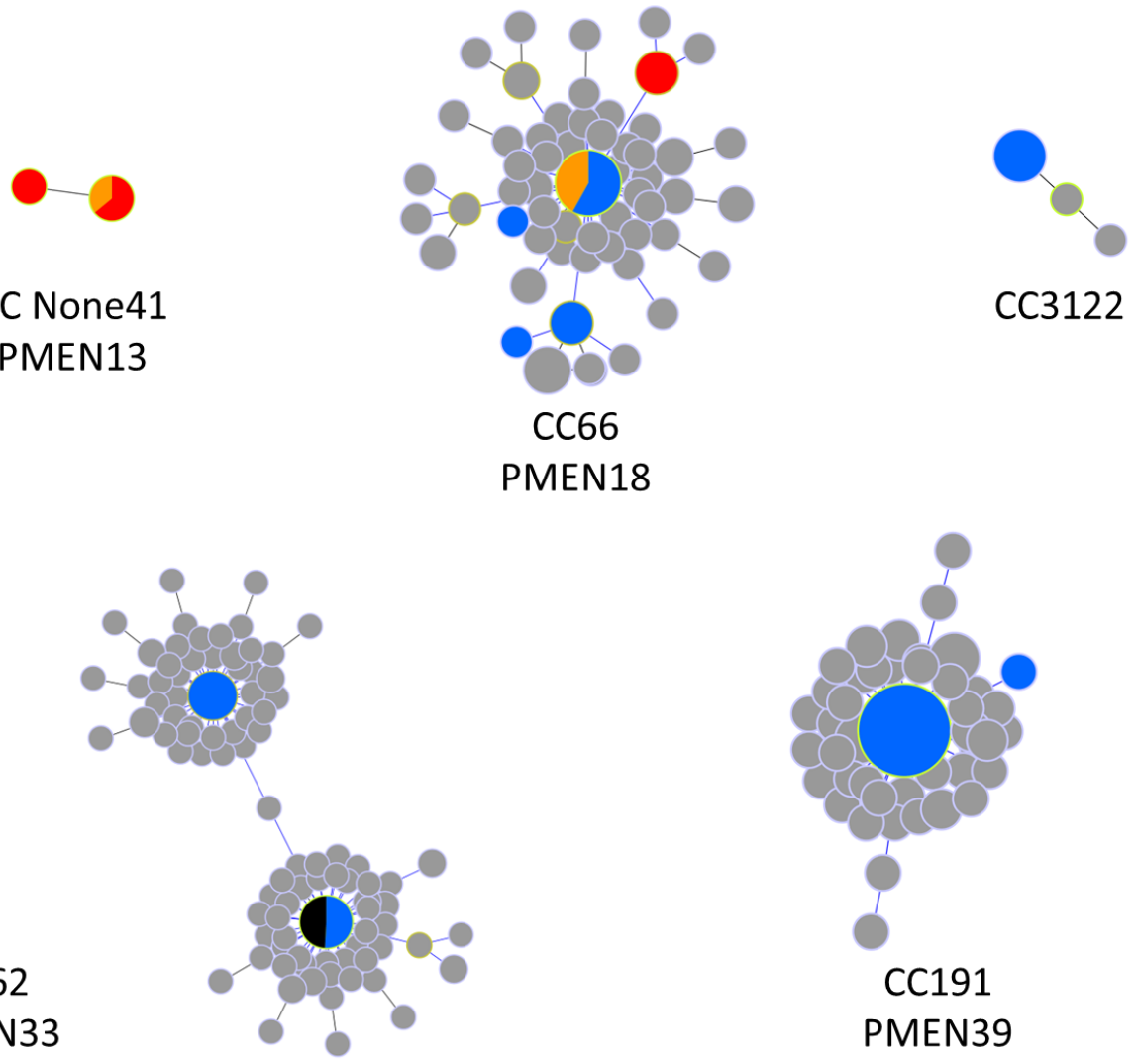


CC15
PMEN9,
PMEN10

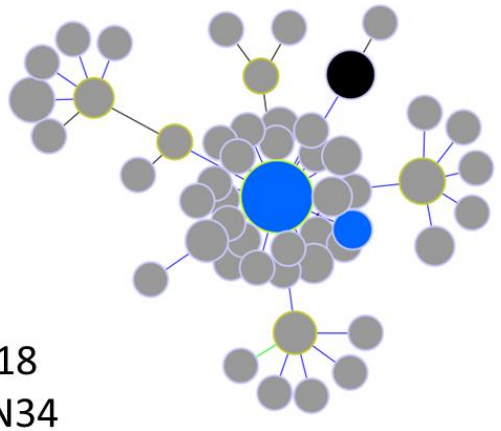


CC113
PMEN36

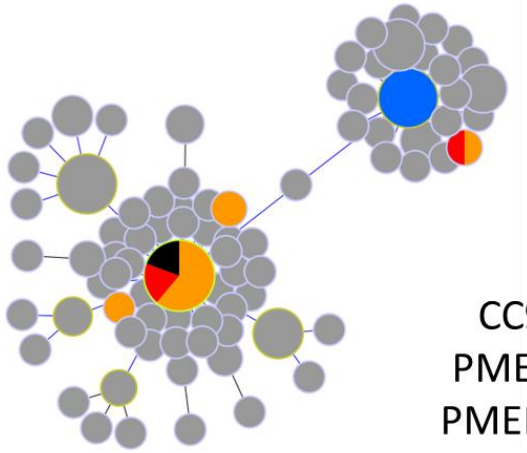
3.1B


CC None41
PMEN13
CC66
PMEN18
CC3122
CC62
PMEN33
CC191
PMEN39

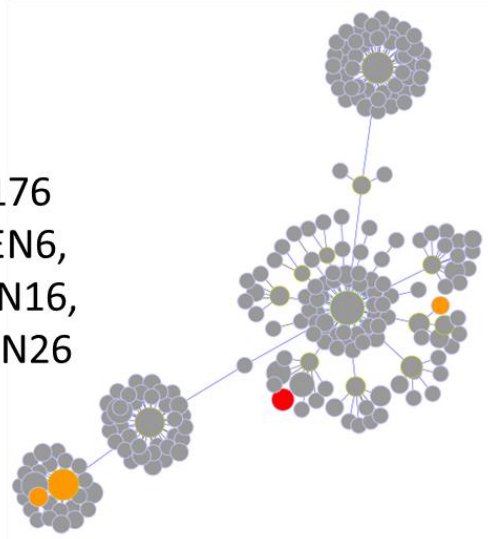
3.1C



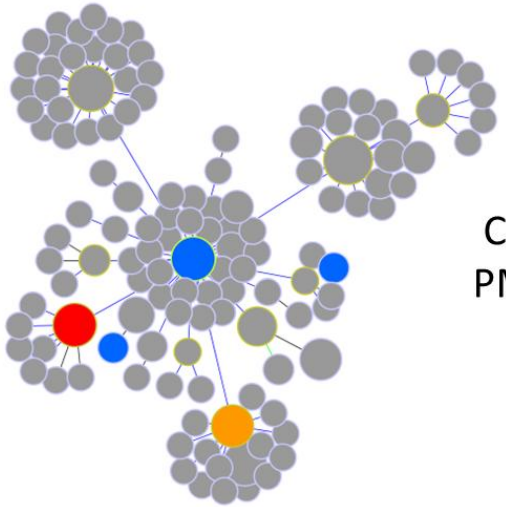
CC218
PMEN34



CC90
PMEN2,
PMEN12,
PMEN22

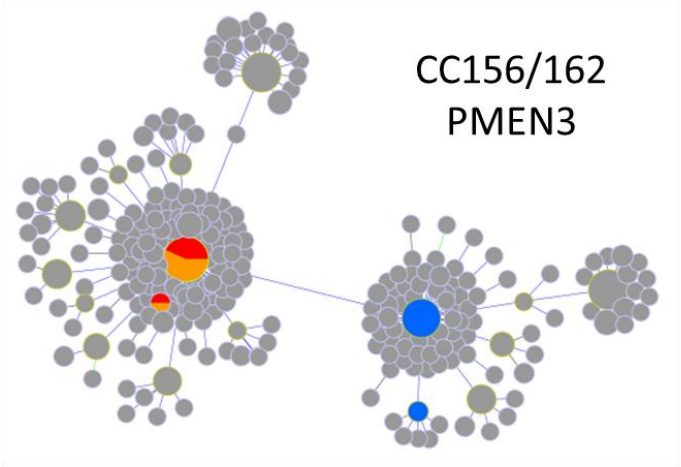


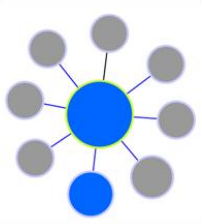
CC176
PMEN6,
PMEN16,
PMEN26



CC439
PMEN4

3.1D





CC447

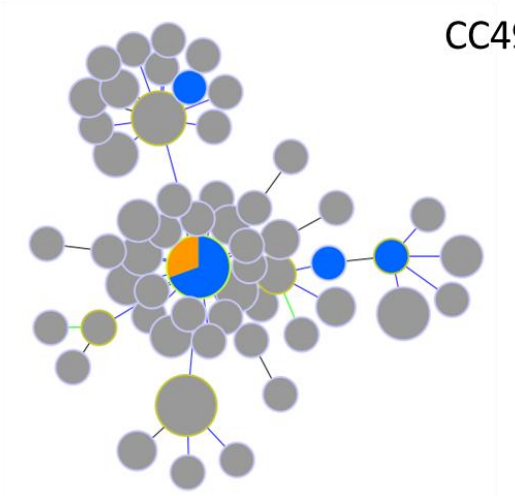


Figure 3.1A – 3.1D (previous four pages) CC structures for CCs represented by >5 isolates in the historical collection.
Structures determined by goeBURST analysis using STs represented within this collection and the international MLST database. Nodes represent STs, edges connect STs which differ at one MLST locus only. Blue, orange and red shading indicate STs represented by penicillin-susceptible, penicillin intermediate-resistant and penicillin-resistant pneumococci in this collection, respectively. Black shading indicates STs represented by isolates of unknown penicillin-susceptibility in this collection. Grey shading indicates STs represented by isolates deposited in the international MLST database but not in this collection.

CC490 was represented by both penicillin-susceptible and penicillin-nonsusceptible pneumococci, isolated across a wide date range and originating from the USA (n = 4), Denmark (n = 2) and Iceland (n = 1). Interestingly, each of the 6 CC490 representatives was associated with a different serotype, perhaps suggesting a high degree of capsule locus variability within this CC (although it should be noted that serotypes 6A, 6B and 6C are encoded by highly structurally similar capsule loci [4]).

The CCs represented by early PNSP (dated prior to 1990) are depicted in Figure 3.2 (note that 8 penicillin-nonsusceptible PMEN reference strains were also dated prior to 1990 but were not included here). Among these pneumococci >50% (55/104) represented CCs associated with PMEN clones. A further ~22% (23/104) were members of CC1094 (mentioned above). The two oldest PNSP (dated 1967 from Australia and 1969 from Papua New Guinea respectively) were not related to any modern pneumococcal clones and represented novel STs. The third and fourth oldest PNSP (dated 1972 from Papua New Guinea and 1976 from Denmark, respectively) each represented a new ST and were assigned to CC277 and CC554 respectively. (It is noteworthy that the 1976 pneumococcus was isolated from a child who was likely of Asian descent. Given that this would have been quite unusual for a Danish citizen at the time, it is likely that the child may have been visiting Denmark from Asia.) CC277

was also represented by two other PNSP in this collection (both dated 1986 from Germany) and a further 59 isolates in the international MLST database. These isolates originated from Africa, Asia, Australia/Oceania, North and South America, Europe and the Middle East. CC554 was not represented by any other isolates in this collection but was represented by 18 isolates in the MLST database, all of which originated from either Europe or Asia.

pbp2x, pbp1a and pbp2b nucleotide sequences

Among the 389 isolates selected for *pbp* sequencing (see Chapter 2) plus the 16 genomes, 401, 389 and 404 *pbp2x, pbp1a* and *pbp2b* sequences, respectively, were obtained (a full complement of sequences could not be obtained for all isolates due to PCR and/or sequencing difficulties most likely arising from sequence divergence in the primer binding regions – see Appendix 2 for more details). Among these sequences 142, 105 and 127 unique *pbp2x, pbp1a* and *pbp2b* alleles were identified respectively, and of these 64 (45.1%), 34 (32.4%) and 59 (46.5%) were divergent (i.e. >1% divergent from the penicillin-susceptible R6 reference strain dated 1964). There was no evidence of PNSP or divergent *pbps* prior to 1967.

pbp nucleotide sequences from 233 of 237 (98.3%) PSP showed ≥99% sequence identity to those of R6. *pbp2x* nucleotide sequences from four PSP showed between 90.4 and 96.3% identity to those of R6. Three of these isolates were PMEN reference strains (PMEN21, PMEN30 and PMEN40) and the fourth was a CC66 representative (USA11).

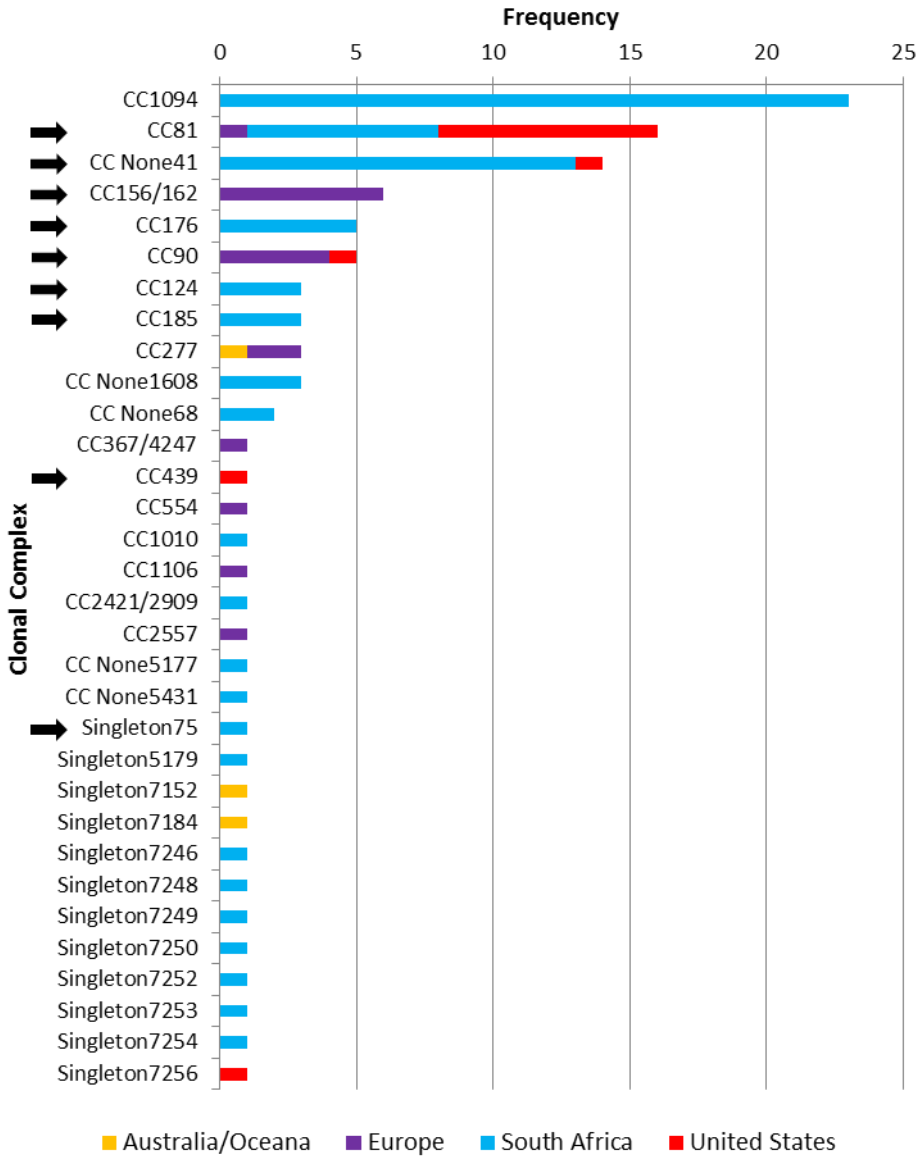


Figure 3.2 CCs represented by early penicillin-nonsusceptible pneumococci (isolated prior to 1990) in the historical collection.
Isolates are stratified by geographical region of origin. CCs were designated by a modified goeBURST method based on MLST data. CCs associated with the internationally recognised PMEN clones are indicated by arrows. PMEN reference strains, for which genotype data have been published previously, are not included.

Comparison of the *pbp* nucleotide sequences of PSP and PNSP representing the same CC supported the hypothesis that *in vivo* development of pneumococcal penicillin-nonsusceptibility results from *pbp* gene recombination rather than mutation. Among CCs for which both historical and modern PSP were identified there was evidence of

the capacity for conservation of *pbp* sequence over many years (e.g. CC191, Figure 3.3) but there was also evidence of *pbp* evolution by both mutation and recombination.

Within CC124 (Figure 3.4) the *pbp2b* sequences of all PSP dated after 1978 had three conserved nucleotide substitutions compared to all PSP dated prior to this time. A single nucleotide substitution was identified within the *pbp2x* sequence of USA6 (dated 1999) and an independent single nucleotide substitution within the same gene was identified in USA7 (also dated 1999). The latter of these *pbp2x* substitutions was also shared by three additional CC124 representatives dated 2001, 2002 and 2002, one of which also had an additional nucleotide substitution in its *pbp2b* gene. Such progression of nucleotide substitutions is suggestive of evolution through point mutation. In contrast, the *pbp2x* sequence of isolate 14/4 contained nucleotide substitutions which were unique to this isolate but also several which were shared with either or both the 9L/2 and 11C/1 isolates. Such a combination of substitutions does not indicate a parsimonious evolutionary route by mutation and so it seems more likely that these substitutions were acquired through one or more recombination events. Among CC218 PSP (Figure 3.3) there was also evidence of *pbp* evolution by recombination. The *pbp2x* and *pbp1a* alleles of CC218, serotype 7F representatives were distinct from those of CC218, serotype 12F representatives but identical to those of CC191, serotype 7F representatives. Independent evolution of identical alleles within two seemingly unrelated lineages is highly unlikely and therefore, change by recombination is the most probable explanation. (This particular recombination event will be explored further in Chapter 5.)

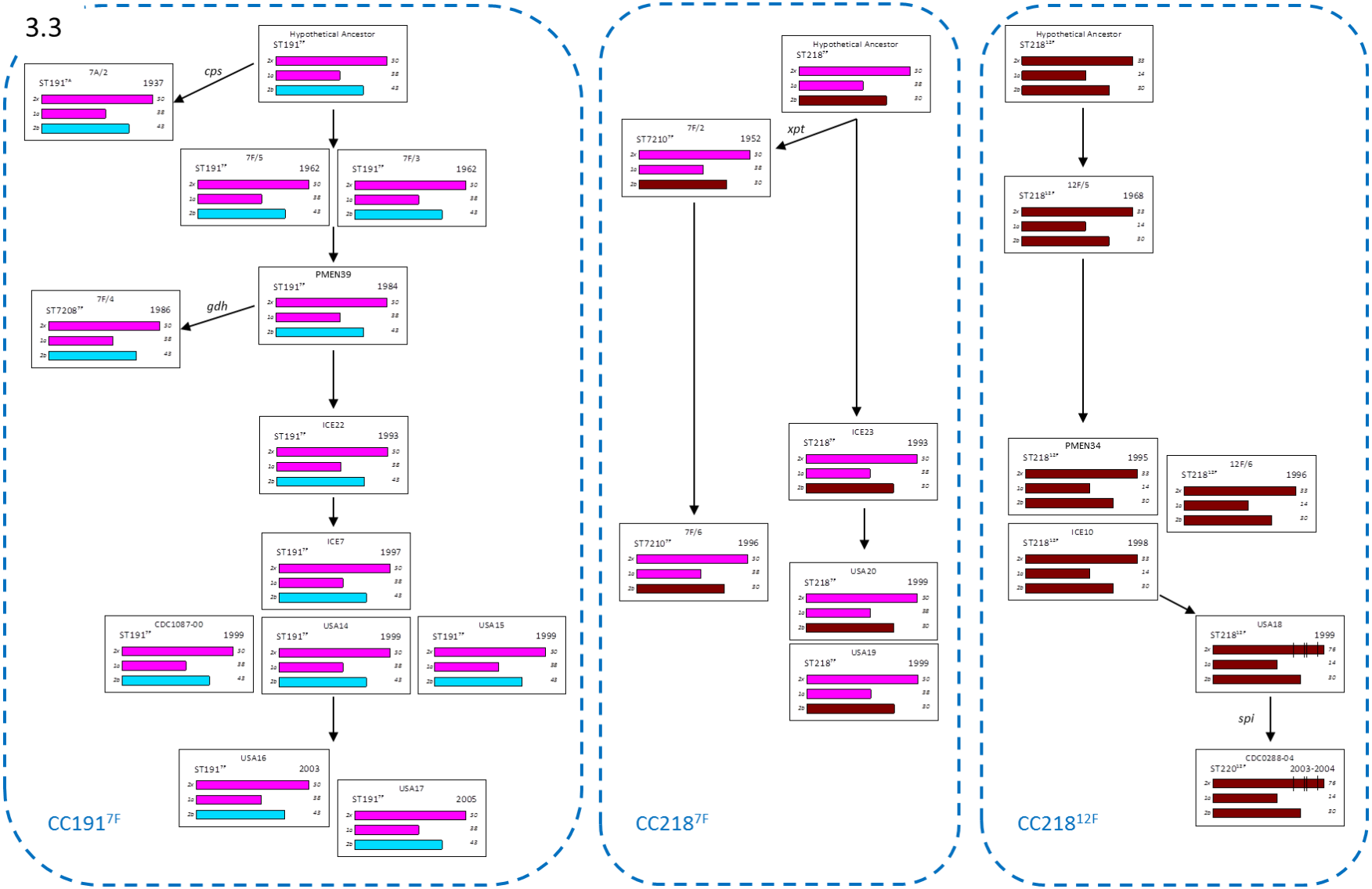
In total 19, 21 and 22 non-divergent *pbp2x, pbp1a* and *pbp2b* alleles respectively were identified among isolates representing >1 CC. Although it is not possible to exclude the possibility that one or more of these alleles were shared between CCs through shared ancestral descent, the data indicate that non-divergent *pbp* alleles are readily transferred between pneumococci through recombination. Interestingly, 9, 14 and 11 of these *pbp2x, pbp1a* and *pbp2b* alleles respectively were identified among isolates representing >1 CC *and* dated prior to 1960.

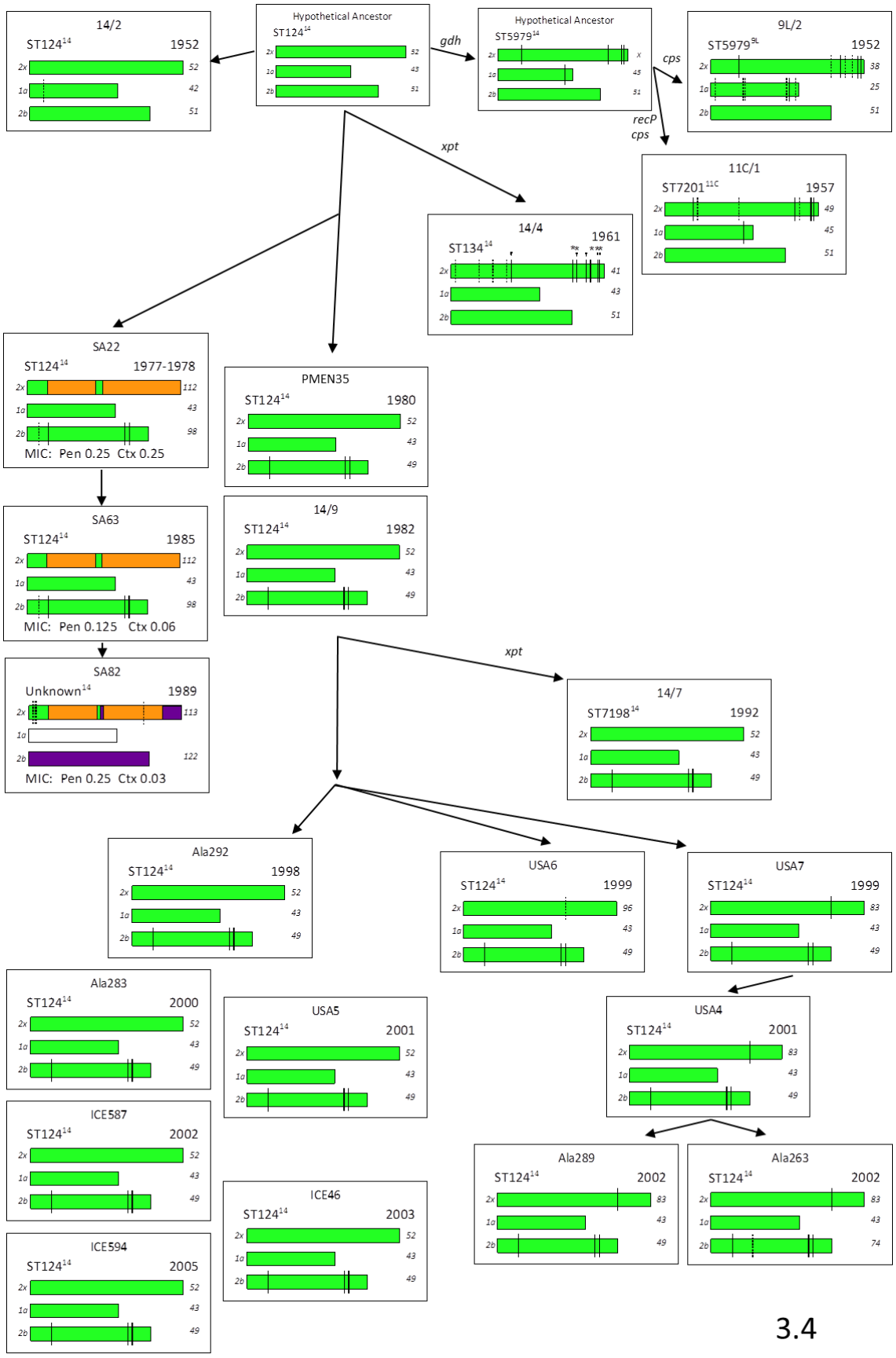
Statistical analyses

Regression analyses indicated that *pbp2x, pbp1a* and *pbp2b* nucleotide divergences were independently each significantly correlated with log(penicillin MIC). (For *pbp2x,* F = 108.8 on 1 and 142 df, $p < 2.2e^{-16}$. For *pbp1a,* F = 161.9 on 1 and 142 df, $p < 2.2e^{-16}$. For *pbp2b,* F = 21.55 on 1 and 142 df, $p = 7.753e^{-06}$.) However, only *pbp1a, pbp2x* and their interaction term were significant in the optimum model (F = 79.39 on 3 and 139 df, *p < 2.2e-16,* R^2 adj = 0.6219).

Figure 3.3 (overleaf) *pbp* nucleotide patterns of CCs 191 and 218.
Each box represents an independent pneumococcal isolate belonging to CC191 (left side) or CC218 (centre and right). Isolate names, ST and year of isolation are listed at the top of the box. All isolates are penicillin and cefotaxime susceptible. Coloured bars represent *pbp* nucleotide sequence mosaic patterns for each of the *pbp2x, pbp1a* and *pbp2b* genes; bar length is proportional to the length of sequenced region. For any given *pbp*, colours represent identical regions shared between multiple strains. Vertical black lines indicate the positions of nucleotide substitutions between otherwise identical sequence blocks. Numbers are *pbp* allele numbers. Arrows indicate proposed lineages of descent and labels are MLST loci/serotypes differing between strains. (Note - proposed lineages of descent are based upon information provided by this data set only.)

3.3


Hypothetical Ancestor
ST191
7A/2
ST191
1937
cps
7F/5
ST191
1962
7F/3
ST191
1962
PMEN39
ST191
1984
7F/4
ST7208
1986
gdh
ICE22
ST191
1993
ICE7
ST191
1997
CDC1087-00
ST191
1999
USA14
ST191
1999
USA15
ST191
1999
USA16
ST191
2003
USA17
ST191
2005
CC191 7F
Hypothetical Ancestor
ST218
7F/2
ST7210
1952
xpt
7F/6
ST7210
1996
ICE23
ST218
1993
USA20
ST218
1999
USA19
ST218
1999
CC218 7F
Hypothetical Ancestor
ST218
12F/5
ST218
1968
PMEN34
ST218
1995
12F/6
ST218
1996
ICE10
ST218
1998
USA18
ST218
1999
spi
CDC0288-04
ST220
2003-2004
CC218 12F


14/2
ST124 14
1952
Hypothetical Ancestor
ST124 14
gdh
Hypothetical Ancestor
ST5979 14
cps
9L/2
ST5979 9L
1952
recP
cps
11C/1
ST7201 11C
1957
xpt
14/4
ST134 14
1961
SA22
ST124 14
1977-1978
MIC: Pen 0.25 Ctx 0.25
PMEN35
ST124 14
1980
SA63
ST124 14
1985
MIC: Pen 0.125 Ctx 0.06
14/9
ST124 14
1982
SA82
Unknown 14
1989
MIC: Pen 0.25 Ctx 0.03
xpt
14/7
ST7198 14
1992
Ala292
ST124 14
1998
USA6
ST124 14
1999
USA7
ST124 14
1999
Ala283
ST124 14
2000
USA5
ST124 14
2001
USA4
ST124 14
2001
ICE587
ST124 14
2002
Ala289
ST124 14
2002
Ala263
ST124 14
2002
ICE46
ST124 14
2003
ICE594
ST124 14
2005

3.4

Figure 3.4 (previous page) *pbp* nucleotide patterns of CC124.
As Figure 3.3, except: 1) Each box represents an independent pneumococcal isolate belonging to CC124. 2) All isolates are penicillin and cefotaxime susceptible unless penicillin (Pen) and cefotaxime (Cef) MICs are indicated (units are μg/ml). 3) Vertical black lines indicate the positions of shared nucleotide substitutions found within a subset of isolates. 4) Vertical dashed black lines indicate the positions of nucleotide substitutions found in single isolates only. 5) The *pbp2x* sequence of 14/4 is marked with numerous nucleotide substitutions, some of which were shared with 9L/2 (asterisk) or 11C/1 (arrow head). The combination of shared substitutions does not indicate a parsimonious evolutionary route by mutation.

3.5 Discussion

Within this study genotype and *pbp* nucleotide sequence data were generated for a unique collection of PSP and PNSP dating from 1937 to 2007. This is likely the largest available collection of historical pneumococci and the largest number of early-PNSP ever included in a single study, providing a unique opportunity to explore pneumococcal evolution. Given that the majority of this sample was comprised solely of those isolates which were still available (and viable), it cannot be considered as a complete representation of historical pneumococcal populations. Nevertheless, the results presented here provide an interesting and important insight into pneumococcal history.

Identification of 163 independent CCs suggested that a large amount of genotypic diversity was represented. Approximately 43% (100/232) of the identified STs had not been previously recognised. Most of the isolates representing new STs were dated prior to 1980 (92/115, 80%). Since the majority of isolates previously genotyped by MLST dated from 1990 onwards, many of the newly identified STs likely represent historical genotypes which are rare or absent from modern pneumococcal populations. It is not possible to know whether these genotypes were also rare among historical

populations or if they were once prevalent and have subsequently been replaced by alternative genotypes, either through stochastic changes or competitive exclusion. A combination of these effects is the most probable explanation.

The data revealed that several PMEN clones have been established within pneumococcal populations for many decades. Notably, the oldest isolates related to the PMEN36 and PMEN39 clones were dated 1939 and 1937 respectively, and the oldest isolate related to the PMEN4 clone was dated 1945. As mentioned above, it is not possible to know the prevalence of these clones amongst older pneumococcal populations, but their continued presence through time is clear. Pneumococcal populations are not static [3, 7] and so the fact that these lineages have survived whilst others have not supports the hypothesis that the PMEN clones have a selective advantage over other pneumococci. Such selective advantages presumably also contributed to the global spread of these clones.

The majority of *pbp* sequences from PSP were ≤1% different from those of the penicillin-susceptible R6 reference strain, which is consistent with the findings of previous studies [13, 18]. The *pbp2x* nucleotide sequences of four PSP were >1% different from those of R6. Possession of divergent *pbp* sequences without increased penicillin MIC has not been noted previously for pneumococci. One possible explanation for this is that the penicillin MIC was not increased due to the lack of altered *pbp1a* and/or *pbp2b* sequences and/or favourable alleles at other loci believed to play a role in penicillin-resistance (see below). Alternatively, *pbp2x* sequence divergence may have followed acquisition and recombination of DNA from penicillin-

susceptible viridans streptococci, resulting in sequence mosaics lacking the appropriate nucleotide changes to confer resistance. The *pbp2x* sequence of one of these PSP (PMEN40) was identical to that of a PNSP (PMEN32), but the *pbp1a* and *pbp2b* sequences were different. Thus, at least in this case it seems that the former hypothesis is more likely.

Despite the lack of *pbp* nucleotide divergence among PSP, there was a considerable amount of *pbp* allelic diversity. Comparison of the *pbp* alleles of clonally-related PSP showed that in some cases these genes have been highly conserved through time; however, there was also evidence of *pbp* gene evolution through both mutation and recombination, even between pneumococci isolated prior to the widespread use of oral penicillins. These results indicate that recombination of *pbp* genes is an inherent aspect of pneumococcal evolution. However, there was no evidence of the acquisition of divergent *pbp* sequences prior to 1967, despite the fact that such sequences must have been present among viridans streptococci at this time. (*pbp* nucleotide sequences of both penicillin-susceptible and –resistant viridans streptococci are highly divergent from those of PSP [12, 45].) Combined with the knowledge that penicillin-nonsusceptible pneumococci were not detected until the 1960s, these data imply that, prior to the 1960s, some sort of mechanistic or selective barrier prevented pneumococcal acquisition, or at least detectable acquisition, of divergent *pbp* sequences.

Perhaps the recombination mechanisms among older pneumococci were such that divergent *pbp* sequences could not be successfully incorporated in the chromosome?

The RecA protein and HexAB mis-match repair system are known to act in such a way that the relative frequency of recombination is reduced with decreasing donor nucleotide sequence similarity (compared to the recipient genome) [35b, 43a]. RecA mediates the search for homologous donor and recipient sequence regions prior to heteroduplex formation. The extent to which homologous sequences are recognised is dependent upon the number of mis-matches between the sequences [43a]. Presumably, heteroduplex formation and subsequent recombination are prevented when mis-match frequencies are sufficiently high.

HexA has been shown to recognise mis-matched DNA during heteroduplex formation. Single base mis-matches are corrected by the HexAB system but largely mis-matched sequences may be removed, thus preventing recombination [7a]. However, the extent to which recombination is restricted by HexAB (or its homologs) is known to vary [35a, 40a]. Perhaps then the restriction of recombination due to the HexAB and/or RecA variants of older pneumococci was greater than that of more modern isolate variants, and was sufficient to prevent acquisition of divergent *pbp* genes.

An alternative, and more likely, explanation for the apparent lack of divergent *pbp* acquisition prior to the 1960s, is that such recombinants suffered a fitness cost and were subsequently not able to survive. Indeed studies have shown that, in the absence of antibiotics, PSP transformed with divergent *pbp* genes suffered a competitive fitness cost relative to wild-type strains [1, 50]. Such a fitness cost may be partially alleviated by the development of compensatory mechanisms, similar to those

described for the compensation of structural protein changes among other bacterial species [35, 46].

PBPs are implicated in the creation of cell wall peptidoglycan cross-links, thus it is possible that other proteins involved with the construction of, or interacting with, the cell wall might compensate for the effects of divergent *pbp* gene expression. As such, several non-*pbp* loci have also been implicated in the expression of pneumococcal penicillin-nonsusceptibility: Cell walls of PNSP include a greater proportion of branched muropeptides compared to the cell walls of PSP, but deletion of the *murMN/fibAB* operon results in a reduction of muropeptide branching [15, 51]. Concordantly, expression of *murMN/fibAB* was shown to be required for maintenance of elevated penicillin MIC [16, 51]. Expression of a second operon, *ciaRH,* encoding a two-component signal transduction system, was shown to be required to maintain normal growth characteristics in the presence of divergent *pbp2x* genes [36]. Furthermore, the heat shock induced ClpL chaperone protein was shown to interact with PBP2x, and changes in the expression levels of the *clpL* gene lead to changes in cell wall thickness and penicillin MIC [49]. Thus it is possible that the early development of pneumococcal penicillin-nonsusceptibility was hampered by the requirement of sufficient time for cells which acquired divergent *pbp* genes to also acquire appropriate compensatory mechanisms.

Another important factor in determining the timing of the development of penicillin-nonsusceptibility, and its subsequent spread, is the strength of the selective pressure imposed by antibiotic use. Resistant cells can only persist within a population if the

selective force imposed by antibiotic use is greater than the cost of resistance [33]. The selective pressure associated with lower antibiotic consumption prior to the 1960s may not have been strong enough to overcome these costs, meaning that emergent PNSP were outcompeted by PSP. Similarly, lower levels of antibiotic consumption throughout the 1960s and 1970s compared to that of later years likely contributed to the delay in the spread of penicillin-nonsusceptibility. PNSP could likely only disseminate when antibiotic consumption had reached a level great enough to significantly reduce the transmission of competing PSP (data indicating that antibiotic use was associated with a decreased probability of PSP carriage, and subsequent increased probability of PNSP carriage, support this hypothesis [43b]).

A second factor acting to delay the spread of PNSP may have been the general genetic background and "fitness" of the earliest PNSP representatives. MLST data indicated that the four oldest PNSP in this collection (dated 1967 to 1976) were not related to any of the PMEN clones, and furthermore, none of their STs or associated *pbp* alleles were identified amongst modern-PNSP. In contrast, the MLST data for 100 PNSP dated between 1977 and 1990 (at the time that penicillin-resistance began to rapidly spread) indicated that these PNSP represented the same or highly similar genotypes as those of PNSP circulating today, including several of the PMEN clones. These data could be interpreted to suggest that the earliest-PNSP were in some way less "fit", in terms of survival or transmissibility, than those which emerged later (including the PMEN clones). However, this interpretation must be considered with caution given that the isolates described here do not represent a systematic sample and thus may not be fully representative of historical pneumococcal populations.

Regression analyses indicated that there was a statistically significant relationship between *pbp* nucleotide divergence and penicillin MIC but the adjusted R^2 value suggested that only ~62% of the variation in the data was explained by the model. The optimum model comprised *pbp1a* nucleotide divergence, *pbp2x* nucleotide divergence and their interaction term. Although *pbp2b* divergence was independently significantly correlated with penicillin MIC, it was not significant in the optimum model. This is likely the result of a statistical effect whereby the variation explained by *pbp2b* divergence can also be explained by the other model terms.

Consistent with the results of the statistical analyses presented here, it has been shown previously that divergent *pbp2x* is the first to be acquired on the path to β-lactam resistance [5, 10] and that subsequent acquisition of divergent *pbp1a* is associated with high-level resistance [47]. Additionally, *in vitro* transformation experiments showed that acquisition of divergent *pbp2x* by the R6 laboratory strain was associated with an increase of penicillin MIC from 0.03 μg/ml to 0.06 μg/ml. Subsequent acquisition of divergent *pbp2b* was associated with an MIC increase to 0.25 μg/ml. Furthermore, acquisition of divergent *pbp1a* by the R6-2x/2b mutant was associated with a final MIC increase to 1.5 μg/ml [47].

The variation in the data which was not explained by the model likely results from a combination of several effects, the first of which is the variation in the susceptibility test itself. Secondly, all nucleotide changes were treated equally in this model, but many were synonymous, and thus would not be expected to result in PBP structural

changes or to affect penicillin MIC. Furthermore, it has been shown that some PBP amino acid changes (resulting from non-synonymous *pbp* nucleotide substitutions) have a greater effect on penicillin MICs than others: amino acid substitutions at or around the PBP active sites are those most commonly associated with elevated penicillin MICs [3a, 38a, 43c] and are thus considered the most important.

In vitro transformation of the R6 laboratory strain with divergent *pbp2x,* followed by site-directed mutagenesis of this *pbp2x* gene further indicated differences in the effects of non-synonymous *pbp2x* substitutions on β-lactam MICs [5]. In this study only 6 of 41 tested PBP2x amino acid substitutions were shown to affect cefotaxime susceptibility (which is closely correlated with penicillin susceptibility). A similar study found that penicillin and cefotaxime MICs were altered by just five of 43 tested PBP1a amino acid substitutions [47a].

Aside from those described above, a further source of unexplained variation in the statistical model could be variation at other genetic loci which play a role in the development of penicillin-nonsusceptibility, e.g. *murMN/fibAB, ciaRH* and *clpL* described above.

This study is the first step towards a better understanding of the history of pneumococcal genetic evolution and penicillin-resistance. It has highlighted the longevity of several of the internationally recognised PMEN clones and the complexity of the evolution of pneumococcal penicillin-resistance. Whilst *pbp* gene alterations are clearly important and statistically significant factors, they do not fully explain

differences in penicillin susceptibility between isolates. Furthermore, *pbp* gene recombination between pneumococci was occurring many years before the first PNSP were identified, but *pbp* gene recombination between pneumococci and other streptococci apparently was not (or was at least rare). There does not seem any reason to believe that historical pneumococci had less contact with viridans streptococci than modern pneumococci do. Consequently, the lack of detectable inter-species *pbp* transfer must have resulted from a mechanistic or selective barrier. Comparisons of other regions of the genomes of both historical and modern strains may shed some light on the role of non-*pbp* loci in the evolution and spread of penicillin-resistance. A better understanding of these processes would greatly aid efforts to restrict the spread of resistance within this major global pathogen.

3.6 References

1. **Albarracín Orio, A. G., G. E. Piñas, P. R. Cortes, M. B. Cian, and J. Echenique.** 2011. Compensatory evolution of *pbp* mutations restores the fitness cost imposed by β-lactam resistance in *Streptococcus pneumoniae.* PLoS Pathog. **7:**e100200.
2. **Appelbaum, P. C., J. N. Scragg, A. J. Bowen, A. Bhamjee, A. F. Hallett, and R. C. Cooper.** 1977. *Streptococcus pneumoniae* resistant to penicillin and chloramphenicol. Lancet **2:**995-997.
3. **Beall, B. W., R. E. Gertz, R. L. Hulkower, C. G. Whitney, M. R. Moore, and A. B. Brueggemann.** 2011. Shifting genetic structure of invasive serotype 19A pneumococci in the United States. J Infect Dis. **203:**1360-1368.

3a. **Biçmen, M., Z. Gulay, S. V. Ramaswamy, D. M. Musher, and D. Gür.** 2006. Analysis of mutations in the *pbp* genes of penicillin-non-susceptible pneumococci from Turkey. Clinical Microbiological Infection **12:**150-155.

4. **Bratcher, P. E., I. H. Park, M. B. Oliver, M. Hortal, R. Camilli, S. K. Hollingshead, T. Camou, and M. H. Nahm.** 2011. Evolution of the capsular gene locus of *Streptococcus pneumoniae* serogroup 6. Microbiol. **157:**189-198.
5. **Carapito, R., L. Chesnel, T. Vernet, and A. Zapun.** 2006. Pneumococcal β-Lactam resistance due to a conformational change in penicillin-binding protein 2x. J Biol Chem. **281:**1771-1777.
6. **Chi, F., O. Nolte, C. Bergmann, M. Ip, and R. Hakenbeck.** 2007. Crossing the barrier: evolution and spread of a major class of mosaic *pbp2x* in *Streptococcus pneumoniae*, *S. mitis* and *S. oralis*. Int J Med Microbiol. **297:**503-512.
7. **Chiou, A.-C., S. S. Andrade, S. C. G. Almeida, R. C. Zanella, A.-L. Andrade, and M.-C. de Cunto Brandileone.** 2008. Molecular assessment of invasive *Streptococcus pneumoniae* serotype 1 in Brazil: evidence of clonal replacement. J Med Microbiol. **57:**839-844.

7a. **Claverys, JP., and S. A. Lacks.** 1986. Heteroduplex deoxyribonucleic acid base mismatch repair in bacteria. Microbiol Rev. **50:**133-165.

8. **Corso, A., E. P. Severina, V. F. Petruk, Y. R. Mauriz, and A. Tomasz.** 1998. Molecular characterisation of penicillin-resistant *Streptococcus pneumoniae* isolates causing respiratory disease in the United States. Microb Drug Resist. **4:**325-337.
9. **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, B. Pichon, S. Baker, C. M. Parry, L. M. Lambertsen, D. Shahinas, D. R. Pillai, T. J. Mitchell, G. Dougan, A. Tomasz, K. P. Klugman, J. Parkhill, W. P. Hanage, and S. D. Bentley.** 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331:**430-434.
10. **Dias, R., D. Felix, and M. Canica.** 2008. Diversity of penicillin binding proteins among clinical *Streptococcus pneumoniae* strains from Portugal. Antimicrob Agents Chemother. **52:**2693-2695.
11. **Dixon, J. M. S., A. E. Lipinski, and M. E. Graham.** 1977. Detection and prevalence of pneumococci with increased resistance to penicillin. CMAJ. **117:**1159-1161.
12. **Dowson, C. G., T. J. Coffey, C. M. Kell, and R. A. Whiley.** 1993. Evolution of penicillin resistance in *Streptococcus pneumoniae*; the role of *Streptococcus mitis* in the formation of a low affinity PBP2B in *S. pneumoniae*. Mol Microbiol. **9:**635-643.
13. **Dowson, C. G., A. Hutchison, J. A. Brannigan, R. C. George, H. David, J. Liñares, A. Tomasz, J. M. Smith, and B. G. Spratt.** 1989. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. Proc Natl Acad Sci U S A. **86:**8842-8846.
14. **Felmingham, D., A. R. White, M. R. Jacobs, P. C. Appelbaum, J. Poupard, L. A. Miller, and R. N. Gruneberg.** 2005. The Alexander Project: the benefits from a decade of surveillance. J Antimicrob Chemother. **56:**ii3-ii21.
15. **Filipe, S. R., E. Severina, and A. Tomasz.** 2001. Functional analysis of *Streptococcus pneumoniae* MurM reveals the region responsible for its specificity in the synthesis of branched cell wall peptides. J Biol Chem. **276:**39618-39628.
16. **Filipe, S. R., E. Severina, and A. Tomasz.** 2001. The role of the *murMN* operon in penicillin resistance and antibiotic tolerance of *Streptococcus pneumoniae*. Microb Drug Resist. **7:**303-316.
17. **Granger, D., G. Boily-Larouche, P. Turgeon, K. Weiss, and M. Roger.** 2006. Molecular characteristics of *pbp1a* and *pbp2b* in clinical *Streptococcus pneumoniae* isolates in Quebec, Canada. J Antimicrob Chemother. **57:**61-70.
18. **Hakenbeck, R., T. Briese, L. Chalkley, H. Ellerbrok, R. Kalliokoski, C. Latorre, M. Leinonen, and C. Martin.** 1991. Variability of penicillin-binding proteins from penicillin-sensitive *Streptococcus pneumoniae*. J Infect Dis. **164:**307-312.
19. **Hakenbeck, R., T. Grebe, D. Zahner, and J. B. Stock.** 1999. β-lactam resistance in *Streptococcus pneumoniae*: penicillin-binding proteins and non-penicillin-binding proteins. Mol Microbiol. **33:**673-678.
20. **Hansman, D., and M. M. Bullen.** 1967. A resistant pneumococcus. Lancet **2:**264-265.
21. **Hansman, D., L. Devitt, H. Miles, and I. Riley.** 1974. Pneumococci relatively insensitive to penicillin in Australia and New Guinea. Med J Aust. **2:**353-356.
22. **Hansman, D., L. Devitt, and I. Riley.** 1973. Pneumococci with increased resistance to penicillin. BMJ. **3:**405-405.
23. **Hicks, L. A., Y. Chien, T. H. Taylor Jr, M. Haber, K. P. Klugman, and the Active Bacterial Core Surveillance (ABCs) Team.** 2011. Outpatient antibiotic prescribing and nonsusceptible *Streptococcus pneumoniae* in the United States, 1996-2003. Clin Infect Dis. **53:**631-639.
24. **Howes, V. J., and R. G. Mitchell.** 1976. Meningitis due to relatively penicillin-resistant pneumococcus. BMJ. **1:**996-996.
25. **Huebner, R. E., A. D. Wasas, and K. P. Klugman.** 2000. Trends in antimicrobial resistance and serotype distribution of blood and cerebrospinal fluid isolates of *Streptococcus pneumoniae* in South Africa, 1991-1998. Int J Infect Dis. **4:**214-218.
26. **Jacobs, M. R., H. J. Koornhof, R. M. Robins-Browne, C. M. Stevenson, Z. A. Vermaak, I. Freiman, B. Miller, M. A. Witcomb, M. Isaäcson, J. I. Ward, and R. Austrian.** 1978. Emergence of multiply resistant pneumococci. N Engl J Med. **299:**735-740.
27. **Kislak, J. W., L. M. B. Razavi, A. K. Daly, and M. Finland.** 1965. Susceptibility of pneumococci to nine antibiotics. Am J Med Sci. **250:**261-268.
28. **Krumwiede, E.** 1949. Penicillin resistance of nonhemolytic streptococci from rheumatic children receiving prophylactic penicillin. Pediatrics **4:**634-642.

29. **Lambertsen, L., M. Brendstrup, H. Friis, and J. J. Christensen.** 2010. Molecular characterisation of invasive penicillin non-susceptible *Streptococcus pneumoniae* from Denmark, 2001 to 2005. Scand J Infect Dis. **42:**333-340.
30. **Lauer, B. A., and L. B. Reller.** 1980. Serotypes and penicillin susceptibility of pneumococci isolated from blood. J Clin Microbiol. **11:**242-244.
31. **Lefèvre, J. C., M. A. Bertrand, and G. Faucon.** 1995. Molecular analysis by pulsed field gel electrophoresis of penicillin-resistant *Streptococcus pneumoniae* from Toulouse, France. Eur J Clin Microbiol Infect Dis. **14:**491-497.
32. **Liñares, J., C. Ardanuy, R. Pallarés, and A. Fenoll.** 2010. Changes in antimicrobial resistance, serotypes and genotypes in *Streptococcus pneumoniae* over a 30-year period. Clin Microbiol Infect. **16:**402-410.
33. **Lipsitch, M.** 2001. The rise and fall of antimicrobial resistance. Trends Microbiol. **9:**438-444.
34. **Mace, J. W., D. S. Janik, R. L. Sauer, and J. J. Quilligan.** 1977. Penicillin-resistant pneumococcal meningitis in an immunocompromised infant. J Pediatr. **91:**506-507.
35. **Maisnier-Patin, S., J. R. Roth, A. Fredriksson, T. Nyström, O. G. Berg, and D. I. Andersson.** 2005. Genomic buffering mitigates the effects of deleterious mutations in bacteria. Nat Genet. **37:**1376-1379.
35a. **Majewski, J., and F. M. Cohan.** 1998. The effect of mismatch repair and heteroduplex formation on sexual isolation in *Bacillus.* Genetics **148:**13-18.
35b. **Majewski, J., P. Zawadzki, P. Pickerill, F. M. Cohan, and C. G. Dowson.** 2000. Barriers to genetic exchange between bacterial species: *Streptococcus pneumoniae* transformation. J Bacteriol. **182:**1016-1023.
36. **Mascher, T., M. Heintz, D. Zahner, M. Merai, and R. Hakenbeck.** 2006. The CiaRH system of *Streptococcus pneumoniae* prevents lysis during stress induced by treatment with cell wall inhibitors and by mutations in *pbp2x* involved in β-lactam resistance. J Bacteriol. **188:**1959-1968.
37. **McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefévre, A. Tomasz, and K. P. Klugman.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. J Clin Microbiol. **39:**2565-2571.
38. **Modde, H. K.** 1978. *Streptococcus pneumoniae* isolates relatively insensitive to penicillin G recovered from patients in Switzerland. Chemotherapy. **24:**227-230.
38a. **Nagai, K., T. A. Davies, M. R. Jacobs, and P. C. Appelbaum.** 2002. Effects of amino acid alterations in penicillin-binding proteins (PBPs) 1a, 2b, and 2x on PBP affinities of penicillin, ampicillin, amoxicillin, cefditoren, cefuroxime, cefprozil, and cefaclor in 18 clinical isolates of penicillin-susceptible, -intermediate, and -resistant pneumococci. Antimicrob Agents Chemother. **46:**1273-1280.
39. **Naraqi, S., K. G. P, and S. Kabins.** 1974. Relapsing pneumococcal meningitis: isolation of an organism with decreased susceptibility to penicillin G. J Pediatr. **85:**671-673.
40. **Paredes, A., L. H. Taber, M. D. Yow, D. Clark, and W. Nathan.** 1976. Prolonged pneumococcal meningitis due to an organism with increased resistance to penicillin. Pediatrics. **58:**378-381.
40a. **Rayssiguier, C., D. S. Thaler, and M. Radman.** 1989. The barrier to recombination between *Escherichia coli* and *Salmonella typhimurium* is disrupted in mismatch-repair mutants. Nature **342:**396-401.
41. **Reichmann, P., E. Varon, E. Günther, R. R. Reinert, R. Lüttiken, A. Marton, P. Geslin, J. Wagner, and R. Hakenbeck.** 1995. Penicillin-resistant *Streptococcus pneumoniae* in Germany: genetic relationship to clones from other European countries. J Med Microbiol. **43:**377-385.
42. **Reinert, R. R., J. M. R, P. C. Appelbaum, S. Bajaksouzian, S. Cordeiro, M. van der Linden, and A. Al-Lahham.** 2005. Relationship between the original multiply resistant South African isolates of *Streptococcus pneumoniae* from 1977 to 1978 and contemporary international resistant clones. J Clin Microbiol. **43:**6035-6041.
43. **Riedel, S., S. E. Beckmann, K. P. Helimann, S. S. Richter, J. Garcia-de-Lomas, M. Ferech, H. Goosens, and G. V. Doern.** 2007. Antimicrobial use in Europe and antimicrobial resistance in *Streptococcus pnuemoniae.* Eur J Clin Microbiol Infect Dis. **26:**485-490.
43a. **Sagi, D., T. Tlusty, and J. Stavans.** 2006. High fidelity of RecA-catalyzed recombination: a watchdog of genetic diversity. Nucleic Acids Res. **34:**5021-5031.

43b. **Samore, M. H., M. Lipsitch, S. C. Alder, B. Haddadin, G. Stoddard, J. Williamson, K. Sebastian, K. Carroll, O. Ergonul, Y. Carmeli, and M. A. Sande.** 2006. Mechanisms by which antibiotics promote dissemination of resistant pneumococci in human populations. Am J Epidemiol. **163:**160-170.

43c. **Sanbongi, Y., T. Ida, M. Ishikawa, Y. Osaki, H. Kataoka, T. Suzuki, K. Kondo, F. Ohsawa, and M. Yonezawa.** 2004. Complete sequences of six penicillin-binding protein genes from 40 *Streptococcus pneumoniae* clinical isolates collected in Japan. Antimicrob Agents Chemother. **48:**2244-2250.

44. **Shi, Z.-Y., M. C. Enright, P. Wilkinson, D. Griffiths, and B. G. Spratt.** 1998. Identification of three major clones of multiply antibiotic-resistant *Streptococcus pneumoniae* in Taiwanese hospitals by multilocus sequence typing. J Clin Microbiol. **36:**3514-3519.

45. **Sibold, C., J. Henrichsen, A. Konig, C. Martin, L. Chalkley, and R. Hakenbeck.** 1994. Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. Mol Microbiol. **12:**1013-1023.

46. **Silva, R. F., S. M. Mendonça, L. M. Carvalho, A. M. Reis, I. Gordo, S. Trindade, and F. Dionisio.** 2011. Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. PLoS Genet. **7:**e1002181.

47. **Smith, A. M., and K. P. Klugman.** 1998. Alterations in PBP 1A essential for high-level penicillin resistance in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **42:**1329-1333.

47a. **Smith, A. M., and K. P. Klugman.** 2003. Site-specific mutagenesis analysis of PBP1A from a penicillin-cephalosporin-resistant pneumococcal isolate. Antimicrob Agents Chemother. **47:**387-389.

48. **Tempest, B., J. Carney, and B. Eberle.** 1974. Distribution of the sensitivities to penicillin of types of *Diplococcus pneumoniae* in an American Indian population. J Infect Dis. **130:**67-69.

49. **Tran, T. D., H. Kwon, E. Kim, K. Kim, D. E. Briles, S. Pyo, and R. Dong-Kwon.** 2011. Decrease in penicillin susceptibility due to heat shock protein ClpL in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **55:**2714-2728.

50. **Trzciński, K., C. M. Thompson, A. M. Gilbey, C. G. Dowson, and M. Lipsitch.** 2004. Incremental increase in fitness cost with increased β-lactam resistance in pneumococci evaluated by competition in an infant rat nasal colonization model. J Infect Dis. **193:**1296-1303.

51. **Weber, B., K. Ehlert, A. Diehl, P. Reichmann, H. Labischinski, and R. Hakenbeck.** 2000. The *fib* locus in *Streptococcus pneumoniae* is required for peptidoglycan crosslinking and the PBP-mediated β-lactam resistance. FEMS Microbiol Lett. **188:**81-85.

4. THE MULTIDRUG-RESISTANT PMEN1 PNEUMOCOCCUS: A PARADIGM FOR GENETIC SUCCESS

This chapter has been prepared for submission to a peer-reviewed journal. The following investigators contributed to this work:

Kelly L. Wyres[1], Lotte L. Lambertsen[2], Nicholas J. Croucher[3], Lesley McGee[4], Anne von Gottberg[5], Josefina Liñares[6], Michael R. Jacobs[7], Karl G. Kristinsson[8], Bernard W. Beall[4], Keith P. Klugman[9], Julian Parkhill[3], Regine Hakenbeck[10], Stephen D. Bentley[3] and Angela B. Brueggemann[1].

[1]University of Oxford, Oxford, United Kingdom; [2]Statens Serum Institut, Copenhagen, Denmark; [3]Wellcome Trust Sanger Institute, Hinxton, United Kingdom; [4]Centers for Disease Control and Prevention, Atlanta, Georgia, USA; [5]National Institute for Communicable Diseases, Gauteng, South Africa; [6]Hospital Universitari de Bellvitge, Barcelona, Spain; [7]Case Western Reserve University, Cleveland, Ohio, USA; [8]Landspitali University Hospital and University of Iceland, Reykjavik, Iceland; [9]Emory University, Atlanta, Georgia, USA; [10]University Kaiserslautern, Kaiserslautern, Germany.

The data described in this chapter include those described in the previous chapter. Additionally, 96 pneumococcal isolates were selected for whole-genome sequencing by the Illumina method. I extracted DNA from 48 of these isolates. DNA was extracted from the remaining 48 isolates by LML. Library construction and sequencing itself was completed by collaborators at the Wellcome Trust Sanger Institute, Hinxton, UK. I assembled raw sequence data to contigs, under the guidance of NJC.

4.1 Abstract

Penicillin resistance rapidly increased amongst pneumococci over the past 30 years, and one particular multidrug-resistant clone (PMEN1) became highly prevalent globally. We studied a large isolate collection dating from 1937 to understand the evolution of penicillin resistance among pneumococci. We show that PMEN1 is closely related to one of the oldest penicillin-nonsusceptible isolates. Following acquisition of new penicillin-resistance genes, colonisation- and virulence-related genes, the new "PMEN1" clone, became highly successful in propagating itself. PMEN1 was also uniquely promiscuous with its DNA, donating penicillin-resistance genes *and* sometimes many other genes associated with antimicrobial resistance, virulence and cell adherence, to many unrelated pneumococci. The impact on the genomic evolution of these pneumococci was considerable and is highly relevant to the use of antibiotics to treat pneumococcal disease.

4.2 Introduction

Worldwide, over 1.6 million deaths annually are attributed to the bacterial pathogen, *Streptococcus pneumoniae,* the "pneumococcus" [55]. This bacterium is a leading cause of otitis media, sinusitis, pneumonia and meningitis, and is associated with high mortality rates in the developing world [17, 55]. While many pneumococcal diseases can be successfully treated with antibiotics such as penicillin (and other β-lactams), resistance is a major global problem. The primary pneumococcal penicillin resistance determinants are three of the six penicillin-binding protein (*pbp*) genes, *pbp2x, pbp1a* and *pbp2b.* In penicillin-nonsusceptible pneumococci (PNSP) these genes contain

mosaic sequence blocks differing from those of penicillin-susceptible pneumococci (PSP) by up to 25% nucleotide divergence [23, 30]. It is believed that these mosaic blocks were acquired from closely-related viridans streptococci via homologous recombination [23].

PNSP were first reported in the late 1960s (in the USA and Australia) [24, 28]. In the late 1970s the first high-level penicillin-resistant pneumococci were reported in South Africa [1, 26]. Throughout the 1980s and 1990s global penicillin nonsusceptibility rates rocketed, surpassing 50% in some regions [32]. The selective pressure of antibiotic use was enormous at this time, as antibiotics were widely used, even overused, in many countries. This selective pressure meant that pneumococci with some level of resistance to penicillin or other antibiotics had a distinct advantage over susceptible strains, which may have indirectly contributed to a change in the distribution of pneumococcal capsular types (or serotypes, defined by the antigenic polysaccharide capsule surrounding the cell) [18]. Following the introduction of a 7-valent pneumococcal conjugate vaccine (PCV7) in the year 2000 in the USA, there was a substantial reduction in PNSP levels [54]. Use of PCV7 in children provides efficacious protection against 7 of the 94 known pneumococcal serotypes [55], including those which were most commonly associated with penicillin resistance. Pneumococcal disease caused by PCV7 type pneumococci has significantly declined in countries that have introduced PCV7. Subsequently, there has been an increase in pneumococcal disease due to non-PCV7 type pneumococci [53], many of which are now also penicillin nonsusceptible [3]. 10- and 13- valent pneumococcal vaccines, containing additional

serotypes, have recently been introduced, and continued surveillance over the next few years will be essential to determine their effects.

Alongside the increase of PNSP there has been an emergence of predominant and widely-distributed pneumococcal clones. 43 important disease-causing clones, many of which are multidrug-resistant, have been identified by the Pneumococcal Molecular Epidemiology Network (PMEN) [36]. The first of these clones was Spain^{23F}-1, here called PMEN1. This multidrug-resistant pneumococcus predominantly circulates as a vaccine serotype 23F, multilocus sequence type 81 clone (ST81^{23F}); however, it has also been associated with several other serotypes, including both vaccine types and non-vaccine types [9, 11]. The first PMEN1 representative was isolated in Spain in 1984 and shortly thereafter in the USA [39], the UK [9], South Africa [29, 47], Hungary [44] and South America [50]. By the late 1990s it was estimated that ~40% of PNSP circulating in the USA were members of this clone [10] and it continues to circulate among pneumococcal populations today.

A recent study of ~240 pneumococci representing PMEN1 ST81 and related clonal variants (all belonging to clonal complex (CC) 81) showed that there is a considerable amount of genetic diversity within this lineage. This diversity, which largely results from hundreds of recombination events, indicates rapid genomic evolution and presumably allowed rapid response to selective pressures such as those imposed by vaccine and antibiotic use [11].

Here we studied a large, genetically diverse, historical and global pneumococcal isolate collection with the aim to better understand the evolution of penicillin resistance within this species. We uncovered the important contribution of the PMEN1 clone to the emergence and spread of penicillin-resistant pneumococci and revealed its genetically highly promiscuous nature. Furthermore, we have been able to confidently identify the ancestor of this globally and historically important clone as being closely related to one of the oldest known PNSP.

4.3 Results

PMEN1 pbp promiscuity

Pneumococcal study isolates (Table 4.1) were assigned to CCs based on multilocus sequence type (MLST) data. Isolates within CCs are considered to have shared ancestry, while those in different CCs are more distantly related. Within our collection, identical or highly similar, 'nonsusceptible' *pbp* alleles and/or sequence regions were identified repeatedly among pneumococci belonging to different CCs. These *pbp* sequences were those associated with the PMEN1 reference strain (hereafter simply called 'PMEN1').

Two isolates, CGSP14 and the PMEN3 reference strain (hereafter called 'PMEN3'), representing independent CCs (Table 4.2), shared identical *pbp2x, pbp1a* and *pbp2b* alleles with PMEN1. Analysis of the genomic regions flanking these *pbp* genes showed that in fact much larger regions of the CGSP14 and PMEN3 genomes (5,362 – 22,104 bp, Figure 4.1) were identical or highly similar to that of PMEN1, suggesting these regions were acquired from a PMEN1 representative(s) in both cases. The PMEN2

reference strain, representing another CC (Table 4.2), also shared identical regions (225 - 1034 bp in length) of all three of its *pbps* with PMEN1. A total of 12 distinct CCs were represented by isolates sharing identical *pbp* sequences in whole or part (119 – 1,376 bp) in one or two of their *pbps* to PMEN1. Nine additional CCs were represented by isolates sharing *pbp2x* and/or *pbp2b* regions (199 – 1,436 bp) that differed from those of PMEN1 by only one or two nucleotide substitutions (see Supplementary Tables 4.4-4.6 for *pbp* allele and CC information). No other *pbp* alleles were as widely distributed as those of PMEN1.

Table 4.1 Isolates included in this study.
See Chapter 2, Table 2.1.

Table 4.2 Pneumococcal isolates discussed in Chapter 4 text.

Isolate†	Other Name(s)	Year	Serotype	Sequence Type	Clonal Complex	Penicillin Susceptibility*
PMEN1	ATCC 700669	1984	23F	81	81	I
CGSP14		2004-2005	14	15	15	R
14/5		1967	14	15	15	S
PMEN9	ATCC 700676	1993	14	9	15	S
ICE13		1998	14	9	15	S
ICE570		2002	14	9	15	S
PMEN3	ATCC 700671; SP195	1993	9V	156	156/162	I
9A/1		1962	9A	312	156/162	S
9V/5		1991	9V	162	156/162	S
9V/6		1994	9V	162	156/162	S
PMEN2	ATCC 700670	1988	6B	90	90	I
23F/4		1967	23F	7184	Singleton7184	I
PMEN18	ATCC BAA-340	1997	14	67	66	R
CDC3059-06		2005	19A	199	199	I
CDC0288-04		Unknown	12F	220	218	S
USA18		1999	12F	218	218	S

† PMEN refers to Pneumococcal Molecular Epidemiology Network reference strains.
*S = MIC ≤ 0.06 μg/ml; I = MIC 0.12 - 1 μg/ml; R = MIC ≥ 2.0 μg/ml.

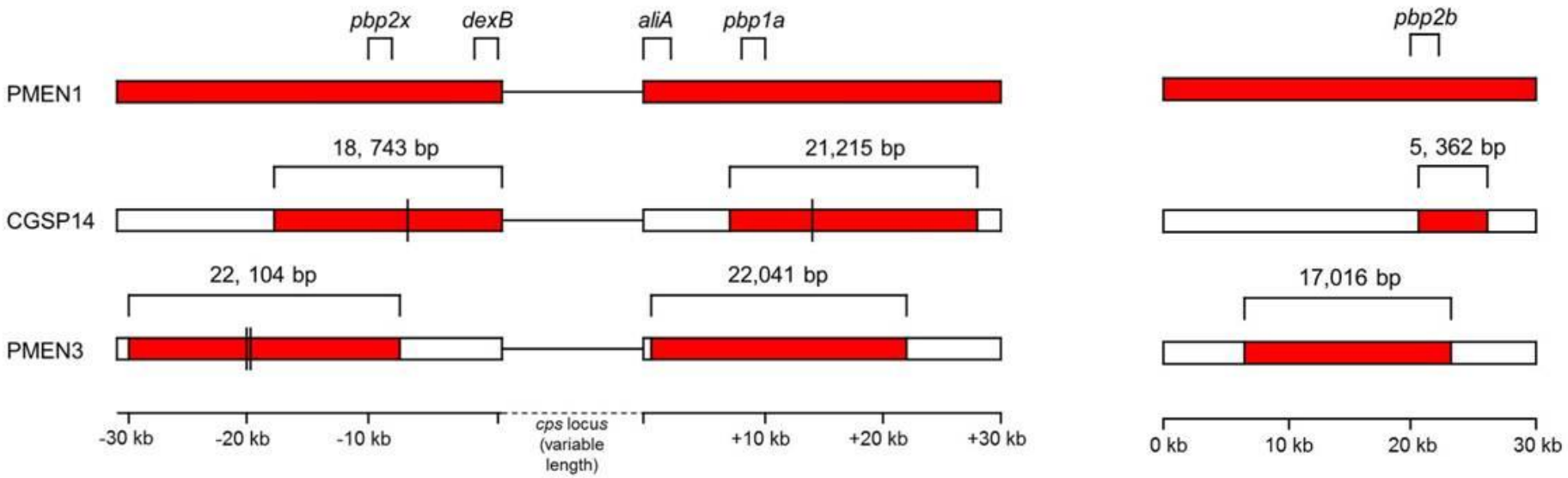


Figure 4.1 Schematic representation of nucleotide sequences at the *pbp* loci and flanking regions of the PMEN1 reference, CGSP14 and PMEN3 reference genomes.

Red blocks indicate identical nucleotide sequence with substitutions or single insertions marked by black lines. The relative positions of the *pbp2x, pbp1a,* and *pbp2b* loci are shown, as are the two loci flanking the capsular (*cps*) locus (*dexB* and *aliA*).

Extensive PMEN1 genomic recombination

We used whole-genome sequence data generated on the Illumina platform to assess whether other regions of the PMEN1 genome were also shared by CGSP14 and PMEN3. We found that in fact 15 regions (1,663 bp – 32,312 bp, totalling ~9.5% of the genome) of the CGSP14 chromosome were identical to those of PMEN1 or differed by only 1 - 2 nucleotides across a minimum of 7,537 bp (Figure 4.2, Supplementary Table 4.7), and were not identical in any of four CGSP14 ancestral representative genomes. The ancestral representatives, 14/5, ICE13, ICE570 and the PMEN9 reference strain (Table 4.2) were chosen from our collection because they represented older and/or penicillin-susceptible members of the CGSP14 CC and so likely resembled the true ancestor of CGSP14. Importantly, this allowed us to identify and exclude identical sequence that was shared by descent, whereas regions of the CGSP14 genome that differed from its ancestral representatives but were identical to those of PMEN1 were most likely acquired from PMEN1 through the process of recombination. The 14/5, ICE13, ICE570 and PMEN9 reference sequences were generally all identical or highly similar to each other at the regions in question; sequences from these isolates were also highly similar to CGSP14 in the regions flanking the putative recombination imports, confirming their suitability as ancestral representatives.

Following a similar comparison of PMEN3 and its ancestral representatives, 9A/1, 9V/5 and 9V/6 (Table 4.2), we identified 9 PMEN3 genomic regions (4,662 bp – 22,104 bp, totalling ~5.3% of the genome) that were identical to sequences from PMEN1 or differed by only two nucleotides across 22,104 bp, and were different from those of the ancestral representatives (Figure 4.2, Supplementary Table 4.8). We suggest that

these regions were also acquired by PMEN3 from PMEN1 through the process of recombination.

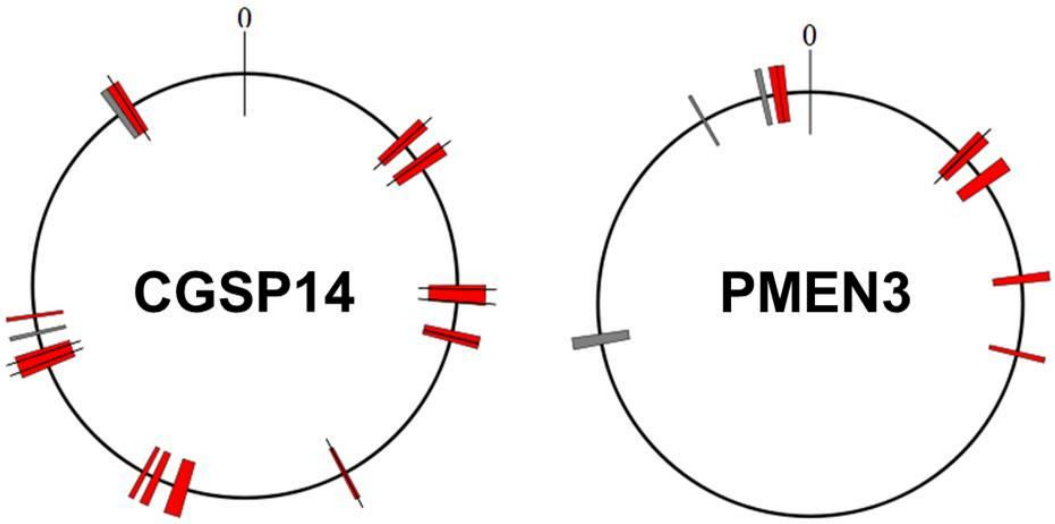


Figure 4.2 PMEN1 reference genome regions putatively acquired by CGSP14 (left) and the PMEN3 reference (right) through the process of recombination.
Putative recombination regions marked by colored blocks and mapped to PMEN1 reference genome coordinates. Red blocks: regions identical or highly similar to that of the PMEN1 reference only. Grey blocks: regions identical or highly similar to those of the PMEN1 reference *and* one or more CC66 representative(s). Single nucleotide substitutions or insertions are marked with black lines. Origin of replication marked as '0'.

The CGSP14 and PMEN3 putative recombinogenic regions contained 213 and 108 whole or part coding sequences respectively (Supplementary Tables 4.7-4.8). In addition to the *pbp2x, pbp1a* and *pbp2b* genes, the following coding sequences associated with enhanced virulence or colonisation potential were included: The CGSP14 *bgaA, clpX,* SPN23F_05810 and SPN23F_20890 genes associated with enhanced host colonisation ability [37]; the CGSP14 virulence-associated *cbpJ, msmK*, chloramphenicol acetyltransferase and zeta toxin/epsilon antitoxin genes [5, 31, 43] (the latter two of which are situated on the CGSP14 conjugative transposon); and the PMEN3 virulence-associated *cbpD, cbpJ* and *clpC* genes [43].

The possibility that the putative imported regions were acquired by CGSP14 and PMEN3 from non-PMEN1 representative pneumococci cannot be ignored. Where the CGSP14 and PMEN3 putative import regions overlapped it is possible that these sequences were, in fact transferred between these pneumococci rather than acquired from a PMEN1 representative. However, in no instance was the full length of a CGSP14 putative import matched by the overlapping PMEN3 import or vice versa (see the regions represented in Figure 4.1 as an example). Thus if it were the case that the overlapping sections were transferred between these strains, the non-overlapping sections must have been acquired independently through additional recombination event(s). In this case the most parsimonious solution is that the regions were acquired as a whole from a PMEN1 representative.

The possibility that the CGSP14 and PMEN3 putative import regions were acquired from other, unrelated pneumococci should also be addressed; however, the majority, 13 and 6 respectively, of regions described here were not identical among any of 96 additional pneumococcal genomes, representing a diverse range of 31 independent CCs. Two and three of the regions respectively (Figure 4.2), were identical or highly similar among isolates representing CC66, and so it is possible that those regions may have been acquired from a CC66 representative (also see Supplementary Information). Although our collection cannot capture the full complement of pneumococcal genetic diversity, given these results and the extremely high similarity of the putative import regions to the PMEN1 genome we believe that they originated from a PMEN1 representative. Furthermore, the high global prevalence of the PMEN1 clone

compared to those of other clones presumably increased its level of contact with unrelated pneumococci, providing greater opportunity for genetic exchange and increasing the probability that a PMEN1 representative(s) was the true donor.

Ancestors of PMEN1

The Genome Comparator tool was used to screen 104 pneumococcal genomes, representing a very diverse range of CCs and serotypes, for regions of DNA (aside from the *pbps*) that were identical to PMEN1. 1990 coding sequences were identified from the PMEN1 Genbank accession NC_011900.1 file. Comparable coding sequences in the query genomes were identified if they shared ≥70% nucleotide sequence identity (to allow for the identification of variable sequences among conserved protein classes [27] and to avoid bias towards sequence variants specific to PMEN1) and could be aligned at 100% length. Consequently, coding sequences spanning the ends of assembly contigs could not be identified.

Surprisingly and uniquely, the analyses showed that 23F/4, the first PNSP isolated from Australia (Table 4.2), was very likely to be an ancestral representative of PMEN1, since ~95% of the identified coding sequences were ≥98% similar to those of PMEN1. ~75% of identified coding sequences in 23F/4 were identical to PMEN1 and over one-third of the coding sequences that were not identical differed by only one or two nucleotide substitutions. (It is worth acknowledging here that some of these apparent substitutions could be the result of sequencing and/or assembly error, although errors are considered to be extremely rare. If some of these substitutions were the result of error, the 23F/4 genome would show an even greater similarity to that of PMEN1,

strengthening our conclusions.) The *pbp2x, pbp1a, pbp2b* genes and 3 of the 7 MLST loci were not identical. Additionally, a comparison of the PMEN1 and 23F/4 genomes, performed using the Artemis Comparison Tool (Figure 4.3), clearly showed three large regions of the PMEN1 genome that were partially or completely missing from the 23F/4 genome. These were the ICE*Sp*23FST81 element, the ϕMMI phage and a Na^+ dependent ATPase island, all associated with increased colonisation and virulence potential [12, 33, 38, 41].

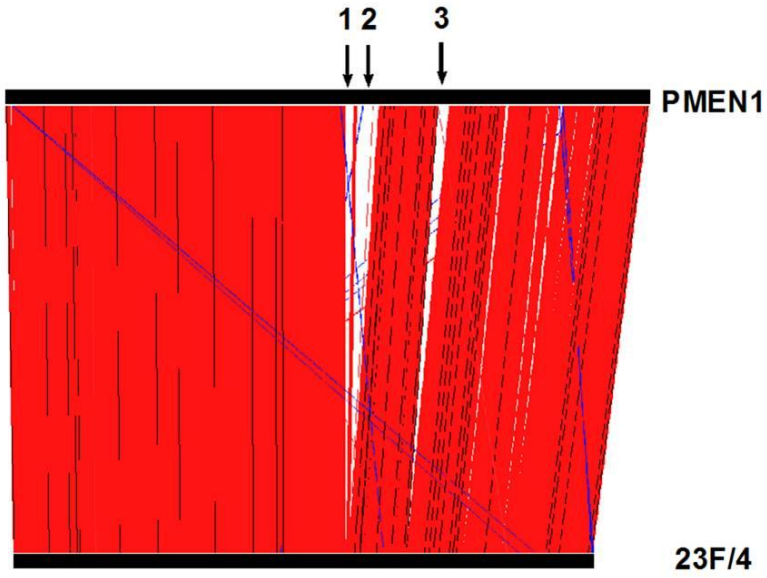


Figure 4.3 Artemis Comparison Tool comparison of the PMEN1 reference and 23F/4 genomes.
Red lines indicate homologous nucleotide sequence matches. Blue lines indicate reverse complement nucleotide sequence matches. Black arrows indicate large regions present in the PMEN1 reference genome but absent from the 23F/4 genome: (1) Na^+ dependent ATPase island; (2) ICE*Sp*23FST81 element; and (3) ϕMMI phage.

The Genome Comparator analyses also revealed that the majority (87, 83.6%) of genomes analysed – representing 31 unique CCs and 30 serotypes – shared 15.1% ± 2 standard deviations (excluding outliers with >30% identical coding sequences; cut-off determined by visual inspection of the data) of the identified coding sequences identically with PMEN1 (Figure 4.4). We suggest that these "baseline" coding

sequences were shared through ancestral descent, i.e. they were identical in the most recent common ancestor of these clones and have not undergone subsequent change.

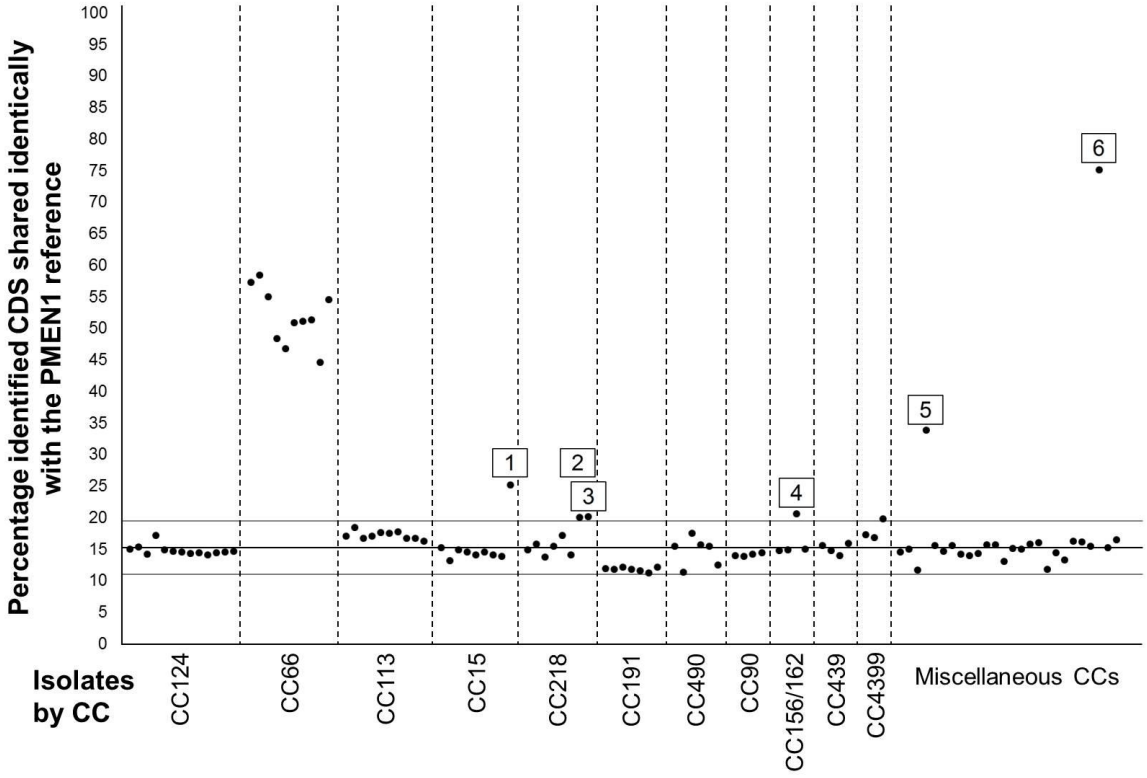


Figure 4.4 Percentage PMEN1 reference coding sequences identified identically within 104 pneumococcal genomes.
Points represent independent pneumococcal isolates. Isolates representing CCs for which n ≥ 3 are labelled by CC. Isolates representing CCs for which n < 3 are labelled as 'Miscellaneous CCs'. Black line represents mean percentage identical coding sequences ± 2 standard deviations (grey lines). Isolates referenced in text are labelled: (1) CGSP14; (2) USA18; (3) CDC2088-04; (4) PMEN3; (5) CDC3059-06; and (6) 23F/4.

Isolates belonging to CC66 shared between 44.5% and 58.4% coding sequences identically with PMEN1. These isolates included PSP and dated from 1952, preceding the estimated origin of PMEN1. Since the proportion of identical coding sequences between CC66 and PMEN1 was much greater than between those of PMEN1 and the other CCs represented here, we conclude that PMEN1 was ancestrally more closely

related to CC66 than to other CCs, which supports the findings of Donati and colleagues [14] (also see Supplementary Information).

Isolate CDC3059-06 (CC199, Genbank accession ABGG01000001-ABGG01000033) shared 33.8% identical coding sequences with PMEN1, but without additional CC199 isolates or ancestral representatives we were unable to make any further inferences in this study. CGSP14 and PMEN3 shared 25.1% and 20.6% identical coding sequences with PMEN1, respectively, which was higher than their respective ancestral representatives and consistent with the findings detailed above. CDC0288-04 and USA18 both shared 20.0% identical coding sequences with PMEN1, potentially also due to recombination with PMEN1, although we cannot be certain as identical regions also matched those of CC66 representatives.

4.4 Discussion

To our knowledge, our collection represents the largest historical pneumococcal collection ever compiled and includes the greatest number of early (1960s - 1980s) PNSP in a single study. This has provided us with a unique opportunity to study pneumococcal evolution. Within our diverse isolate collection, whole or part *pbp* alleles that were identical or highly similar to those of the PMEN1 reference strain were identified from pneumococci representing a total of 24 distinct CCs. Two isolates, CGSP14 and PMEN3, shared *pbp2x, pbp1a* and *pbp2b* sequences identically with PMEN1, and shared highly similar sequences in the *pbp* flanking regions. The full extent of sharing all three of these PMEN1 *pbp* sequences has never before been uncovered (nor have the large regions around all three *pbps*), although previous

authors have independently noted *pbp* similarities between CGSP14, PMEN3, PMEN2 and PMEN1 [9, 13, 34, 46, 48]. Highly similar *pbp2x* nucleotide sequences have also been identified among numerous pneumococcal and viridans streptococcal clones [8, 46].

It is very unlikely that these identical or highly similar sequences were assembled independently in different CCs, but rather that the PMEN1-like *pbp* sequences have been transferred, as whole or part alleles, between unrelated pneumococcal clones at least once for each independent CC. Note that the establishment of defined CCs must have occurred prior to acquisition of these *pbps*, since we also identified penicillin-susceptible members of these clones (i.e. $CC15^{14}$, $CC156/162^{9A/V}$).

Our data suggest that the PMEN1 reference *pbp2x, pbp1a and pbp2b* sequence combination may provide an evolutionary selective advantage over other resistance conferring *pbps,* which is consistent with recent experimental data suggesting certain combinations of *pbp2x* and *pbp1a* alleles are more advantageous than others [57]. A study of genomic evolution within the PMEN1 lineage also provided evidence consistent with the hypothesis that the *pbp2x* and *pbp1a* alleles were positively selected [11]. *pbp2x* is upstream and *pbp1a* is downstream of the capsular (*cps*) locus in the pneumococcal genome. The authors noted that the boundaries of recombination events affecting the *cps* locus appeared to be restricted by the nearby *pbp2x* and *pbp1a* loci. It is known that both of these loci and the *cps* locus can be transferred simultaneously between pneumococci [6], and thus the restriction within the PMEN1 lineage is likely not entirely mechanistic.

In addition to the *pbp* loci and flanking regions, CGSP14 and PMEN3 acquired multiple genomic regions, totalling ~9.5% and ~5.3% of the genome respectively, from PMEN1 representative(s). This is consistent with two recent studies, which demonstrated that multiple fragment recombination had taken place *in vivo,* between pneumococci isolated from a single patient [25] and among vaccine escape strains isolated in the USA [20]. The suggestion that a significant proportion of the genome may be acquired from a single pneumococcal donor was recently supported mechanistically by *Attaeich* and colleagues [2]. These authors showed that cellular levels of pneumococcal SsbB protein were sufficient to maintain an intracellular pool of internalised DNA of up to 1.15 Mb of length (~50% of the genome). Such a pool of genetic material, potentially originating from a single donor cell, could subsequently be used for successive rounds of recombination.

We cannot know for certain whether the CGSP14 and PMEN3 imported regions were acquired simultaneously in a single event or via successive recombination events. Either way, the data suggest that they were acquired from a PMEN1 representative. It should also be noted that recombination events additional to those described here, and involving acquisition of DNA from other pneumococcal clones or other species, may also have altered the CGSP14 and PMEN3 genomes.

The genomic regions acquired from the PMEN1 donor may also have contributed to the virulence and host colonisation ability of the recipients. CGSP14 was multidrug-resistant and isolated from a child suffering from necrotising pneumonia complicated

with haemolytic uremic syndrome [13]. PMEN3 represents a globally-distributed multidrug-resistant clone [36]. The putative recombinogenic regions acquired by both of these pneumococci included several genes associated with virulence or host colonisation potential. Interestingly, large sections of the CGSP14 transposon, which was shown previously to contain regions of homology to those of the PMEN1 ICE*Sp*23FST81 element [13], were in fact identical to those of PMEN1. Two interpretations of the regions of similarity within the ICE*Sp*23FST81 element are possible. One is convergent evolution between PMEN1 and CGSP14, with both strains independently acquiring similar, recently-diverged elements. Alternatively, a more divergent element may have been acquired by CGSP14 that was subsequently modified through recombination with DNA from PMEN1. The mosaic nature of these variable conjugative elements [11] makes distinguishing these two hypotheses difficult.

We also showed that one of the first reported PNSP (23F/4) represented the ancestor of PMEN1. The year of isolation of 23F/4 (1967) correlated with the predicted date of emergence of the PMEN1 lineage (~1970), although the country of isolation (Australia) did not match the predicted region of PMEN1 origin (Europe) [11]. Despite its high overall genomic similarity to PMEN1, the 23F/4 ST, which differed from that of PMEN1, has not been identified amongst modern pneumococci. Perhaps the differences between these two genomes could explain the differences in disseminative success of these two STs. The acquisition of more favourable *pbp2x, pbp1a* and *pbp2b* alleles in the PMEN1 genome provides a partial explanation; however, other changes and/or acquisition of other loci have also likely played an important role. The PMEN1

ICE*Sp*23FST81 element, Na^+ dependent ATPase island and ϕMMI phage were absent from the 23F/4 genome. The ICE*Sp*23FST81 element carries both tetracycline and chloramphenicol resistance determinants [12] plus a *umuDC*-like gene shown to provide protection against UV damage [38]. The Na^+ dependent ATPase island and a ϕMMI phage closely related to that of PMEN1 have been associated with an increased incidence of invasive pneumococcal disease [41] and increased cell adherence capabilities [33], respectively.

Any or all of these regions may have provided PMEN1 with a selective advantage, allowing it to persist among the greater pneumococcal population for the past three decades. Alternatively, the combination of these regions with the PMEN1 reference *pbp* genes and the 23F/4 genetic background may have been the most important factor. In any case, the subsequent global dissemination of PMEN1 likely aided the spread of its *pbp* alleles and other regions of its genome to numerous unrelated pneumococci. Some of the recipient pneumococci might subsequently have transferred these PMEN1-like sequences to additional pneumococcal clones.

Recently, through comparison of multiple representatives within the PMEN1 CC, Croucher *et al.* showed that the PMEN1 lineage contained a vast amount of genetic diversity, largely resulting from horizontal acquisition of DNA via recombination [11]. Through comparison of genomic sequences *across* a genetically diverse collection of pneumococci, our study has greatly increased our understanding of the emergence of the PMEN1 clone and its subsequent contribution to the spread of penicillin resistance determinants among other pneumococci. One key finding was that we discovered the

ancestor of PMEN1, which was unknown at the time of the Croucher *et al.* study and fills a key gap in the PMEN1 story. Secondly, we revealed that the ancestor lacked the PMEN1-reference *pbp* alleles and several other key virulence-associated regions of the PMEN1-reference genome, which we argue contributed to the epidemiological success of PMEN1.

Our third key finding was that members of the PMEN1 lineage have subsequently donated their DNA to numerous unrelated pneumococci from many different CCs, i.e. pneumococci not representing the PMEN1 lineage, in a way that has promoted penicillin-nonsusceptibility and the spread of other colonisation and virulence-associated genes among pneumococci. It is well understood that many bacterial species, particularly the pneumococcus, frequently undergo horizontal gene transfer; however, this sort of directional genetic promiscuity, from a single epidemiologically successful clone to numerous unrelated clones, has not been demonstrated before in pneumococci, or other bacterial species as far as we are aware.

Antibiotics are among the most influential global public health successes, but impose selective pressures that drive bacterial genomic evolution. PMEN1 is an excellent example of a bacterium that has become resistant to multiple antibiotics and that has evolved to become very successful in colonisation, transmission, and causing disease. Moreover, PMEN1 has subsequently shared its advantageous DNA with other unrelated pneumococci. Studies such as this one may help to identify ways to counteract these biological changes and preserve the value of antibiotics for the future.

4.5 Methods

Strain Collection

A unique collection of 426 global PSP and PNSP dated 1937 - 2007 was studied, as summarised in Table 4.1. Isolates included one of the first PNSP reported from Australia [24], the first high level PNSP reported from South Africa [1, 26] and the 43 PMEN reference strains [36]: 211 pneumococci dated 1937 - 1996 were provided by the Statens Serum Insititut (SSI), Denmark. These isolates included both PSP (n = 203), PNSP (n = 7) and an isolate of unknown penicillin susceptibility (n = 1). Following the analyses of MLST and *pbp* sequence data from these isolates, the 43 PMEN clones (including both PSP (n = 18) and PNSP (n = 25)) and additional modern (1990s - 2000s) PSP (n = 38) and PNSP (n = 9) that had been recently genotyped by ABB/LM and represented CCs of interest, were selected for inclusion. In addition, a literature search was performed to identify studies that reported early PNSP. The authors of those papers were contacted and isolates were requested; the majority of isolates were still available and were added to our collection. The available isolates comprised PNSP dated 1969 and 1972 from Papua New Guinea (n = 2), PSP (n = 8) and PNSP (n = 71) dated 1977 - 1989 from South Africa, PSP (n = 1) and PNSP (n = 26) dated 1981 - 1991 from the USA, PNSP dated 1987 - 1989 (n = 9) from Spain and PNSP dated 1986 - 1994 from Germany (n = 8). Finally, genomic sequence data for 16 isolates was retrieved from the Genbank database. These isolates included both PSP (n = 5), PNSP (n = 4) and 7 pneumococci of unknown penicillin susceptibility. Serotype and penicillin susceptibility data are summarised by decade in Supplementary Table 4.1. Where not

previously published, isolates were serotyped by Quellung reaction and penicillin minimum inhibitory concentrations were obtained by Etest (bioMérieux) as per manufacturer's guidelines.

DNA extraction

Single pneumococcal colonies were cultured on Columbia agar with sheep blood (Oxoid) and incubated overnight at 37°C + 5% CO_2. 3 - 4 sweeps of growth were subcultured onto tryptic soy agar (Oxoid) + 1000 U/ml catalase (Oxoid) and spread with a swab before further overnight incubation at 37°C + 5% CO_2 (at Oxford); or cultured in serum broth for 7 hrs (at SSI). Pneumococcal growth was suspended in 1 ml phosphate buffered saline (pH 7.4, Fisher Bioreagents) and centrifuged at 7,600 rpm for 10 min. Extractions were completed using the Qiagen DNeasy kit as per manufacturer's instructions.

MLST and clonal complex (CC) designation

Where not previously published (n = 352) multilocus sequence types were determined for all isolates [16]. MLST data for all of our isolates and those deposited in the pneumococcal MLST database (www.mlst.net) were analysed by the goeBURST method to predict CC group and sub-group founders. Isolates were assigned to CCs so that only single/double locus variants of the group founder and any additional single locus variants of large (n ≥ 5 STs) sub-group founders were included. CCs were named according to the predicted group founder(s) ST(s). When goeBURST could not distinguish a group founder the group was designated NoneX, where X was the ST of lowest numerical value in the group. When isolates could not be assigned to a CC due

to a lack of closely related STs, they were designated SingletonX, where X was the isolate ST.

pbp profiling

pbp2x (1,673 bp), *pbp1a* (901 bp) and *pbp2b* (1,272 bp) nucleotide sequences for 389 selected isolates, were retrieved from Genbank (n = 36 sequences) or determined by conventional PCR and Sanger sequencing (n = 1131 sequences). 174/211 isolates from the SSI were chosen for *pbp* nucleotide sequencing because they represented a wide range of serotypes, STs and isolation dates. All isolates from each of the other sources were also included. PCR reactions containing 1 µl DNA extract, 13.5 µl Hotstar Taq DNA Polymerase master mix (Qiagen), 0.5 µl each of the forward and reverse primers (100 µM) (Supplementary Table 4.2) were prepared to a final volume of 25 µl. Cycling parameters were as follows: Initial denaturation of 95°C for 180 s, followed by 35 (40 for *pbp2x*) cycles of 95°C for 30 s, 55°C (58°C for *pbp1a*) for 30 s and 72°C for 60 s followed by a final extension period of 72°C for 10 min. Products were precipitated by 20% polyethylene glycolate/2.5M NaCl, washed in 70% ETOH and rehydrated in 10 - 20 µl nuclease-free H_2O (Sigma). Sequencing reactions containing 2 µl rehydrated PCR product, 1.75 µl 5x sequencing buffer, 4 µl primer (1 µM) (Supplementary Table 4.2) and 0.5 µl Taq FS (Big Dye) were prepared in a final volume of 10 µl. Cycling parameters were as follows: 25 cycles of 96°C for 10 s, 50°C for 5 s and 60°C for 2 min. Sequencing products were precipitated with 5 M sodium acetate (Sigma) solution, washed with 70% ETOH and subsequently separated and detected with an ABI Prism 3730 automated DNA sequencer (Applied Biosystems). Sequences were aligned by

MUSCLE [15] and imported to MEGA5 [49] for visual comparison. Unique *pbp* sequences were assigned unique allele numbers.

Whole-genome sequencing

96 isolates were selected for whole-genome sequencing (Supplementary Table 4.3) using the following criteria:

1) Isolates that were not closely related to PMEN1 based on MLST data but shared identical *pbp* alleles or partial *pbp* alleles to those of the PMEN1-reference strain. The selected isolates were distributed through time, by geographic location and by genotype in order to capture the greatest possible amount of genetic diversity.
2) Isolates representing CCs of interest (e.g. major international CCs) for which both historical and modern isolate representatives were available. Inclusion of the historical representatives was crucial, as it allowed us to differentiate identical/highly similar sequence imports obtained exogenously versus identical sequence due to common ancestry. Within these CCs the selected isolates were distributed through time, by geographic location, by serotype and by *pbp* allele combination, in order to capture the greatest possible amount of genetic diversity.

DNA was extracted to a final concentration of 20 ng/µl as described above. Sequencing was via the Illumina platform. As per standard protocols, multiplex library construction with a 200 bp insertion size was followed by 54 nucleotide paired end sequencing. Illumina reads were assembled to contigs using the Velvet software [56]. Data for 7 genomes were rejected based on technical problems. K-mer lengths were

optimised to the maximum length achieving an average of ≥20x read coverage across assemblies. Resulting contigs were between 41 bp and 328,742 bp length (mean = 3,797 bp). Contigs were deposited in a BIGS database [27]. Contigs were ordered to the PMEN1 (Genbank accession NC_011900.1) reference genome using Abacas and viewed with the Artemis Comparison Tool [7], allowing extraction of selected sequence regions spanning >1 contig.

Identification of recombination fragments from PMEN1

The BIGS database Genome Comparator tool was used to compare PMEN1 coding sequences to those of CGSP14 (Genbank accession NC_010582.1), PMEN3 (Genbank accession ABGE00000000.1) and their respective predicted ancestral representatives. Parameters were as follows; minimum % identity = 70%, minimum % alignment length = 100%, BLASTN word size = 15. Potential recombination regions were indicated by coding sequences shared identically by PMEN1 and CGSP14 or PMEN3, but not by any of the respective ancestral representatives. These sequence regions were extracted, aligned and imported into MEGA5 [49] to produce variable site maps for visual inspection (see Supplementary Figure 4.1). Recombination regions were classified as the maximum length of sequence over which CGSP14 or PMEN3 were identical to PMEN1 but different from their ancestral representatives.

Identification of PMEN1 ancestral representatives

Genome comparator was used to screen 104 isolates, representing 34 CCs and 31 serotypes for coding sequences identical to those of the PMEN1 reference (parameters as above). 23F/4 was identified as sharing a high percentage of coding sequences

identically with PMEN1 and was thus subjected to further analysis. PMEN1 coding sequences which differed in 23F/4 were extracted from the PMEN1 genome sequence and BLASTed against the 23F/4 Velvet contigs in the BIGS database. Sequences were aligned and percentage nucleotide difference was calculated. 23F/4 contigs were ordered against the PMEN1 chromosome by Abacas and viewed with the Artemis Comparison Tool to identify genomic regions present in PMEN1 but absent from 23F/4.

4.6 Acknowledgements

We thank Dalia Denapaite, Irma Ochigava, and Shwan Rachid for DNA sequencing and MLST analysis, and Roberta Fisher for comments on the manuscript. Funding was provided by the Wellcome Trust (to ABB), Stiftung Rheinland Pfalz für Innovation (to RH), and Ciber de Enfermedades Respiratorias (Instituto de Salud Carlos III, Madrid Spain; (to JL)). ABB is a Wellcome Trust Career Development Fellow. MLST, *pbp* sequence and Illumina data have been deposited at http://zoo-talisker.zoo.ox.ac.uk/perl/bigsdb/bigsdb.pl?db=streptococci. Authors declare no conflicts of interest.

4.7 References

1. **Appelbaum, P. C., J. N. Scragg, A. J. Bowen, A. Bhamjee, A. F. Hallett, and R. C. Cooper.** 1977. *Streptococcus pneumoniae* resistant to penicillin and chloramphenicol. Lancet **2:**995-997.
2. **Attaiech, L., A. Olivier, I. Mortier-Barrière, A. L. Soulet, C. Granadel, B. Martin, P. Polard, and C. J. P.** 2011. Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. PLoS Genet. **7:**e1002156.
3. **Beall, B. W., R. E. Gertz, R. L. Hulkower, C. G. Whitney, M. R. Moore, and A. B. Brueggemann.** 2011. Shifting genetic structure of invasive serotype 19A pneumococci in the United States. J Infect Dis. **203:**1360-1368.
4. **Bentley, S. D., D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabbinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail, G. Samuel, I. C. Skovsted, M. S. Kaltoft, B. Barrell, P. R. Reeves, J. Parkhill, and B. G. Spratt.** 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. **2:**262-269.

5. **Brown, J. S., S. M. Gilliland, B. G. Spratt, and D. W. Holden.** 2004. A locus contained within a variable region of pneumococcal pathogenicity island 1 contributes to virulence in mice. Infect Immun. **72:**1587-1593.
6. **Brueggemann, A. B., R. Pai, D. Crook, and B. Beall.** 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog. **3:**e168.
7. **Carver, T. J., K. M. Rutherford, M. Berriman, R. MA, B. G. Barrell, and J. Parkhill.** 2005. ACT: the Artemis comparison tool. Bioinformatics **21:**3422-3423.
8. **Chi, F., O. Nolte, C. Bergmann, M. Ip, and R. Hakenbeck.** 2007. Crossing the barrier: evolution and spread of a major class of mosaic *pbp2x* in *Streptococcus pneumoniae*, *S. mitis* and *S. oralis*. Int J Med Microbiol. **297:**503-512.
9. **Coffey, T. J., C. G. Dowson, M. Daniels, J. Zhou, C. Martin, B. G. Spratt, and J. M. Musser.** 1991. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthesis genes, in natural populations of *Streptococcus pneumoniae*. Mol Microbiol. **5:**2255-2260.
10. **Corso, A., E. P. Severina, V. F. Petruk, Y. R. Mauriz, and A. Tomasz.** 1998. Molecular characterisation of penicillin-resistant *Streptococcus pneumoniae* isolates causing respiratory disease in the United States. Microb Drug Resist. **4:**325-337.
11. **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, B. Pichon, S. Baker, C. M. Parry, L. M. Lambertsen, D. Shahinas, D. R. Pillai, T. J. Mitchell, G. Dougan, A. Tomasz, K. P. Klugman, J. Parkhill, W. P. Hanage, and S. D. Bentley.** 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331:**430-434.
12. **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, S. D. Bentley, and T. J. Mitchell.** 2009. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae*^{Spain23F} ST81. J Bacteriol. **191:**1480-1489.
13. **Ding, F., P. Tang, M.-H. Hsu, P. Cui, S. Hu, J. Yu, and C.-H. Chiu.** 2009. Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. BMC Genet. **10:**158.
14. **Donati, C., N. L. Hiller, H. Tettelin, A. Muzzi, N. J. Croucher, S. V. Angiuoli, M. Oggioni, J. C. Dunning Hotopp, F. Z. Hu, D. R. Riley, A. Covacci, T. J. Mitchell, S. D. Bentley, M. Kilian, G. D. Ehrlich, R. Rappuoli, E. R. Moxon, and V. Masignani.** 2010. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. Genome Biol. **11:**R107.
15. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792-1797.
16. **Enright, M. C., and B. G. Spratt.** 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology. **144:**3049-3060.
17. **Fedson, D. S., J. Anthony, and G. Scott.** 1999. The burden of pneumococcal disease among adults in developed and developing countries: what is and is not known. Vaccine **17:**S11-S18.
18. **Feikin, D. R., and K. P. Klugman.** 2002. Historical changes in pneumococcal serogroup distribution: Implications for the era of pneumococcal conjugate vaccines. Clin Infect Dis. **35:**547-555.
19. **Fernández, A., C. Cabellos, F. Tubau, J. Liñares, P. F. Viladrich, and F. Gudiol.** 2001. Relationship between penicillin and cephalosporin resistance of *Streptococcus pneumoniae* strains and its inflammatory activity in the experimental model of meningitis. Med Microbiol Immunol. **190:**135-138.
20. **Golubchik, T., A. B. Brueggemann, T. Street, R. E. Gertz Jr, C. C. A. Spencer, T. Ho, E. Giannoulatou, R. Link-Gelles, R. M. Harding, B. Beall, T. E. A. Peto, M. R. Moore, P. Donnelly, D. W. Crook, and R. Bowden.** 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. Nat Genet. **44:**352-355.
21. **Hakenbeck, R., T. Briese, L. Chalkley, H. Ellerbrok, R. Kalliokoski, C. Latorre, M. Leinonen, and C. Martin.** 1991. Antigenic variation of penicillin-binding proteins from penicillin-resistant clinical strains of *Streptococcus pneumoniae.* J Infect Dis. **164:**313-319.
22. **Hakenbeck, R., T. Briese, L. Chalkley, H. Ellerbrok, R. Kalliokoski, C. Latorre, M. Leinonen, and C. Martin.** 1991. Variability of penicillin-binding proteins from penicillin-sensitive *Streptococcus pneumoniae*. J Infect Dis. **164:**307-312.

23. **Hakenbeck, R., T. Grebe, D. Zahner, and J. B. Stock.** 1999. β-lactam resistance in *Streptococcus pneumoniae*: penicillin-binding proteins and non-penicillin-binding proteins. Mol Microbiol. **33:**673-678.
24. **Hansman, D., and M. M. Bullen.** 1967. A resistant pneumococcus. Lancet **2:**264-265.
25. **Hiller, N. L., A. Ahmed, E. Powell, D. P. Martin, R. Eutsey, J. Earl, B. Janto, R. J. Boissy, J. Hogg, K. Barbadora, R. Sampath, S. Lonergan, J. C. Post, F. Z. Hu, and G. D. Ehrlich.** 2010. Generation of genetic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal paediatric infection. PLoS Pathog. **6:**e1001108.
26. **Jacobs, M. R., H. J. Koornhof, R. M. Robins-Browne, C. M. Stevenson, Z. A. Vermaak, I. Freiman, B. Miller, M. A. Witcomb, M. Isaäcson, J. I. Ward, and R. Austrian.** 1978. Emergence of multiply resistant pneumococci. N Engl J Med. **299:**735-740.
27. **Jolley, K. A., and M. C. Maiden.** 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. **10:**595.
28. **Kislak, J. W., L. M. B. Razavi, A. K. Daly, and M. Finland.** 1965. Susceptibility of pneumococci to nine antibiotics. Am J Med Sci. **250:**261-268.
29. **Klugman, K. P., T. J. Coffey, A. Smith, A. Wasas, M. Meyers, and B. G. Spratt.** 1994. Cluster of an erythromycin-resistant variant of the Spanish multiply resistant 23F clone of *Streptococcus pneumoniae* in South Africa. Eur J Clin Microbiol Infect Dis. **13:**171-174.
30. **Lambertsen, L., M. Brendstrup, H. Friis, and J. J. Christensen.** 2010. Molecular characterisation of invasive penicillin non-susceptible *Streptococcus pneumoniae* from Denmark, 2001 to 2005. Scand J Infect Dis. **42:**333-340.
31. **Lau, G. W., S. Haataja, M. Lonetto, S. E. Kensit, A. Marra, A. P. Bryant, D. McDevitt, D. A. Morrison, and D. W. Holden.** 2001. A functional genomic analysis of type 3 *Streptococcus pneumoniae* virulence. Mol Microbiol. **40:**555-571.
32. **Liñares, J., C. Ardanuy, R. Pallarés, and A. Fenoll.** 2010. Changes in antimicrobial resistance, serotypes and genotypes in *Streptococcus pneumoniae* over a 30-year period. Clin Microbiol Infect. **16:**402-410.
33. **Loeffler, J. M., and V. A. Fischetti.** 2006. Lysogeny of *Streptococcus pneumoniae* with MM1 phage: Improved adherence and other phenotypic changes. Infect Immun. **74:**4486-4495.
34. **Martin, C., C. Sibold, and R. Hakenbeck.** 1992. Relatedness of penicillin-binding protein 1a genes from different clones of *Streptococcus pneumoniae* isolated in South Africa and Spain. EMBO J. **11:**3831-3836.
35. **McDougal, L. K., J. K. Rasheed, J. W. Biddle, and F. C. Tenover.** 1995. Identification of multiple clones of extended-spectrum cephalosporin-resistant *Streptococcus pneumoniae* isolates in the United States. Antimicrob Agents Chemother. **39:**2282-2288.
36. **McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefévre, A. Tomasz, and K. P. Klugman.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. J Clin Microbiol. **39:**2565-2571.
37. **Muñoz-Elías, E. J., J. Marcano, and A. Camilli.** 2008. Isolation of *Streptococcus pneumoniae* biofilm mutants and their characterization during nasopharyngeal colonization. Infect Immun. **76:**5049-5061.
38. **Munoz-Najar, U., and M. N. Vijayakumar.** 1999. An operon that confers UV resistance by evoking the SOS mutagenic response in streptococcal conjugative transposon Tn*5252.* J Bacteriol. **181:**2782-2788.
39. **Muñoz, R., T. J. Coffey, M. Daniels, C. G. Dowson, G. Laible, J. Casal, R. Hakenbeck, M. Jacobs, J. M. Musser, B. G. Spratt, and A. Tomasz.** 1991. Intercontinental spread of a multiresistant clone of serotype 23F *Streptococcus pneumoniae.* J Infect Dis. **164:**302-306.
40. **Muñoz, R., C. G. Dowson, M. Daniels, T. J. Coffey, C. Martin, R. Hakenbeck, and B. G. Spratt.** 1992. Genetics of resistance to third-generation cephalosporins in clinical isolates of *Streptococcus pneumoniae*. Mol Microbiol. **6:**2461-2465.
41. **Obert, C., J. Sublett, D. Kaushal, E. Hinojosa, T. Barton, E. I. Tuomanen, and C. J. Orihuela.** 2006. Identification of a candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. Infect Immun. **74:**4766-4777.
42. **Pallarés, R., J. Liñares, M. Vadillo, C. Cabellos, F. Manresa, P. F. Viladrich, R. Martin, and F. Gudiol.** 1995. Resistance to penicillin and cephalosporin and mortality from severe pneumococcal pneumonia in Barcelona, Spain. N Engl J Med. **333:**474-480.

43. **Polissi, A., A. Pontiggia, G. Feger, M. Altieri, H. Mottl, L. Ferrari, and D. Simon.** 1998. Large-scale identification of virulence genes from *Streptococcus pneumoniae.* Infect Immun. **66:**5620-5629.

44. **Reichmann, P., E. Varon, E. Günther, R. R. Reinert, R. Lüttiken, A. Marton, P. Geslin, J. Wagner, and R. Hakenbeck.** 1995. Penicillin-resistant *Streptococcus pneumoniae* in Germany: genetic relationship to clones from other European countries. J Med Microbiol. **43:**377-385.

45. **Reinert, R. R., J. M. R, P. C. Appelbaum, S. Bajaksouzian, S. Cordeiro, M. van der Linden, and A. Al-Lahham.** 2005. Relationship between the original multiply resistant South African isolates of *Streptococcus pneumoniae* from 1977 to 1978 and contemporary international resistant clones. J Clin Microbiol. **43:**6035-6041.

45a. **Shi, Z.-Y., M. C. Enright, P. Wilkinson, D. Griffiths, and B. G. Spratt.** 1998. Identification of three major clones of multiply antibiotic-resistant *Streptococcus pneumoniae* in Taiwanese hospitals by multilocus sequence typing. J Clin Microbiol. **36:**3514-3519.

46. **Sibold, C., J. Henrichsen, A. Konig, C. Martin, L. Chalkley, and R. Hakenbeck.** 1994. Mosaic *pbpX* genes of major clones of penicillin-resistant *Streptococcus pneumoniae* have evolved from *pbpX* genes of a penicillin-sensitive *Streptococcus oralis*. Mol Microbiol. **12:**1013-1023.

47. **Sibold, C., J. F. Wang, J. Henrichsen, and R. Hakenbeck.** 1992. Genetic relatedness of penicillin-susceptible and -resistant *Streptococcus pneumoniae* strains isolated on different continents. Infect Immun. **60:**4119-4126.

48. **Stanhope, M. J., T. Lefébure, S. L. Walsh, J. A. Becker, P. Lang, P. D. Pavinski Bitar, L. A. Miller, M. J. Italia, and H. Amrine-Madsen.** 2008. Positive selection in penicillin-binding proteins 1a, 2b, and 2x from *Streptococcus pneumoniae* and its correlation with amoxicillin resistance development. Infect Genet Evol. **8:**331-339.

49. **Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar.** 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. **28:**2731-2739.

50. **Tomasz, A., A. Corso, Members of the PAHO/Rockefeller University Workshop:, E. P. Severina, G. Echániz-Aviles, M. C. de Cunto Brandileone, T. Camou, E. Castañeda, O. Figueroa, A. Rossi, and J. L. di Fabio.** 1998. Molecular epidemiologic characterisation of penicillin-resistant *Streptococcus pneumoniae* invasive paediatric isolates recovered in six Latin-American countries: An overview. Microb Drug Resist. **4:**195-207.

51. **Viladrich, P. F., C. Cabellos, R. Pallarés, F. Tubau, J. Martinez-Lacasa, J. Liñares, and F. Gudiol.** 1996. High doses of cefotaxime in treatment of adult meningitis due to *Streptococcus pneumoniae* with decreased susceptibilities to broad spectrum cephalosporins. Antimicrob Agents Chemother. **40:**218-220.

52. **Viladrich, P. F., F. Gudiol, J. Liñares, R. Pallarés, I. Sabaté, G. Rufí, and J. Ariza.** 1991. Evaluation of vancomycin for therapy of adult pneumococcal meningitis. Antimicrob Agents Chemother. **35:**2467-2472.

53. **Weinberger, D. M., R. Malley, and M. Lipsitch.** 2011. Serotype replacement in disease after pneumococcal vaccination. Lancet **378:**1962-1973.

54. **Whitney, C. G., M. M. Farley, J. Hadler, L. H. Harrison, N. M. Bennett, R. Lynfield, A. Reingold, P. R. Cieslak, T. Pilishvili, D. Jackson, R. R. Facklam, J. H. Jorgensen, and A. Schuchat.** 2003. Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. N Engl J Med. **348:**1737-1746.

54a. **Wolf, B., L. C. Rey, S. Brisse, L. B. Moreira, D. Milatovic, A. Fleer, J. J. Roord, and J. Verhoef.** 2000. Molecular epidemiology of penicillin-resistant *Streptococcus pneumoniae* colonizing children with community-acquired pneumonia and children attending day-care centres in Fortaleza, Brazil. J Antimicrob Chemother. **46:**757-765.

55. **World Health Organisation.** 2007. Pneumococcal conjugate vaccine for childhood immunisation - WHO position paper. Wkly Epidemiol Rec. **82:**93-104.

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

57. **Zerfass, I., R. Hakenbeck, and D. Denapaite.** 2009. An important site in PBP2x of penicillin-resistant clinical isolates of *Streptococcus pneumoniae*: mutational analysis of Thr338. Antimicrob Agents Chemother. **53:**1107-1115.

4.8 Supplementary Information

Limitations of the isolate sample and genome collection

As discussed in chapter 2, the nature of the historical isolate collection imposed limitations on the analyses. Of particular importance to the findings discussed here (in chapter 4) are those relevant to the representation of *pbp* gene diversity. The possibility that *pbp* nucleotide sequences other than those associated with PMEN1, were similarly transferred between older PNSP which were not represented here, or modern PNSP representatives from less-well sampled geographic locations (e.g. Asia and South America) cannot be excluded. However, the latter possibility is unlikely given that other studies have demonstrated a predominance of internationally disseminated pneumococcal clones, representatives of which were included here, among PNSP from Asia and South America [45a, 54a]. Even if the former possibility were true, the contribution of such *pbp* sequences in the dissemination of penicillin-nonsusceptibility amongst modern pneumococcal clones does not seem to have been as important as that of PMEN1.

Another important limitation which should be considered is the extent to which the genome collection described here represents the genetic diversity of the wider pneumococcal population. This point is particularly relevant when considering the certainty with which genomic regions acquired by CGSP14 and PMEN3 were acquired from PMEN1 (also see below). Isolates were selected for genomic sequencing, in a way that was thought to maximise the amount of genetic diversity that was represented. The selected isolates represented numerous CCs and were distributed

through time / by geographic location (as much as possible), whilst also including sufficient isolates to allow intra-CC comparisons (e.g. for the analyses presented in chapters 5 and 6). It would be impossible for any collection of genomes to represent the full extent of pneumococcal genetic diversity because it is not possible to collect data for all pneumococci. The sample described here is diverse, as evidenced by the number of serotypes and CCs which it represents. Nevertheless, had it been possible to do so, the sample would have been improved by inclusion of a greater number of isolates from diverse geographic locations and pneumococci isolated from nasopharyngeal carriage in addition to those isolated from IPD.

MLST and pbp diversity

232 unique STs and 163 CCs were represented by isolates in this study. (Note that after repeated attempts using multiple primer sets, the *ddl* sequence for isolate SA82 (1989, CC124[14]) could not be determined, presumably due to sequence divergence within the primer binding region.) Among the 389 isolates selected for *pbp* sequencing plus the 16 genomes, 401, 389 and 404 *pbp2x, pbp1a* and *pbp2b* sequences, respectively, were obtained (a full complement of sequences could not be obtained for all isolates due to PCR and/or sequencing difficulties most likely also arising from sequence divergence in the primer binding regions). Among these sequences 142, 105 and 127 unique *pbp2x, pbp1a* and *pbp2b* alleles were identified, respectively, and of these 64 (45.1%), 34 (32.4%) and 59 (46.5%) were altered (i.e. >1% divergent from the penicillin-susceptible R6 reference strain dated 1964). There was no evidence of PNSP or altered *pbps* prior to 1967. 18/64 *pbp2x*, 6/34 *pbp1a* and 17/59 *pbp2b* altered alleles contained regions which spanned ≥1 active site coding region and were

identical or highly similar (≤2 nucleotide substitutions) to those of the PMEN1 reference strain (Supplementary Tables 4.4-4.6).

pbp sequences from 233 of 237 (98.3%) PSP were ≤1% divergent from those of R6. *pbp2x* sequences from four PSP were between 3.7% and 9.6% divergent from those of R6. Three of these isolates were PMEN reference strains (PMEN21, PMEN30 and PMEN40) and the fourth was a CC66[14] representative (USA11). To our knowledge, possession of altered *pbp* sequences without increased penicillin MIC has not been noted previously for pneumococci. One possible explanation for this is that the penicillin MIC was not increased for these pneumococci due to the lack of altered *pbp1a* and/or *pbp2b* sequences and/or favourable alleles at other loci believed to play a role in penicillin resistance. Alternatively, *pbp2x* sequence divergence may have followed acquisition and recombination of DNA from penicillin-susceptible viridans streptococci, resulting in sequence mosaics lacking the appropriate nucleotide changes to confer resistance.

pbp diversity within the PMEN1 lineage

Within the PMEN1 lineage itself there was some variation of *pbp* alleles: one, six and two of the 28 isolates in our collection possessed alternative *pbp2x, pbp1a* and/or *pbp2b* alleles respectively (i.e. these alleles were not identical to those of the PMEN1 reference strain). However, all but one of these alleles (the *pbp2x* allele) did contain region(s) which were identical or highly similar to those of the reference strain alleles.

CGSP14 and PMEN3 reference imported regions: search for identical regions among other pneumococci.

In order to assess whether the putatively imported regions of the CGSP14 and PMEN3 reference genomes could have in fact been acquired from a non-PMEN1 representative pneumococcus, we searched for these regions among our collection of pneumococcal genomes. While it is not possible to represent all potential pneumococcal genetic diversity, our collection does represent a diverse range of CCs including many of the major global clones. Genome Comparator was used to identify coding sequences identical to those of the PMEN1 reference genome, the subsequent output was ordered by position in the PMEN1 reference chromosome and viewed in a spread sheet format. None of the putative imported regions were identical or highly similar across their whole length among other pneumococci in our collection, excluding isolate 23F/4 and CC66 representatives. In some cases similarities over parts of the regions could be identified, and we cannot exclude the possibility that these 'sub-regions' could have been acquired by CGSP14 or PMEN3 from a non-PMEN1 representative pneumococcus. However, if this were true the remaining 'sub-region(s)' would need to have been acquired independently via secondary recombination event(s). In such an instance the most parsimonious solution, which we favour, is that the region was acquired as a single whole from a PMEN1 representative.

It is also not possible to exclude the possibility that other non-PMEN1 representative pneumococci, which were not represented within our collection, were the true donors of the CGSP14 and PMEN3 putative imported regions. However, unless such a pneumococcus was very closely related to PMEN1, it seems unlikely that it would

possess highly similar nucleotide sequence across all of the CGSP14 and PMEN3 imported genome regions, particularly given that the genome comparator analyses suggested a mean percentage PMEN1-identical coding sequence level of only ~15% among unrelated pneumococci. Any contribution from other pneumococci would thus be minor in comparison to that from PMEN1 representatives, and would not greatly alter our conclusions.

Isolate 23F/4 is an ancestral representative of PMEN1 and thus it is not surprising that some of the CGSP14 and PMEN3 putative imported regions may be highly similar in the 23F/4 genome (Supplementary Figure 4.2). The 23F/4 ST has not been reported among other pneumococci, suggesting that it is now rare. In contrast, following its emergence the PMEN1 ST became highly prevalent. CGSP14 and PMEN3 were isolated in 2004/5 and 1993, respectively, and were therefore more likely to have come into contact with, and acquired DNA from, a PMEN1 representative than a 23F/4 representative. Furthermore, the PMEN1, CGSP14 and PMEN3 *pbp* genes were not identical to those of 23F/4, demonstrating that these genomic regions cannot have been acquired from 23F/4.

CC66 was considered to be closely related to PMEN1 (see below) and thus it is also not surprising that regions of the genomes of CC66 members were highly similar to the CGSP14 and PMEN3 putative imported regions. However, when the nucleotide sequences of the regions in question were viewed, the PMEN1 reference was clearly the best match in all but two and three cases respectively, supporting the hypothesis that the majority of the described regions were most likely acquired from a PMEN1

representative. Regarding the remaining five regions, we cannot be certain whether they were acquired from PMEN1 or CC66, but since these are the minority we do not believe that this result greatly affects our conclusions.

Finally, to assess whether our analyses could be biased by use of PMEN1 as the Genome Comparator reference, we compared PMEN1, CGSP14, PMEN3 and all available members of their respective CCs using a CC66 representative (JJA) as the reference. We identified 104 and 40 coding sequences within the CGSP14 and PMEN3 genomes, respectively, which were identical to those of JJA but differed from those of other members of their respective CCs. 12 and 7 of these coding sequences, respectively, also differed from those of PMEN1. Using the same output we identified 197 and 109 coding sequences within the CGSP14 and PMEN3 genomes, respectively which were identical to those of PMEN1 but differed from those of other members of their respective CCs. 105 and 76 of these, respectively also differed from those of JJA. Therefore, the extent to which CGSP14 and PMEN3 may have acquired DNA from CC66 representatives is likely much less than that acquired from PMEN1 representatives, and supports our previous findings.

A high proportion of coding sequences identified from CC66 representatives were identical to those of the PMEN1 reference strain.

Nine isolates from our collection represented CC66 and were selected for whole-genome sequencing. The genome of an additional CC66 isolate, JJA (unknown year of isolation but presumed from 1990 onwards, ST66[14]), was retrieved from Genbank (accession CP000919.1). Together, these isolates represented five independent

serotypes (7B (n = 1), 23F (n = 1), 9N (n = 2), 19F (n = 2) and 14 (n = 4)), and five different STs (67 (n = 1), 7180 (n = 1), 8119 (n = 1), 71 (n = 2) and 66 (n = 6)). Among the CC66 isolate coding sequences identified by Genome Comparator, between 44.5% and 58.4% were shared identically with the PMEN1 reference. The three highest percentages (57.3%, 58.4% and 55.0%, respectively) were associated with the three oldest isolates (7B/2 (1952, ST7180^{7B}), 9N/6 (1960, ST71^{9N}) and 19F/11 (1972, ST71^{19F}), respectively. Two of these isolates pre-dated the first identified PMEN1 representative (from 1984) and the predicted emergence of the PMEN1 clone (~1970), thus this observation is consistent with our suggestion that the similarities between these clones are due to ancestral descent rather than recombinogenic transfer of genetic material from PMEN1. However, PMEN18, a member of CC66, possessed *pbp2b* sequence regions highly similar to those of PMEN1 and these *pbp2b* regions were more likely to have been acquired through recent recombination events, owing to their absence among older CC66 isolates.

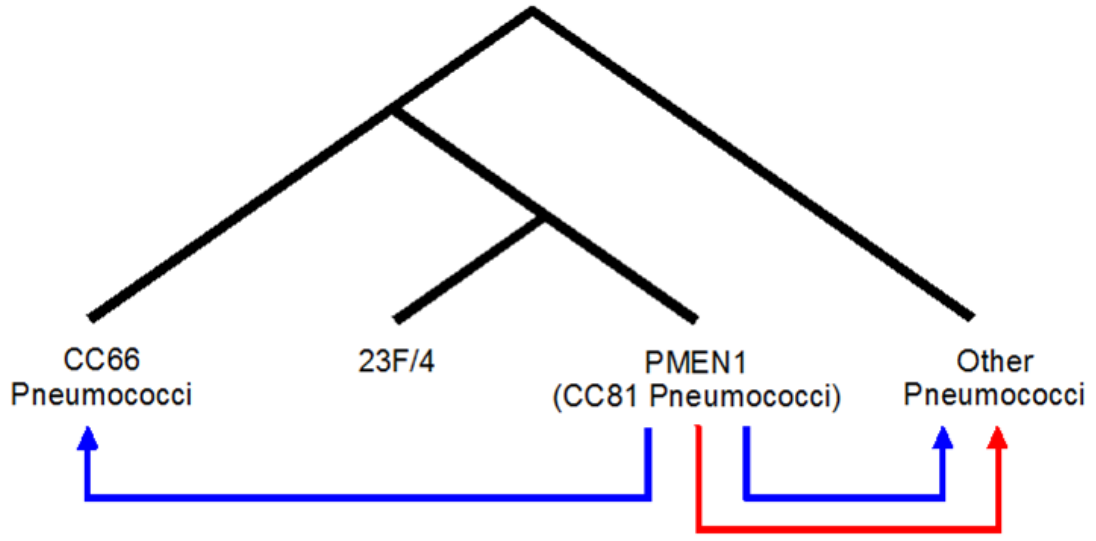
An alternative explanation for the increased number of identical coding sequences between PMEN1 and CC66 compared to other CCs, could be that the ancestor of these CC66 isolates acquired identical copies of the large PMEN1 reference-associated mobile genetic elements from an independent source. Such an acquisition would artificially inflate the percentages of identical coding sequences reported here; however, we did not find any evidence in support of this theory. The largest of the PMEN1 reference-associated elements was ICE*Sp*ST8123F, which consisted of 73 coding sequences. Only 5 and 11 of these coding sequences were identified identically, from isolates USA11 (Unknown year of isolation but presumed from 1990

onwards, ST66[14]) and PMEN18 (1997, ST67[14]), respectively. The other large PMEN1 reference mobile elements included the ϕMM1 phage (50 coding sequences) and the *psrP* element (20 coding sequences). The ϕMM1 phage was also largely absent from the CC66 isolates, with only one, three and six identical coding sequences identified from USA11, 7B/2 and USA12 (2001, ST66[23F]), respectively. Furthermore, only one to four PMEN1 reference *psrP* element identical coding sequences were identified from 9 of the 10 CC66 isolates, with none identified from the tenth isolate.

```
                          1111122222 2222222222 2222233333 3333333333 3333344444 4555555555
      3444444556 6666666679 0333400468 8888888999 9999900000 0011112233 3333414555 8233444457
      7345667351 2256788809 6669424121 1378899012 5667744456 6755573711 1167394189 7556235985
      3265476314 0398403958 8035356281 7511567432 9271637842 6303671735 6779204087 3640074361
PMEN3 CCTAAACCAC TTTGCGACAA AGACGCAATC TCGGAACTGT TGTAAAACGT CTGCCTCTGG TTGTGTTAAC TTGTACGTTT
9A/1  .......... .......... .CGTATGGCT CTAACGACTC CACTGGCAAC TCATTCTCAA AATCACCGGT CCACGTAGCC
9V/5  .......... .......... .CGTA-GGCT CTAAC.ACTC CACTGGCAAC TCATTCTCAA AATCACCGGT CCACGTAGCC
9V/6  .......... .......... .CGTATGGCT CTAAC.ACTC CACTGGCAAC TCATTCTCAA AATCACCGGT CCACGTAGCC
PMEN1 TTCTGGAAGT GACATATTGG G......... .......... .......... .......... .......... ..........

                                                                                  111111 1111111111
      5555555555 5555666666 6666666666 6777777777 7777777777 7788888899 9999011111 1111112222
      7788889999 9999113344 4445559999 9000000011 1111111222 2215888923 4588914444 4558880011
      7902771234 5569172334 4480032388 9011223324 4467778134 4655699347 6007616666 6347770047
      8383486903 3760086510 3681337647 9824692350 3640392932 5870639581 2991180135 8130175752
PMEN3 GGTTAAAGCA CGTCCTATAC ATTGATTAAA CTTAATCATC ACGTTAGGCA AAGCGATGGC AGATCGGCGT CATACCGTTT
9A/1  AACGCCGAGC ATCATCCCTT TCGATCCCTT TCCCTCACGT GAACCTAATT GGT....... .......... ..........
9V/5  AACGCCGAGC ATCATCCCTT TCGATCCCTT TCCCTCACGT GAACCTAATT GGT....... .......... ..........
9V/6  AACGCCGAGC ATCATCCCTT TCGATCCCTT TCCCTCACGT GAACCTAATT GGT....... .......... ..........
PMEN1 .......... .......... .......... .......... .......... ...TAGCTAA GAGCTACATA GGAGAAACCC
```

Supplementary Figure 4.1 MEGA5 variable site output showing the PMEN3 reference strain putative imported region 4 and flanking regions.
Regions from the genomes of the PMEN1 reference strain, the PMEN3 reference strain and the PMEN3 ancestral representatives, 9A/1, 9V/5 and 9V/6 are shown. Numbers indicate variable site positions within the alignment. Periods (".") indicate bases identical to the PMEN3 reference at a given site. Dashes ("-") indicate missing bases at a given site. Red brackets indicate the putative imported region.



Supplementary Figure 4.2 Idealised representation of the pneumococcal lineage relationships inferred from this study.
Blue arrows represent horizontal transfer of *pbp* nucleotide sequences between lineages. Red arrows represent horizontal transfer of other genomic regions between the PMEN1 lineage and alternate pneumococcal lineages (CC15 and CC156/162 described as examples in text).

Supplementary Table 4.1 Summary of isolates included in this study stratified by era of isolation.

		Period of Isolation									
		1916	**1937-39**	**1940s**	**1950s**	**1960s**	**1970s**	**1980s**	**1990s**	**2000s**	**Unknown**	**Total**
Serotype	1			4	1	1			5	1		12
	2	1		1	2							4
	3					5	2	1		1		9
	4				1	4			2			7
	5					1	1	2	1		1	6
	6						1					1
	6A				1		13	7	3	1		25
	6B		1					18	6			25
	6C									1		1
	7A		1			1						2
	7B				1				1			2
	7C		1				1					2
	7F				1	2		2	9	2		16
	8				1	3		1				5
	9A					1						1
	9L			1	1	1						3
	9N		1		1	3			1	1		7
	9V		1			1		6	3			11
	10A		1									1
	10B							1				1
	10C							1				1
	10F				1							1
	11A		1					1			1	3
	11B			1								1
	11C				1							1
	11D							1				1
	11F				1							1
	12A					12						12

12B						1				1
12F				3		1	4		1	9
13			1							1
14	1		1	3	5	7	19	13	1	50
15A	1						1			2
15B	1					1	1			3
15C				1						1
15F			1	1						2
16A						1				1
16F			1							1
17A					1					1
17F	1		1	2						4
18A			1							1
18B		1								1
18C	1	1		1		1	4	2		10
18F		1		1						2
19A	1		1	1	16	15	1	1		36
19B					1					1
19C	1									1
19F			3	5	1	4	5	1		19
20	1	1		1						3
21			1	1			1			3
22A	2									2
22F		1				1				2
23A		1								1
23B		1					1			2
23F		1		1	1	22	16	1		42
24A		1								1
24B		1								1
24F		1	1							2
25A						2				2

25F				1	2					3
27				1						1
28A			1							1
28F				1						1
29				1			3			4
31				2	1					3
32A			1							1
32F				1						1
33A	1		1							2
33B					1					1
33C			1							1
33D						2				2
33F	1									1
34				1	1	2				4
35A	1									1
35B	1				1	1		1		4
35C			2							2
35F	1									1
36	1					1		2		4
37			1		1	4		1		7
38	1						1			2
39			1							1
40				1						1
41A						2				2
41F			1							1
42						1				1
43			1	1						2
44				1						1
45			2	2						4
46				1						1
47A						1				1

	47F						1					1
	48				3	2						5
	NT					1		2	1			4
Penicillin	S		24	30	41	63	26	24	45	19	1	273
Susceptibility*	I					2	7	57	24	4		94
	R						25	21	19	2		67
	Unknown	1			1	1		1	1		3	8
Grand Total		1	**24**	**30**	**42**	**66**	**58**	**103**	**89**	**25**	**4**	**442†**

* S = Susceptible, MIC ≤ 0.06 µg/ml, I = Intermediately resistant, MIC 0.12 - 1 µg/ml, R = Resistant, MIC ≥ 2.0 µg/ml

† Including strains for which genomic sequences were retrieved from Genbank (n = 16).

Supplementary Table 4.2 Primers for PCR and conventional sequencing of regions of *pbp2x, pbp1a* and *pbp2b.*

See Chapter 2, Table 2.3.

Supplementary Table 4.3 Isolates for which whole-genome sequences were analysed in this study.

See Appendices 1.1 and 1.2.

Supplementary Tables 4.4-4.6 *pbp* allele sequence regions identical or highly similar to those of the PMEN1 reference strain.

pbp2x				
Allele #	PMEN1 Matching Regions		Nucleotide Substitution Positions (bp)	CC Representatives
	Start (bp)	End (bp)		
67	NA (PMEN1)		NA	15, 81, 156/162
75	298	End	None	18
80	298	1331	None	90
85	317	1526	None	1094, None41, Singleton7248
88	343	End	None	90
89	310	End	None	242
86	253	End	291	271/320
107	169	1604	618	558
117	298	1331	1132	90
119	298	1604	735	22
121	298	1331	596	90
68	259	1514	735, 753	87/88
77	253	End	618, 1199	490, 2090
99	138 335	284 End	None 617, 1199	384
100	442	End	618, 1199	230, 304
118	439	End	1369, 1374	Singleton7256
129	169	End	618, 1199	1094
141	181	End	618, 1662	1106

pbp1a				
Allele #	PMEN1 Matching Regions		Nucleotide Substitution Positions (bp)	CC Representatives
	Start (bp)	End (bp)		
72	NA (PMEN1)		NA	15, 81, 156/162
77	62 464	225 End	None None	18, 81, 156/162
78	Start	349	None	2090
79	Start 464	225 End	None None	81, 90
101	Start 464	225 669	None None	384
86	71 635	349 End	222 None	90
109	62 464	225 End	None 876	81

pbp2b				
Allele #	PMEN1 Matching Regions		Nucleotide Substitution Positions (bp)	CC Representatives
	Start (bp)	End (bp)		
62	NA (PMEN1)		NA	15, 81, 156/162
61	Start	232	None	87/88
71	Start	232	None	Singleton175
73	Start	217	None	15, 384
106	Start	232	None	90
111	Start	232	None	1106
130	Start	232	None	1094
76	Start	232	74	2090
79	Start	232	74	490
81	Start	163	74	242
85	Start	232	74	230
86	Start 392	232 514	74 None	271/320
87	Start 394	232 514	74 None	66
95	Start	232	74	344
96	Start 396	232 514	74 None	558
104	Start 396	232 514	74 None	81
105	Start	199	74	22, 490
129	50	232	74	5431

Supplementary Table 4.7 Regions of the CGSP14 genome identified as putatively recombinogenic and acquired from PMEN1.

Region	PMEN1 reference Coordinates		Size (bp)	Included CDS	CDS Predicted Function	CDS in Whole or Part
	Start	End				
1	285184	303926	18743	*agaW*	N-acetylgalactosamine-specific phosphotransferase system (PTS), IIC component	Part
				agaD	N-acetylgalactosamine-specific phosphotransferase system (PTS), IID component	Whole
				SPN23F_03000	Hypothetical protein	Whole
				SPN23F_03010	Heparinase II/III-like protein	Whole
				SPN23F_03020	LacI family transcriptional regulator	Whole
				SPN23F_03030	Membrane protein	Whole
				SPN23F_03040	Hypothetical protein	Whole
				SPN23F_03050	DNA-binding protein	Whole
				mraW	S-adenosyl-methyltransferase MraW	Whole
				ftsL	Cell division protein	Whole
				pbp2x	Penicillin-binding protein 2x	Whole
				mraY	Phospho-N-acetylmuramoyl-pentapeptide-transferase	Whole (1 nucleotide substitution at position 295551)
				clpL	ATP-dependent protease ATP-binding subunit ClpL	Whole
				SPN23F_03120	Group II intron maturase	Whole
				luxS	S-ribosylhomocysteinase	Whole
				SPN23F_03140	Hypothetical protein	Whole
				dexB	Glucan 1,6-alpha-glucosidase	Whole
				wzg | cpsA	Capsule biosynthesis integral membrane regulatory protein Wzg	Part
2	330194	351408	21215	SPN23F_03400	Cell wall surface anchored protein	Part
				pbp1A	Penicillin-binding protein 1A	Whole
				recU	Holliday junction-specific endonuclease	Whole
				SPN23F_03430	Hypothetical protein	Whole
				SPN23F_03440	DivIVA protein	Whole

				SPN23F_03450	RNA methylase family protein	Whole
				SPN23F_03460	Hypothetical protein	Whole (1 nucleotide substitution at position 338114)
				gnd	6-phosphogluconate dehydrogenase	Whole
				csrR	Response regulator protein	Whole
				cbpJ	Choline binding protein J	Whole
				SPN23F_03500	Major facilitator superfamily protein	Whole
				mvaK1	Mevalonate kinase	Whole
				mvaD	Mevalonate diphosphate decarboxylase	Whole
				mvaK2	Phosphomevalonate kinase	Whole
				fni	Isopentenyl pyrophosphate isomerase	Whole
				SPN23F_03570	Hypothetical protein	Whole
				SPN23F_03580	Sensor histidine kinase	Whole
				SPN23F_03590	Response regulator protein	Whole
3	554668	584068	29401	*rplK*	50S ribosomal protein L11	Whole
				rplA	50S ribosomal protein L1	Whole
				SPN23F_05720	Hypothetical protein	Whole
				SPN23F_05730	Hypothetical protein	Whole
				SPN23F_05740	ABC transporter ATP-binding protein	Whole
				SPN23F_05750	Hypothetical protein	Whole
				SPN23F_05760	Hypothetical protein	Whole (1 nucleotide substitution at position 560786)
				SPN23F_05770	Hypothetical protein	Whole
				SPN23F_05780	Hypothetical protein	Whole
				SPN23F_05790	Surface-anchored serine protease	Whole
				SPN23F_05800	Sugar phosphotransferase system (PTS), IIA component	Whole
				SPN23F_05810	Sugar phosphotransferase system (PTS), IIB component	Whole
				SPN23F_05820	Sugar phosphotransferase system (PTS), galactitol-specific family, IIC component	Whole

				bgaA	Surface anchored beta-galactosidase	Whole
				SPN23F_05840	Hypothetical protein	Whole
				SPN23F_05850	Hypothetical protein	Whole
				SPN23F_05860	Hypothetical protein	Whole
				SPN23F_05890	rRNA methylase	Whole
				SPN23F_05900	Hypothetical protein	Whole (1 nucleotide insertion in CGSP14 at position 581754)
				SPN23F_05910	Sodium hydrogen exchange transporter	Whole
4	631008	639926	8919	*lctO*	L-lactate oxidase	Whole
				SPN23F_06400	Coenzyme	Whole
				thiM	Hydroxyethylthiazole kinase	Whole
				thiE	Thiamine-phosphate pyrophosphorylase	Whole
				SPN23F_06440	Hypothetical protein	Whole
				SPN23F_06450	ABC transporter ATP-binding protein	Whole
				SPN23F_06460	Hypothetical protein	Whole
				SPN23F_06470	Coenzyme	Whole
				SPN23F_06480	Hypothetical protein	Whole
				SPN23F_06490	Hydroxyethylthiazole kinase	Part
5	640607	642269	1663	*thiE*	Thiamine-phosphate pyrophosphorylase	Part
				thiD	Phosphomethylpyrimidine kinase	Whole
6	642471	649059	6589	SPN23F_06520	Regulator	Part
				SPN23F_06530	Hypothetical protein	Whole
				SPN23F_06540	Metal transporting P-type ATPase	Whole
				spxB	Pyruvate oxidase	Whole
				SPN23F_06560	Glyoxalase family protein	Whole
7	934640	942176	7537	SPN23F_09650	Hypothetical protein	Part
				SPN23F_09660	Alpha-amylase	Whole
				SPN23F_09670	Hypothetical protein	Whole
						1 nucleotide substitution

						at position 938232
				SPN23F_09700	Epsilon antitoxin	Whole
				SPN23F_09710	Zeta-toxin	Whole
				SPN23F_09731	Hypothetical protein	Whole
				SPN23F_09740	*Tn5252*, ORF 10 protein	Whole
				SPN23F_09750	*Tn5252*, ORF 9 protein	Part
8	1207734	1231897	24164	SPN23F_12410	Integrase	Whole
				SPN23F_12420	Hypothetical protein	Whole
				SPN23F_12430	Relaxase	Whole
				SPN23F_12440	Mobilisation protein	Whole
				SPN23F_12450	Mobilisation protein	Whole
				SPN23F_12460	Hypothetical protein	Whole
				SPN23F_12470	DNA helicase II, UvrD	Whole
				SPN23F_12480	Hypothetical protein	Whole
				SPN23F_12490	Hypothetical protein	Whole
				SPN23F_12500	Hypothetical protein	Whole
				SPN23F_12510	NTPase protein	Whole
				SPN23F_12520	Hypothetical protein	Whole
				SPN23F_12530	Phosphoserine phosphatase	Whole
				SPN23F_12540	Hypothetical protein	Whole
				SPN23F_12550	Hypothetical protein	Whole
				SPN23F_12560	Hypothetical protein	Whole
				SPN23F_12570	Hypothetical protein	Whole
				rep	Replication protein	Whole
				SPN23F_12590	Chloramphenicol acetyltransferase	Whole
				SPN23F_12600	Hypothetical protein	Whole
				SPN23F_12610	Hypothetical protein	Whole
				SPN23F_12620	Zeta toxin	Whole
				SPN23F_12630	Epsilon antitoxin	Whole
9	1257543	1270330	12788	SPN23F_12840	Conjugal transfer protein	Part
				SPN23F_12850	Conjugal transfer protein	Whole

				SPN23F_12860	Hypothetical protein	Whole
				SPN23F_12870	Hypothetical protein	Whole
				SPN23F_12880	Hypothetical protein	Whole
				traG	Conjugal transfer protein TraG	Whole
				SPN23F_12900	Hypothetical protein	Whole
				SPN23F_12910	Hypothetical protein	Whole
				SPN23F_12920	Hypothetical protein	Whole
				SPN23F_12930	Hypothetical protein	Whole
				SPN23F_12940	Conjugative transposon protein	Whole
				SPN23F_12950	Conjugative transposon protein	Whole
				SPN23F_12960	Conjugative transposon FtsK/SpoIIIE-family protein	Whole
				SPN23F_12970	Conjugative transposon replication initiation factor	Part
10	1280998	1288962	7965	SPN23F_13060	Conjugative transposon regulatory protein	Part
				SPN23F_13061	Hypothetical protein	Whole
				SPN23F_13070	Conjugative transposon regulatory protein	Whole
				xis	Excisionase	Whole
				SPN23F_13110	CAAX amino terminal protease	Whole
				SPN23F_13120	Hypothetical protein	Whole
				SPN23F_13130	Hypothetical protein	Whole
				SPN23F_13140	Hypothetical protein	Whole
				SPN23F_13150	DNA methylase	Whole
				SPN23F_13160	Replication initiator protein	Whole
				SPN23F_13170	Hypothetical protein	Whole
				rplL	50S ribosomal protein L7/L12	Part
11	1526952	1559263	32312	SPN23F_15810	Hypothetical protein	Whole
				SPN23F_15820	Hypothetical protein	Whole
				SPN23F_15830	Hypothetical protein	Whole
				SPN23F_15840	Endoribonuclease L-PSP family protein	Whole
				engB	Ribosome biogenesis GTP-binding protein YsxC	Whole
				clpX	ATP-dependent protease ATP-binding subunit ClpX	Whole
				SPN23F_15870	Hypothetical protein	Whole

dyr	Dihydrofolate reductase	Whole
dpr	Dps-like peroxide resistance protein Dpr	Whole
SPN23F_15900	Exported choline-binding glycosyl hydrolase	Whole
tpiA	Triosephosphate isomerase	Whole (1 nucleotide substitution at position 1536099)
SPN23F_15920	DNA replication protein DnaD	Whole
metA	Homoserine O-succinyltransferase	Whole
apt	Adenine phosphoribosyltransferase	Whole
SPN23F_15950	Methyltransferase	Whole
SPN23F_15960	Hypothetical protein	Whole
msmK	Multiple sugar-binding transport ATP-binding protein	Whole
SPN23F_15990	Isochorismatase family protein	Whole
codY	Transcriptional repressor CodY	Whole
SPN23F_16010	DEAD/DEAH box helicase	Whole
SPN23F_16020	Major facilitator superfamily protein	Whole
SPN23F_16030	Pyridine nucleotide-disulfide oxidoreductase	Whole
SPN23F_16040	Mur ligase family protein	Whole
SPN23F_16050	Aminotransferase	Whole
pepQ	Xaa-Pro dipeptidase	Whole
SPN23F_16070	GNAT family acetyltransferase	Whole
SPN23F_16080	IS3-Spn1 ORF B	Whole (1 nucleotide substitution at position 1551575)
SPN23F_16100	Hypothetical protein	Whole
thiD	Phosphomethylpyrimidine kinase	Whole
truA	tRNA pseudouridine synthase A	Whole
SPN23F_16140	Major facilitator superfamily protein	Whole
SPN23F_16150	Hypothetical protein	Whole
SPN23F_16160	PhnA protein	Whole
cmk	Cytidylate kinase	Whole

				SPN23F_16180	Hypothetical protein	Whole
				SPN23F_16190	Hypothetical protein	Whole
				SPN23F_16200	Membrane glycosyl transferase	Part
12[ψ]	1584029	1589382	5354	SPN23F_16480	Metallo-beta-lactamase superfamily protein	Whole
				pepO	Endopeptidase O	Whole
				mtsB	Metal cation ABC transporter ATP-binding protein	Whole
				mtsC	Metal cation ABC transporter membrane protein	Whole
				mtsA	Metal ABC transporter substrate-binding lipoprotein precursor	Part
13	1612934	1618295	5362	*pbp2b*	Penicillin-binding protein 2b	Part
				SPN23F_16750	Transcription regulator	Whole
				SPN23F_16760	ROK family protein	Whole
				nanH	N-acetylneuraminate lyase	Whole
				SPN23F_16780	Hypothetical protein	Whole
14[ψ]	1994357	2005122	10766	*tkt*	Transketolase	Part
				SPN23F_20520	L-ascorbate 6-phosphate lactonase	Whole
				SPN23F_20530	Transcriptional antiterminator	Whole
				araD	L-ribulose-5-phosphate 4-epimerase	Whole
				ulaD	3-keto-L-gulonate-6-phosphate decarboxylase	Whole
				SPN23F_20580	Sugar phosphotransferase system (PTS), IIA component	Whole
				SPN23F_20590	Sugar phosphotransferase system (PTS), lactose/cellobiose-specific family, IIB subunit protein	Whole
				SPN23F_20600	PTS system ascorbate-specific transporter subunit IIC	Whole
				SPN23F_20610	Hypothetical protein	Whole
				SPN23F_20620	ssDNA-binding protein	Whole
				SPN23F_20630	Membrane protein OxaA 1 precursor	Whole
				rnpA	Ribonuclease P	Part
15	2005748	2024932	19185	*ackA*	Acetate kinase	Part
				SPN23F_20670	Hypothetical protein	Whole
				SPN23F_20700	Hypothetical protein	Whole
				SPN23F_20710	Hypothetical protein	Whole
				SPN23F_20720	Hypothetical protein	Whole

SPN23F_20730	Competence protein	Whole
comYC	Competence protein	Whole
SPN23F_20750	Competence protein	Whole
SPN23F_20760	Competence protein	Whole
SPN23F_20770	Hypothetical protein	Whole
SPN23F_20780	Zinc-binding alcohol dehydrogenase	Whole
SPN23F_20790	N-acetylglucosamine-6-phosphate deacetylase	Whole
SPN23F_20800	Acyltransferase family protein	Whole
		1 nucleotide substitution at position 2017038
tgt	Queuine tRNA-ribosyltransferase	Whole
SPN23F_20820	Hypothetical protein	Whole
SPN23F_20830	Pyrrolidone-carboxylate peptidase	Whole
SPN23F_20840	Hypothetical protein	Whole
SPN23F_20850	MarR family transcriptional regulator	Whole
SPN23F_20870	Hypothetical protein	Whole
SPN23F_20880	Haloacid dehalogenase-like hydrolase	Whole
SPN23F_20890	Multi antimicrobial extrusion (MATE) family transporter	Whole
SPN23F_20900	Threonine synthase	Part

CDS = coding sequences. $^{\Psi}$ Region also identical or highly similar in one or more CC66 representative(s).

Supplementary Table 4.8 Regions of the PMEN3 genome identified as putatively recombinogenic and acquired from PMEN1.

Region	PMEN1-reference Coordinates		Size (bp)	Included CDS	Predicted Function	CDS in Whole or Part
	Start	End				
1	272645	294748	22104	SPN23F_02870	Glycosylhydrolase	Part
				SPN23F_02880	Glutathione peroxidase	Whole
				SPN23F_02890	Hyaluronate lyase	Whole
				kgdA	Keto-hydroxyglutarate-aldolase/keto-deoxy- phosphogluconate aldolase	Whole
				SPN23F_02920	PfkB family carbohydrate kinase	Whole
				SPN23F_02930	Hypothetical protein	Whole
				idnO	Gluconate 5-dehydrogenase	Whole
						1 nucleotide substitutions at positions 282302 1 nucleotide insertion in PMEN3 at position 282417
				SPN23F_02950	N-acetylgalactosamine-specific phosphotransferase system (PTS), IIA component	Whole
				ugl	Unsaturated glucuronyl hydrolase	Whole
				agaV	N-acetylgalactosamine-specific phosphotransferase system (PTS), IIB component	Whole
				agaW	N-acetylgalactosamine-specific phosphotransferase system (PTS), IIC component	Whole
				agaD	N-acetylgalactosamine-specific phosphotransferase system (PTS), IID component	Whole
				SPN23F_03000	Hypothetical protein	Whole
				SPN23F_03010	Heparinase II/III-like protein	Whole
				SPN23F_03020	LacI family transcriptional regulator	Whole
				SPN23F_03030	Membrane protein	Whole
				SPN23F_03040	Hypothetical protein	Whole
				SPN23F_03050	DNA-binding protein	Whole
				mraW	S-adenosyl-methyltransferase MraW	Whole

				ftsL	Cell division protein	Whole
				pbp2x	Penicillin-binding protein 2x	Whole
				mraY	Phospho-N-acetylmuramoyl-pentapeptide- transferase	Part
2	324107	346147	22041	*aliA*	Extracellular oligopeptide-binding protein	Part
				SPN23F_03400	Cell wall surface anchored protein	Whole
				pbp1A	Penicillin-binding protein 1A	Whole
				recU	Holliday junction-specific endonuclease	Whole
				SPN23F_03430	Hypothetical protein	Whole
				SPN23F_03440	DivIVA protein	Whole
				SPN23F_03450	RNA methylase family protein	Whole
				SPN23F_03460	Hypothetical protein	Whole
				gnd	6-phosphogluconate dehydrogenase	Whole
				csrR	Response regulator protein	Whole
				cbpJ	Choline binding protein J	Whole
				SPN23F_03500	Major facilitator superfamily protein	Whole
				mvaK1	Mevalonate kinase	Whole
				mvaD	Mevalonate diphosphate decarboxylase	Part
3	507199	520268	13070	*pheT*	Phenylalanyl-tRNA synthetase subunit beta	Part
				SPN23F_05260	Integral membrane protein (possible nuclease activity)	Whole
				SPN23F_05280	Hypothetical protein	Whole
				metE	5-methyltetrahydropteroyltriglutamate-- homocysteine S-methyltransferase	Whole
				metF	5,10-methylenetetrahydrofolate reductase	Whole
				pnpA	Polynucleotide phosphorylase/polyadenylase	Whole
				cysE	Serine acetyltransferase	Whole
				SPN23F_05330	GNAT family acetyltransferase	Whole
				cysS	Cysteinyl-tRNA synthetase	Whole
				SPN23F_05350	Ribonuclease	Whole
				SPN23F_05360	Leucine-rich protein	Part
4	634685	642400	7716	*thiE*	Thiamine-phosphate pyrophosphorylase	Part
				SPN23F_06440	Hypothetical protein	Whole
				SPN23F_06450	ABC transporter ATP-binding protein	Whole

				SPN23F_06460	Hypothetical protein	Whole
				SPN23F_06470	Coenzyme	Whole
				SPN23F_06480	Hypothetical protein	Whole
				SPN23F_06490	Hydroxyethylthiazole kinase	Whole
				thiE	Thiamine-phosphate pyrophosphorylase	Whole
				thiD	Phosphomethylpyrimidine kinase	Whole
				SPN23F_06520	Regulator	Part
5‡	1598816	1615831	17016	*ileS*	Isoleucyl-tRNA synthetase	Part
				SPN23F_16620	Cell-division protein DivIVA	Whole
				SPN23F_16630	Hypothetical protein	Whole
				SPN23F_16640	Hypothetical protein	Whole
				SPN23F_16650	Hypothetical protein	Whole
				SPN23F_16660	Hypothetical protein	Whole
				ftsZ	Cell division protein FtsZ	Whole
				ftsA	Cell division protein	Whole
				SPN23F_16690	Hypothetical protein	Whole
				SPN23F_16700	MutT/NUDIX hydrolase family protein	Whole
				murF	UDP-N-acetylmuramoyl-tripeptide--D-alanyl-D- alanine ligase	Whole
				ddl	D-alanyl-alanine synthetase A	Whole
				recR	Recombination protein RecR	Whole
				pbp2b	Penicillin-binding protein 2b	Whole
				SPN23F_16750	Transcription regulator	Whole
				SPN23F_16760	ROK family protein	Part
6ψ	2033600	2038261	4662	*gltX*	Glutamyl-tRNA synthetase	Whole
				pgi	Glucose-6-phosphate isomerase	Whole
				SPN23F_20970	Peptidase	Whole
				SPN23F_20980	ABC transporter ATP-binding protein	Part
7ψ	2139058	2147857	8800	SPN23F_21910	Exported glycosyl hydrolase	Part
				SPN23F_21920	Hypothetical protein	Whole
				SPN23F_21930	PTS system, IId component	Whole
				SPN23F_21940	PTS system, IIc component	Whole

				SPN23F_21950	PTS system, mannose-specific IIAB component	Whole
				SPN23F_21960	PTS system, IIa component	Whole
				fucU	Fuscose operon FucU protein	Whole[†]
				fucA	L-fuculose phosphate aldolase	Whole
				fucK	Fuculokinase	Whole
				SPN23F_22000	Fucose phosphotransferase system repressor	Whole
8	2159726	2171248	11523	SPN23F_22150	Hypothetical protein	Whole
				SPN23F_22160	Hypothetical protein	Whole
				glpF1	Glycerol uptake facilitator protein 1	Whole
				glpO	Alpha-glycerophosphate oxidase	Whole
				glpK	Glycerol kinase	Whole
				SPN23F_22210	DNA-binding protein	Whole
				hslO	Hsp33-like chaperonin	Whole
				SPN23F_22230	tRNA-dihydrouridine synthase	Whole
				cbpA	Choline-binding surface protein A	Part
9	2171249	2182005	10757	*cbpA*	Choline-binding surface protein A	Part
				SPN23F_22250	Isoprenylcysteine carboxyl methyltransferase (ICMT) family protein	Whole
				SPN23F_22260	Two-component system, sensor histidine kinase	Whole
				SPN23F_22270	Two-component system, response regulator	Whole
				clpC	Stress response-related Clp ATPase	Whole
				ctsR	Transcriptional regulator	Whole
				SPN23F_22300	ABC transporter ATP-binding protein	Whole
				SPN23F_22310	ABC transporter substrate-binding protein	Whole
				SPN23F_22320	ABC transporter permease	Whole
				SPN23F_22330	Hypothetical protein	Whole
				cbpD	Choline binding protein D	Part

CDS = coding sequences. [Ψ] Region also identical or highly similar in one or more CC66 representative. [‡] One or more CC66 representatives are also identical over the first ~14 kb of this region. † A duplication of 359 bp of the *fucU* gene was identified directly 5' of the full version of this gene in the PMEN3 genome. The sequence was identical to the equivalent region of the PMEN1 reference genome and differed from that of the PMEN3 ancestral representatives, suggesting that this duplication occurred subsequent to the import of this region from PMEN1.

5. PNEUMOCOCCAL CAPSULAR SWITCHING: AN HISTORICAL PERSPECTIVE

This chapter has been prepared for submission to a peer-reviewed journal. The following investigators contributed to this work:

Kelly L. Wyres[1], Lotte L. Lambertsen[2], Nicholas J. Croucher[3], Lesley McGee[4], Anne von Gottberg[5], Josefina Liñares[6], Michael R. Jacobs[7], Karl G. Kristinsson[8], Bernard W. Beall[4], Keith P. Klugman[9], Julian Parkhill[3], Regine Hakenbeck[10], Stephen D. Bentley[3], Angela B. Brueggemann[1*]

[1]University of Oxford, Oxford, United Kingdom, [2]Statens Serum Institut, Copenhagen, Denmark, [3]The Wellcome Trust Sanger Institute, Hinxton, United Kingdom, [4]Centers for Disease Control and Prevention, Atlanta, Georgia, USA, [5]National Institute for Communicable Diseases, Gauteng, South Africa, [6]Hospital Universitari de Bellvitge, Barcelona, Spain, [7]Case Western Reserve University, Cleveland, Ohio, USA, [8]Landspitali University Hospital and University of Iceland, Reykjavik, Iceland, [9]Emory University, Atlanta, Georgia, USA, [10]University Kaiserslautern, Kaiserslautern, Germany.

The data described in this chapter were extracted from the previously mentioned historical isolate genome sequences, an additional 131 pneumococcal genomes and 88 *cps* locus reference sequences retrieved from Genbank. I retrieved 30 of the genomes and all of the *cps* locus reference sequences. The remaining 111 genomes were retrieved by A. van Tonder, who is currently working as a research assistant in Dr. Brueggemann's laboratory.

5.1 Abstract

Changes in serotype prevalence among post-conjugate vaccine pneumococcal populations threaten the effectiveness of vaccine programs, and result from both serotype replacement and serotype (capsular) switching. Temporal changes in pneumococcal serotype distributions in the absence of pneumococcal vaccination are also well documented, but the historical contribution of capsular switching to such epidemiological change is unknown. Similarly, it is not clear to what extent vaccine-induced selective pressures drive capsular switching or modification of capsular switching mechanisms.

We collected serotype and multilocus sequence type data for an historical collection of 426 pneumococcal isolates dated from 1937 to 2007. We identified capsular switch variants representing at least 36 independent switching events. Whole-genome sequence data for selected isolates were used to characterise the genetic changes resulting in 18 events. Recombination fragment lengths were estimated for 11 events and ranged from ~19.0 kb to ≥58.2 kb in length. No temporal patterns in recombination fragment size were identified. In two cases, both dated no later than 1960, the imported fragments included the capsular locus *and* the nearby penicillin-binding protein genes, *pbp2x* and *pbp1a*.

These data show that capsular switching has been a regular occurrence throughout the last 7 decades of pneumococcal history. Recombination of large DNA fragments (>30 kb), sometimes including the capsular locus *and* one or two primary penicillin binding-protein genes, pre-dated both vaccine introduction and widespread

antibiotic use. We conclude that this type of recombination has likely been an intrinsic feature throughout the history of pneumococcal evolution.

5.2 Introduction

Streptococcus pneumoniae is an asymptomatic coloniser of the human nasopharynx, most commonly in children <5 years of age. This bacterium, the "pneumococcus", is also a major cause of otitis media, sinusitis, pneumonia and meningitis, resulting in ~14.5 million annual serious disease episodes globally [50]. The primary pneumococcal virulence determinant is its polysaccharide capsule, of which there are at least 94 types (serotypes) separated into 48 serologic groups [7, 13, 13a, 37, 51]. The capsule provides protection against phagocytosis during invasive pneumococcal disease (IPD) and may also play a role in prevention of clearance during nasopharyngeal colonisation [30, 49]. Ninety-one of the known capsule types are synthesised and exported through the 'Wzy-dependent' pathway, whereby extracellular polymerisation of component lipid-linked repeat units is preceded by 'flippase-mediated' transfer across the cell membrane [7, 39]. The proteins responsible for this synthesis and export are encoded by the capsule polysaccharide synthesis (*cps*) genes, which are located between the *dexB* and *aliA* loci on the pneumococcal chromosome. The remaining two capsule types are synthesised through independent biochemical pathways.

Analysis of the nucleotide sequences of 88 reference *cps* loci indicated the presence of a range of genes and variable sequence length from ~10 kb to ~30 kb [7]. The *cps* genes encode the synthases, transferases and polymerases specific to capsule production. Some of these genes (the serotype non-specific genes) were present

among all Wzy-dependent capsule types, whereas others (the serotype-specific genes) were only found among one or a subset of types. The analyses also indicated that the distinct serotypes have evolved through a combination of mutation and inter-/intra-species recombination [7, 58]. Separate analyses of multiple serogroup 6 and 19 representatives indicated that further diversity *within* these groups is also generated through both mutation and recombination [9, 21, 44].

Pneumococcal capsule type is a key determinant of invasive disease potential and prevalence; certain serotypes are more commonly associated with carriage and others more commonly with IPD [10, 54]. There is considerable variation between host age, geographic location [34] and time period (see below). A 7-valent pneumococcal protein conjugate polysaccharide vaccine (PCV7), was introduced into the childhood vaccination program in the USA in 2000 [8]. Subsequently, this vaccine was also introduced into many other countries. The 7 serotypes (4, 6B, 9V, 14, 18C, 19F and 23F) were chosen because, at the time, they represented those most commonly causing paediatric IPD in the USA. More recently 10- and 13- valent vaccines (PHiD-CV and PCV13, respectively), providing protection against the original 7 serotypes plus serotypes 1, 5 and 7F (in both vaccines) and 3, 6A and 19A (in PCV13 only), have been developed and have replaced PCV7 [23, 53].

Pneumococcal population surveillance following PCV7 introduction showed a decline of both disease and carriage of vaccine type (VT) pneumococci; however, there was also an increase of disease and carriage due to non-vaccine type (NVT) pneumococci [59]. Such epidemiological changes can be attributed to the phenomena of “serotype

replacement" and/or "serotype (capsular) switching" [6]. Serotype replacement describes the expansion of pre-existing NVT pneumococcal clones in the absence of competition from VT pneumococci. Serotype switching describes a change of serotype of a single clone by alteration or exchange of its *cps* locus. (This phenomenon was first described by Griffith in 1928 [32] and was the basis of the experiments that lead to the discovery of DNA by Avery and colleagues in 1944 [4].) The relative importance of serotype switching compared to that of serotype replacement is yet to be fully determined, and it should be remembered that these effects are not completely independent, i.e. new clones generated by capsular switching can subsequently expand within a population. Both phenomena can be studied by comparison of the pneumococcal serotypes and genotypes present within pre- and post-vaccine populations. Pneumococcal genotypes, as defined by multilocus sequence typing (MLST) [22], are known to show predominant serotype-specific associations [10, 36]. Any isolate exemplifying a new or unusual genotype/serotype combination is therefore considered to represent a capsular switch variant. Such variants are assumed to have arisen by recombination at the *cps* locus, and studies of a limited number of such strains have indicated that the range of recombination fragment sizes varies from ~21.9 kb to ~56.5 kb [11, 31, 43, 52]. In some cases the fragments also included part or all of the *pbp2x* and *pbp1a* genes (two of the three primary penicillin-resistance determinant genes which are located ~8 kb upstream and ~7 kb downstream of the *cps* locus, respectively).

Vaccine-induced selective pressure is contributing to the post-vaccination changes of pneumococcal serotype epidemiology, but natural fluctuations in serotype prevalence

also play a role in driving these changes. Many studies have documented pre-vaccine temporal changes in relative pneumococcal serotype prevalence [5, 12, 15, 24-28, 33, 41, 42, 48, 55]; however, all but three of these studies [24, 27, 33] included pneumococci isolated no earlier than 1969. Furthermore, only two studies [15, 28] provided any genotype data to compare the relative importance of serotype switching to that of serotype replacement.

Here we studied a large, unique, genetically diverse, collection of historical and modern pneumococci spanning 7 decades, with the aim to better understand the mechanisms and role of capsular switching in pneumococcal evolution. We analysed serotype and MLST genotype data to identify capsular switch variants among strains dating from 1937 to 2007. Variants from 7 independent ancestral lineages were studied in further detail by analysis of their *cps* locus and flanking region nucleotide sequences, which were extracted from whole-genome assemblies. Capsular switch events were classified by whether they most likely resulted from mutation or recombination. Recombination breakpoints and putative donor isolate representatives were identified where possible. Finally, we compared the *cps* loci of isolates expressing the same serotype (including those from the seven previously mentioned lineages and one additional lineage) to assess within-serotype temporal *cps* variability.

5.3 Results

Indication of serotype switching

Amongst our collection of 426 pneumococci, 21 of 163 unique clonal complexes (CCs) were represented by isolates of ≥2 serotypes. 12, 5, 1, 2 and 1 CCs were represented by isolates of 2, 3, 4, 5, and 6 unique serotypes, respectively. CCs are groups of closely-related genotypes; genotypes are defined by MLST, whereby the nucleotide sequences of regions of 7 housekeeping genes are determined and unique sequences are assigned allele numbers. Each combination of allele numbers is assigned a sequence type (ST) [22]. Isolates that belonged to the same CC but which had different serotypes were considered to represent capsular switch variants. At least 36 independent changes of serotype within pneumococcal CCs were represented in our collection, 34 of which pre-dated the introduction of pneumococcal conjugate vaccines, since the isolates were dated prior to 2000. In fact four of these *cps* changes occurred prior to 1950 and 11, 17, 22 and 28 cumulative independent changes occurred prior to 1960, 1970, 1980 and 1990, respectively (Figure 5.1).

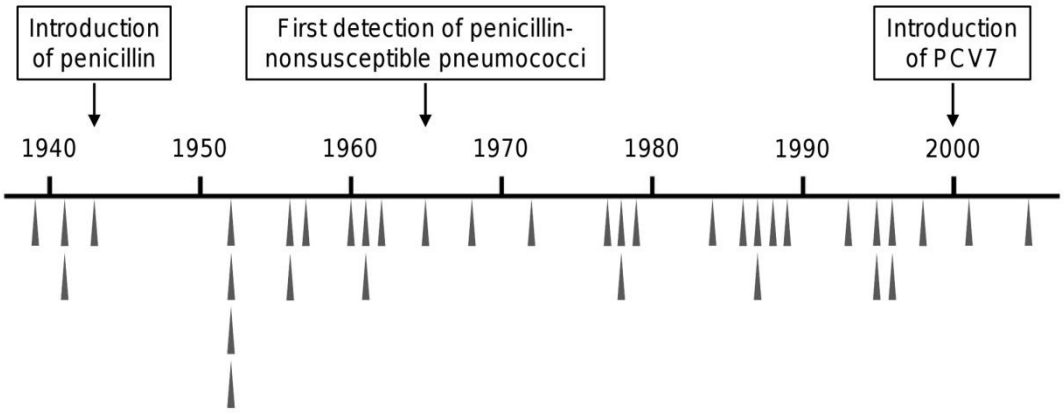


Figure 5.1 Temporal distribution of capsular switch variants among the historical isolate collection.
Grey triangles indicate isolation dates for the oldest recombinant representatives of 36 independent capsular switch events identified among the historical pneumococcal collection.

Table 5.1 Newly characterised capsular switching events.

CC	Ancestral Serotype	New Serotype	Recombinant Year(s)	Change by Point Mutation or Recombination	Recombination Donor	Recombination Breakpoints	Recombination Region Size (~kb)
66	7B	9N	1960, 2001	Recombination	CC None3782[9N]	≥ 8901 bp 5' *pbp2x* - 7415 bp 3' *pbp1a*	≥54.3
	9N	19F	1972	Recombination	Unknown[19F]	2538 bp 5' *dexB* - 34 bp 5' *aliA*	≤23.7
	9N	19F	2005	Recombination	Unknown[19F]	115 bp 3' *dexB*- bp 1173 *aliA*	≤20.6
	9N	14	1995, 1999	Recombination	CC15[14]	2158 bp 5' *dexB* - between bp 88 *aliA* and 785 bp 3' *aliA*	23.1 - 25.8
	9N	14	1997	Recombination	CC15[14]	Undetermined	Undetermined
	9N	14	Unknown	Recombination	CC124[14]	≥ 2520 bp 5' *dexB* – 4062 bp 3' *aliA*	≥29.9
	9N	23F	2001	Recombination	Unknown[23F]	1964 bp 5' *dexB* – 3013 bp 3' *aliA*	≤30.4
113	18C	18B	1941†	Point Mutation	-	-	-
	18C	35C	1941, 1943†	Recombination	Unknown[35C]	144 bp 3' *dexB* – 76 bp 5' *aliA*	≤19.0
	35C	17F	1939	Recombination	CC574[17F]	bp 549 *wze* – 3278 bp 3' *pbp1a*	30.9
	18C	9V	1968†	Recombination	Unknown[9V]	Undetermined	Undetermined
124	14	11C	1957†	Recombination	Unknown[11C]	Undetermined	Undetermined
	14	9L	1952	Recombination	Unknown[9L]	Undetermined	Undetermined
156/162	9V	9A	1962†	Point Mutation	-	-	-
	9V	19A	2005, 2006	Recombination	CC199[19A]	≥ 5701 bp 5' *dexB* - ≥ 6022 bp 3' *aliA*	≥ 50.4
191	7F	7A	1937†	Point Mutation	-	-	-
218	12F	7F	1952, 1993, 1999	Recombination	CC191[7F]	≥ 9045 bp 5' *pbp2x* – 4782 bp 3' *pbp1a*	≥ 58.2
574	17F†	2	1956†	Recombination	CCNone128[2]	Between bp 326 *dexB* and 639 bp 3' *dexB* – 5229 bp 3' *pbp1a*	34.9 - 36.8

† This isolate was used as the *cps* reference strain in reference [2].

Within CC cps diversity

CC15: Ten CC15[14] (CC15, serotype 14) representative *cps* nucleotide sequences were available for analysis. Among these there were 7 unique sequences (Supplementary Table 5.1). Interestingly, two of the CC15[14] representative *cps* sequences (isolates ICE13 and ICE50) shared *wciY* locus nucleotide substitutions with four CC124[14] representatives (isolates 14/9, USA6, Ala289 and ICE594), suggesting that a possible recombination event occurred at some point during the history of these clones (although it is not possible to determine which of these CCs was most likely the recipient and which, if either, was most likely the donor). Additionally, in comparison to the CC124[14] representatives, the CC15[14] representatives showed a deletion of a 432 bp internal fragment of the *wciY* gene. This deletion resulted in a predicted truncation of the WciY protein, reducing its length from 360 to 110 amino acid residues. One isolate, CGSP14 (Genbank accession NC_010582.1 [19]), also seemed to have an additional 52 bp deletion further 3' within the *wciY* gene. If this is a true deletion event rather than a sequencing/assembly error the resultant protein would be further reduced to 38 amino acid residues.

CC66: Isolates assigned to CC66 (n = 10) were dated from 1952 to 2005 and represented five different serotypes, the *cps* locus structures for each of which were sufficiently different to suggest change by recombination rather than point mutation. The oldest isolate was the sole representative of serotype 7B. The next oldest isolate was dated 1960 and was assigned serotype 9N. An ~53.4 kb region of the chromosome of this isolate (the recombinant representative), including *pbp2x,* the *cps* locus and *pbp1a*, differed from that of the CC66[7B] isolate (the recipient representative)

and was highly similar to that of a CC3782^{9N} isolate dated 1952 (the donor representative, see Table 5.1, Supplementary Table 5.2 and Figure 5.2). Thus it seems that at some time no later than 1960 a large recombination import was acquired by a CC66^{7B} isolate from a CC3782^{9N}–like isolate and resulted in a serotype switch event. (Note that while the data strongly suggest that a CC3782^{9N} pneumococcus was the donor in this case, it is possible that another clone which was not represented in this sample was in fact the true donor. As such we use the term, "CCX^{X}–like" when describing putative donor representatives.) In the CC66^{9N} genome, the region directly 3' of the putative recombination import was similar to the CC66^{7B} genome; however, the region directly 5' of the putative recombination import was not. It is possible that an independent recombination event may have affected this 5' region of the chromosome in one or more of the isolates considered here, masking the true boundaries of the recombination import. Alternatively, the true donor and/or recipient strains may have differed in this region from their respective representatives. Across the *cps* locus synthesis-related regions (see section 5.5, *Methods*) a second CC66^{9N} isolate differed from the oldest CC66^{9N} isolate by only five nucleotide substitutions (Supplementary Table 5.1).

Among our whole-genome sequences there were two CC66^{19F} representatives. Analysis of the *cps* locus and flanking nucleotide sequences of these genomes suggested that they represented two independent capsular switch events. There were no potential donor representatives amongst the other serotype 19F isolates in our collection. The putative recombination imports resulting in these two capsular switching events were estimated as the regions over which these genomes clearly

differed from those of CC66^{9N}. Independent 5' and 3' recombination breakpoints were indicated in both cases (Table 5.1, Supplementary Table 5.2 and Figure 5.2), but without suitable donor representatives it is not possible to speculate whether the differences between the CC66^{19F} and CC66^{9N} genomes were the result of a single or multiple recombination events.

There were four CC66^{14} representatives within our genome collection, the analysis of which indicated three independent capsular switching events. Based on nucleotide sequence similarity, the donor isolates for two of these events were likely CC15^{14}-like pneumococci, while the third donor was most likely a CC124^{14}-like pneumococcus. The recombination breakpoints for one of the events could not be determined due to extensive diversity in the *cps* flanking regions of the donor, recipient and recombinant representative genomes. The recombination breakpoints for the two remaining events were estimated (Table 5.1, Supplementary Table 5.2 and Figure 5.2). Interestingly the 5' breakpoints for the latter two serotype 14 switches and one of the aforementioned serotype 19F switches were all between 2.2 and 2.5 kb 5' of the *dexB* locus, a region corresponding to the *clpL* locus in the ATCC 700669 reference genome (Genbank accession NC_011900.1 [18]). The *clpL* locus has been shown to be involved in modulating the expression of virulence-related genes [40] and in the development of penicillin-nonsusceptibility [57].

Finally, there was one CC66^{23F} representative in our genome collection. The genome of this isolate differed from those of the CC66^{9N} isolates from ~2 kb 5' of the *dexB* locus to ~3 kb 3' of the *aliA* locus, suggesting a potential recombination import of ~30.4 kb

length. However, no suitable donor representative could be identified and thus it was not possible to assess whether the import was acquired during a single or multiple recombination events.

CC113: There were 11 CC113 isolates in our genome collection, representing serotypes 18C (n = 6), 35C (n = 2), 9V (n = 1), 17F (n = 1) and 18B (n = 1). Serotype 18C was assumed to be the ancestral type and was also represented by one of the two oldest isolates, both dated 1939 (the second was serotype 17F). Analysis of the CC113 *cps* loci indicated that the change from serotype 18C to 18B was the result of a single nucleotide substitution within the *wciX* gene (as previously described [7]), while each of the other serotype changes was associated with a recombination event. Both $CC113^{35C}$ isolates appear to have arisen from a single capsular switch event (Table 5.1, Supplementary Table 5.2 and Figure 5.2). Interestingly, the change to serotype 17F appears to have involved the acquisition of DNA from a $CC574^{17F}$-like pneumococcus by a $CC113^{35C}$ pneumococcus. The 5' region of the $CC113^{17F}$ *cps* locus was distinct from that of three other serotype 17F isolates in our collection, but differed by only a single nucleotide substitution from the $CC113^{35C}$ representatives. The $CC113^{17F}$ representative genome was highly similar to that of a $CC574^{17F}$ representative between position 549 of the *wze* locus and ~3 kb 3' of the *pbp1a* locus. Directly 3' of this region the $CC113^{17F}$ genome resembled that of the $CC113^{18C}$ isolates.

Among the majority of $CC113^{18C}$ representatives there was little variation at the *cps* locus. The loci of three isolates were identical and those of two further isolates each differed by only single nucleotide substitutions (Supplementary Table 5.1). A final

CC113^{18C} isolate differed by a single nucleotide substitution in the *wchA* gene and by a putative recombination import of ~1.1 kb length spanning part of the *rmlA* and *rmlC* genes. The nucleotide sequence corresponding to this putative import region differed from those of the other CC113^{18C} representatives by 44 nucleotide substitutions (Figure 5.4A).

CC124: Our genome collection included 14 CC124 representatives, 12 of which were assigned serotype 14 and represented 8 unique, but highly similar *cps* sequences (Supplementary Table 5.1). The remaining two isolates were assigned serotypes 9L and 11C, and represented independent capsular switch recombination events. No suitable donor representatives were identified amongst our collection and the *cps* flanking regions of both isolates were highly divergent from those of the CC124^{14} representatives for at least 30 kb in either direction. Thus the putative recombination breakpoints could not be estimated.

CC156/162: Seven isolates dated 1952 to 2006 and representing three serotypes were assigned to CC156/162. The most common serotype was 9V (n = 4), although the oldest isolate was assigned serotype 9A. The CC156/162^{9A} *cps* locus differed from that of the oldest CC156/162^{9V} locus by a single nucleotide substitution in the *wzh* gene and a single bp deletion in the *wcjE* gene as previously described (Reference [7], Table 5.1, Supplementary Table 5.2 and Figure 5.2). Three of the four CC156/162^{9V} isolates were identical at the *cps* locus and the final isolate differed by only a single nucleotide substitution.

The remaining two CC156/162 representatives (Genbank accessions AGOR01000001 - AGOR01000024 and AGQA01000001 - AGQA01000007) were assigned serotype 19A. Analysis of the *cps* loci and genomic flanking regions indicated that these isolates represented a single capsular switch recombination event. Approximately 50.4 kb regions of the CC156/162^{19A} genomes were highly similar to those of a CC199^{19A} representative (Table 5.1, Supplementary Table 5.2 and Figure 5.2). These putative import regions included *pbp2x,* the *cps* locus and *pbp1a*. The regions directly flanking the putative import did not resemble those of the CC156/162^{9V} representatives and thus the true recombination breakpoints could not be estimated; however, we deduced that the original import was at least 50.4 kb in length.

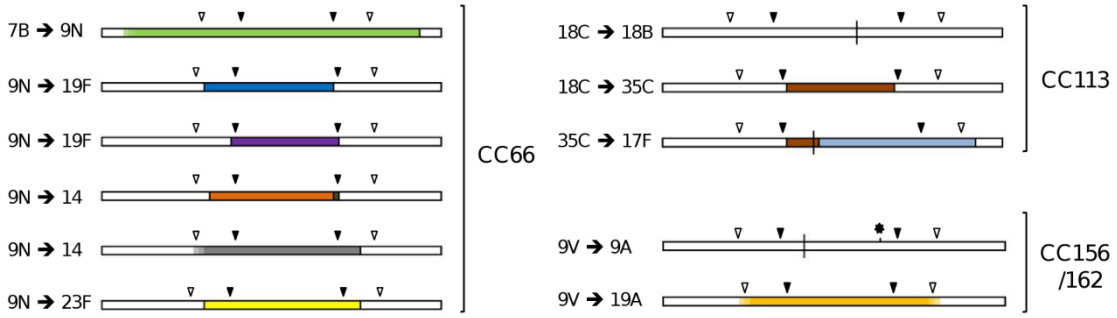


Figure 5.2 Genetic changes leading to capsular switching within CCs 66, 113 and 156/162. Bars represent the *cps* locus and flanking regions. Filled arrows indicate the relative positions of the *dexB* (left) and *aliA* (right) loci. Open arrows indicate the relative positions of the *pbp2x* (left) and *pbp1a* (right) loci. Vertical bisecting lines indicate single nucleotide substitutions. "*" indicates a single bp insertion/deletion. Recombination fragments are indicated by colouring. Identical sequences are indicated by identical colouring. Regions bounded by solid lines are depicted at their maximum length. Regions not bounded by solid lines are depicted at their minimum length (see Table 5.1 and Supplementary Table 5.2). Schematic approximately to scale.

CC191: All but one of the CC191 representatives in our genome collection were assigned serotype 7F (n = 7). The remaining isolate was the oldest CC191 representative and was serotype 7A. As previously described [7], the serotype 7F and

7A *cps* loci differed by two nucleotide substitutions (one each in the *wzd* and *wcwF* genes) and a single bp insertion within the *wcwD* gene (in the 7A *cps* locus). Within the *wcwC* gene there was a repeated motif (5'-CTA AGA TGA ATA-3') for which the number of repeats differed between isolates. Three variations of total motif copy number were identified and were 3 (n = 3), 4 (n = 2) and 6 (n = 3, including the serotype 7A locus). (It should be noted that both of the latter two 7F loci were split across two assembly contigs, and in both cases the first contig ended two bp short of the end of the third motif copy). The nature of next-generation sequencing technologies and assembly algorithms mean that repeat motifs such as these are difficult to sequence accurately, and so we cannot make any further speculations about changes in this region of the *cps* locus. Aside from the repeat region, the CC191^{7F} *cps* sequences each differed by a maximum of four nucleotide substitutions (Supplementary Table 5.1, Figure 5.4B).

CC218: Eight CC218 isolates were included in our genome collection, representing serotypes 12F (n = 5) and 7F (n = 3). The CC218^{7F} representative genomes differed from that of the CC218^{12F} genomes and more closely resembled the CC191^{7F} genomes between ~9.0 kb 5' of the *pbp2x* locus and ~4.8 kb 3' of the *pbp1a* locus (Table 5.1, Supplementary Table 5.2 and Figure 5.3). However, numerous nucleotide substitutions differentiating the CC218^{7F} representatives from the CC191^{7F} representatives were identified in this region; 67 nucleotide substitutions were present in the oldest CC218^{7F} representative (dated 1952) compared to the oldest CC191^{7F} representative (dated 1962). Within the *cps* locus these nucleotide differences included one substitution in the *wcwA* gene and one substitution in the

wcwD gene. In addition, 39 nucleotide substitutions were identified within a 197 bp region spanning the *wzg* and *wzh* genes. Five further nucleotide substitutions and two single bp insertions within an ~1 kb region of the *glf* locus were also identified. Such clustering of nucleotide substitutions is indicative of recombination events rather than individual point mutations (Figure 5.4B). It is not possible to know whether these recombination events occurred prior to the *cps* switch (in the true donor genome) or subsequent to the switch (in the recombinant genome). We note also that the *cps* 5' flanking region was not contiguous within the serotype 7F representative genomes. A region of ~585 bp in length was missing between the *pbp2x* and *dexB* loci for both the CC218 and CC191 representatives compared above.

The *cps* loci of the two remaining CC218^{7F} isolates each differed from that of the oldest isolate by two or three nucleotide substitutions (Supplementary Table 5.1). There were also differences in the number of copies of the *wcwC* repeat motif described above; two CC218^{7F} representatives each had a total of four copies of this motif and the third isolate had a total of six copies. Among the five CC218^{12F} representatives a total of four unique *cps* sequences were identified (Supplementary Table 5.1).

CC574: Two CC574 isolates were included in our genome collection. These were a serotype 17F isolate dated 1952 and a serotype 2 isolate dated 1956. The genomes of these isolates differed from each other from position 326 of the *dexB* locus to ~5 kb 3' of the *pbp1a* locus. Within the CC574^{2} representative genome, the majority of this region (totalling ~34.9 kb, from 639 bp 3' of the *dexB* locus to ~5 kb 3' of the *pbp1a* locus) was highly similar to that of a CCNone128^{2} representative dated 1916. These

data suggest that the change from serotype 17F to 2 was the result of an import of at least 34.9 kb of DNA from a CCNone128[2]-like donor. An additional recombination event prior to, or following this change could explain the region of unknown sequence between bp 326 of the *dexB* locus and 639 bp 3' of the end of this locus. Alternatively, the true recombination import may have included this region, with the true donor showing sequence diversity compared to the CCNone128[2] representative studied here.

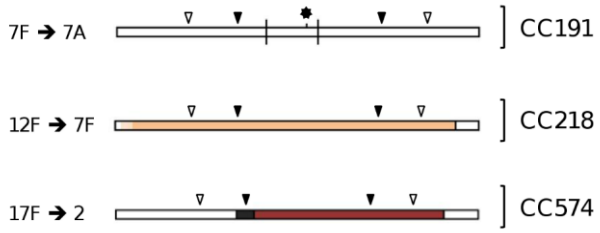


Figure 5.3 (left) Genetic changes leading to capsular switching within CCs 191, 218 and 574.
As Figure 5.2.

A

```
                                1111111 1111111111 1111111111 1111111111 11111111
                               2342556666 6666666666 6666666666 6666666666 66666667
                               2463891111 1111112222 2222222222 2233333555 66888890
                               0728682333 3446790022 3345566777 9922346114 78566850
                               3822145247 8034910614 0651809258 6903875121 05103465
18C/2 (1939)                   AGCGGCACGG TCTGCGATTT GTAACATTGT CCAAGAGTCC AAAACTGT
18B/2 (1939)                   ...T...... .......... .......... .......... ........
18C/1 (1940)                   G......... .......... .......... .......... ........
18C/3 (1968)                   .A........ .......... .......... .......... ........
Netherlands18C-36 (1980)       .......... .......... .......... .......... ........
USA2 (1999)                    ..T.TTGTCT CTACTAGCCA ACCGTGACAC TTCTAGCCAT CCGGTATC
ICE501 (2002)                  .......... .......... .......... .......... ........
```

B

```
                                                                          1111 11222222
                               111111111 1111111111 1111111111 1111111111 25668890012 99000000
                               000111111 1111111111 1122222222 2222222222 33480605776 49144444
                              1899011255 6667778889 9900011111 2233556777 76959072183 80833457
                              6328739228 1450392580 1208923689 1746127259 59941641217 61556103
7A/2 (1937)                   AAAAACACGG TCCAAGACAC ACTCGCAATA CAACACAGCG GCCCCTCTTCA TCCCATAA
7F/3 (1962)                   .......... .......... .......... .......... A....-.A... ...-----
7F/5 (1962)                   G......... .......... .......... .......... A....-.A... ........
Netherlands7F-39 (1984)       .......... .......... .......... .......... A....-.A... ...-----
7F/4 (1986)                   .......... .......... .......... .......... A....-.A... ...-----
ICE22 (1993)                  .......... .......... .......... .......... A....-.A... ...TGCG.
CDC1087-00 (1999)             .......... .......... .......... .......... A....-.AC.. ........
USA16 (2003)                  .......... .......... .......... .......... A....-.A... ........
7F/2 (1952)*                  .CTGCACTCA ATAGGTTAGT TTCGTTCCAG AGCTGAGAGA A..AT-.A... GTT.G..G
ICE23 (1993)*                 .CTGCACTCA ATAGGTTAGT TTCGTTCCAG AGCTGAGAGA AATAT-.A..C GTT.G..G
USA20 (1999)*                 .CTGCACTCA ATAGGTTAGT TTCGTTCCAG AGCTGAGAGA A..AT-TA.G. GTT-----
```

Figure 5.4 (previous page) Variable site maps for the *cps* loci of serogroups 18 and 7.
Rows represent the nucleotide sequences of independent isolates and are labelled as *Isolate Name (Year).* Nucleotide differences relative to row 1 are shown. "." indicates an identical base, "-" indicates a missing base. Numbers above each column indicate the nucleotide position within the alignment. Nucleotide substitutions marking putative recombination regions are marked in bold and underlined. (A) CC113 isolates include one of serotype 18B (18B/2) and six of serotype 18C. (B) CC191 isolates include one of serotype 7A (7A/2) and seven of serotype 7F. (B*) CC218 isolates all represent serotype 7F. The arrow indicates the position of a single bp insertion in the 7A/2 sequence compared to the other serogroup 7 sequences.

5.4 Discussion

Within our collection of historical and modern pneumococci (dated from 1937 to 2007) we identified at least 36 independent capsular switching events. The majority (~94%) of the resultant capsular switch variants were isolated prior to the introduction of PCV7 and were approximately evenly distributed through time (Figure 5.1). The study collection was not designed to be able to infer a capsular switching rate; nevertheless these data imply that this phenomenon has been a regular occurrence i.e. there is evidence of capsular switching within a diverse range of CCs, every decade throughout (at least) the last 7 decades.

Analysis of the *cps* loci of 10 representatives of CC66 indicated multiple independent changes to the same serotype. Within this CC an initial capsular switch from serotype 7B to 9N (no later than 1952) was followed by at least two independent changes to serotype 19F (no later than 1972 and 2005, respectively) and at least three independent changes to serotype 14 (each no later than the mid-1990s). This is consistent with the findings of previous authors who indicated multiple changes of serotype within the same CC, e.g. from 9V to 14 in CC156/162 [2, 16, 45], from 23F to 19F and 19A in CC81 [17] and from 4 to 19A in CC695 [11, 31]. This finding also

supports the notion that capsular switching may occur regularly amongst pneumococci.

We studied 18 capsular switching events in further detail. Three of these events were the result of nucleotide substitution, a finding consistent with previous analyses of the *cps* locus reference sequences [7]. (Note that the CC191$^{7A/7F}$, CC156/162$^{9V/9A}$ and CC113$^{18B/18C}$ *cps* loci studied here included the reference sequences, but that the MLST genotypes of these isolates were not known at the time of the original study.) The remaining 15 capsular switching events were presumed to be the result of change by recombination. The recombination breakpoints for 11 of these events could be estimated by comparison of the *cps* locus and flanking sequences of the putative ancestor, donor and recombinant representatives. Putative donor representatives were identified as isolates for which the *cps* locus nucleotide sequence differed from that of the recombinant by a maximum of three nucleotide substitutions (excluding clusters of closely linked substitutions which were considered to have arisen through recombination within the *cps* locus itself). Subsequently, MLST data was used to infer the CC most likely represented by the true donor isolate (the same as that of the donor representative). Figure 5.5 depicts the exchange of c*ps* loci between pneumococcal CCs, as inferred by these data. It should be noted that we cannot exclude the possibility that the true donors actually represented different CCs. It is possible that the isolates identified as donor representatives from our collection share a high level of *cps* locus sequence identity with the true donors (which were not included in this sample), either through ancestral descent or an additional serotype switch event. However, even if this were the case our conclusions about the number of independent

capsular switches, and the range of approximate recombination fragment sizes represented by our sample would remain unchanged.

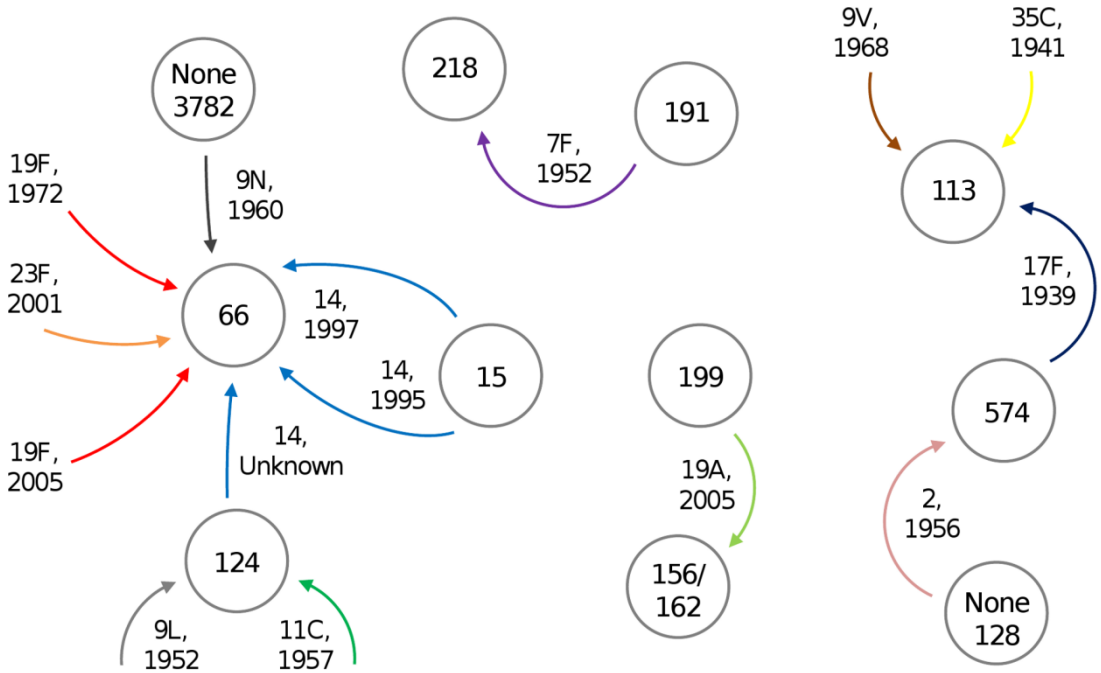


Figure 5.5 Inter-CC *cps* locus transfers inferred in this study.
Circles represent pneumococcal CCs as named. Arrows indicate transfer of *cps* loci between CCs, as inferred from nucleotide sequence analyses for 15 capsular switching events. Donor CCs could not be predicted for a total of 7 events because donor representative isolates were not present among this collection. Arrow labels indicate serotypes associated with the transferred *cps* loci and isolation year of the oldest recombinant representative pneumococcus in this collection.

In general, our analyses did not indicate any recombination breakpoint hotspots around the *cps* locus, which had also not been indicated by any previous studies that characterised such events [11, 17, 31, 52]. It is not possible to know whether the putative imports were acquired through single or multiple recombination events, particularly when donor representatives were not available or sequences were not contiguous; however, the most parsimonious solution would suggest a single event in each case. Subsequently, the putative recombination import lengths were estimated

to range from ~19.0 kb to ≥58.2 kb and did not show any trends towards increasing/decreasing lengths through time.

The CC218$^{12F\rightarrow 7F}$ and CC66$^{7B\rightarrow 9N}$ switch events were characterised by very large (>50 kb) recombination import fragments. These events were particularly interesting as they must have taken place no later than 1952 and 1960, given that the recombinant representative strains were isolated in these years, respectively. Large-scale recombination between pneumococci has therefore clearly been occurring for decades, though the technology capable of detecting and detailing such events has only recently become available [11, 17, 31, 35]. Additionally, both of these putative recombination imports included the *pbp2x* + *cps* + *pbp1a* loci, as has been reported for other capsular switching events [11, 52] and another newly characterised event in this study (CC156/162$^{9V\rightarrow 19A}$).

Given the fact that previously reported *pbp2x* ± *cps* ± *pbp1a* recombination events were detected soon after widespread vaccination began in the USA, no such recombinants had been reported before in nature, and the imported *pbp2x* and *pbp1a* nucleotide sequences conferred penicillin-nonsusceptibility, it was theorised that vaccine- and/or antibiotic-induced selective pressures may play a role in driving these genetic changes [3, 11]. Penicillin was introduced in the 1940s; oral penicillins were not available until the mid-1950s and their use was limited in the early years. Consequently, our analyses question the above theory since both of the CC218$^{12F\rightarrow 7F}$ and CC66$^{7B\rightarrow 9N}$ events were associated with penicillin-susceptible pneumococci, and took place prior to the introduction of PCV7 *and* the widespread use of penicillins.

Instead, our new data suggest that recombination of the *cps* locus and flanking regions might be occurring as part of "normal" pneumococcal biology processes, the evolution of which has likely been influenced by naturally-occurring immunity and other selective pressures. Presumably vaccine-induced immune pressures, and/or the pressure of antibiotic use, subsequently influence the spread and maintenance of advantageous genes (and/or alleles) by selecting those recombinant isolates that are best able to survive. The potential negative effects associated with such events should not be underestimated – e.g. the serotype 4 to 19A so-called "vaccine escape" capsular switches reported after widespread vaccination in the USA were first recovered from young children with invasive disease only three years after PCV7 was introduced, and two years later were the third-most common serotype 19A CC causing invasive disease among all age groups [6]. These strains showed decreased susceptibility to penicillin, owing to the simultaneous acquisition of altered *pbp2x* and *pbp1a* genes. As a consequence, the increase in prevalence of these strains likely contributed to the post-vaccine increase in pneumococcal penicillin-nonsusceptibility in the USA [6, 47].

Another very interesting finding was that a CC199^{19A}-like representative was the most probable donor of the 19A *cps* locus and flanking *pbps* to the CC156/162^{19A} isolates described in this study: CC199^{19A} representatives were the donors of the 19A *cps* locus ± *pbps* to the previously reported vaccine escape progeny [11, 31] and an ST320$^{19F\rightarrow19A}$ *cps* locus switch [52]. Future studies will attempt to determine whether there is a mechanistic explanation for why CC199^{19A} representatives appear to be particularly "good" *cps* locus donors.

Our analyses also allowed us to identify recombination at the *cps* locus which did not result in capsular switching. Two such examples were identified among isolates belonging to the same ancestral lineages (CC113^{18C} and CC191^{7F}) and two examples were identified by comparison of isolates belonging to different ancestral lineages (CC191^{7F} vs. CC218^{7F} and CC15^{14} vs. CC124^{14}). Similar recombination events have been noted previously within the CC81^{23F} lineage [43] and amongst serogroup 6 and 19 isolates [9, 21, 44]. A comparison of the *cps* loci of CC15^{14} and CC124^{14} representatives also indicated a deletion of part of the CC15^{14} *wciY* gene, which encodes a putative glycerol phosphotransferase. This deletion resulted in a predicted truncation of the WciY amino acid sequence. Notwithstanding the deletion, the CC15^{14} isolates were successfully serotyped by the Quellung reaction, indicating successful capsule production. Seemingly the WciY protein (or at least the deleted portion) is not required for serotype 14 capsule synthesis. Indeed, an *in silico* analysis of the predicted *cps* protein coding regions failed to identify a specific reaction catalysed by the serotype 14 version of this protein [1].

Given the overall capacity for *cps* locus recombination demonstrated above, the documented associations between genotypes and serotypes are puzzling, as is our evidence of a high level of *cps* sequence conservation within some CCs (e.g. among CC191^{7F} and CC218^{12F} pneumococci). Perhaps there is some synergism between serotype and the genetic background of a strain which conveys an advantage to certain combinations over others. This sort of synergistic effect has been invoked to explain changes in the pneumococcal population in South Korea, where there was a pre-vaccine reduction of CC271/320^{19F} pneumococci and replacement by CC271/320^{19A}

pneumococci over several years [15]. Antibiotic consumption in this region was high but representatives of both genotype/serotype combinations were nonsusceptible to penicillin, cefotaxime and erythromycin, suggesting that differences in drug susceptibility did not play a role in these epidemiological changes. A report from Brazil detailing temporal genotype changes among antibiotic-susceptible serotype 1 pneumococci also supports this hypothesis [14].

Our collection of pneumococcal isolates has provided a unique opportunity to study evolution at the *cps* locus over ~70 years. We show that capsular switching has occurred regularly throughout the last 7 decades and prior to both PCV7 introduction and widespread antibiotic use. Identification of *cps* locus recombination breakpoints did not indicate any hotspots for recombination. The development of the capacity for large scale recombination, particularly that involving simultaneous import of the *pbp2x* ± *cps* locus ± *pbp1a* loci, appears to have been an intrinsic feature of pneumococcal biology for several decades.

In light of the mounting evidence of post-PCV7 changes to pneumococcal serotype distributions, it is highly likely that the proliferation of newly generated NVT capsular switch variants will continue to be favoured by PHiD-CV (PCV10) and PCV13 vaccination programs. Similarly, variants expressing NVT in combination with penicillin-nonsusceptibility will have an even greater advantage. Although the magnitude of these selective forces relative to those favouring established genotype/VT associations remains unclear, the implementation of vaccine programs

across the globe will most likely favour the intercontinental spread of NVT and penicillin-nonsusceptible pneumococci.

5.5 Methods

Strains and genome sequencing

A global collection of 426 pneumococci dated from 1937 to 2007 were previously serotyped by the Quellung reaction and genotyped by MLST. Isolates were assigned to CCs by a modified goeBURST [29] method, as previously described [Chapter 2]. CCs were named after the predicted founder ST. When no single founder ST could be determined, CCs were named NoneX, where X was the ST of lowest numeric value within the CC.

Our isolate collection, which was previously assembled and described by Wyres and colleagues [Chapter 2], included 211 global pneumococci (dated 1937 – 1996) collected by the Statens Serum Insititut (SSI), Denmark. These isolates included both penicillin-susceptible (n = 203) and penicillin-nonsusceptible pneumococci (n = 7), and an isolate of unknown penicillin susceptibility (n = 1). The *cps* loci reference isolates, for which the full *cps* locus nucleotide sequence (from *dexB* to *aliA*) has been published and deposited in the Genbank public database [7], were among those provided by the SSI. The 43 Pneumococcal Molecular Epidemiology Network reference clones [46], which represent internationally disseminated, disease causing clones, were added to the collection. Forty-seven modern pneumococcal isolates from Iceland and the USA (dated 1990s - 2000s and including 38 penicillin-susceptible and 25 penicillin-nonsusceptible pneumococci), which had been previously genotyped and were closely

related to STs represented by numerous older pneumococci, were also included. Given that the older pneumococcal isolates were predominantly penicillin-susceptible pneumococci, a literature search was performed to identify studies that reported early penicillin-nonsusceptible isolates. The authors of those papers were contacted and where possible isolates were added to our collection. The available isolates comprised 116 penicillin-nonsusceptible pneumococci and 9 penicillin-susceptible pneumococci dated from 1969 to 1994 and originating from Germany, Papua New Guinea, South Africa, Spain and the USA.

A genetically diverse set of 96 isolates from our collection were selected for whole-genome sequencing on the Illumina platform as previously described [Chapter 4]. Raw sequence data was assembled to contigs using Velvet [60] and contigs were deposited in a BIGS database (BIGSdb) [38]. Sequence data were also deposited in the European Nucleotide Archive, accession: ERR025254 - ERR028719.

Nucleotide sequence analysis

CCs 15, 66, 113, 124, 156/162, 191, 218 and 574 were represented by isolates subjected to whole-genome sequencing and were selected for further *cps* locus analysis. With the exception of CC15, each of the selected CCs included capsular switch variant representatives. Within each CC the ancestral serotype was assumed to be that most commonly associated with members of the CC, or that represented by the oldest isolate. Change of serotype events resulting from point mutation(s) were differentiated from those resulting from recombination events by comparison of the published *cps* locus structures. When there was discordance between the serotype of

the oldest representative and the most commonly represented serotype (n = 4 CCs), the ancestral type was inferred following further study. In two cases (CC156/162 serotype 9V and 9A, CC191 serotype 7F and 7A) the differences between the *cps* loci in question included nucleotide insertion/deletion whereby deletion events were associated with frame-shift mutations. Consequently the ancestral serotypes were assumed to be those which did not contain the deletion (i.e. those that retained a full-length coding sequence). Where the structure of the *cps* loci in question indicated that a change from one to the other was the result of a recombination event, we sought to infer the ancestral serotype by comparison of the *cps* locus and flanking genomic regions from these isolates and other representatives of the same serotypes but different CCs (methodology as below). There were two such cases (CC66 serotype 7B and 9N, CC218 serotype 12F and 7F), in both of which a candidate *cps* donor representative for one of the serotypes could be identified from our genome collection. Consequently it was possible to infer which of the two serotypes in question were most likely to represent a recombinant (i.e. not the ancestral type), because the *cps* locus and flanking genomic regions were highly similar to those of an isolate representing a different CC, plus at least one of the adjacent genomic regions were highly similar to that of isolates representing the alternative serotype in the same CC (the ancestral type). In these cases, the combination of ancestor/recombinant assignments that would explain the observed nucleotide sequence patterns by the simplest combination of recombination events (i.e. a single event) was assumed to be the most likely.

Serotype changes presumed to be the result of recombination at the *cps* locus were further studied as follows: *cps* locus nucleotide sequence comparison was used to identify potential *cps* locus donor representatives among our whole-genome sequenced isolates, the *cps* locus reference sequences, and an additional 131 pneumococcal genomes retrieved from the Genbank public database and deposited in the BIGSdb. Regions spanning the synthesis-related genes of the reference *cps* loci (see below) were BLASTed against the sequences within the BIGSdb. Matching nucleotide sequences were extracted and aligned using the web-based MUSCLE alignment algorithm [20]. Aligned sequences were imported to MEGA5 [56] for visual comparison of variable sites. *cps* locus sequences which were highly similar to the capsular switch representative in question (the putative recombinant) and which had been assigned to a different CC were considered as putative donor representatives. The putative donor *cps* locus sequences differed from the recombinant sequences by 0 - 3 nucleotide substitutions across the entire locus, with the exception of putative within *cps* locus recombination import regions as described for CC218[7F]. The *cps* locus genomic flanking regions for the putative ancestor, donor and recombinant representatives were retrieved from the BIGSdb by BLASTing and extracting each of the *dexB* and *aliA* loci plus flanking sequences (within the BIGSdb the user can request extraction of a specified length of sequence flanking the given BLAST matches). Sequences were aligned by MUSCLE and imported to MEGA5 as described above. Recombination regions were identified as the minimum region over which the recombinant representative differed from the ancestral representative *and* was identical or highly similar (>99.7% sequence identity) to the donor representative. Recombination breakpoints were determined by visual inspection and identified as the

most 5' and 3' variable nucleotide positions at which the recombinant representative differed from the ancestral representative and was identical to the donor representative. Where no suitably matched donor representative could be identified, the maximum recombination region was estimated as the maximum length over which the recombinant representative differed from the ancestral representative. Given the nature of next-generation sequencing technologies it was not possible to obtain completely contiguous sequence over the entire putative import length for all isolates. Missing sequence regions estimated at >100 bp length are stated in the text (estimated by comparison to the ATCC 700669 reference genome; Genbank accession NC_011900.1 [18]). Additionally, the transposase genes, which flank the functional coding genes of most *cps* loci, are notoriously difficult to sequence/assemble and as such were partially or completely missing for numerous isolates.

Serotype changes resulting from point mutation(s) and within serotype, within CC *cps* evolution were studied as follows: Regions of the reference *cps* locus nucleotide sequences spanning from the first to the last synthesis-related gene (i.e. excluding any flanking transposases, Supplementary Table 5.3) were BLASTed against the appropriate genomes within the BIGSdb. These genes included the 6 universally conserved genes (*wzg, wzh, wzd*, *wze, wzx* and *wzy*) and all of the serotype-specific genes. An exception to this rule was the serotype 14 *cps* for which the 3'-most serotype-specific gene, *lrp*, contains a set of repeat regions which do not sequence/assemble well. Consequently, the *lrp* gene was excluded from the serotype 14 *cps* analyses. The *cps* locus sequences were extracted from the BIGSdb, aligned by MUSCLE and imported to MEGA5 as above. Variable sites were identified and their positions within the alignment were

mapped to coding loci using the Genbank reference annotations. Single bp insertions or deletions within poly-A/T sequence regions were assumed to result from sequence/assembly errors and were not subjected to further study. Isolates for which >2 such insertions or deletions were identified were excluded from the analyses, as were isolates for which the *cps* locus sequence was split across >2 assembly contigs.

5.6 Acknowledgements

We thank Dalia Denapaite, Irma Ochigava, and Shwan Rachid for DNA sequencing and MLST analysis; and Andries van Tonder for retrieval of genomic sequence data from Genbank. We also gratefully acknowledge the clinicians, microbiologists, and investigators of the Active Bacterial Core surveillance program of the Emerging Infections Program Network, USA.

5.7 References

1. **Aanensen, D. M., A. Mavroldi, S. D. Bentley, P. R. Reeves, and B. G. Spratt.** 2007. Predicted functions and linkage Specificities of the products of the *Streptococcus pneumoniae* capsular biosynthetic loci. J Bacteriol. **189:**7856-7876.
2. **Albarracín Orio, A. G., P. R. Cortes, M. Tregnaghi, G. E. Piñas, Argentinean Network Pneumococcus Study Group, and J. R. Echenique.** 2008. A new serotype 14 variant of the pneumococcal Spain^{9V}-3 international clone detected in the central region of Argentina. J Med Microbiol. **57:**992-999.
3. **Ansaldi, F., P. Canepa, D. de Florentiis, R. Bandettini, P. Durando, and G. Icardi.** 2011. Increasing incidence of *Streptococcus pneumoniae* serotype 19A and emergence of two vaccine escape recombinant ST695 strains in Liguria, Italy, 7 years after implementation of the 7-valent conjugated vaccine. Clin Vaccine Immunol. **18:**343-345.
4. **Avery, O. T., C. M. Macleod, and M. McCarty.** 1944. Studies on the chemical nature of the substance inducing transformation of pneumococcal types: Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J Exp Med. **79:**137-158.
5. **Barry, M. A., D. E. Craven, and M. Finland.** 1984. Serotypes of *Streptococcus pneumoniae* isolated from blood cultures at Boston City Hospital between 1979 and 1982. J Infect Dis. **149:**449-452.
6. **Beall, B. W., R. E. Gertz, R. L. Hulkower, C. G. Whitney, M. R. Moore, and A. B. Brueggemann.** 2011. Shifting genetic structure of invasive serotype 19A pneumococci in the United States. J Infect Dis. **203:**1360-1368.

7. **Bentley, S. D., D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabbinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail, G. Samuel, I. C. Skovsted, M. S. Kaltoft, B. Barrell, P. R. Reeves, J. Parkhill, and B. G. Spratt.** 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. **2:**262-269.
8. **Black, S., H. Shinefield, B. Fireman, E. Lewis, P. Ray, J. R. Hansen, L. Elvin, K. M. Ensor, J. Hackell, G. Siber, F. Malinoski, D. Madore, I. Chang, R. Kohberger, W. Watson, R. Austrian, and K. Edwards.** 2000. Efficacy, safety and immunogenicity of heptavalent pneumococcal conjugate vaccine in children. Pediatr Infect Dis J. **19:**187-195.
9. **Bratcher, P. E., I. H. Park, M. B. Oliver, M. Hortal, R. Camilli, S. K. Hollingshead, T. Camou, and M. H. Nahm.** 2011. Evolution of the capsular gene locus of *Streptococcus pneumoniae* serogroup 6. Microbiol. **157:**189-198.
10. **Brueggemann, A. B., D. T. Griffiths, E. Meats, T. E. A. Peto, D. W. Crook, and B. G. Spratt.** 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.
11. **Brueggemann, A. B., R. Pai, D. Crook, and B. Beall.** 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog. **3:**e168.
12. **Butler, J. C., R. F. Breiman, H. B. Lipman, J. Hofmann, and R. R. Facklam.** 1995. Serotype distribution of *Streptococcus pneumoniae* infections among preschool children in the United States, 1978-1994: Implications for development of a conjugate vaccine. J Infect Dis. **171:**885-889.
13. **Calix, J. J., and M. H. Nahm.** 2010. A new pneumococcal serotype, 11E, has a variably inactivated *wcjE* gene. J Infect Dis. **202:**29-38.
13a. **Calix, J. J., R. J. Porambo, A. M. Brady, T. R. Larson, J. Yother, C. Abeygunwardana, and M. H. Nahm.** 2012. Biochemical, genetic and serological characterization of two capsule subtypes among *Streptococcus pneumoniae* serotype 20 strains: discovery of a new pnuemococcal serotype. J Biol Chem. ***In press***.
14. **Chiou, A.-C., S. S. Andrade, S. C. G. Almeida, R. C. Zanella, A.-L. Andrade, and M.-C. de Cunto Brandileone.** 2008. Molecular assessment of invasive *Streptococcus pneumoniae* serotype 1 in Brazil: evidence of clonal replacement. J Med Microbiol. **57:**839-844.
15. **Choi, E. H., S. H. Kim, B. W. Eun, S. J. Kim, N. H. Kim, J. Lee, and H. J. Lee.** 2008. *Streptococcus pneumoniae* serotype 19A in children, South Korea. Emerg Infect Dis. **14:**275-281.
16. **Coffey, T. J., M. Daniels, M. C. Enright, and B. G. Spratt.** 1999. Serotype 14 variants of the Spanish penicillin-resistant serotype 9V clone of *Streptococcus pneumoniae* arose by large recombinational replacements of the *cpsA-pbp1a* region. Microbiology. **145:**2023-2031.
17. **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, B. Pichon, S. Baker, C. M. Parry, L. M. Lambertsen, D. Shahinas, D. R. Pillai, T. J. Mitchell, G. Dougan, A. Tomasz, K. P. Klugman, J. Parkhill, W. P. Hanage, and S. D. Bentley.** 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331:**430-434.
18. **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, S. D. Bentley, and T. J. Mitchell.** 2009. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae*^{Spain23F} ST81. J Bacteriol. **191:**1480-1489.
19. **Ding, F., P. Tang, M.-H. Hsu, P. Cui, S. Hu, J. Yu, and C.-H. Chiu.** 2009. Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. BMC Genet. **10:**158.
20. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792-1797.
21. **Elberse, K., S. Witteveen, H. van der Heide, I. van de Pol, C. Schot, A. van der Ende, G. Berbers, and L. Schouls.** 2011. Sequence diversity within the capsular genes of *Streptococcus pneumoniae* serogroup 6 and 19. PLoS One. **6:**e25018.
22. **Enright, M. C., and B. G. Spratt.** 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology. **144:**3049-3060.
23. **Esposito, S., S. Tansey, A. Thompson, A. Razmpour, J. Liang, T. R. Jones, G. Ferrera, A. Maida, G. Bona, C. Sabatini, L. Pugni, E. A. Emini, W. C. Gruber, D. A. Scott, and N. Principi.** 2010.

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:**1017-1026.

24. **Feikin, D. R., and K. P. Klugman.** 2002. Historical changes in pneumococcal serogroup distribution: Implications for the era of pneumococcal conjugate vaccines. Clin Infect Dis. **35:**547-555.
25. **Fenoll, A., J. J. Granizo, L. Aguilar, M. J. Giménez, L. Aragoneses-Fenoll, G. Hanquet, J. Casal, and D. Tarragó.** 2007. Temporal trends of invasive *Streptococcus pneumoniae* serotypes and antimicrobial resistance patterns in Spain from 1979 to 2007. J Clin Microbiol. **47:**1012-1020.
26. **Fenoll, A., I. Jado, D. Vicioso, A. Pérez, and J. Casal.** 1998. Evolution of *Streptococcus pneumoniae* serotypes and antibiotic resistance in Spain: Update (1990 to 1996). J Clin Microbiol. **36:**3447-3454.
27. **Finland, M., and M. W. Barnes.** 1977. Changes in occurence of capsular serotypes of *Streptococcus pneumoniae* at Boston City Hospital during selected years betwen 1935 and 1974. J Clin Microbiol. **5:**154-166.
28. **Foster, D., K. Knox, A. S. Walker, D. T. Griffiths, H. Moore, E. Haworth, T. Peto, A. B. Brueggemann, and D. W. Crook on behalf of the Oxford Invasive Pneumococcal Surveillance Group.** 2008. Invasive pneumococcal disease: epidemiology in children and adults prior to implementation of the conjugate vaccine in the Oxfordshire region, England. J Med Microbiol. **57:**480-487.
29. **Francisco, A. P., M. Bugalho, M. Ramirez, and J. A. Carriço.** 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics. **10:**152.
30. **García, E., and R. López.** 1997. Molecular biology of the capsular genes of *Streptococcus pneumoniae*. FEMS Microbiol Lett. **149:**1-10.
31. **Golubchik, T., A. B. Brueggemann, T. Street, R. E. Gertz Jr, C. C. A. Spencer, T. Ho, E. Giannoulatou, R. Link-Gelles, R. M. Harding, B. Beall, T. E. A. Peto, M. R. Moore, P. Donnelly, D. W. Crook, and R. Bowden.** 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. Nat Genet. **44:**352-355.
32. **Griffith, F.** 1928. The significance of pneumococcal types. J Hyg (Lond) **27:**113-159.
33. **Harboe, Z. B., T. L. Benfield, P. Valentiner-Branth, T. Hjuler, L. Lambertsen, M. Kaltoft, K. Krogfelt, H. C. Slotved, J. J. Christensen, and H. B. Konradsen.** 2010. Temporal trends in invasive pneumococcal disease and pneumococcal serotypes over 7 decades. Clin Infect Dis. **50:**329-337.
34. **Hausdorff, W. P., A. B. Brueggemann, J. G. Hackell, and J. A. G. Scott.** 2008. Pneumococcal serotype epidemiology., *In* G. Siber, K. P. Klugman, and P. H. Mäkelä (ed.), Pneumococcal vaccines: the impact of conjugate vaccine. ASM Press, Washington D.C.
35. **Hiller, N. L., A. Ahmed, E. Powell, D. P. Martin, R. Eutsey, J. Earl, B. Janto, R. J. Boissy, J. Hogg, K. Barbadora, R. Sampath, S. Lonergan, J. C. Post, F. Z. Hu, and G. D. Ehrlich.** 2010. Generation of genetic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal paediatric infection. PLoS Pathog. **6:**e1001108.
36. **Jefferies, J. M. C., A. Smith, S. C. Clarke, C. Dowson, and 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:**5681-5688.
37. **Jin, P., F. Kong, M. Xiao, S. Oftadeh, F. Zhou, C. Liu, F. Russell, and G. L. Glibert.** 2009. First report of putative *Streptococcus pneumoniae* serotype 6D among nasopharyngeal isolates from Fijian children. J Infect Dis. **200:**1375-1380.
38. **Jolley, K. A., and M. C. Maiden.** 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. **10:**595.
39. **Kolkman, M. A. B., W. Wakarchuk, P. J. M. Nuijten, and B. A. M. van der Zeijst.** 1997. Capsular polysaccharide synthesis in *Streptococcus pneumoniae* serotype 14: molecular analysis of the complete *cps* locus and identification of genes encoding glycosyltransferases required for the biosynthesis of the tetrasaccharide subunit. Mol Microbiol. **26:**197-208.
40. **Kwon, H., S. Kim, M. Choi, A. D. Ogunniyi, J. C. Paton, S. Park, S. Pyo, and D. Rhee.** 2003. Effect of heat shock and mutations in ClpL and ClpP on virulence gene expression in *Streptococcus pneumoniae.* Infect. Immun. **71:**3757-3765.

41. **Kyaw, M. H., S. Clarke, G. F. S. Edwards, I. G. Jones, and H. Campbell.** 2000. Serotypes/groups distribution and antimicrobial resistance of invasive pneumococcal isolates: implications for vaccine strategies. Epidemiol Infect. **125:**561-572.
42. **Lagos, R., A. Muñoz, O. San Martin, A. Maldonado, J. C. Hormazabal, W. C. Blackwelder, and M. M. Levine.** 2008. Age- and serotype-specific pediatric invasive pneumococcal disease: insights from systematic surveillance in Santiago, Chile, 1994-2007. J Infect Dis. **198:**1809-1817.
43. **Marttinen, P., W. P. Hanage, N. J. Croucher, T. R. Connor, S. R. Harris, S. D. Bentley, and J. Corander.** 2012. Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res **40:**e6.
44. **Mavroidi, A., D. Godoy, D. M. Aanensen, D. A. Robinson, S. K. Hollingshead, and B. G. Spratt.** 2004. Evolutionary genetics of the capsular locus of serogroup 6 pneumococci. J Bacteriol. **186:**8181-8192.
45. **McEllistrem, M. C., A. C. Noller, S. Visweswaran, J. M. Adams, and L. H. Harrison.** 2004. Serotype 14 variants of the France $9V^{-3}$ clone from Baltimore, Maryland, can be differentiated by the *cpsB* gene. J Clin Microbiol. **42:**250-256.
46. **McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefévre, A. Tomasz, and K. P. Klugman.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. J Clin Microbiol. **39:**2565-2571.
47. **Mera, R. M., L. A. Miller, H. Amrine-Madsen, and D. F. Sahm.** 2009. The impact of the pneumococcal conjugate vaccine on antimicrobial resistance in the United States since 1996: Evidence for a significant rebound by 2007 in many classes of antibiotics. Microb Drug Resist. **15:**261-268.
48. **Mufson, M. A., and R. J. Stanek.** 1999. Bacteremic pneumococcal pneumonia in one American city: a 20-year longitudinal study, 1978-1997. Am J Med. **107:**34S-43S.
49. **Nelson, A. L., A. M. Roche, J. M. Gould, K. Chim, A. J. Ratner, and J. N. Weiser.** 2007. Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance. Infect. Immun. **75:**83-90.
50. **O'Brien, K. L., L. J. Wolfson, J. P. Watt, E. Henkle, M. Deloria-Knoll, N. McCall, E. Lee, K. Mulholland, O. S. Levine, and T. Cherian.** 2009. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet **374:**893-902.
51. **Park, I. H., D. G. Pritchard, R. Cartee, A. Brandao, M. C. C. Brandileone, and M. H. Nahm.** 2007. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. J Clin Microbiol. **45:**1225-1233.
52. **Pillai, D. R., D. Shahinas, A. Buzina, R. A. Pollock, R. Lau, K. Khairnar, A. Wong, D. J. Farrell, K. Green, A. McGeer, and D. E. Low.** 2009. Genome-wide dissection of globally emergent multi-drug resistant serotype 19A *Streptococcus pneumoniae*. BMC Genomics **10:**642.
53. **Prymula, R., and L. Schuerman.** 2009. 10-valent pneumococcal nontypeable *Haemophilus influenzae* PD conjugate vaccine: Synflorix™. Expert Rev Vaccines **8:**1479-1500.
54. **Sandgren, A., K. Sjöstrom, B. O. Olsson-Liljequist, B. Christensson, A. Samuelsson, G. Kronvall, and B. Henriques Normark.** 2004. Effect of clonal and serotype-specific properties on the invasive capacity of *Streptococcus pneumoniae.* J Infect Dis. **189:**785-796.
55. **Scheifele, D., S. Halperin, L. Pelletier, J. Talbot, and Members of the Canadian Paediatric Society/Laboratory Centre for Disease Control Immunization Monitoring Program.** 2000. Invasive pneumococcal infections in Canadian children, 1991-1998: Implications for new vaccination strategies. Clin Infect Dis. **31:**58-64.
56. **Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar.** 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. **28:**2731-2739.
57. **Tran, T. D., H. Kwon, E. Kim, K. Kim, D. E. Briles, S. Pyo, and R. Dong-Kwon.** 2011. Decrease in penicillin susceptibility due to heat shock protein ClpL in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **55:**2714-2728.
58. **Varvio, S., K. Auranen, E. Arjas, and P. H. Mãkela.** 2009. Evolution of the capsular regulatory genes in *Streptococcus pneumoniae.* J Infect Dis. **200:**1144-1151.
59. **Weinberger, D. M., R. Malley, and M. Lipsitch.** 2011. Serotype replacement in disease after pneumococcal vaccination. Lancet **378:**1962-1973.

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

5.8 Supplementary Information

SupplementaryTable 5.1 Genetic evolution at the *cps* locus among isolates representing the same pneumococcal serotypes and CCs.

CC[Serotype]	Isolate	Year	Nucleotide Substitution(s) Compared to Ancestor (*gene*)	Recombination Regions
15[14]	14/5	1967	NA	
	PMEN10	1987	A8695G (*wchM*) Note: missing bp 11509 - 11525 and missing bp 12423 – end*	
	SPnINV200	1995	C12439T (*wciY*)	
	ICE13	1998	C12466A and C12467A (*wciY*)	
	Ala243	1998	C12439T (*wciY*)	
	GA13338	1999	C12439T (*wciY*)	
	Ala317	2001	None	
	ICE50	2003	C12439T, C12466A, C12467A and G12502A (*wciY*) Note: 22 bp insertion at position 11509	
	CGSP14	2004 - 2005	C1010T (*wzg*), C12439T (*wciY*) Note: missing bp 11488 – 11539	*6684 bp 5' pbp2x - cps* bp 367‡
	SP14-BS292	Unknown	T1976A (*wzh*), C12439T (*wciY*)	
66[9N]	9N/6	1960	NA	
	USA8	2001	A4399G (*wchA*), A14000G, T14014G, C14037T, T14042C (between *ugd* and IS1381 putative transposase)	
113[18C]	18C/2	1939	NA	
	18C/1†	1940	A2203G (*wzd*)	
	18C/3	1968	G3478A (*wze*)	
	Netherlands[18C]-36	1980	None	
	USA2	1999	C4622T (*wchA*)	*cps* bp 15861 - 17005
	ICE501	2002	None	

124[14]	14/2†	1952	NA	
	14/4	1961	A3109G and G3478A (*wze*), A5616G and C5617T (*wchK)*, T6662A (*wzy*)	*cps* bp 12330 - 12467
	PMEN35	1980	None. Note: missing bp 12437 – end*	
	14/9	1982	T11491C, A11524G, C12466A, C12467A, G12495A and A12505T (*wciY*)	
	14/7	1992	A4141G (*wchA*)	
	Ala292	1998	C3816A and G4087A (*wchA*)	
	USA6	1999	C43A (*wzg*), A12403G, C12466A, C12467A and G12502A (*wciY*)	
	Ala263	2002	None	
	Ala289	2002	C12466A, C12467A and G12502A (*wciY*)	
	ICE46	2003	None	
	ICE594	2005	C12466A, C12467A and G12502A (*wciY*)	
	CCR11974		G2455T (*wzd*)	
156/162[9V]	9V/5	1991	NA	
	PMEN3	1993	None	
	GA08780	1997	T5230C (*wchO*)	
	SP9-BS68	Unknown	None	
191[7F]	7F/3	1962	NA Note: missing bp 20430 – end*	
	7F/5†	1962	A16G (*wzg*)	
	Netherlands[7F]-39	1984	None Note: missing bp 20434 – end*	
	7F/4	1986	None Note: missing bp 20434 – end*	
	ICE22	1993	C20435T, A20436G, T20441C and A20450G (*glf*)	
	CDC1087-00	1999	T10697C (*wchF*)	
	USA16	2003	None	

218[12F]	12F/5	1988	NA
	Denmark[12F]-34	1995	G16593T (between *fnlA* and *fnlB*)
	12F/6	1996	G16593T (between *fnlA* and *fnlB*), C18121T (*fnlC*)
	USA18	1999	G16593T (between *fnlA* and *fnlB*)
	CDC0288-04	2003-2004	C7397T (*wzy*), G16593T (between *fnlA* and *fnlB*), C18121T (*fnlC*)
218[7F]	7F/2	1952	NA
	ICE23	1993	C5369A (*wchF*), C6499T (*wcwA*), A12622C (*wcwH*)
	USA20	1999	C9059T (*HG140*), C11766G (*wcwH*) Note: missing bp 20431 – end*

NA = Not applicable. * Likely due to sequencing/assembly failure. ‡ See chapter 4. † This isolate was used as the *cps* reference strain in reference [2].

Supplementary Table 5.2 Genetic changes leading to capsular switching events within seven independent pneumococcal CCs.

CC	Ancestral Serotype	New Serotype	Change by Point Mutation or Recombination	Point Mutations	Recombination Donor	*cps* Point Mutations Compared to Donor Representative‡
66	7B	9N	Recombination	-	CC None3782^{9N}	None*
	9N	19F	Recombination	-	Unknown^{19F}	-
	9N	19F	Recombination	-	Unknown^{19F}	-
	9N	14	Recombination	-	CC15^{14}	A8695G, G10976T, C12439T
	9N	14	Recombination	-	CC15^{14}	A4486G, A8695G, G10461A$^{\psi}$
	9N	14	Recombination	-	CC124^{14}	A1704G, T12439C
	9N	23F	Recombination	-	Unknown^{23F}	-
113	18C	18B	Point Mutation	G12382T	-	-
	18C	35C	Recombination	-	Unknown^{35C}	-
	35C	17F	Recombination	-	CC574^{17F}	A8794G
	18C	9V	Recombination	-	Unknown^{9V}	-
124	14	11C	Recombination	-	Unknown^{11C}	-
	14	9L	Recombination	-	Unknown^{9L}	-
156/162	9V	9A	Point Mutation	G1543A, Deletion bp 16996	-	-
	9V	19A	Recombination	-	CC199^{19A}	None
191	7F	7A	Point Mutation	A2375G, A10521T, Single nucleotide insertion at bp 8606	-	-
218	12F	7F	Recombination	-	CC191^{7F}	C6854A, C8089T and 2 putative recombination events marked by 39 nucleotide substitutions bp 1083 – 1279 and 5 nucleotide substitutions plus 2 insertions bp 19486 - 20473
574	17F†	2	Recombination	-	CCNone128^{2}	G3294T, T3339C, G3956T

‡ Calculated over the region spanning the synthesis-related genes only - see section 5.5 *Methods*. * Missing bp 14006 - 15542 (end) likely due to sequencing/assembly failure. $^{\psi}$ Missing bp 12437 - 12516 (end) likely due to sequencing/assembly failure. † This isolate was used as the *cps* reference strain in reference [2].

Supplementary Table 5.3 *cps* loci included in this study.

Serotype	*cps* Reference Length (bp)	Length Included in Analyses (bp)	Genes Included in Analyses (5' - 3')
2	20420	17500	*wzg, wzh, wzd, wze, wchA, wchF, wchG, wchH, wzy, wchI, wzx, ugd, glf, rmlA, rmlC, rmlB, rmlD*
7A	23837	20475	*wzg, wzh, wzd, wze, wchA, wchF, wcwA, wcwC, wcwD, HG140, wcwF, wcwG, wcwH, wzy, wzx, rmlA, rmlC, rmlB, rmlD, glf*
7B	20947	19545	*wzg, wzh, wzd, wze, wchA, wchF, wcwI, wcwL, wcwK, wcxU, wzy, rbsF, wzx, rmlA, rmlC, rmlB, rmlD, glf*
7F	23811	~20450*	*wzg, wzh, wzd, wze, wchA, wchF, wcwA, wcwC, wcwD, HG140, wcwF, wcwG, wcwH, wzy, wzx, rmlA, rmlC, rmlB, rmlD, glf*
9A	20356	17308	*wzg, wzh, wzd, wze, wchA, wchO, wcjA, mnaA, wzy, wcjB, wzx, wcjC, wcjD, tnp, ugd, wcjE*
9L	17433	15541	*wzg, wzh, wzd, wze, wchA, wchO, wcjA, mnaA, wzy, wcjB, wzx, wcjC, ugd, tnp, wcjE*
9N	17434	15542	*wzg, wzh, wzd, wze, wchA, wchO, wcjA, mnaA, wzy, wcjB, wzx, wcjC, ugd, tnp, wcjE*
9V	20671	17309	*wzg, wzh, wzd, wze, wchA, wchO, wcjA, mnaA, wzy, wcjB, wzx, wcjC, wcjD, tnp, ugd, wcjE*
11C	18006	13383	*wzg, wzh, wzd, wze, wchA, wchJ, wchK, wcyK, wcwR, wcrL, wzy, wcwT, wcwU, wzx, gct*
12F	23596	19081	*wzg, wzh, wzd, wze, wciI, wciJ, wcxB, wzy, wcxD, wcxE, wcxF, wzx, mnaB, mnaA, fnlA, fnlB, fnlC*
14	19736	12516	*wzg, wzh, wzd, wze, wchA, wchJ, wchK, wzy, wchL, wchM, wchN, wzx, wciY*
17F	22714	19846	*wzg, wzh, wzd, wze, wchA, wchF, wcxG, abp1, abp2, wciP, wcrT, wcrU, wzy, wcrV, wzx, rmlA, rmlC, rmlB, rmlD, glf*
18B	21637	20252	*wzg, wzh, wzd, wze, wchA, wchF, wciU, wciV, wciW, wzx, wzy, wciX, wciY, gct, HG94, rmlA, rmlC, rmlB, rmlD, glf*
18C	21637	20252	*wzg, wzh, wzd, wze, wchA, wchF, wciU, wciV, wciW, wzx, wzy, wciX, wciY, gct, HG94, rmlA, rmlC, rmlB, rmlD, glf*
19A	18121	14949	*wzg, wzh, wzd, wze, wchA, wchO, wchP, wchQ, wzy, wzx, mnaA, rmlA, rmlC, rmlB, rmlD*
19F	19616	14820	*wzg, wzh, wzd, wze, wchA, wchO, wchP, wchQ, wzy, wzx, mnaA, rmlA, rmlC, rmlB, rmlD*
23F	21813	18653	*wzg, wzh, wzd, wze, wchA, wchF, wzy, wchV, wchW, wzx, wchX, gtp1, gtp2, gtp3, rmlA, rmlC, rmlB, rmlD*
35C	19218	17648	*wzg, wzh, wzd, wze, wchA, wciB, wzy, wcrI, wcrJ, wcrK, mnp1, wcrH, mnp2, wzx, wciG, glf, wcjE*

* See text for details.

6. EXPLORING HISTORICAL NON-BETA-LACTAM RESISTANCE DETERMINANTS AND RESISTANT-DETERMINANT ELEMENT VARIATION AMONG CLONALLY RELATED PNEUMOCOCCI

This chapter describes the study of mobile genetic elements which contain antimicrobial resistance determinants. The genomes of 75 pneumococci from the historical collection (described in Chapters 2 and 4) and an additional 24 pneumococcal genomes retrieved from Genbank (as described in Chapter 5) were included in analyses.

6.1 Abstract

Over the past two/three decades pneumococcal resistance to the non-β-lactam drugs has greatly increased. Macrolide, tetracycline and chloramphenicol resistance are generally conferred by acquisition of the *ermB* and/or *mefA/E*, *tetM* and *cat* genes, respectively. These resistance determinants are associated with mobile genetic elements such as MEGA, Tn*1207.1* and those of the Tn*916* family. The first macrolide, tetracycline and chloramphenicol resistant pneumococci were detected between 1962 and 1970. Tn*916* was described in 1981.

Here the genomes of 38 pneumococcal isolates dated prior to 1980 were probed for the presence of *ermB*, *mefA/E, tetM*, *cat* and *int* (integrase) to indicate the presence of Tn*916*-like, MEGA and Tn*1207.1* elements. Additionally, the presence and structures of elements among the genomes of 80 pneumococci representing globally-disseminated and temporally-persistent clones were compared.

Two Tn*916*-like elements were identified among pneumococci dated 1967 and 1968, respectively. Four Tn*916*-like, three Tn*1207.1* and 14 MEGA elements were identified among isolates dated from 1980 onwards. One Tn*916* family element included a ~31 kb region of similarity to *Streptococcus equi* ICE*Se*2. Among some CCs there was evidence of multiple, independent acquisition of elements, whilst among other CCs there was no evidence of acquisition of any such elements.

6.2 Introduction

Pneumococcal resistance to non-β-lactam antimicrobials is of increasing concern, as it is for many other pathogenic bacterial species. In 2009, 39.2% of pneumococcal isolates in the USA were macrolide-nonsusceptible, >24% were tetracycline-nonsusceptible and 8.2% were chloramphenicol-nonsusceptible [35]. In 2010 rates of erythromycin-nonsusceptible pneumococci in Europe were slightly lower (up to 30%) but tetracycline and chloramphenicol-nonsusceptibility rates in some European countries were higher than those reported for the USA. At that time, 35.1% of pneumococci isolated in Spain and 42.3% of those isolated in France were nonsusceptible to tetracycline, while 26.1% and 20.3% respectively, were chloramphenicol-nonsusceptible [23]. In some Asian countries nonsusceptibility rates were even higher, with up to 96.4% of isolates nonsusceptible to erythromycin, 84.8% nonsusceptible to tetracycline and 66.8% nonsusceptible to chloramphenicol [33, 36].

Nonsusceptibility to erythromycin and other macrolides is most commonly conferred by acquisition of the *ermB* or *mefA/E* genes [7, 25, 43]. Among susceptible pneumococci, macrolides bind the 50s ribosomal subunit thereby preventing peptide elongation [1]. The *ermB* gene encodes an N-methyltransferase which methylates an adenine residue in the V domain of 23s rRNA, subsequently preventing macrolide binding [53]. The macrolide ribosomal binding site overlaps with those of lincosamides and streptogramin B and thus acquisition of *ermB* also confers resistance to these antimicrobials (known as the MLS_B phenotype). *mefA/E* encode macrolide efflux pumps specific to the 14- and 15- membered macrolides [51] (including erythromycin,

clarithromycin and azithromycin [1]). Active efflux allows the resistant cell to reduce its internal macrolide concentration and eliminate the detrimental effects of these drugs. Unlike *ermB*, possession of *mefA/E* does not confer resistance to lincosamides or streptogramin B and is associated with the M-phenotype.

Chloramphenicol and tetracycline nonsusceptibility are conferred through acquisition of the *cat* and *tet* genes (most commonly *tetM*), respectively [21, 55]. Among susceptible cells, chloramphenicol also binds the 50s ribosomal subunit to prevent peptide elongation. The *cat* gene encodes a chloramphenicol acetyltransferase which catalyses the conversion of chloramphenicol into non-ribosomal-binding derivatives [1]. Tetracyclines bind to the 30s ribosomal subunit and inhibit protein synthesis by blocking the interaction of aminoacyl-tRNA with the ribosomal acceptor site. *tetM* encodes a ribosomal protection protein which has been shown to initiate reversal of tetracycline-ribosome binding in a GTP-dependent manner [6].

Any or all of the above-named resistance determinants may be passed from parent to daughter cells during replication. Alternatively, these genes may be acquired through the processes of transformation and/or conjugation, often as part of a mobile genetic element. Among pneumococci *mefA* is associated with a genetic element known as Tn*1207.1* [46]. This 7,244 bp element comprises 8 open reading frames, including those putatively required for self-mobilisation [46]. *mefE* is associated with the macrolide efflux genetic assembly (MEGA) element, which is ~5.5 kb in length and comprises five open reading frames. Unlike Tn*1207.1*, the MEGA element does not

contain the genes necessary for self-mobilisation [27]. Furthermore, MEGA elements may be integrated into the pneumococcal genome at a variety of locations, but Tn*1207.1* is almost always inserted at the same position [16, 17, 46].

In recent years it has become apparent that the majority of chloramphenicol and/or tetracycline nonsusceptible pneumococci harbour genetic elements known as integrative conjugative elements (ICEs), and that many of these are members of the Tn*916* family [10, 30, 40, 48]. Tn*916* itself was first described in 1981 from an *Enterococcus faecalis* isolate [26], but there is evidence suggesting that this or similar element(s) have been present among pneumococci since the 1970s [11, 15, 49]. To date, Tn*916*-like elements have been identified among >35 bacterial genera [45].

Tn*916* contains 24 open reading frames arranged into functional modules, each of which is associated with either conjugal transfer, recombination, transcriptional regulation or accessory functions, such as antimicrobial resistance (see Figure 6.1 [44, 45]). Tn*916*, which contains *tetM*, has the ability to transpose intracellularly and to mediate its own intercellular transfer via conjugation [44]. Successful conjugal transfer between pneumococci and into pneumococci from other species is a strain dependent property [9, 47]. Transfer between pneumococci by transformation has also been documented [18].

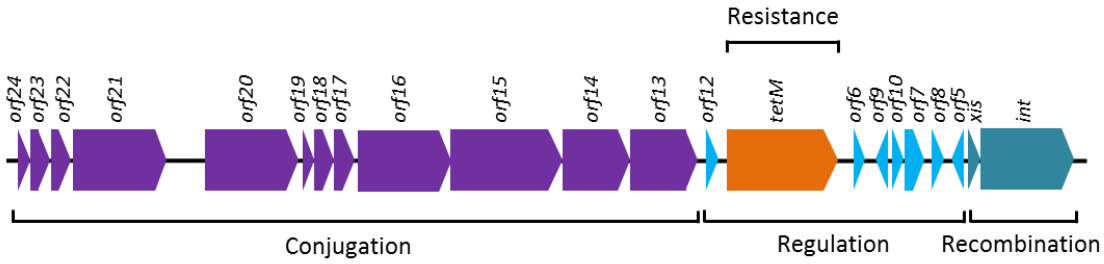


Figure 6.1 Organisation of open reading frames within Tn*916*.
Open reading frames within the ICE, Tn*916*, are organised into functional modules as marked. The accessory gene, *tetM*, conferring tetracycline resistance is inserted within the transcriptional regulation module (Regulation). The recombination module includes the excisionase (*xis*) and integrase (*int*) genes.

Aside from Tn*916*, other members of this family may contain alternate accessory genes and/or feature an independent genetic element integrated within the Tn*916* genetic background. Additionally, Tn*916*-like elements themselves may be integrated within independent genetic elements. Commonly, such composite elements are formed by association of a Tn*916*-like element with a Tn*5252*-like element [3, 12, 13, 19, 30, 40], containing *cat* and *umuDC* (shown to provide protection from UV damage [41]). Among pneumococci, several members of the Tn*916* family and composite elements have been identified (Table 6.1).

The ICEs described in Table 6.1 demonstrate the dynamic variability of the Tn*916* family. Such variability has also been demonstrated by comparative studies of ICEs identified from independent pneumococcal isolates [12, 30, 38, 40, 42].

Table 6.1 Pneumococcal Tn*916* family and composite ICEs.

Element	Features	Resistance Determinants	Reference(s)
Tn*916*	24 open reading frames, including *tetM*.	*tetM*	[45]
Tn*1545*	Tn*916* with insertion sequence IS1239 integrated between *orf13* and *orf12* and a MAS element (macrolide-amini-glycoside-streptothricin element containing *ermB, sat4* and *aphA-3* which confer resistance to macrolides, streptothricin and kanamycin, respectively) integrated between *orf20* and *orf19*.	*tetM, ermB, sat4, aph-3*	[10]
Tn*2008*	Tn*916*-like element integrated into a *cat* containing Tn*5252*-like element. The Tn*916*-like element itself contains a MAS element in the place of the *tetM* gene.	*cat*, *ermB, sat4, aph-3*	[19]
Tn*2009*	Tn*916* with a MEGA element inserted within *orf6*.	*tetM*, *mefE*	[18]
Tn*2010*	Tn*916* with an erythromycin resistance element inserted between *orf20* and *orf19,* and a MEGA element inserted within *orf6*.	*tetM, ermB, mefE*	[38]
Tn*3872*	Tn*916* with a Tn*917* element (contains *ermB*) integrated between *tetM* and *orf6*.	*tetM, ermB*	[10, 19]
Tn*5251*	Highly similar to Tn*916*, differing by 79 nucleotide substitutions (69/79 located within *tetM*).	*tetM*	[47]
Tn*5253*	Tn*5251* integrated into Tn*5252*.	*cat, tetM*	[40]
Tn*6002*	Tn*916* with a Tn*917* element integrated between *orf20* and *orf19*.	*tetM, ermB*	[10]
Tn*6003*	Tn*916* with a MAS element integrated between *orf20* and *orf19*.	*tetM, ermB*	[10]
Tn*6058*	Tn*6002* integrated into Tn*5252*.	*cat, tetM, ermB*	[40]
ICE*Sp*23FST81	Tn*916*-like element integrated into a Tn*5252*-like element, containing a lantibiotic synthesis/export gene cluster.	*cat, tetM*	[13]
ICE*Spn*11876	Tn*916*-like element (containing an erythromycin resistance element) integrated into a Tn*5252*-like element lacking *cat*. The Tn*916*-like element is reverse oriented and integrated further 5' within the Tn*5252*-like element than that of ICE*Sp*23FST81.	*tetM, ermB*	[12]
ICE*Spn*11930	Tn*916*-like element (containing an erythromycin resistance element and Tn*917*) integrated into a Tn*5252*-like element. The Tn*916*-like element is reverse oriented and integrated further 5' within the Tn*5252*-like element than that of ICE*Sp*23FST81.	*cat, tetM, ermB*	[12]

Pneumococcal tetracycline resistance was first identified in 1962 [24]. Pneumococcal resistance to erythromycin and chloramphenicol were recognised in 1967 [20] and 1970 [14], respectively. Therefore, given the important role of the Tn*916* family, MEGA and Tn*1207.1* elements in the dissemination of resistance determinants, it seems likely that they or similar elements were present among pneumococcal populations from at least as early as the 1960s. The aim of this study was to search for, characterise and compare such older elements among pneumococci isolated prior to 1980, and thus pre-dating the earliest descriptions of Tn*916*, MEGA and Tn*1207.1* [26, 27, 46]. Variation in the presence/absence and structures of these elements among clonally related pneumococci, representing internationally-disseminated and temporally-persistent clones is also described.

6.3 Methods

96 pneumococci were subjected to whole-genome sequencing on the Illumina platform and assembled to contigs using Velvet [56] as described in Chapters 2 and 4. A further 131 pneumococcal genomes were retrieved from the Genbank public database as described in Chapter 5. All data were deposited in a BIGS database [34]. All isolates were assigned to clonal complexes (CCs) based on multilocus sequence type data as described in Chapter 2.

The BIGS database BLAST tool was used to search for the *ermB*, *mefA/E*, *cat* and *tetM* resistance determinants within the genomes of isolates dated prior to 1980 (n = 38, Table 6.2) and isolates that were assigned to CCs represented by >7 isolates, at least

one of which was dated prior to 1980 (n = 80 isolates representing 7 CCs, Table 6.3). The criterion for inclusion of CCs with >7 representatives was chosen to ensure a reasonable number of representatives from each of several diverse CCs were included for comparison. Note that the D39 and R6 strains were dated 1916 and 1964 respectively, but they were not included among the 38 pre-1980 isolates studied here since their genomes have been studied in detail elsewhere [31, 37].

Reference nucleotide sequences for each of the above named resistance determinants were retrieved from Genbank. Additionally, integrase (*int*) gene sequences were retrieved from all uniquely designated pneumococcal ICEs identified by interrogation of Genbank using the following search term combinations: Organism = *Streptococcus pneumoniae* + Title = *Transposon*, Organism = *Streptococcus pneumoniae* + Title = *Conjug**, Organism = *Streptococcus pneumoniae* + Title = *element*. Three of the 9 elements each contained two independent *int* genes. A total of 12 *int* nucleotide sequences were aligned by MUSCLE [22] and imported to MEGA5 [52] for assignment to clusters (by visual comparison), and sequence identity calculation. A single representative sequence from each cluster was BLASTed against the genomes of isolates listed in Tables 6.2 and 6.3 as a probe for the presence of ICEs.

Table 6.2 Pneumococcal isolates dated prior to 1980 for which genomic sequence data was included in analyses.

Isolate	Serotype	Year	Country	ST	CC
7A/2 *	7A	1937	Denmark	191	191
14/3	14	1939	Denmark	875	1106
17F/2 *	17F	1939	Denmark	123	113
18C/2 *	18C	1939	Denmark	4706	113
22A/2	22A	1939	Denmark	7181	490
18B/2 *	18B	1941	Denmark	4706	113
35C/2 *	35C	1941	Denmark	7196	113
2/3	2	1943	Denmark	3744	490
35C/3 *	35C	1943	Denmark	5989	113
23A/2	23A	1945	Denmark	439	439
7B/2 *	7B	1952	USA	7180	66
7F/2 *	7F	1952	USA	7210	218
9L/2 *	9L	1952	USA	5979	124
9N/2	9N	1952	USA	7205	None3782
14/2 *	14	1952	USA	124	124
17F/1	17F	1952	USA	574	574
19F/8	19F	1952	Denmark	7229	4399
2/2	2	1956	USA	574	574
10F/2	10F	1956	USA	7186	490
11C/1 *	11C	1957	USA	7201	124
9N/6 *	9N	1960	Denmark	71	66
12F/3	12F	1961	Denmark	7228	4399
14/4 *	14	1961	Denmark	134	124
18F/1	18F	1961	USA	7182	490
7F/3 *	7F	1962	Denmark	191	191
9A/1 *	9A	1962	USA	312	156/162
12F/2	12F	1962	USA	7194	Singleton7194
17F/3	17F	1962	Denmark	392	392
17F/4	17F	1962	Denmark	392	392
19F/5	19F	1962	Denmark	7229	4399
19F/10	19F	1963	Denmark	7230	Singleton7230
14/5 *	14	1967	Denmark	15	15
23F/4	23F	1967	Australia	7184	Singleton7184
9V/4 *	9V	1968	Denmark	123	113
18C/3 *	18C	1968	Denmark	7195	113
19F/11 *	19F	1972	Denmark	71	66
14/8	14	1976	Denmark	7206	554
23F/5	23F	1979	Germany	439	439

* Also included in Table 6.3.

Table 6.3 Identification of the *ermB, mefA/E, cat* and *tetM* resistance determinants among 80 pneumococcal genomes representing 7 unique CCs.

Isolate	Accession*	Serotype	Year	Country	ST	CC	*ermB*	*mefA*	*mefE*	*cat*	*tetM*
PMEN9		14	1993	England	9	15		x			
SPnINV200	FQ312029.1	14	1995	England	9	15		x			
ICE13		14	1998	Iceland	9	15		x			
Ala243		14	1998	USA	13	15			x		
GA13338	AGOX01000001-15	14	1999	USA	13	15			x		
Ala317		14	2001	USA	13	15			x		
ICE50		14	2003	Iceland	13	15			x		
GA41688	AGPP01000001-11	14	2004	USA	13	15			x		
BS397	NZ_ABWC01000001-22	NT	Unknown	Unknown	13	15			x		
BS457	NZ_ABWB01000001-97	NT	Unknown	Unknown	13	15			x		
BS458	NZ_ABWA01000001-109	14	Unknown	Unknown	13	15			x		
SP14-BS292	NZ_ABWQ01000001-25	14	Unknown	Unknown	13	15			x		
14/5		14	1967	Denmark	15	15					x
CGSP14	NC_010582.1	14	2004-2005	Taiwan	15	15	x			x	x
CSR14-10		14	1987	CSR	20	15	x			x	x
BS455	NZ_ADHN01000001-113	NT	Unknown	Unknown	2011	15					
JJA	CP000919.1	14	1995	Brazil	66	66					
USA9		14	1999	USA	66	66					
USA12		23F	2001	USA	66	66					
USA13		19F	2005	USA	66	66					
USA11		14	Unknown	USA	66	66					x
PMEN18		14	1997	USA	67	66	x				x
9N/6		9N	1960	Denmark	71	66					
19F/11		19F	1972	Denmark	71	66					
7B/2		7B	1952	USA	7180	66					
USA8		9N	2001	USA	8119	66					

PMEN36		18C	1980	Netherlands	113	113				
USA2		18C	1999	USA	113	113				
ICE501		18C	2002	Iceland	113	113				
17F/2		17F	1939	Denmark	123	113				
9V/4		9V	1968	Denmark	123	113				
GA13637	AGPA01000001-21	18C	1999	USA	3060	113				
18C/2		18C	1939	Denmark	4706	113				
18B/2		18B	1941	Denmark	4706	113				
35C/3		35C	1943	Denmark	5989	113				
18C/3		18C	1968	Denmark	7195	113				x
35C/2		35C	1941	Denmark	7196	113				
14/2		14	1952	USA	124	124				
PMEN35		14	1980	Netherlands	124	124				
14/9		14	1982	Denmark	124	124				
Ala292		14	1998	USA	124	124				
USA6		14	1999	USA	124	124				
Ala263		14	2002	USA	124	124				
Ala289		14	2002	USA	124	124				
ICE46		14	2003	Iceland	124	124				
ICE594		14	2005	Iceland	124	124				
CCRI1974	NZ_ABZC01000001-121	14	Unknown	Unknown	124	124				
SP14-BS69	NZ_ABAD01000001-49	14	Unknown	Unknown	124	124				
14/4		14	1961	Denmark	134	124				
GA13494	AGOZ01000001-15	14	1999	USA	656	124	x	x	x	
9L/2		9L	1952	USA	5979	124				
14/7		14	1992	Denmark	7198	124				
11C/1		11C	1957	USA	7201	124				

PMEN3		9V	1993	France	156	156/162	
GA08780	AGQJ01000001-11	9V	1997	USA	156	156/162	x
GA54644	AGQA01000001-7	19A	2008	USA	156	156/162	x
GA17570	AFGB01000001-9	9V	Unknown	Unknown	156	156/162	x
9V/5		9V	1991	Denmark	162	156/162	
9V/6		9V	1994	Denmark	162	156/162	
9A/1		9A	1962	USA	312	156/162	
SP9-BS68	NZ_ABAB01000001-61	9	Unknown	Unknown	1269	156/162	
GA47388	AGPU01000001-7	19A	2006	USA	4026	156/162	x
GA44511	AGOR01000001-24	19A	2005	USA	4464	156/162	x
7A/2		7A	1937	Denmark	191	191	
7F/3		7F	1962	Denmark	191	191	
PMEN39		7F	1984	Netherlands	191	191	
ICE22		7F	1993	Iceland	191	191	
CDC1087-00	ABFT01000001-75	7F	1999	USA	191	191	
USA16		7F	2003	USA	191	191	
GA47283	AGPR01000001-16	7F	2006	USA	191	191	
7F/4		7F	1986	Scotland	7208	191	
12F/5		12F	1988	Denmark	218	218	
ICE23		7F	1993	Iceland	218	218	
PMEN34		12F	1995	Canada	218	218	
12F/6		12F	1996	Denmark	218	218	
USA18		12F	1999	USA	218	218	
USA20		7F	1999	USA	218	218	
CDC0288-04	ABGF01000001-38	12F	2003-2004	USA	220	218	
GA47439	AGPV01000001-11	7F	Unknown	USA	1176	218	
7F/2		7F	1952	USA	7210	218	

* Accession numbers for genomic sequences retrieved from Genbank.

Isolates positive for any resistance determinant or *int* gene (n = 24) were subjected to further study by extraction of the relevant Velvet contig from the BIGS database and comparison to the genomes of the reference strains R6 and PMEN1 (Genbank accessions AE007317.1 and NC_011900.1, respectively). Sequences were visually compared using the Artemis Comparison Tool (ACT) [8]. BLAST comparison files were generated using the web-based DoubleACT, hosted by the Health Protection Agency (http://www.hpa-bioinfotools.org.uk/pise/double_act.html). Candidate genetic element regions were identified as those which contained the resistance determinants and/or *int* genes, and were not similar to any region of the R6 genome. Candidate Tn*916*-like elements were further identified as those which showed similarity to the Tn*916*-like region of the PMEN1 reference strain ICE, ICE*Sp*23FST81. To ensure that the entire length of any Tn*916*-like element was identified, all contigs corresponding to each of the Tn*916*-like positive genomes were ordered to the PMEN1 reference using Abacas, and viewed in ACT [8]. This allowed identification of putative Tn*916*-like element regions which were split across multiple assembly contigs. Candidate genetic elements were subsequently extracted from the query contigs/ordered sequences.

Genetic elements identified among isolates assigned to the same CC were compared using DoubleACT and ACT [8]. Putative open reading frames within Tn*916*-like elements were predicted using Prodigal [32]. Where possible, putative functions were assigned by BLAST match to sequences deposited in Genbank. Comparable sequence regions (as indicated by positive DoubleACT BLASTn matches) were aligned by MUSCLE [22] and imported to MEGA5 [52] for visual inspection / sequence identity calculation.

6.4 Results

Nucleotide comparison of pneumococcal ICE int genes

The *int* gene nucleotide sequence alignment showed four clearly defined sequence clusters designated below. Sequences assigned to the same cluster shared >98% sequence identity.

1) int_{Tn916}, comprising *int* from Tn*1311*, Tn*1545*, Tn*2010*, Tn*5251*, Tn*5253*(a), Tn*6003*, ICE*Spn*11876(a) and ICE*Spn*11930(a) (all of which represented Tn*916*-like elements or were known to carry Tn*916*-like *int*, Genbank accessions FN667862.2, X61025.1, AB426620.1, FJ711160.1, EU351020.1, AM410044.5, FR671404.1 and FR671403.1, respectively).

2) $int_{ICESp23FST81}$, comprising *int* from ICE*Spn*11876(b) and ICE6094 (both highly similar to *int* from ICE*Sp*23FST81, Genbank accessions FR671404.1 and FR670347.2, respectively).

3) $int_{ICESpn8140}$, comprising *int* from ICE*Spn*8140 only (Genbank accession FR671412.1).

4) $int_{ICESpn11930}$, comprising *int* ICE11930(b) and Tn*5253*(b) (Genbank accessions FR671403.1 and EU351020.1, respectively).

Resistance determinants and ICE int genes among pneumococci isolated prior to 1980

In total, 18 unique CCs and 21 unique serotypes were represented by the 38 isolates dated prior to 1980 (Table 6.2, isolated between 1937 and 1979, inclusive). Two of

these isolates, 14/5 (1967) and 18C/3 (1968), were positive for both int_{Tn916} and *tetM*. Within the 14/5 genome these genes were located on a Tn*916* element, the entire length of which was structurally identical (i.e shared the same putative open reading frame content / order and was identified as a BLASTn match) to the Tn*916* region of ICE*Sp*23FST81 (Figure 6.2). Within the 18C/3 genome, int_{Tn916} and *tetM* were located on a Tn*916* element, itself associated with a bacteriophage showing similarity to regions of the *Streptococcus* phage 040922 (Genbank accession FR671406.1, Figure 6.3). The 18C/3 Tn*916*-like region was structurally similar, but not identical (i.e. shared the same putative open reading frames in the same order but was not a complete BLASTn match), to those of 14/5 and ICESp23FST81. The structure of the 18C/3 Tn*916*-like region differed from those of 14/5 and ICESp23FST81 by a 155 bp deletion between *orf12* and *tetM.* No other isolates dated prior to 1980 were positive for any of the *ermB*, *mefA/E*, *cat*, *tetM* or *int* genes.

The nucleotide sequence of the 14/5 element differed from that of the ICE*Sp*23FST81 Tn*916*-like region by only 101 substitutions and a 5 bp insertion within *orf14*. 94 of the nucleotide differences between these sequences were located within the *tetM* gene. The 18C/3 element nucleotide sequence was highly similar to that of the 14/5 element differing by only two nucleotide substitutions (and the deletion described above).

Comparison of resistance determinants and ICE int genes among isolates representing the same CC

No resistance determinants or ICE *int* genes were identified among isolates assigned to CCs 191 (n = 8) or 218 (n = 9). One of 11 CC113 isolates, 18C/3 (1968), was positive for

both int_{Tn916} and *tetM* as described above (Table 6.3). One of 16 CC124 isolates, GA13494 (1999), was positive for *mefE*, *cat*, *tetM* and int_{Tn916} (Table 6.3). These genes were located on a genetic element which was structurally highly similar to ICE*Sp*23FST81. Unlike ICE*Sp*23FST81, the GA13494 ICE contained a region adjacent to the *tetM* locus representing a MEGA element, and as such containing the *mefE* gene (Figure 6.3). No other CC113 or CC124 isolates were positive for any of the resistance determinants of *int* genes.

Two of 10 CC66 isolates were positive for ≥1 resistance determinant and/or *int* genes. These were PMEN18 (1997) and USA11 (date unknown, but presumed later than 1990). PMEN18 was positive for *ermB*, *tetM* and int_{Tn916} while USA11 was positive only for *tetM* (Table 6.3). In both cases the genes were located on Tn*916*-like composite elements. The PMEN18 element contained regions of similarity to those of ICE*Sp*23FST81 (Figure 6.4). There was no region of similarity to the 5'-most section of ICE*Sp*23FST81, which contained $int_{ICESp23FST81}$, *cat* and the lantibiotic synthesis/export gene cluster. Instead the 5'-most region of the PMEN18 element contained two open reading frames which were absent from ICE*Sp*23FST81 and were predicted to encode hypothetical proteins. Two further such putative hypothetical protein-encoding open reading frames and an erythromycin resistance element were also present in the PMEN18 element, but were absent from ICE*Sp*23FST81. The 2,855 bp PMEN18 erythromycin resistance element, located between *orf20* and *orf19*, contained an open reading frame predicted to encode a hypothetical protein, the *ermB* gene and a transposase. This element showed 100% identity to regions of the Tn*916*-like ICE of CC81 strain H03400032 (Genbank accession RR671414.1), *S. pneumoniae* Tn*2010*

(Genbank accession AB426623.1), *Streptococcus cristatus* Tn*6002* (Genbank accession AY898750) and regions of the genomes of *S. pneumoniae* strains 670 (Genbank accession CP002176.1) and PMEN6 (Genbank accession NC_010380.1).

The USA11 element contained a region which was structurally similar to Tn*916*, with the exception that the int_{Tn916} and *xis* (excisionase) genes were absent. The USA11 element also contained a large region (31,067 bp) which was not similar to ICE*Sp*23FST81 and contained 20 putative open reading frames (Figure 6.4). This region showed similarity to part of ICE*Se*2 described for *Streptococcus equi* subsp. equi strain 4047 (from just 5' of the SEQ1267 locus to just 3' of the SEQ1250 locus, Genbank accession AM909652.1). Comparison of the nucleotide sequences of the PMEN18 and USA11 Tn*916*-like regions showed that these sequences differed by ~1.7% nucleotide divergence.

Five of 10 CC156/162 isolates, were positive for the *mefE* gene (Table 6.3). In all cases this gene was located on a MEGA element. Four of these elements (from isolates dated 1997, 2005, 2006 and 2008 respectively) were each 5,412 bp in length and were located at the same position within the genome (equivalent to nucleotide position 175,318 of the R6 genome). Nucleotide sequence comparison showed that these elements shared >99.9% sequence identity.

The fifth CC156/162 MEGA element was identified within the genome of GA17570 (date unknown but presumed later than 1990) and was located at nucleotide position 101,134 relative to the R6 genome. This 5,520 bp element contained an additional 92

bp between *mefE* and the adjacent *mel* gene compared to those of the other CC156/162 representatives; however, in comparison to the other CC156/162 MEGA elements, the GA17570 sequence also harboured a 16 bp deletion at nucleotide positions 942-957 and 6 single bp deletions between positions 457 and 5,336 (five of these were associated with polyA/T regions and thus may have resulted from sequence and/or assembly error).

Nine of 16 CC15 isolates were positive for *mefE*. All of these isolates were multilocus sequence type (ST) 13. Five ST13 isolates were dated between 1998 and 2004. The remaining ST13 isolates were of unknown date, but presumed later than 1990. The *mefE* genes from these pneumococci were located on 5,412 bp MEGA elements, which shared >99.9% nucleotide sequence identity. These MEGA elements also shared the same degree of nucleotide identity and genome insertion sites as the first four CC156/162 isolates described above.

A further three CC15 isolates, all assigned ST9 and dated between 1993 and 1998, were positive for *mefA.* In all three cases the *mefA* genes were located on a 7,244 bp Tn*1207.1* element inserted at nucleotide position 852,275 relative to the R6 genome. Nucleotide sequence comparison showed that only a single substitution at bp 1,787 differentiated these elements (G ➔ A for isolate PMEN9).

The CC15 14/5 isolate (1967) was positive for *tetM* and int_{Tn916} as described above. CGSP14 (2004-2005) was positive for *ermB*, *tetM*, *cat* and $int_{ICESp23FST81}$. The genome of

this isolate included an ICE similar to that of ICE*Sp*23FST81 and named Tn*2008* [19]. CGSP14 also contained a second ICE (Tn*3872*-like) as previously described [19].

The PMEN10 isolate (1987) was also positive for *ermB*, *cat* and *tetM*; however, unlike CGSP14, this isolate was positive for int_{Tn916} but not $int_{ICESp23FST81}$. The PMEN10 element included a Tn*916*-like region in the reverse orientation compared to those of ICE*Sp*23FST81 and Tn*2008* (Figure 6.2). Additionally, the PMEN10 Tn*916*-like region was re-arranged compared to ICE*Sp*23FST81 and the other Tn*916*-like elements described here, whereby the *orf7* – int_{Tn916} region was located at the 5' end of the element upstream of the remaining Tn*916*-like region and the other regions which were similar to those of ICE*Sp*23FST81. A 5,267 bp erythromycin resistance element was also inserted within the PMEN10 Tn*916*-like sequence immediately 3' to the *orf7* – int_{Tn916} region. This erythromycin resistance element showed >99.8% sequence identity to the Tn*917* region of the CGSP14 Tn*3872*-like ICE, and the 3'-most region of the erythromycin resistance element inserted in Tn*2008* (Figure 6.5). In contrast, the PMEN10 erythromycin resistance element differed from that described above for PMEN18, by the presence of different transposon genes (Figure 6.5).

In addition to the erythromycin resistance element, the PMEN10 element contained three open reading frames which were absent from ICE*Sp*23FST81; these open reading frames were associated with a hypothetical protein, a neopullulanase and an ABC transporter. Like Tn*2008* of CGSP14, the lantibiotic synthesis/export gene cluster was absent from the PMEN10 element.

Nucleotide sequence comparison indicated that the comparable regions of the 14/5 and CGSP14 Tn*2008* elements were highly similar and separated only by 5 substitutions and an insertion of 3 bp length in *orf13* of CGSP14. The 14/5 and PMEN10 sequences were also highly similar across their comparable regions, excluding the *tetM* gene which differed by 65 substitutions between these isolates. Outside of the *tetM* locus, the 14/5 and PMEN10 Tn*916*-like nucleotide sequences differed only by 5 substitutions and two insertions each of 3 bp length in *orf13* and *orf9* of PMEN10.

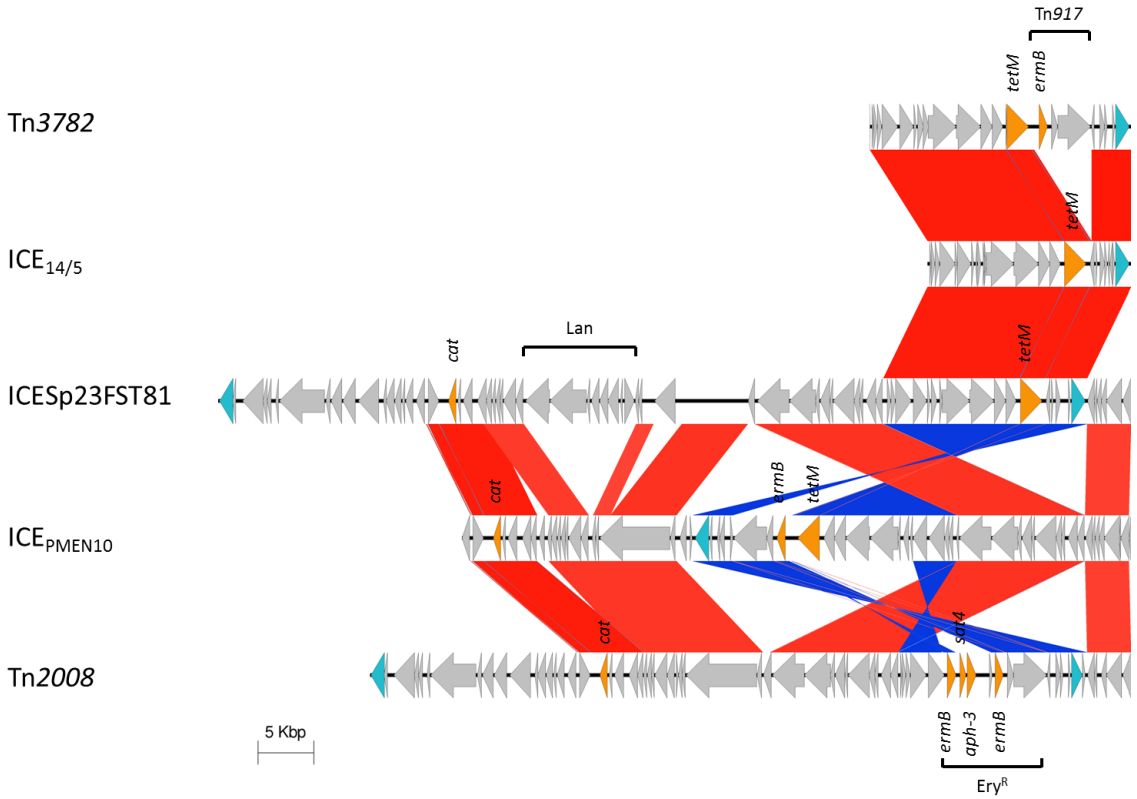


Figure 6.2 Comparison of the Tn*916*-like ICEs of PMEN1 and isolates assigned to CC15.
The PMEN1 ICE*Sp*23FST81 reference ICE is compared to the ICEs of 14/5 ($ICE_{14/5}$), PMEN10 (ICE_{PMEN10}) and the two previously described CGSP14 ICEs (Tn*3782* and Tn*2008*). Putative open reading frames are marked by arrows. Cyan arrows represent *int* genes. Orange arrows represent named antibiotic resistance determinant genes. The Tn*917* element, lantibiotic synthesis/export gene cluster (Lan) and erythromycin resistance element (Ery^R) are marked. BLASTn sequence matches are connected by red bars. Reverse oriented BLASTn sequence matches are connected by blue bars. Figure produced using the Easyfig software available at http://easyfig.sourceforge.net [50].

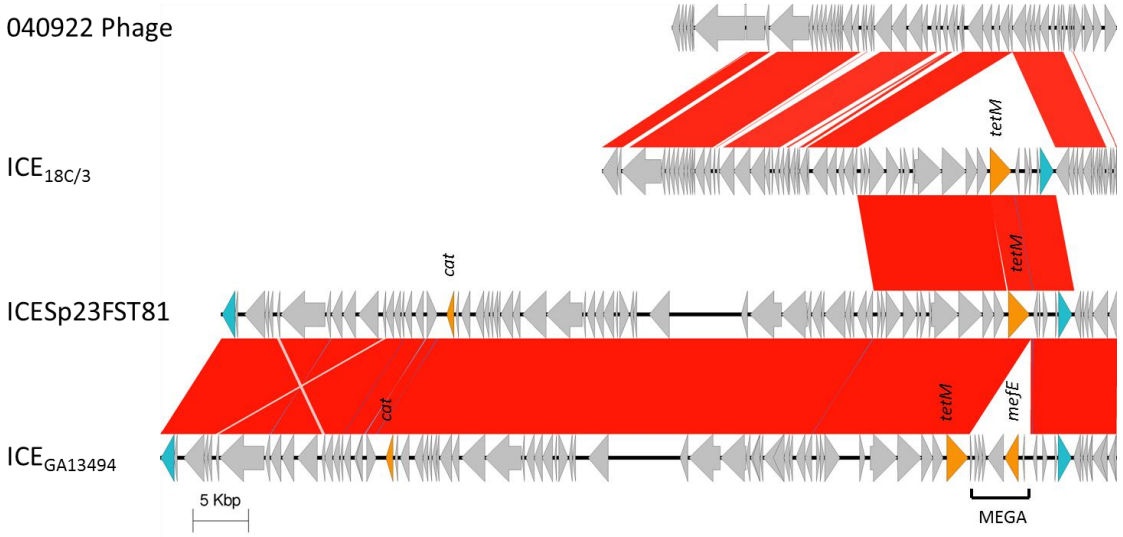


Figure 6.3 Comparison of the Tn*916*-like ICEs of PMEN1, 18C/3 (CC113) and GA13494 (CC124).
The PMEN1 ICE*Sp*23FST81 reference ICE is compared to the ICEs of 18C/3 ($ICE_{18C/3}$) and GA13494 ($ICE_{GA13494}$). The *Streptococcus* 040922 phage, which shows similarity to regions of $ICE_{18C/3}$ is also shown. Putative open reading frames are marked by arrows. Cyan arrows represent *int* genes. Orange arrows represent named antibiotic resistance determinant genes. The MEGA element is marked. BLASTn sequence matches are connected by red bars. Reverse oriented BLASTn sequence matches are connected by blue bars. Figure produced using the Easyfig software available at http://easyfig.sourceforge.net [50].

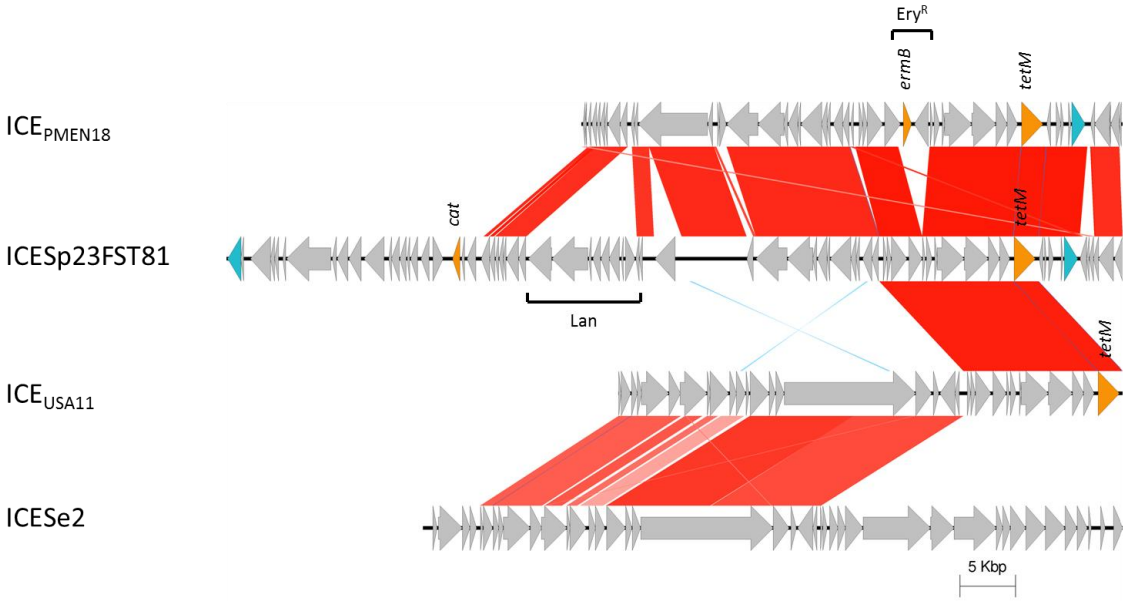


Figure 6.4 Comparison of the ICEs of PMEN1, pneumococcal isolates assigned to CC66 and *S. equi* strain 4047.
The PMEN1 ICE*Sp*23FST81 reference ICE is compared to the ICEs of PMEN18 (ICE_{PMEN18}), USA11 (ICE_{USA11}) and *S. equi* strain 4047 (ICE*Se*2). Putative open reading frames are marked by arrows. Cyan arrows represent *int* genes. Orange arrows represent named antibiotic resistance determinant genes. The lantibiotic synthesis/export gene cluster (Lan) and erythromycin resistance element (Ery^R) are marked. BLASTn sequence matches are connected by red bars. Reverse oriented BLASTn sequence matches are connected by blue bars. Figure produced using the Easyfig software available at http://easyfig.sourceforge.net [50].

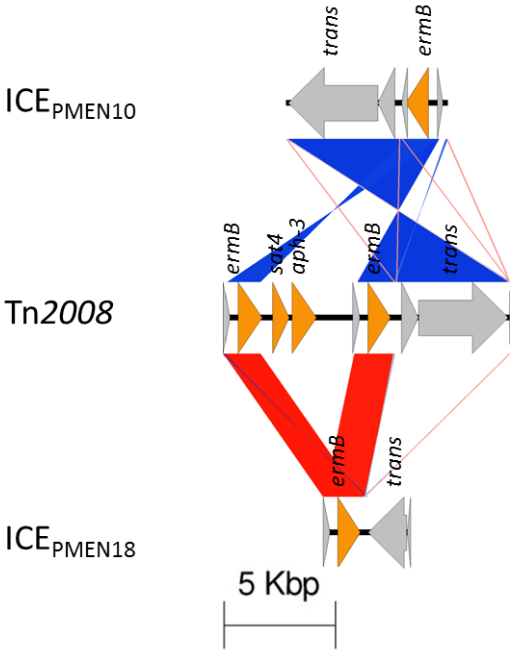


Figure 6.5 Comparison of erythromycin resistance elements of PMEN10, PMEN18 and the CGSP14 reference.
Erythromycin resistance determinants inserted in the ICEs present in the PMEN10 (ICE_{PMEN10}) and PMEN18 (ICE_{PMEN18}) genomes are shown in comparison to that inserted within the Tn*2008* element of CGSP14. Putative open reading frames are marked by arrows. Orange arrows represent named antibiotic resistance determinant genes. Transposons are also marked (*trans*). BLASTn sequence matches are connected by red bars. Reverse oriented BLASTn sequence matches are connected by blue bars. Figure produced using the Easyfig software available at http://easyfig.sourceforge.net [50].

6.5 Discussion

This study of an historical collection of pneumococci has led to the discovery of two of the earliest known representatives of the Tn*916* family of ICEs. These genetic elements, from isolates dated 1967 and 1968, were structurally highly similar to, and shared a high level of sequence identity to, the Tn*916*-like region of ICE*Sp*23FST81 (from the PMEN1 reference strain, dated 1984 [13]). Both of these newly identified ICEs contained the tetracycline resistance determinant, *tetM*, but did not contain any other resistance determinant genes. Identification of such elements among

pneumococci isolated in the late 1960s confirms their existence from the first decade within which pneumococcal tetracycline-resistance was reported [24].

The Tn*916*-like element dated 1968 (identified within the genome of a CC113 isolate) was inserted within a phage showing similarity to regions of the *Streptococcal* phage 040922. This phage was not previously associated with any Tn*916*-like ICE [12]. Bacteriophage are known to mediate horizontal gene transfer between bacteria through the process of transduction [4], although it is not clear to what extent this process occurs amongst pneumococci, or whether the 040922 phage has retained the necessary genes to allow this. Pneumococcal Tn*916* family ICEs have not generally been associated with phage [10, 12, 13, 16-19, 30, 38, 40, 42, 47] and no such other phage-associated representatives are present in the Genbank database. Thus, it seems unlikely that phage mediated transfer has played an important role in the dissemination of these elements.

Aside from the *tetM* genes contained within the two Tn*916*-like elements described above, no other *tetM, ermB, mefA/E* or *cat* resistance determinants were identified among 38 pneumococci isolated prior to 1980. Additionally, there was no evidence of any other ICEs within the genomes of these isolates. However, four additional Tn*916* family ICEs were identified among pneumococci dated from 1980 onwards, and each of these was unique. One, from the genome of a CC124 isolate, was structurally highly similar to ICE*Sp*23FST81, with the addition of a MEGA element within the Tn*916*-like region. Two further ICEs, one from the genome of a CC15 isolate and one from the genome of a CC66 isolate, also showed similarity to regions of ICE*Sp*23FST81; however,

compared to the PMEN1 reference ICE, both of those described here were truncated at the 5' end and did not include the lantibiotic synthesis/export gene cluster. Similarly truncated ICEs have also been described previously for members of the PMEN1 CC (CC81) [12].

The final ICE described in this study (from a CC66 isolate) was a composite element comprising a Tn*916*-like element in association with a region of similarity to ICE*Se*2, described for *S. equi* strain 4047 [29]. No previously published report describing such an ICE has been found. The region of similarity to ICE*Se*2 contained putative conjugation-associated genes but did not include the site-specific integration genes or the ICE*Se*2 equibactin locus, which is associated with iron sequestration. Interestingly, int_{Tn916} and the Tn*916 xis* genes were also absent from this composite element, meaning that it would be unable to mediate its own integration into a new host genome.

Fourteen isolates dated from 1980 onwards harboured a MEGA element which was not associated with a Tn*916*-like element. These isolates represented CCs 15 and 156/162. Thirteen of the elements shared a high degree of sequence identity and were all inserted at the same position of the pneumococcal genome. The fourteenth MEGA element contained an additional 92 bp between the *mefE* and adjacent *mel* gene, and had an alternate genome integration site. All of these findings are consistent with those of previous authors [16, 17, 27].

Three isolates dated from 1980 onwards harboured *mefA* containing Tn*1207.1* elements. These isolates represented CC15 and serotype 14, a combination previously associated with possession of *mefA* [16, 54]. All of these Tn*1207.1* elements were integrated into the pneumococcal genome at the same site as those described previously [17, 27].

Among isolates representing CCs 15, 66 and 156/162, there was evidence of multiple, independent acquisition of ICEs or MEGA elements by the identification of nucleotide sequence, structure and/or integration site variations. Previous authors have also demonstrated multiple, independent acquisition of such elements by isolates representing CC15 [42, 54], CC81 [12], CC90 [54], CC242 [54] and CC236/271. Furthermore, intra-clonal variation of *tetM* nucleotide sequence, indicating multiple independent acquisition at least of this gene, has been demonstrated for CC18 [21, 39].

There was variation in the prevalences of the *ermB, mefA/E, tetM* and *cat* resistance determinants among different pneumococcal CCs in this study. No CC191 or CC218 representatives were positive for any of these genes. One isolate from each of CC113 and CC124 were positive for ≥1 of these genes. Two CC66 isolates were positive for the *tetM* gene. In contrast, 5 of 10 CC156/162 isolates were positive for *mefE* and 15 of 16 CC15 isolates were positive for ≥1 resistance determinant gene. The historical isolate collection does not represent a sample suitable for statistical analysis; inclusion of historical isolates was dependent on their availability rather than by systematic sampling. It is therefore possible that the inter-CC differences in resistant determinant

prevalence reported here do not reflect true differences across the wider pneumococcal population. However, support for these findings is indicated by inter-CC antimicrobial susceptibility variation among pneumococcal isolates for which data are deposited in the international MLST database (www.mlst.net).

The reasons for the putative differences described above are unclear at this time. It is likely that the prevalence of different genotypes within the population will play an important role. Isolates representing genotypes which are commonly carried in the nasopharynx presumably have greater opportunity to acquire resistance determinants from co-colonising bacteria than do isolates representing genotypes which are less commonly carried. Such an effect, however, cannot explain the reasons that some internationally prevalent genotypes are readily associated with resistance, whilst others are not (e.g. ST156 and ST191, respectively [39a].

Perhaps then certain features associated with some genotypes provide a barrier to acquisition and/or integration of ICEs or other resistance determinant-containing elements. This theory is consistent with the evidence that successful conjugal transfer of ICEs between or to pneumococci was strain-dependent (although an association with CC was not tested) [9, 47]. Such features, should they exist, could include clonal-specific variants of proteins shown to interact with internalised DNA, such as the pneumococcal SsbB protein which was recently shown to antagonise plasmid transfer [2]. Given that transfer of ICEs by conjugation may be a similar process to that of plasmid transfer [53a], it is conceivable that the SsbB protein may have an effect.

Alternatively, the primary mechanism of transfer of resistance-determinant containing elements among pneumococci could be via the transformation process. As such, clonal-differences in the prevalence of these elements could result from differences in transformation ability. Indeed, clonal-differences in transformation-associated recombination rates have previously been demonstrated among pneumococci, and did seem to correlate with differences in antimicrobial susceptibility [28].

It is clear that any differences in the propensity of clones to acquire ICEs and other resistance-determinant elements warrant further study. The first step in which should be to confirm clonal differences in the possession of such elements by the study of a larger and suitably sampled isolate collection. The phenotypic affects associated with possession of these genetic elements have an important impact on the ability to treat pneumococcal disease. In this context, understanding the processes driving and/or inhibiting the spread of such elements is of the utmost importance.

6.6 References

1. **Ambrose, K., and D. S. Stephens.** 2004. Macrolide, quinolone, and other non-β-lactam antibiotic resistance in *Streptococcus pneumoniae*., *In* E. I. Tuomanen (ed.), The Pneumococcus. ASM Press, Washington D. C.
2. **Attaiech, L., A. Olivier, I. Mortier-Barrière, A. L. Soulet, C. Granadel, B. Martin, P. Polard, and C. J. P.** 2011. Role of the single-stranded DNA-binding protein SsbB in pneumococcal transformation: maintenance of a reservoir for genetic plasticity. PLoS Genet. **7:**e1002156.
3. **Ayoubi, P., A. O. Kilic, and M. N. Vijayakumar.** 1991. Tn5253, the pneumococcal omega (cat tet) BM6001 element, is a composite structure of two conjugative transposons, Tn5251 and Tn5252. J Bacteriol. **173:**1617-1622.
4. **Brüssow, H., C. Canchaya, and W.-D. Hardt.** 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. Microbiol Mol Biol Rev. **68:**560-602.
6. **Burdett, V.** 1996. *Tet(M)*-promoted release of tetracycline from ribosomes is GTP dependent. J Bacteriol. **178:**3246-3251.
7. **Calatayud, L., C. Ardanuy, E. Cercenado, A. Fenoll, E. Bouza, R. Pallares, R. Martin, and J. Liñares.** 2007. Serotypes, clones, and mechanisms of resistance of erythromycin-resistant

Streptococcus pneumoniae isolates collected in Spain. Antimicrob Agents Chemother. **51:**3240-3246.

8. **Carver, T. J., K. M. Rutherford, M. Berriman, R. MA, B. G. Barrell, and J. Parkhill.** 2005. ACT: the Artemis comparison tool. Bioinformatics **21:**3422-3423.
9. **Celli, J., C. Poyart, and P. Trieu-Cuot.** 1997. Use of an excision reporter plasmid to study the intracellular mobility of the conjugative transposon Tn*916* in gram-positive bacteria. Microbiology. **143:**1253-1261.
10. **Cochetti, I., E. Tili, M. Mingoia, P. E. Varaldo, and M. P. Montanari.** 2008. *erm*(B)-carrying elements in tetracycline-resistant pneumococci and correspondence between Tn*1545* and Tn*6003.* Antimicrob Agents Chemother. **52:**1285-1290.
11. **Cooksey, R. C., J. M. Swenson, N. C. Clark, and C. Thornsberry.** 1989. DNA hybridization studies of a nucleotide sequence homologous to transposon Tn*1545* in the "Minnesota" strain of multiresistant *Streptococcus pneumoniae* isolated in 1977. Diagn Microbiol Infect Dis. **12:**13-16.
12. **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, B. Pichon, S. Baker, C. M. Parry, L. M. Lambertsen, D. Shahinas, D. R. Pillai, T. J. Mitchell, G. Dougan, A. Tomasz, K. P. Klugman, J. Parkhill, W. P. Hanage, and S. D. Bentley.** 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331:**430-434.
13. **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, S. D. Bentley, and T. J. Mitchell.** 2009. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae*^{Spain23F} ST81. J Bacteriol. **191:**1480-1489.
14. **Cybulska, J., J. Jeljaszewicz, E. Lund, and A. Munksgaard.** 1970. Prevalence of types of *Diplococcus pneumoniae* and their susceptibility to 30 antibiotics. Chemotherapy. **15:**304-316.
15. **Dang-Van, A., G. Tiraby, J. F. Acar, W. V. Shaw, and D. H. Bouanchaud.** 1978. Chloramphenicol resistance in *Streptococcus pneumoniae*: enzymatic acetylation and possible plasmid linkage. Antimicrob Agents Chemother. **13:**577-583.
16. **Del Grosso, M., R. Camilli, F. Iannelli, G. Pozzi, and A. Pantosti.** 2006. The *mef*(E)-carrying genetic element (mega) of *Streptococcus pneumoniae*: Insertion sites and association with other genetic elements. Antimicrob Agents Chemother. **50:**3361-3366.
17. **Del Grosso, M., F. Iannelli, C. Messina, M. Santagati, N. Petrosillo, S. Stefani, G. Pozzi, and A. Pantosti.** 2002. Macrolide efflux genes *mef*(A) and *mef*(E) are carried by different genetic elements in *Streptococcus pneumoniae.* J Clin Microbiol. **40:**774-778.
18. **Del Grosso, M., A. Scotto d'Abusco, I. F, P. G, and A. Pantosti.** 2004. Tn*2009*, a Tn*916*-like element containing *mef*(E) in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **48:**2037-2042.
19. **Ding, F., P. Tang, M.-H. Hsu, P. Cui, S. Hu, J. Yu, and C.-H. Chiu.** 2009. Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. BMC Genet. **10:**158.
20. **Dixon, J. M. S.** 1967. Pneumococcus resistant to eryhthromycin and lincomycin. The Lancet.**:**573.
21. **Doherty, N., K. Trzcinski, P. Pickerill, P. Zawadzki, and C. G. Dowson.** 2000. Genetic diversity of the *tet(M)* Gene in tetracycline-resistant clonal lineages of *Streptococus pneumoniae*. Antimicrob Agents Chemother. **44:**2979-2984.
22. **Edgar, R. C.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32:**1792-1797.
23. **European Centre for Disease Prevention and Control (ECDC)** 2012. European Antimicrobial Resistance Surveillance Network (EARS-Net) database. European Centre for Disease Prevention and Control (ECDC). [Online.]
24. **Evans, W., and D. Hansman.** 1963. Tetracycline-resistant pneumococcus. The Lancet.**:**451.
25. **Farrell, D. J., K. P. Klugman, and M. Pichichero.** 2007. Increased antimicrobial resistance among nonvaccine serotypes of *Streptococcus pneumoniae* in the pediatric population after the introduction of 7-valent pneumococcal vaccine in the United States. Pediatr Infect Dis J. **26:**123-128.
26. **Franke, A. E., and D. B. Clewell.** 1981. Evidence for a chromosome-borne resistance transposon (Tn916) in *Streptococcus faecalis* that is capable of "conjugal" transfer in the absence of a conjugative plasmid. J Bacteriol. **145:**494-502.

27. **Gay, K., and D. S. Stephens.** 2001. Structure and dissemination of a chromosomal insertion element encoding macrolide efflux in *Streptococcus pneumoniae.* J Infect Dis. **184:**56-65.
28. **Hanage, W. P., C. Fraser, J. Tang, T. R. Connor, and J. Corander.** 2009. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. Science **324:**1454-1457.
29. **Heather, Z., M. T. G. Holden, K. F. Steward, J. Parkhill, L. Song, G. L. Challis, C. Robinson, N. Davis-Poynter, and A. S. Waller.** 2008. A novel streptococcal integrative conjugative element involved in iron acquisition. Mol Microbiol. **70:**1274-1292.
30. **Henderson-Begg, S. K., A. P. Roberts, and L. M. C. Hall.** 2009. Diversity of putative Tn*5253*-like elements in *Streptococcus pneumoniae.* Int J Antimicrob Agents. **33:**364-367.
31. **Hoskins, J., W. E. Alborn, J. Arnold, L. C. Blaszczak, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D. J. Fu, W. Fuller, C. Geringer, R. Gilmour, J. S. Glass, H. Khoja, A. R. Kraft, R. E. Lagace, D. J. LeBlanc, L. N. Lee, E. J. Lefkowitz, J. Lu, P. Matsushima, S. M. McAhren, M. McHenney, K. McLeaster, C. W. Mundy, T. I. Nicas, F. H. Norris, M. O'Gara, R. B. Peery, G. T. Robertson, P. Rockey, P. M. Sun, M. E. Winkler, Y. Yang, M. Young-Bellido, G. S. Zhao, C. A. Zook, R. H. Baltz, S. R. Jaskunas, P. R. Rosteck, P. L. Skatrud, and J. I. Glass.** 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J Bacteriol. **183:**5709-5717.
32. **Hyatt, D., G. L. Chen, P. F. LoCascio, M. L. Land, F. W. Larimer, and L. J. Hauser.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. **11:**119.
33. **Jacobs, M. R., D. Felmingham, P. C. Appelbaum, R. N. Gruneberg, and The Alexander Project Group.** 2003. The Alexander Project 1998-2000: susceptibility of pathogens isolated from community-acquired respiratory tract infection to commonly used antimicrobial agents. J Antimicrob Chemother. **52:**229-246.
34. **Jolley, K. A., and M. C. Maiden.** 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. **10:**595.
35. **Jones, R. N., and H. S. Sader.** 2010. Declining antimicrobial susceptibility of *Streptococcus pneumoniae* in the United States: report from the SENTRY antimicrobial surveillance program (1998-2009). Diagn Microbiol Infect Dis. **68:**334-336.
36. **Kim, S. H., J. Song, D. R. Chung, V. Thamlikitkul, Y. Yang, H. Wang, M. Lu, T. M. So, P. R. Hsueh, R. M. Yasin, C. C. Carlos, H. V. Pham, M. K. Lalitha, N. Shimono, J. Perera, A. M. Shibl, J. Y. Baek, C. I. Kang, K. S. Ko, and K. R. Peck.** 2012. Changing trends in antimicrobial resistance and serotypes of *Streptococcus pneumoniae* isolates in Asian countries: an Asian Network for Surveillance of Resistant Pathogens (ANSORP) study. Antimicrob Agents Chemother. **56:**1418-1426.
37. **Lanie, J. A., W. L. Ng, K. M. Kazmierczak, T. M. Anrzejewski, T. M. Davidson, K. J. Wayne, H. Tettelin, J. I. Glass, and M. E. Winkler.** 2007. Genome sequence of Avery's virulent serotype 2 strain D39 of *Steptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. J Bacteriol. **189:**38-51.
38. **Li, Y., H. Tomita, Y. Lv, J. Liu, F. Xue, B. Zheng, and Y. Ike.** 2011. Molecular characterization of *erm*(B)- and *mef*(E)-mediated erythromycin-resistant *Streptococcus pneumoniae* in China and complete DNA sequence of Tn*2010*. J Appl Microbiol. **110:**254-265.
39. **Marimón, J. M., E. Pérez-Trallero, M. Ercibengoa, A. Gonzalez, A. Fenoll, and o. b. o. t. S. P. I. S. N. (G03/103).** 2006. Molecular epidemiology and variants of the multidrug-resistant *Streptococcus pneumoniae* Spain^{14}-5 international clone among Spanish clinical isolates. J Antimicrob Chemother. **57:**654-660.
39a. **McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefévre, A. Tomasz, and K. P. Klugman.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. J Clin Microbiol. **39:**2565-2571.
40. **Mingoia, M., E. Tili, E. Manso, P. E. Varaldo, and M. P. Montanari.** 2011. Heterogeneity of Tn*5253*-like composite elements in clinical *Streptococcus pneumoniae* isolates. Antimicrob Agents Chemother. **55:**1453-1459.
41. **Munoz-Najar, U., and M. N. Vijayakumar.** 1999. An operon that confers UV resistance by evoking the SOS mutagenic response in streptococcal conjugative transposon Tn*5252.* J Bacteriol. **181:**2782-2788.

42. **Quintero, B., M. Araque, C. van der Gaast-de Jongh, and P. W. Hermans.** 2011. Genetic diversity of Tn*916*-related transposons among drug-resistant *Streptococcus pneumoniae* isolates colonizing healthy children in Venezuela. Antimicrob Agents Chemother. **55:**4930-4932.
43. **Rantala, M., S. Huikko, P. Huovinen, and J. Jalava.** 2005. Prevalence and molecular genetics of macrolide resistance among *Streptococcus pneumoniae* isolates collected in Finland in 2002. Antimicrob Agents Chemother. **49:**4180-4184.
44. **Roberts, A. P., and P. Mullany.** 2009. A modular master on the move: the Tn*916* family of mobile genetic elements. Trends Microbiol. **17:**251-258.
45. **Roberts, A. P., and P. Mullany.** 2011. Tn*916*-like genetic elements: a diverse group of modular mobile elements conferring antibiotic resistance. FEMS Microbiol. Rev. **35:**856-871.
46. **Santagati, M., F. Iannelli, M. R. Oggioni, S. Stefani, and G. Pozzi.** 2000. Characterization of a genetic element carrying the macrolide efflux gene *mef*(A) in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **44:**2585-2587.
47. **Santoro, F., M. R. Oggioni, G. Pozzi, and I. F.** 2010. Nucleotide sequence and functional analysis of the *tet*(M)-carrying conjugative transposon Tn*5251* of *Streptococcus pneumoniae.* FEMS Microbiol Lett **308:**150-158.
48. **Shen, X., H. Yang, S. Yu, K. Yao, Y. Wang, L. Yuan, and Y. Yang.** 2008. Macrolide-resistance mechanisms in *Streptococcus pneumoniae* isolates from Chinese children in association with genes of *tetM* and integrase of conjugative transposons 1545. Microbial Drug Res. **14:**155-161.
49. **Shoemaker, N. B., M. D. Smith, and W. R. Guild.** 1979. Organization and transfer of heterologous chloramphenicol and tetracycline resistance genes in pneumococcus. J Bacteriol. **139:**432-441.
50. **Sullivan, M. J., N. K. Petty, and S. A. Beatson.** 2011. Easyfig: a genome comparison visualizer. Bioinformatics. **27:**1009-1010.
51. **Tait-Kamradt, A., J. Clancy, M. Cronan, F. Dib-Hajj, L. Wondrack, W. Yuan, and J. Sutcliffe.** 1997. *mefE* is necessary for the erythromycin-resistant M phenotype in *Streptococcus pneumoniae.* Antimicrob Agents Chemother. **41:**2251-2255.
52. **Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar.** 2011. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. **28:**2731-2739.
53. **Weisblum, B.** 1995. Erythromycin resistance by ribosome modification. Antimicrob Agents Chemother. **39:**577-585.
53a. **Wozniak, R. A. F., and M. K. Waldor.** 2010. Integrative and conjugative elements: mosaic mobile genetic elements enabling dynamic lateral gene flow. Nat Rev Microbiol. **8:**552-563.
54. **Xu, X., L. Cai, M. Xiao, F. Kong, S. Oftadeh, F. Zhou, and G. L. Gilbert.** 2010. Distribution of serotypes, genotypes, and resistance determinants among macrolide-resistant *Streptococcus pneumoniae* isolates. Antimicrob Agents Chemother. **54:**1152-1159.
55. **Yao, J. D. C., and R. C. Moellering Jr.** 1995. Antimicrobial Agents, *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. H. Yolken (ed.), Manual of Clinical MIcrobiology, 6th ed. ASM Press, Washington D. C.
56. **Zerbino, D. R., and E. Birney.** 2008. Velvet: Algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res. **18:**821-829.

7. CONCLUSIONS AND FUTURE WORK

The pneumococcus is a globally important cause of disease and is associated with substantial health and economic costs [17, 25]. Clinical interventions such as vaccine and antibiotic use have been highly successful in reducing the burden of disease and its associated morbidity/mortality [21, 38, 39]; however, in many cases pneumococci have been able to evade the effects of vaccines [1, 3, 26] and/or develop resistance to antibiotics [13, 23]. The evolution and spread of vaccine-escape or drug-resistance traits amongst pneumococci threatens our ability to prevent and treat pneumococcal disease. In order to understand how this bacterium will continue to respond and how it may respond to the selective pressures of new treatments or vaccines, we must first gain a better understanding of its evolutionary history. The data presented in this thesis help to achieve this goal. The principle findings are summarised below:

Chapter 3. Pneumococcal history since 1937: multilocus sequence types and penicillin-binding protein genes.

1) Several of the now internationally recognised PMEN clones have been circulating within pneumococcal populations for decades. Many non-PMEN associated genotypes which were also identified among older pneumococci have not been found among modern isolate collections, suggesting that they are now rare or extinct.
2) The evolution of pneumococcal penicillin resistance is complex. *pbp* gene alterations are important and statistically significant factors but do not fully explain inter-isolate differences in penicillin susceptibility.

3) *pbp* gene transfer between penicillin-susceptible pneumococci was occurring many years before the first PNSP were identified, but there was no evidence of 'divergent' *pbp* transfer at this time, despite the fact that such *pbps* were likely present among viridans streptococci. This suggests that there may have been some sort of historical selective or mechanistic barrier which once prevented pneumococcal uptake of 'divergent' *pbp*s.

Chapter 4. The multidrug-resistant PMEN1 pneumococcus: A paradigm for genetic success.

1) Analyses of whole-genome sequence data showed that one of the earliest identified PNSP (from Australia in 1967) represented the ancestor of the PMEN1 clone.
2) The PMEN1 clone has been highly promiscuous with its DNA, donating its *pbp* genes to numerous unrelated pneumococci. The extensive donation of DNA from the PMEN1 clone to other pneumococci has likely contributed to the spread of penicillin resistance within this species.
3) CGSP14 and the PMEN3 reference strain each acquired *pbp2x, pbp1a* and *pbp2b* gene sequences from PMEN1 representative(s), along with many other genes including those associated with virulence or resistance to non-β-lactams. In total the regions acquired from PMEN1 were equivalent to ~9.5% and ~5.3% of the genome, respectively. The PMEN1-like sequences were acquired in multiple fragments and appeared to be randomly distributed around the genome.

Chapter 5. Pneumococcal capsular switching: An historical perspective.

1) At least 36 independent capsular switching events were detected among 426 isolates in the historical pneumococcal collection. Capsular switch representatives were isolated throughout the last 7 decades, suggesting that capsular switching may have been a regular occurrence throughout this time.
2) Recombination at the *cps* locus was not associated with any temporal patterns of recombination breakpoints. Recombination of large DNA fragments (>30 kb), sometimes including the capsular locus *and pbp2x* ± *pbp1a*, was occurring prior to both vaccine introduction and widespread antibiotic use.

Chapter 6. Exploring historical non-β-lactam resistance determinants and resistant-determinant element variation among clonally related pneumococci.

1) The *tetM* and int_{Tn916} genes were identified among the genomes of two pneumococci isolated in 1967 and 1968, respectively. In both cases, these genes were located on a Tn*916*-like ICE which was highly similar to the Tn*916*-like region of the PMEN1 reference genome. No other *tetM*, *int, ermB, mefA/E,* or *cat* genes were identified among the genomes of 38 pneumococci isolated prior to 1980.
2) Four Tn*916* family ICEs were newly identified among the genomes of pneumococci dated from 1980 onwards. One of these ICEs was a composite element comprising a Tn*916*-like element and a region of

similarity to part of *S. equi* ICE*Se*2. Three Tn*1207.1* and 14 MEGA elements were identified among pneumococci dated from 1980 onwards.

3) There was evidence of nucleotide sequence, structural and integration site diversity of Tn*916*-like and MEGA elements among pneumococci representing the same CC. Such diversity suggested multiple, independent acquisition of resistance determinant-containing elements.

Since the advent of molecular typing techniques, it has become apparent that there are differences in the geographic distribution of pneumococcal clones, and that some are more widely distributed than others [2, 4, 8, 18, 22, 31, 33, 40]. An important finding described here, which could only have been elucidated through the study of an historical isolate collection, was that there also seem to be differences in the longevity of pneumococcal clones (Chapter 3), with some having persisted for at least 7 decades.

The reasons for the prolonged survival and transmission of some clones compared to that of others, which have only been identified once or on a small number of occasions, are uncertain. Antibiotic-resistance traits may play a role in successful survival and transmission among modern pneumococcal populations, but cannot explain the success of susceptible clones. Thus, unless clonal transmission and survival is a purely stochastic process (which seems unlikely), variation in other traits must be driving these epidemiological patterns.

Perhaps some pneumococci are better competitors than others? It is known that multiple strains of pneumococci can co-colonise the nasopharynx [32, 35], which

presumably results in competition for finite resources. Pneumococci produce antimicrobial peptides known as bacteriocins, which target non-related cells and may be important during intra-species competition [19]. Production and secretion of these peptides involves a two-component secretory system similar to that described for the fratricide pathway. Secretion of the BlpMN bacteriocin has been shown to account for the *in vitro* bacteriocidal activity of pneumococcal strains against other pneumococci, and to provide a competitive advantage in a mouse model of colonisation [10]. It is not yet known whether secretion of BlpMN is a clonal property, although the presence/absence of *blpMN* alleles did not seem to show any clonal associations ([11], see Appendix 3). An *in vitro* assay similar to that described by Dawid and colleagues [10] could be used to test for any associations between CC and bacteriocin secretion, and/or activity spectrum, which could affect competitive ability.

Aside from competitive ability, a second property which may contribute to the success of a clone is its ability to adapt to changing environments. One way to assess the potential adaptability of a clone could be to measure its rate of genomic evolution. The analysis by Croucher and colleagues [9] indicated a rapid rate of genomic evolution within the PMEN1 lineage, but it is not clear how this rate compares to that of other clones. Similar analyses could be completed for a range of clones in order to address this question. A problem with this approach could be that it may not be possible to collect a sufficient number of representatives of "less fit" clones, given that they are less common or now extinct (as is possible for some of the historical clones). A potential solution could be to group the "less fit" clones and calculate an average measure of genomic evolution to use as a reference value.

An important factor affecting the rate of genomic evolution is the rate of recombination. Differences in pneumococcal lineage-specific recombination rates have been shown previously by analysis of MLST data [15]. Acquisition of mobile genetic elements, potentially via transformation and recombination, may also vary between clones (Chapter 6).

For many years *S. pneumoniae* has been considered as highly recombinogenic [12] but only recently has the true scale of recombination within this species been recognised (here and in [3, 9, 14, 16]). The work described in this thesis has shown for the first time that exchange of large regions of DNA (>30 kb) between pneumococci has been occurring for decades. Additionally, exchange of the *cps* locus and *pbp* loci (which can be associated with vaccine escape and the development of penicillin resistance, respectively) was likely a common occurrence even before the introduction of vaccines and antibiotics, and highlights the intrinsic role that such recombination plays in pneumococcal biology.

Although it is clear that some regions of the pneumococcal genome are more likely to be affected by recombination than others [9], it is possible that recombination is generally a random process. Indeed, the multi-fragment genomic recombination described in Chapter 4 and by Golubchik and colleagues [14] did not appear to show any particular pattern (with the exception of consistent transfer of *resistance-conferring pbp* genes (Chapter 4) and the *cps* locus [14]). A random process of

recombination could explain the reasons for *pbp* exchange among susceptible pneumococci, and prior to the wide-spread use of antibiotics.

The general evolutionary purpose and maintenance of recombination among bacteria are the subject of much debate among micro- and evolutionary- biologists [20, 24, 36, 37]. Two of the theories which attempt to explain the evolution of recombination, here called the "Repair Hypothesis" and the "Advantageous Trait Hypothesis," are compatible with the theory of random recombination. The former hypothesis states that recombination mechanisms have evolved as a pathway for repair of damaged DNA. In this case, horizontally acquired, non-damaged, nucleotide sequences provide a template for the repair of homologous, damaged sequences, which may otherwise be detrimental to the survival fitness of the recipient cell [6]. Among pneumococci, support for this theory is the evidence that competence can be induced by the DNA damaging agent, mitomycin C [29], and by protein mis-folding during protein synthesis [34], the underlying causes of which could include mis-coding due to genetic damage.

The Advantageous Trait Hypothesis is the theory that recombination can enhance the rate at which bacterial populations adapt to their environment. For example, recombination can allow the generation of new, potentially advantageous, alleles at a faster rate than can mutation alone. Additionally, recombination can allow advantageous alleles/genes to spread horizontally through the population at a rate which is greater than multiple independent generation by mutation. Consequently, individuals can rapidly accumulate advantageous genes/alleles and greatly increase the probability of their survival. This theory is intuitive but difficult to test experimentally;

however, Perron and colleagues have demonstrated that, when exposed to antibiotic-selective pressures, *Acintobacter baylyi* populations develop *in vitro* multi-drug resistance more quickly through recombination than by mutation alone [27].

Regardless of the reasons for the evolution of recombination among pneumococci, its effects on the genomic evolution of this organism are extremely important, and play a role in determining the effective duration of pneumococcal vaccines and treatments. As a consequence it is essential that we better understand the mechanisms which drive this process and the factors which influence recombination rates.

It has been shown that the majority of pneumococci possess one of two *comC* variants [5, 7, 28, 30], corresponding to the production of two different CSP variants. At this time however, little or no information is known about the extent of variation within other competence- and/or recombination-associated genes. Given the recent increase in generation and availability of pneumococcal genomic data, it should be relatively easy to gain an overview of this variation (especially when using a platform such as that available in BIGSdb). Variable loci of interest could be selected for further study, perhaps employing conventional PCR and sequencing to explore the diversity among a suitably designed sample population. CC-specific allelic variant or diversity associations could also be tested. Two studies, [7] and [5], previously compared *comC* sequence and genotype data for 88 and 483 pneumococci, respectively. Interestingly, the former study found no association between genotype and CSP-variant, whilst the latter study did find an association. The reasons for these differences are not clear and warrant further investigation. Clonal differences in transformation rates could also be

measured *in vitro*, perhaps using a protocol similar to that described by Yano and colleagues [41]. Appropriate genetically-modified isolates could also be included to assess the effects of transformation/competence gene allelic variation on transformation rate.

As next generation sequencing technologies continue to advance, the quantity and quality of available pneumococcal genomic data will increase rapidly. This vast volume of data will allow further and more detailed studies of pneumococcal recombination, and other processes driving genomic evolution. A full understanding of these processes will greatly aid the development of optimum vaccine and treatment strategies in the fight against pneumococcal disease.

7.1 References

1. **Ansaldi, F., P. Canepa, D. de Florentiis, R. Bandettini, P. Durando, and G. Icardi.** 2011. Increasing incidence of *Streptococcus pneumoniae* serotype 19A and emergence of two vaccine escape recombinant ST695 strains in Liguria, Italy, 7 years after implementation of the 7-valent conjugated vaccine. Clin Vaccine Immunol. **18:**343-345.
2. **Beall, B., M. C. McEllistrem, R. E. Gertz, S. Wedel, D. J. Boxrud, A. L. Gonzalez, M.-J. Medina, R. Pai, T. A. Thompson, L. H. Harrison, L. McGee, C. G. Whitney, and Active Bacterial Core Surveillance Team.** 2006. Pre- and postvaccination clonal compositions of invasive pneumococcal serotypes for isolates collected in the United States in 1999, 2001, and 2002. J Clin Microbiol. **44:**999-1017.
3. **Brueggemann, A. B., R. Pai, D. Crook, and B. Beall.** 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog. **3:**e168.
4. **Brueggemann, A. B., and B. G. Spratt.** 2003. Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. J Clin Microbiol. **41:**4966-4970.
5. **Carrolo, M., F. R. Pinto, J. Melo-Cristino, and M. Ramirez.** 2009. Pherotypes are driving genetic differentiation within *Streptococcus pneumoniae.* BMC Microbiol. **9:**191-200.
6. **Claverys, J., M. Prudhomme, and B. Martin.** 2006. Induction of competence regulons as a general response to stress in gram-positive bacteria. Annu Rev Microbiol. **60:**451-475.
7. **Cornejo, O. E., L. McGee, and D. E. Rozen.** 2010. Polymorphic competence peptides do not restrict recombination in *Streptococcus pneumoniae.* Mol Biol Evol. **27:**694-702.
8. **Corso, A., E. P. Severina, V. F. Petruk, Y. R. Mauriz, and A. Tomasz.** 1998. Molecular characterisation of penicillin-resistant *Streptococcus pneumoniae* isolates causing respiratory disease in the United States. Microb Drug Resist. **4:**325-337.
9. **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, B. Pichon, S. Baker, C. M. Parry, L. M. Lambertsen, D. Shahinas, D. R. Pillai, T. J. Mitchell, G. Dougan, A. Tomasz, K. P. Klugman, J. Parkhill, W. P.**

Hanage, and S. D. Bentley. 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331:**430-434.

10. **Dawid, S., A. M. Roche, and J. N. Weiser.** 2007. The *blp* bacteriocins of *Streptococcus pneumoniae* mediate intraspecies competition both *in vitro* and *in vivo.* Infect. Immun. **75:**443-451.
11. **Ek, K. L., K. L. Wyres, K. G. Kristinsson, and A. B. Brueggemann.** 2010. Prevalence and sequence diversity of bacteriocin genes *blpMN* in *Streptococcus pneumoniae*., 7th International Symposium on Pneumococci and Pneumococcal Diseases, Tel Aviv, Israel.
12. **Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt.** 2000. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. Genetics. **154:**1439-1450.
13. **Felmingham, D., A. R. White, M. R. Jacobs, P. C. Appelbaum, J. Poupard, L. A. Miller, and R. N. Gruneberg.** 2005. The Alexander Project: the benefits from a decade of surveillance. J Antimicrob Chemother. **56:**ii3-ii21.
14. **Golubchik, T., A. B. Brueggemann, T. Street, R. E. Gertz Jr, C. C. A. Spencer, T. Ho, E. Giannoulatou, R. Link-Gelles, R. M. Harding, B. Beall, T. E. A. Peto, M. R. Moore, P. Donnelly, D. W. Crook, and R. Bowden.** 2012. Pneumococcal genome sequencing tracks a vaccine escape variant formed through a multi-fragment recombination event. Nat Genet. **44:**352-355.
15. **Hanage, W. P., C. Fraser, J. Tang, T. R. Connor, and J. Corander.** 2009. Hyper-recombination, diversity, and antibiotic resistance in pneumococcus. Science **324:**1454-1457.
16. **Hiller, N. L., A. Ahmed, E. Powell, D. P. Martin, R. Eutsey, J. Earl, B. Janto, R. J. Boissy, J. Hogg, K. Barbadora, R. Sampath, S. Lonergan, J. C. Post, F. Z. Hu, and G. D. Ehrlich.** 2010. Generation of genetic diversity among *Streptococcus pneumoniae* strains via horizontal gene transfer during a chronic polyclonal paediatric infection. PLoS Pathog. **6:**e1001108.
17. **Huang, S. S., K. M. Johnson, G. T. Ray, P. Wroe, T. A. Lieu, M. R. Moore, E. R. Zell, J. A. Linder, C. G. Grijalva, J. P. Metlay, and J. A. Finkelstein.** 2011. Healthcare utilization and cost of pneumococcal disease in the United States. Vaccine **29:**3398-3412.
18. **Huebner, R. E., A. D. Wasas, and K. P. Klugman.** 2000. Trends in antimicrobial resistance and serotype distribution of blood and cerebrospinal fluid isolates of *Streptococcus pneumoniae* in South Africa, 1991-1998. Int J Infect Dis. **4:**214-218.
19. **Johnsborg, O., T. Blomqvist, M. Kilian, and H. L. Sigve.** 2007. Biologically active peptides in Streptococci., *In* R. Hakenbeck and S. Chhatwal (ed.), Molecular Biology of Streptococci. Horizon Scientific Press, Norwich, UK.
20. **Johnsborg, O., V. Eldholm, and L. S. Havärstein.** 2007. Natural genetic transformation: prevalence, mechanisms and function. Res Microbiol. **158:**767-778.
21. **Kellner, J. D., O. G. Vanderkooi, J. MacDonald, D. L. Church, G. J. Tyrrell, and D. W. Scheifele.** 2009. Changing epidemiology of invasive pneumococcal disease in Canada, 1998-2007: update from the Calgary-area *Streptococcus pneumoniae* research (CASPER) study. Clin Infect Dis. **49:**205-212.
22. **Lefèvre, J. C., M. A. Bertrand, and G. Faucon.** 1995. Molecular analysis by pulsed field gel electrophoresis of penicillin-resistant *Streptococcus pneumoniae* from Toulouse, France. Eur J Clin Microbiol Infect Dis. **14:**491-497.
23. **McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefévre, A. Tomasz, and K. P. Klugman.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. J Clin Microbiol. **39:**2565-2571.
24. **Michod, R. E., H. Bernstein, and A. M. Nedelcu.** 2008. Adaptive value of sex in microbial pathogens. Infect Genet Evol. **8:**267-285.
25. **O'Brien, K. L., L. J. Wolfson, J. P. Watt, E. Henkle, M. Deloria-Knoll, N. McCall, E. Lee, K. Mullholland, O. S. Levine, and T. Cherian.** 2009. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet **374:**893-902.
26. **Pelton, S. I., H. Huot, J. A. Finkelstein, C. J. Bishop, K. K. Hsu, J. Kellenberg, S. S. Huang, R. Goldstein, and W. P. Hanage.** 2007. Emergence of 19A as virulent and multidrug resistant pneumococcus in Massachusetts following universal immunization of infants with pneumococcal conjugate vaccine. Pediatr Infect Dis J. **26:**468-472.

27. **Perron, G. G., A. E. G. Lee, Y. Wang, W. E. Huang, and T. G. Barraclough.** 2012. Bacterial recombination promotes the evolution of multi-drug-resistance in functionally diverse populations. Proc Biol Sci. **279:**1477-1484.
28. **Pozzi, G., L. Masala, F. Iannelli, R. Manganelli, L. S. Hävarstein, L. Piccoli, D. Simon, and D. A. Morrison.** 1996. Competence for genetic transformation in encapsulated strains of *Streptococcus pneumoniae*: two allelic variants of the peptide pheromone. J Bacteriol. **178:**6087-6090.
29. **Prudhomme, M., L. Attaiech, G. Sanchez, B. Martin, and J.-P. Claverys.** 2006. Antibiotic stress induces genetic transformability in the human pathogen *Streptococcus pneumoniae.* Science **313:**89-92.
30. **Ramirez, M., D. A. Morrison, and A. Tomasz.** 1997. Ubiquitous distribution of the competence related genes *comA* and *comC* among isolates of *Streptococcus pneumoniae.* Microb Drug Resist. **3:**39-52.
31. **Reichmann, P., E. Varon, E. Günther, R. R. Reinert, R. Lüttiken, A. Marton, P. Geslin, J. Wagner, and R. Hakenbeck.** 1995. Penicillin-resistant *Streptococcus pneumoniae* in Germany: genetic relationship to clones from other European countries. J Med Microbiol. **43:**377-385.
32. **Rivera-Olivero, I. A., M. Blommaart, D. Bogaert, P. W. M. Hermans, and J. H. de Waard.** 2009. Multiplex PCR reveals a high rate of nasopharyngeal pneumococcal 7-valent conjugate vaccine serotypes co-colonizing indigenous Warao children in Venezuela. J Med Microbiol. **58:**584-587.
33. **Shi, Z.-Y., M. C. Enright, P. Wilkinson, D. Griffiths, and B. G. Spratt.** 1998. Identification of three major clones of multiply antibiotic-resistant *Streptococcus pneumoniae* in Taiwanese hospitals by multilocus sequence typing. J Clin Microbiol. **36:**3514-3519.
34. **Stevens, K. E., D. Chang, E. E. Zwack, and M. E. Sebert.** 2011. Competence in *Streptococcus pneumoniae* is regulated by the rate of ribosomal decoding errors. Mbio. **2:**e00071-00011.
35. **Vestrheim, D. F., P. Gaustad, I. S. Aaberge, and D. A. Caugant.** 2011. Pherotypes of pneumococcal strains co-existing in healthy children. Infect Genet Evol. **11:**1703-1708.
36. **Vogan, A. A., and P. G. Higgs.** 2011. The advantages and disadvantages of horizontal gene transfer and the emergence of the first species. Biol Direct **6:**1.
37. **Vos, M.** 2009. Why do bacteria engage in homologous recombination? Trends Microbiol. **17:**226-232.
38. **Weinstein, M. P., K. P. Klugman, and R. N. Jones.** 2009. Rationale for revised penicillin susceptibility breakpoints versus *Streptococcus pneumoniae*: Coping with antimicrobial susceptibility in an era of resistance. Clin Infect Dis. **48:**1596-1600.
39. **Whitney, C. G., M. M. Farley, J. Hadler, L. H. Harrison, N. M. Bennett, R. Lynfield, A. Reingold, P. R. Cieslak, T. Pilishvili, D. Jackson, R. R. Facklam, J. H. Jorgensen, and A. Schuchat.** 2003. Decline in invasive pneumococcal disease after the introduction of protein-polysaccharide conjugate vaccine. N Engl J Med. **348:**1737-1746.
40. **Wolf, B., L. C. Rey, S. Brisse, L. B. Moreira, D. Milatovic, A. Fleer, J. J. Roord, and J. Verhoef.** 2000. Molecular epidemiology of penicillin-resistant *Streptococcus pneumoniae* colonizing children with community-acquired pneumonia and children attending day-care centres in Fortaleza, Brazil. J Antimicrob Chemother. **46:**757-765.
41. **Yano, M., S. Gohil, J. R. Coleman, C. Manix, and L. Pirofski.** 2011. Antibodies to *Streptococcus pneumoniae* capsular polysaccharide enhance pneumococcal quorum sensing. MBio. **2:**e00176-00111.

APPENDIX 1.1 Demographic information for the historical isolate collection plus 16 isolates for which genomic sequences were retrieved from Genbank and used throughout this thesis.

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
1	1/2		1	1965	USA	LML		
2	1/3		1	1956	Denmark	LML		
3	1/4	519/43	1	1943	Denmark	LML	[2]	
4	1/5		1	1943	Denmark	LML		
5	1/6		1	1944	Denmark	LML		
6	1/7		1	1991	Denmark	LML		
7	2/2 *	pn2L	2	1956	USA	LML	[2]	ERR025820‡
8	2/3 *		2	1943	Denmark	LML		ERR026715‡
9	2/4		2	1959	Denmark	LML		
10	3/2		3	1973	USA	LML		
11	3/3		3	1971	USA	LML		
12	3/4		3	1961	Denmark	LML		
13	3/5		3	1962	Denmark	LML		
14	3/6	524/62	3	1962	Denmark	LML	[2]	
15	3/7		3	1961	Denmark	LML		
16	3/8		3	1961	Denmark	LML		
17	4/2		4	1952	USA	LML		
18	4/3		4	1962	Denmark	LML		
19	4/4	600/62	4	1962	Denmark	LML	[2]	
20	4/5		4	1962	Denmark	LML		
21	5/1	Ambrose	5	1976	USA	LML	[2]	

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
22	5/2		5	1961	USA	LML		
23	5/3		5	1987	Denmark	LML		
24	5/4		5	1987	Singapore	LML		
25	6A/2	34351	6A	1952	USA	LML	[2]	
26	6B/3	2616/39	6B	1939	Denmark	LML	[2]	
27	7A/2 *	2040/37	7A	1937	Denmark	LML	[2]	ERR025824‡
28	7A/3		7A	1968	USA	LML		
29	7B/2 *	Johnson	7B	1952	USA	LML	[2]	ERR025825‡
30	7B/3		7B	1996	France	LML		
31	7C/2		7C	1939	Denmark	LML		
32	7C/3	Sutcliff	7C	1971	USA	LML	[2]	
33	7F/2 *		7F	1952	USA	LML		ERR026172‡
34	7F/3 *		7F	1962	Denmark	LML		ERR026176‡
35	7F/4 *		7F	1986	Scotland	LML		ERR026177‡
36	7F/5	554/62	7F	1962	Denmark	LML	[2]	
37	7F/6		7F	1996	Sweden	LML		
38	8/2		8	1952	USA	LML		
39	8/3		8	1962	Denmark	LML		
40	8/4	573/62	8	1962	Denmark	LML	[2]	
41	8/5		8	1969	Belgium	LML		
42	9A/1 *	Wilder	9A	1962	USA	LML	[2]	ERR025826‡
43	9L/2 *		9L	1952	USA	LML		ERR026178‡
44	9L/3		9L	1941	Denmark	LML		
45	9L/4	T9233/128/68	9L	1968	Denmark	LML	[2]	
46	9N/2 *		9N	1952	USA	LML		ERR026179‡
47	9N/3		9N	1938	Denmark	LML		

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
48	9N/4	533/62	9N	1962	Denmark	LML	[2]	
49	9N/5		9N	1962	Denmark	LML		
50	9N/6 *		9N	1960	Denmark	LML		ERR026180‡
51	9V/2		9V	1938	Denmark	LML		
52	9V/4 *	980/68	9V	1968	Denmark	LML	[2]	ERR025827‡
53	9V/5 *		9V	1991	Denmark	LML		ERR026181‡
54	9V/6 *		9V	1994	Denmark	LML		ERR026182‡
55	10A/1	10061/38	10A	1938	Denmark	LML	[2]	
56	10B/2	423/82	10B	1982	Denmark	LML	[2]	
57	10C/2	Gro Norge	10C	1983	Norway	LML	[2]	
58	10F/2 *	34355	10F	1956	USA	LML	[2]	ERR025828‡
59	11A/2	1813/39	11A	1939	Denmark	LML	[2]	
60	11B/2	8087/40	11B	1940	Denmark	LML	[2]	
61	11C/1 *	Eddy nr. 53	11C	1957	USA	LML	[2]	ERR025829‡
62	11D/1	70/86	11D	1986	Denmark	LML	[2]	
63	11F/2	34356	11F	1952	Denmark	LML	[2]	
64	12A/2		12A	1966	Denmark	LML		
65	12A/3		12A	1966	Denmark	LML		
66	12A/5	559/66	12A	1966	Denmark	LML	[2]	
67	12A/6		12A	1966	Denmark	LML		
68	12A/7		12A	1966	Denmark	LML		
69	12A/8		12A	1965	Denmark	LML		
70	12A/9		12A	1966	Denmark	LML		
71	12A/10		12A	1966	Denmark	LML		
72	12A/11		12A	1966	Denmark	LML		
73	12A/12		12A	1966	Denmark	LML		

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
74	12A/13		12A	1966	Denmark	LML		
75	12A/14		12A	1966	Denmark	LML		
76	12B/1	Gambia 1/81	12B	1981	Gambia	LML	[2]	
77	12F/2 *		12F	1962	USA	LML		ERR026183‡
78	12F/3 *		12F	1961	Denmark	LML		ERR026195‡
79	12F/4	6312	12F	1968	USA	LML	[2]	
80	12F/5 *		12F	1988	Denmark	LML		ERR026196‡
81	12F/6 *		12F	1996	Denmark	LML		ERR026187‡
82	13/2	34357	13	1952	USA	LML	[2]	
83	14/2 *	34359	14	1952	USA	LML	[2]	ERR025830‡
84	14/3 *		14	1939	Denmark	LML		ERR026719‡
85	14/4 *		14	1961	Denmark	LML		ERR026188‡
86	14/5 *		14	1967	Denmark	LML		ERR026720‡
87	14/6		14	1968	Denmark	LML		
88	14/7 *		14	1992	Denmark	LML		ERR026721‡
89	14/8 *		14	1976	Denmark	LML		ERR026722‡
90	14/9 *		14	1982	Denmark	LML		ERR026723‡
91	15A/2	389/39 (da)	15A	1939	Denmark	LML	[2]	
92	15B/2	7904/39 (da)	15B	1939	Denmark	LML	[2]	
93	15B/3		15B	1996	Denmark	LML		
94	15C/2	553/62 (da)	15C	1962	Denmark	LML	[2]	
95	15F/2		15F	1952	USA	LML		
96	15F/3	688/63 (da)	15F	1963	Denmark	LML	[2]	
97	16A/1	R105 (am)	16A	1985	USA	LML	[2]	
98	16F/2	nr.34361 (am)	16F	1952	USA	LML	[2]	
99	17A/2	nr. 4704 (ge)	17A	1971	Germany	LML	[2]	

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
100	17F/1 *	Rose	17F	1952	USA	LML	[2]	ERR025831‡
101	17F/2 *		17F	1939	Denmark	LML		ERR026724‡
102	17F/3 *		17F	1962	Denmark	LML		ERR026725‡
103	17F/4 *		17F	1962	Denmark	LML		ERR026726‡
104	18A/2	8609/43	18A	1952	Denmark	LML	[2]	
105	18B/2 *	1033/41 (da)	18B	1941	Denmark	LML	[2]	ERR026193‡
106	18C/1	4593/40 (da)	18C	1940	Denmark	LML	[2]	
107	18C/2 *		18C	1939	Denmark	LML		ERR026186‡
108	18C/3 *		18C	1968	Denmark	LML		ERR026716‡
109	18F/1 *	Gethens	18F	1961	USA	LML	[2]	ERR026173‡
110	18F/2		18F	1945	Denmark	LML		
111	19A/2		19A	1939	Denmark	LML		
112	19A/3		19A	1952	Denmark	LML		
113	19A/5	Nr. 141/68	19A	1968	Denmark	LML	[2]	
114	19B/2	nr. 4594 (ge)	19B	1971	Germany	LML	[2]	
115	19C/2	7588/39 (da)	19C	1939	Denmark	LML	[2]	
116	19F/2		19F	1952	USA	LML		
117	19F/3	485/61 (da)	19F	1961	Denmark	LML	[2]	
118	19F/4		19F	1962	Denmark	LML		
119	19F/5 *		19F	1962	Denmark	LML		ERR026717‡
120	19F/6		19F	1968	USA	LML		
121	19F/7 *		19F	1952	Denmark	LML		ERR026189‡
122	19F/8 *		19F	1952	Denmark	LML		ERR026190‡
123	19F/10 *		19F	1963	Denmark	LML		ERR026174‡
124	19F/11 *		19F	1972	Denmark	LML		ERR026175‡
125	19F/12 *		19F	1995	Denmark	LML		ERR026185‡

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
126	19F/14		19F	1996	Spain	LML		
127	20/2		20	1949	USA	LML		
128	20/3	34365 (am)	20	1937	Denmark	LML	[2]	
129	20/4		20	1962	Denmark	LML		
130	21/2		21	1952	USA	LML		
131	21/3	546/62	21	1962	Denmark	LML	[2]	
132	22A/2 *	3405/39(da)	22A	1939	Denmark	LML	[2]	ERR025822‡
133	22A/3		22A	1939	Denmark	LML		
134	22F/2	1772/40 (da)	22F	1940	Denmark	LML	[2]	
135	23A/2 *	1196/45	23A	1945	Denmark	LML	[2]	ERR025821‡
136	23B/1	1039/41 (da)	23B	1941	Denmark	LML	[2]	
137	23B/2		23B	1996	Denmark	LML		
138	23F/2		23F	1940	Denmark	LML		
139	23F/4 *		23F	1967	Australia	LML	[10]	ERR026191‡
140	23F/5 *		23F	1979	Germany	LML		ERR026718‡
141	23F/6		23F	1987	Denmark	LML		
142	23F/10 *	Dr Melchior	23F	1996	Denmark	LML	[2]	ERR026194‡
143	24A/2	2748/40 (da)	24A	1940	Denmark	LML	[2]	
144	24B/2	2236/42 (da)	24B	1942	Denmark	LML	[2]	
145	24F/2	24F L (am)	24F	1952	USA	LML	[2]	
146	24F/3		24F	1943	Denmark	LML		
147	25A/1	tp 25/38; sp 65/81	25A	1981	Singapore	LML	[2]	
148	25A/2		25A	1986	Netherlands	LML		
149	25F/2		25F	1951	USA	LML		
150	25F/3		25F	1962	Denmark	LML		
151	25F/4	601/62 (da)	25F	1962	Denmark	LML	[2]	

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
152	27/2	nr. 34371 (am)	27	1956	USA	LML	[2]	
153	28A/2	1982/45 (da)	28A	1945	Denmark	LML	[2]	
154	28F/2	34372 (am)	28F	1957	USA	LML	[2]	
155	29/2	34373 (am)	29	1952	USA	LML	[2]	
156	31/2		31	1951	USA	LML		
157	31/3	nr. 34374 (am)	31	1952	USA	LML	[2]	
158	31/4		31	1964	Denmark	LML		
159	32A/2	2813/41(da)	32A	1941	Denmark	LML	[2]	
160	32F/2	nr. 34375 (am)	32F	1952	USA	LML	[2]	
161	33A/2	Biehl	33A	1946	USA	LML	[2]	
162	33A/3		33A	1939	Denmark	LML		
163	33B/2	E294 (da)	33B	1962	Denmark	LML	[2]	
164	33C/2	7098/41(da)	33C	1941	Denmark	LML	[2]	
165	33D/2	CSF/79	33D	1979	India	LML	[2]	
166	33D/3		33D	1979	India	LML		
167	33F/2	3084/37 (da)	33F	1937	Denmark	LML	[2]	
168	34/2		34	1959	USA	LML		
169	34/3		34	1968	Malaysia	LML		
170	34/4		34	1972	Denmark	LML		
171	34/5	676/74	34	1974	Denmark	LML	[2]	
172	35A/3	1936/39 (da)	35A	1939	Denmark	LML	[2]	
173	35B/2	4356/39 (da)	35B	1939	Denmark	LML	[2]	
174	35B/3		35B	1968	Malaysia	LML		
175	35B/4		35B	1971	Denmark	LML		
176	35C/2 *		35C	1941	Denmark	LML		ERR026192‡
177	35C/3 *	7765/43 (da)	35C	1943	Denmark	LML	[2]	ERR025823‡

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
178	35F/2	361/39(da)	35F	1939	Denmark	LML	[2]	
179	36/2	1095/39 (da)	36	1939	Denmark	LML	[2]	
180	36/3		36	1979	USA	LML		
181	36/4		36	1992	Denmark	LML		
182	36/5		36	1993	France	LML		
183	37/2		37	1965	USA	LML		
184	37/3		37	1943	Denmark	LML		
185	37/5		37	1972	Denmark	LML		
186	37/6	264/73	37	1973	Denmark	LML	[2]	
187	37/7		37	1973	USA	LML		
188	37/8		37	1973	Denmark	LML		
189	37/9		37	1996	Denmark	LML		
190	38/2	9687/39 (da)	38	1939	Denmark	LML	[2]	
191	39/2	203/40 (da)	39	1940	Denmark	LML	[2]	
192	40/2	Coleman	40	1950	USA	LML	[2]	
193	41A/2	6803 (ge)	41A	1972	Germany	LML		
194	41A/3		41A	1972	Germany	LML		
195	41F/2	8211/40 (da)	41F	1940	Denmark	LML	[2]	
196	42/2	198/71	42	1971	Denmark	LML	[2]	
197	43/3	2427/48(da)	43	1948	Denmark	LML	[2]	
198	43/4		43	1954	USA	LML		
199	44/3	Hammer	44	1956	Germany	LML	[2]	
200	45/2		45	1950	Denmark	LML		
201	45/3		45	1949	Denmark	LML		
202	45/4		45	1949	Denmark	LML		
203	45/5	Eddy nr.72 (am)	45	1954	USA	LML	[2]	

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
204	46/2	Eddy nr. 73 (am)	46	1954	USA	LML	[2]	
205	47A/1	L351	47A	1977	Unknown	LML	[2]	
206	47F/1	Eddy nr. 52 (am)	47F	1977	Unknown	LML	[2]	
207	48/2		48	1962	Denmark	LML		
208	48/3	656/63 (da)	48	1963	Denmark	LML	[2]	
209	48/4		48	1958	Denmark	LML		
210	48/5		48	1957	Denmark	LML		
211	48/6		48	1957	Denmark	LML		
212	Ala243 *	L2003-01243	14	1998	USA	ABB		ERR028708‡
213	Ala263 *	L2003-01263	14	2002	USA	ABB		ERR028712‡
214	Ala283	L2003-01283	14	2000	USA	ABB		
215	Ala287	L2003-01287	14	2001	USA	ABB		
216	Ala289 *	L2003-01289	14	2002	USA	ABB		ERR028713‡
217	Ala291	L2003-01291	14	1998	USA	ABB		
218	Ala292 *	L2003-01292	14	1998	USA	ABB		ERR028714‡
219	Ala317 *	L2003-01317	14	2001	USA	ABB		ERR028715‡
220	GE1	D105	6A	1986	Germany	RH	[8]	
221	GE2	D106	22F	1986	Germany	RH	[8]	
222	GE3	D107	23F	1986	Germany	RH	[8]	
223	GE4	D108	23F	1986	Germany	RH		
224	GE6	D110	NT	1989	Germany	RH		
225	GE7	D111	6B	1989	Germany	RH		
226	GE9	D113	14	1993	Germany	RH		
227	GE10	D121	14	1994	Germany	RH		
228	ICE7 *		7F	1997	Iceland	KGK		ERR025269‡
229	ICE10		12F	1998	Iceland	KGK		

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
230	ICE11 *		6B	1998	Iceland	KGK		ERR025270‡
231	ICE13 *		14	1998	Iceland	KGK		ERR026198‡
232	ICE22 *		7F	1993	Iceland	KGK		ERR028709‡
233	ICE23 *		7F	1993	Iceland	KGK		ERR026202‡
234	ICE27 *		19F	1995	Iceland	KGK		ERR026203‡
235	ICE46 *		14	2003	Iceland	KGK		ERR028710‡
236	ICE50 *		14	2003	Iceland	KGK		ERR026204‡
237	ICE57 *		9N	1994	Iceland	KGK		ERR026205‡
238	ICE186 *		18C	1996	Iceland	KGK		ERR026206‡
239	ICE211 *		23F	1999	Iceland	KGK		ERR026207‡
240	ICE269 *		18C	2005	Iceland	KGK		ERR026208‡
241	ICE501 *		18C	2002	Iceland	KGK		ERR026209‡
242	ICE570 *		14	2002	Iceland	KGK		ERR026199‡
243	ICE587		14	2002	Iceland	KGK		
244	ICE594 *		14	2005	Iceland	KGK		ERR026200‡
245	PMEN1 †	Spain^{23F}-1; ATCC 700669	23F	1984	Spain	LM	[3, 16]	NC_011900.1₴
246	PMEN2 *	Spain^{6B}-2; ATCC 700670	6B	1988	Spain	LM	[16]	ERR025264‡
247	PMEN3 *†	Spain^{9V}-3; ATCC 700671; SP195	9V	1993	France	LM	[16]	ABGE01000001-ABGE01000041₴, ERR025265‡
248	PMEN4 *	Tennessee^{23F}-4; ATCC 51916	23F	1991	USA	LM	[16]	ERR025255‡
249	PMEN5 *	Spain^{14}-5; ATCC 700902	14	1990	Spain	LM	[16]	ERR025256‡
250	PMEN6	Hungary^{19A}-6; ATCC 700673	19A	1989	Hungary	LM	[16]	
251	PMEN7	S.Africa^{19A}-7; ATCC 700674	19A	1989	South Africa	LM	[16]	
252	PMEN8	S.Africa^{6B}-8; ATCC 700675	6B	1990	South Africa	LM	[16]	
253	PMEN9 *	England^{14}-9; ATCC 700676	14	1993	England	LM	[16]	ERR025257‡

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
254	PMEN10 *	CSR^{14}-10; ATCC 700677	14	1987	CSR	LM	[16]	ERR025267‡
255	PMEN11	CSR^{19A}-11; ATCC 700678	19A	1990	CSR	LM	[16]	
256	PMEN12 *	Finland^{6B}-12; ATCC 700903	6B	1987	Finland	LM	[16]	ERR025271‡
257	PMEN13 *	S.Africa^{19A}-13; ATCC 700904	19A	1988	South Africa	LM	[16]	ERR028711‡
258	PMEN14 *	Taiwan^{19F}-14; ATCC 700905	19F	1997	Taiwan	LM	[16]	ERR026201‡
259	PMEN15 *	Taiwan^{23F}-15; ATCC 700906	23F	1997	Taiwan	LM	[16]	ERR025272‡
260	PMEN16	Poland^{23F}-16; ATCC BAA-343	23F	1995	Poland	LM	[16]	
261	PMEN17 *	Maryland^{6B}-17; ATCC BAA-342	6B	1997	USA	LM	[16]	ERR025273‡
262	PMEN18 *	Tennessee^{14}-18; ATCC BAA-340	14	1997	USA	LM	[16]	ERR025274‡
263	PMEN19	Columbia^{5}-19; ATCC BAA-341	5	1995	Columbia	LM	[16]	
264	PMEN20	Poland^{6B}-20; ATCC BAA-612	6B	1996	Poland	LM	[16]	
265	PMEN21	Portugal^{19F}-21; ATCC BAA-657	19F	1996	Portugal	LM	[16]	
266	PMEN22 *	Greece^{6B}-22; ATCC BAA-658	6B	1995	Greece	LM	[16]	ERR025275‡
267	PMEN23	N.Carolina^{6A}-23; ATCC BAA-659	6A	1996	USA	LM	[16]	
268	PMEN24	Utah^{35B}-24; ATCC BAA-660	35B	1994	USA	LM	[16]	
269	PMEN25	Sweden^{15A}-25; ATCC BAA-661	15A	1998	Portugal	LM	[16]	
270	PMEN26	Columbia^{23F}-26; ATCC BAA-662	23F	1998	Portugal	LM	[16]	
271	PMEN27	Sweden^{1}-27; ATCC BAA-1654	1	1992	Sweden	LM		
272	PMEN28	Sweden^{1-}28	1	1997	Sweden	LM		
273	PMEN29	USA^{1}-29	1	1948	USA	LM		
274	PMEN30	Greece^{21}-30	21	1996	Greece	LM		
275	PMEN31	Netherlands^{3}-31	3	1984	Netherlands	LM		
276	PMEN32	Denmark^{14}-32	14	1996	Denmark	LM		
277	PMEN33	Netherlands^{8}-33	8	1989	Netherlands	LM		
278	PMEN34 *	Denmark^{12F}-34	12F	1995	Canada	LM		ERR025276‡

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
279	PMEN35 *	Netherlands^{14}-35	14	1980	Netherlands	LM		ERR025277‡
280	PMEN36 *	Netherlands^{18C}-36	18C	1980	Netherlands	LM		ERR025278‡
281	PMEN37	Netherlands^{15B}-37	15B	1987	Netherlands	LM		
282	PMEN38	Sweden^{4}-38	4	1995	Sweden	LM		
283	PMEN39 *	Netherlands^{7F}-39	7F	1984	Netherlands	LM		ERR025268‡
284	PMEN40	Sweden^{1}-40	1	1995	Sweden	LM		
285	PMEN41	Portugal^{6A}-41	6A	1996	Portugal	LM		
286	PMEN42	Norway^{NT}-42	NT	1996	Portugal	LM		
287	PMEN43	USA^{NT}-43	NT	1980	USA	LM		
288	PN1	Pn12	6	1972	PNG	RH	[13]	
289	PN2	P1	4	1969	PNG	RH	[6]	
290	SA1	203A	19A	1978	South Africa	MRJ	[12, 19]	
291	SA2	223 (19A)	19A	1978	South Africa	MRJ	[12, 19]	
292	SA3	381	19A	1978	South Africa	MRJ	[12, 19]	
293	SA4	107	19A	1978	South Africa	MRJ	[12, 19]	
294	SA5	61857	19A	1978	South Africa	MRJ	[12, 19]	
295	SA6	61783	19A	1978	South Africa	MRJ	[12, 19]	
296	SA7	145	19A	1978	South Africa	MRJ	[12, 19]	
297	SA8	209	19A	1978	South Africa	MRJ	[12, 19]	
298	SA9	205	19A	1978	South Africa	MRJ	[12, 19]	
299	SA10	DURBANT	19A	1977	South Africa	MRJ	[1, 19]	
300	SA11	60292	6A	1978	South Africa	MRJ	[12, 19]	
301	SA12	994	6A	1978	South Africa	MRJ	[12, 19]	
302	SA13	60293	6A	1978	South Africa	MRJ	[12, 19]	
303	SA14	62272	6A	1978	South Africa	MRJ	[12, 19]	
304	SA15	532	6A	1978	South Africa	MRJ	[12, 19]	

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
305	SA16	61155	6A	1978	South Africa	MRJ	[12, 19]	
306	SA17	13811	6A	1978	South Africa	MRJ	[12, 19]	
307	SA18	223 (6A)	6A	1978	South Africa	MRJ	[12, 19]	
308	SA19	61153	6A	1978	South Africa	MRJ	[12, 19]	
309	SA20	61154	6A	1978	South Africa	MRJ	[12, 19]	
310	SA21	135	6A	1978	South Africa	MRJ	[12, 19]	
311	SA22	61855	14	1978	South Africa	MRJ	[12, 19]	
312	SA23	61370	19A	1978	South Africa	MRJ	[12, 19]	
313	SA24	A922	19A	1978	South Africa	MRJ	[12, 19]	
314	SA25	A910	19A	1978	South Africa	MRJ	[12, 19]	
315	SA26	A1002	19A	1978	South Africa	MRJ	[12, 19]	
316	SA27	SA4; 8256	14	1978-1980	South Africa	RH	[8]	
317	SA28	SA5; 272	14	1978-1980	South Africa	RH	[8]	
318	SA31	SA8; 2039	6A	1978-1980	South Africa	RH	[8]	
319	SA32	SA9; 10760	6A	1978-1980	South Africa	RH	[8]	
320	SA34	SA11; 11405	29	1988	South Africa	RH		
321	SA35	SA12; 45607	29	1987	South Africa	RH	[8]	
322	SA36	SA13; 60814	29	1987	South Africa	RH	[8, 9]	
323	SA37	SA14; 56828	19F	1987	South Africa	RH	[8]	
324	SA38	SA18; 60785	6A	1987	South Africa	RH	[8]	
325	SA39	SA19; 63509	19A	1987	South Africa	RH	[8]	
326	SA40	SA20; 205	19A	mid 1980s	South Africa	RH		
327	SA41	SA21; 215	6B	mid 1980s	South Africa	RH		
328	SA42	SA22; 613	19F	mid 1980s	South Africa	RH		
329	SA43	SA23; 20463	19A	mid 1980s	South Africa	RH		
330	SA44	SA24; 39062	6A	mid 1980s	South Africa	RH		

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
331	SA45	SA25; 60292	6A	mid 1980s	South Africa	RH		
332	SA47	97-1	19A	1979	South Africa	KPK, AVG		
333	SA48	48	19A	1979	South Africa	KPK, AVG		
334	SA49	48-2	19A	1982	South Africa	KPK, AVG		
335	SA51	46878	6B	1989	South Africa	KPK, AVG		
336	SA52	13979	6B	1984	South Africa	KPK, AVG		
337	SA53	N383	11A	1983	South Africa	KPK, AVG		
338	SA54	136664	6B	1983	South Africa	KPK, AVG		
339	SA55	31664	19A	1983	South Africa	KPK, AVG		
340	SA56	6415	6B	1984	South Africa	KPK, AVG		
341	SA57	108551	6B	1984	South Africa	KPK, AVG		
342	SA58	19931	19F	1986	South Africa	KPK, AVG		
343	SA59	3630	6A	1985	South Africa	KPK, AVG		
344	SA60	511	19A	1985	South Africa	KPK, AVG		
345	SA61	97	19A	1985	South Africa	KPK, AVG		
346	SA62	28649	6B	1985	South Africa	KPK, AVG		
347	SA63	93343	14	1985	South Africa	KPK, AVG		
348	SA64	6482	19F	1985	South Africa	KPK, AVG		
349	SA65	6076	38	1985	South Africa	KPK, AVG		
350	SA66	6428	6B	1985	South Africa	KPK, AVG		
351	SA67	34023	6B	1985	South Africa	KPK, AVG		
352	SA68	03904	6A	1985	South Africa	KPK, AVG		
353	SA69	34142	19A	1985	South Africa	KPK, AVG		
354	SA70	6130	19A	1985	South Africa	KPK, AVG		
355	SA71	34581	23F	1989	South Africa	KPK, AVG		
356	SA72	34139	6A	1988	South Africa	KPK, AVG		

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
357	SA73	48628	19A	1988	South Africa	KPK, AVG		
358	SA74	13897	23F	1988	South Africa	KPK, AVG		
359	SA75	13619	23F	1987	South Africa	KPK, AVG		
360	SA76	95103	14	1988	South Africa	KPK, AVG		
361	SA77	26940	23F	1987	South Africa	KPK, AVG		
362	SA78	47590	14	1989	South Africa	KPK, AVG		
363	SA79	49171	19A	1989	South Africa	KPK, AVG		
364	SA80	22453	23F	1989	South Africa	KPK, AVG		
365	SA81	23099	23F	1989	South Africa	KPK, AVG		
366	SA82	47085	14	1989	South Africa	KPK, AVG		
367	SA83	46518	6B	1989	South Africa	KPK, AVG		
368	SA84	SA3; 35897	14	1978-1980	South Africa	RH	[9]	
369	SN1	602	6B	1987	Spain	JL	[7, 17, 18]	
370	SN2	629	9V	1987	Spain	JL		
371	SN3	795	9V	1988	Spain	JL	[21]	
372	SN4	946	9V	1988	Spain	JL	[22]	
373	SN5	998	9V	1988	Spain	JL		
374	SN6	1196	6B	1989	Spain	JL		
375	SN7	1064	6B	1989	Spain	JL		
376	SN8	1070	9V	1989	Spain	JL		
377	SN9	1000	9V	1989	Spain	JL		
378	USA1	4793-00	18C	1999	USA	LM, BWB		
379	USA2 *	6223-99	18C	1999	USA	LM, BWB		ERR028716[‡]
380	USA3	6543-99	18C	1999	USA	LM, BWB		
381	USA4	0050-02	14	2001	USA	LM, BWB		
382	USA5	6605-01	14	2001	USA	LM, BWB		

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
383	USA6 *	4347-99	14	1999	USA	LM, BWB		ERR025254‡
384	USA7	2836-99	14	1999	USA	LM, BWB		
385	USA8 *	0680-02	9N	2001	USA	LM, BWB		ERR025258‡
386	USA9 *	PATH#183	14	1999	USA	LM, BWB		ERR025259‡
387	USA10	PATH#184	14	1999	USA	LM, BWB		
388	USA11 *	PATH#286	14	Unknown	USA	LM, BWB		ERR025260‡
389	USA12 *	PATH#678	23F	2001	USA	LM, BWB		ERR025261‡
390	USA13 *	PATH#688	19F	2005	USA	LM, BWB		ERR025262‡
391	USA14	1087-00	7F	1999	USA	LM, BWB		
392	USA15	7102-99	7F	1999	USA	LM, BWB		
393	USA16 *	PATH#224	7F	2003	USA	LM, BWB		ERR028717‡
394	USA17	2008008776	7F	2005	USA	LM, BWB		
395	USA18 *	4621-00	12F	1999	USA	LM, BWB		ERR028718‡
396	USA19	5400-00	7F	1999	USA	LM, BWB		
397	USA20 *	PATH#129	7F	1999	USA	LM, BWB		ERR025263‡
398	USA21	PATH#143	6A	2007	USA	LM, BWB		
399	USA22 *	3140-06	6C	2005	USA	LM, BWB		ERR028719‡
400	USA23	89FL23F	23F	1989	USA	LM	[15]	
401	USA24	89SA23F	23F	1989	South Africa	LM	[15]	
402	USA25	90MI14	14	1990	USA	LM	[15]	
403	USA26	85PA23F	23F	1985	USA	LM	[15]	
404	USA27	85NY23F	23F	1985	USA	LM	[15]	
405	USA28	88FL23F	23F	1988	USA	LM	[15]	
406	USA29	90GA14	14	1990	USA	LM	[15]	
407	USA30	91NC23F	23F	1991	USA	LM	[15]	
408	USA31	83MI23F	23F	1983	USA	LM	[15]	

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
409	USA32	91AZ23F	23F	1991	USA	LM	[15]	
410	USA33	81CO6B	6B	1981	USA	LM	[15]	
411	USA34	85NY19A	19A	1985	USA	LM	[15]	
412	USA35	89NY23F	23F	1989	USA	LM	[15]	
413	USA36	90OH23F	23F	1990	USA	LM	[15]	
414	USA37	91MT23F	23F	1991	USA	LM	[15]	
415	USA38	90TX23F	23F	1990	USA	LM	[15]	
416	USA39	90MO6B	6B	1990	USA	LM	[15]	
417	USA40	90OK23F	23F	1990	USA	LM	[15]	
418	USA41	90IN23F	23F	1990	USA	LM	[15]	
419	USA42	84PA23F	23F	1984	USA	LM	[15]	
420	USA43	90NB23F	23F	1990	USA	LM	[15]	
421	USA44	88AZ23F	23F	1988	USA	LM	[15]	
422	USA45	91CA23F	23F	1991	USA	LM	[15]	
423	USA46	89NB23F	23F	1989	USA	LM	[15]	
424	USA47	85UT23F	23F	1985	USA	LM	[15]	
425	USA48	91GA23F	23F	1991	USA	LM	[15]	
426	USA49	89CA23F	23F	1989	USA	LM	[15]	
427	670 †		6B	1988	Spain			CP002176.1[ǂ]
428	70585 †		5	Unknown	Unknown			CP000918.1[ǂ]
429	CDC0288-04 †		12F	2003-2004	Unknown			ABGF01000001-ABGF01000038[ǂ]
430	CDC1087-00 †		7F	1999	USA			ABFT01000001-ABFT01000075[ǂ]
431	CDC1873-00 †		6A	1999	USA			ABFS01000001-ABFS01000053[ǂ]
432	CDC3059-06 †		19A	2005	USA			ABGG01000001-

#	Isolate	Other Name(s)	Serotype	Year	Country	Contributor	Reference(s)	Genome Accession
								ABGG01000033[ƨ]
433	CGSP14 †		14	2004-2005	Taiwan		[4]	NC_010582.1[ƨ]
434	D39 †		2	1916	Unknown		[14]	CP000410.1[ƨ]
435	JJA †		14	1995	Brazil		[5]	CP000919.1[ƨ]
436	MLV-016 †		11A	Unknown	Unknown			ABGH01000001-ABGH01000126[ƨ]
437	P1031 †		1	2002	China			CP000920.1[ƨ]
438	R6 †		NT	1964	Unknown		[11]	AE007317.1[ƨ]
439	SPnINV104 †		1	1998	England		[5]	FQ312030.1[ƨ]
440	SPnINV200 †		14	1995	England		[5]	FQ312029.1[ƨ]
441	SPnOXC141 †		3	2001	England		[5]	FQ312027.1[ƨ]
442	TIGR4 †		4	1991	Norway		[20]	AE005672.3[ƨ]

* Isolate selected for whole-genome sequencing. † Genome retrieved from Genbank. [‡] European Nucleotide Archive accession. [ƨ] Genbank accession.

A1.1.1 References

1. **Appelbaum, P. C., J. N. Scragg, A. J. Bowen, A. Bhamjee, A. F. Hallett, and R. C. Cooper.** 1977. *Streptococcus pneumoniae* resistant to penicillin and chloramphenicol. Lancet **2:**995-997.
2. **Bentley, S. D., D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabbinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail, G. Samuel, I. C. Skovsted, M. S. Kaltoft, B. Barrell, P. R. Reeves, J. Parkhill, and B. G. Spratt.** 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. PLoS Genet. **2:**262-269.
3. **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, S. D. Bentley, and T. J. Mitchell.** 2009. Role of conjugative elements in the evolution of the multidrug-resistant pandemic clone *Streptococcus pneumoniae*^{Spain23F} ST81. J Bacteriol. **191:**1480-1489.
4. **Ding, F., P. Tang, M.-H. Hsu, P. Cui, S. Hu, J. Yu, and C.-H. Chiu.** 2009. Genome evolution driven by host adaptations results in a more virulent and antimicrobial-resistant *Streptococcus pneumoniae* serotype 14. BMC Genet. **10:**158.
5. **Donati, C., N. L. Hiller, H. Tettelin, A. Muzzi, N. J. Croucher, S. V. Angiuoli, M. Oggioni, J. C. Dunning Hotopp, F. Z. Hu, D. R. Riley, A. Covacci, T. J. Mitchell, S. D. Bentley, M. Kilian, G. D. Ehrlich, R. Rappuoli, E. R. Moxon, and V. Masignani.** 2010. Structure and dynamics of the pan-genome of *Streptococcus pneumoniae* and closely related species. Genome Biol. **11:**R107.

6. **Dowson, C. G., A. Hutchison, J. A. Brannigan, R. C. George, H. David, J. Liñares, A. Tomasz, J. M. Smith, and B. G. Spratt.** 1989. Horizontal transfer of penicillin-binding protein genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. Proc Natl Acad Sci U S A. **86:**8842-8846.
7. **Fernández, A., C. Cabellos, F. Tubau, J. Liñares, P. F. Viladrich, and F. Gudiol.** 2001. Relationship between penicillin and cephalosporin resistance of *Streptococcus pneumoniae* strains and its inflammatory activity in the experimental model of meningitis. Med Microbiol Immunol. **190:**135-138.
8. **Hakenbeck, R., T. Briese, L. Chalkley, H. Ellerbrok, R. Kalliokoski, C. Latorre, M. Leinonen, and C. Martin.** 1991. Antigenic variation of penicillin-binding proteins from penicillin-resistant clinical strains of *Streptococcus pneumoniae.* J Infect Dis. **164:**313-319.
9. **Hakenbeck, R., T. Briese, L. Chalkley, H. Ellerbrok, R. Kalliokoski, C. Latorre, M. Leinonen, and C. Martin.** 1991. Variability of penicillin-binding proteins from penicillin-sensitive *Streptococcus pneumoniae*. J Infect Dis. **164:**307-312.
10. **Hansman, D., and M. M. Bullen.** 1967. A resistant pneumococcus. Lancet **2:**264-265.
11. **Hoskins, J., W. E. Alborn, J. Arnold, L. C. Blaszczak, S. Burgett, B. S. DeHoff, S. T. Estrem, L. Fritz, D. J. Fu, W. Fuller, C. Geringer, R. Gilmour, J. S. Glass, H. Khoja, A. R. Kraft, R. E. Lagace, D. J. LeBlanc, L. N. Lee, E. J. Lefkowitz, J. Lu, P. Matsushima, S. M. McAhren, M. McHenney, K. McLeaster, C. W. Mundy, T. I. Nicas, F. H. Norris, M. O'Gara, R. B. Peery, G. T. Robertson, P. Rockey, P. M. Sun, M. E. Winkler, Y. Yang, M. Young-Bellido, G. S. Zhao, C. A. Zook, R. H. Baltz, S. R. Jaskunas, P. R. Rosteck, P. L. Skatrud, and J. I. Glass.** 2001. Genome of the bacterium *Streptococcus pneumoniae* strain R6. J Bacteriol. **183:**5709-5717.
12. **Jacobs, M. R., H. J. Koornhof, R. M. Robins-Browne, C. M. Stevenson, Z. A. Vermaak, I. Freiman, B. Miller, M. A. Witcomb, M. Isaäcson, J. I. Ward, and R. Austrian.** 1978. Emergence of multiply resistant pneumococci. N Engl J Med. **299:**735-740.
13. **Laible, G., B. G. Spratt, and R. Hakenbeck.** 1991. Interspecies recombinational events during the evolution of altered PBP 2x genes in penicillin-resistant clinical isolates of *Streptococcus pneumoniae*. Mol Microbiol. **5:**1993-2002.
14. **Lanie, J. A., W. L. Ng, K. M. Kazmierczak, T. M. Anrzejewski, T. M. Davidson, K. J. Wayne, H. Tettelin, J. I. Glass, and M. E. Winkler.** 2007. Genome sequence of Avery's virulent serotype 2 strain D39 of *Steptococcus pneumoniae* and comparison with that of unencapsulated laboratory strain R6. J Bacteriol. **189:**38-51.
15. **McDougal, L. K., J. K. Rasheed, J. W. Biddle, and F. C. Tenover.** 1995. Identification of multiple clones of extended-spectrum cephalosporin-resistant *Streptococcus pneumoniae* isolates in the United States. Antimicrob Agents Chemother. **39:**2282-2288.
16. **McGee, L., L. McDougal, J. Zhou, B. G. Spratt, F. C. Tenover, R. George, R. Hakenbeck, W. Hryniewicz, J. C. Lefévre, A. Tomasz, and K. P. Klugman.** 2001. Nomenclature of major antimicrobial-resistant clones of *Streptococcus pneumoniae* defined by the Pneumococcal Molecular Epidemiology Network. J Clin Microbiol. **39:**2565-2571.
17. **Muñoz, R., C. G. Dowson, M. Daniels, T. J. Coffey, C. Martin, R. Hakenbeck, and B. G. Spratt.** 1992. Genetics of resistance to third-generation cephalosporins in clinical isolates of *Streptococcus pneumoniae*. Mol Microbiol. **6:**2461-2465.
18. **Pallarés, R., J. Liñares, M. Vadillo, C. Cabellos, F. Manresa, P. F. Viladrich, R. Martin, and F. Gudiol.** 1995. Resistance to penicillin and cephalosporin and mortality from severe pneumococcal pneumonia in Barcelona, Spain. N Engl J Med. **333:**474-480.
19. **Reinert, R. R., J. M. R, P. C. Appelbaum, S. Bajaksouzian, S. Cordeiro, M. van der Linden, and A. Al-Lahham.** 2005. Relationship between the original multiply resistant South African isolates of *Streptococcus pneumoniae* from 1977 to 1978 and contemporary international resistant clones. J Clin Microbiol. **43:**6035-6041.
20. **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, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser.** 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae.* Science **293:**498-506.

21. **Viladrich, P. F., C. Cabellos, R. Pallarés, F. Tubau, J. Martinez-Lacasa, J. Liñares, and F. Gudiol.** 1996. High doses of cefotaxime in treatment of adult meningitis due to *Streptococcus pneumoniae* with decreased susceptibilities to broad spectrum cephalosporins. Antimicrob Agents Chemother. **40:**218-220.
22. **Viladrich, P. F., F. Gudiol, J. Liñares, R. Pallarés, I. Sabaté, G. Rufí, and J. Ariza.** 1991. Evaluation of vancomycin for therapy of adult pneumococcal meningitis. Antimicrob Agents Chemother. **35:**2467-2472.

APPENDIX 1.2 MLST, *pbp* allele and penicillin/cefotaxime susceptibility data for the historical isolate collection plus 16 isolates for which genomic sequences were retrieved from Genbank and used throughout this thesis.

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
1	1/2	10	31	4	1	6	4	94	615	615	0.016	0.008	26	18	34
2	1/3	10	337	4	1	6	4	94	7241	615	0.023	0.016	26	18	34
3	1/4	12	8	13	5	17	4	8	5316	306	0.012	0.008	4	59	51
4	1/5	12	8	13	5	17	4	8	5316	306	0.012	0.012	4	59	51
5	1/6	12	8	13	5	17	4	8	5316	306	0.023	0.023	4	59	51
6	1/7	17	8	13	5	17	4	28	250	304	0.016	0.016	4	59	60
7	2/2 *	2	1	1	1	6	31	14	574	574	0.016	0.012	26	60	55
8	2/3 *	2	13	4	1	6	6	145	3744	490	0.032	0.023	46	36	57
9	2/4	7	5	1	1	10	7	15	595	None128	0.023	0.012	53	26	29
10	3/2	13	9	15	14	10	16	19	378	378	0.012	0.012	63	33	14
11	3/3	13	9	15	14	311	16	19	7244	378	0.016	0.008	63	33	14
12	3/4	13	9	15	14	10	16	19	378	378	0.023	0.016	63	33	13
13	3/5	1	5	328	11	13	1	14	7169	2021	0.016	0.006	59	61	49
14	3/6	10	336	28	11	17	111	17	7211	1024	0.012	0.006	59	62	10
15	3/7	13	84	15	14	10	16	19	1377	378	0.023	0.012	63	33	14
16	3/8	13	84	15	14	10	16	19	1377	378	0.023	0.016	63	33	14
17	4/2	16	13	330	5	6	440	14	7224	Singleton7224	0.016	0.008	18	18	42
18	4/3	16	13	4	5	6	10	14	247	247	0.016	0.012	21	18	42

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
19	4/4	16	13	4	5	309	10	14	7223	247	0.016	0.016	21	18	42
20	4/5	16	13	20	5	6	10	14	249	247	0.023	0.016	21	63	42
21	5/1	1	103	14	8	6	105	29	4840	None3035	0.012	0.012	41	30	47
22	5/2	1	103	14	8	6	105	29	4840	None3035	0.016	0.012	41	30	47
23	5/3	16	12	9	1	41	33	33	289	289	0.023	0.032	ND	ND	ND
24	5/4	16	12	9	1	295	33	33	7222	289	0.023	0.016	ND	ND	ND
25	6A/2	5	7	4	5	10	1	27	7188	460	0.016	0.012	20	43	48
26	6B/3	7	8	9	1	17	3	20	7199	Singleton7199	0.008	0.006	ND	ND	ND
27	7A/2 *	8	9	2	1	6	1	17	191	191	0.008	0.012	50	38	43
28	7A/3	5	5	27	1	10	1	70	5994	Singleton5994	0.032	0.016	3	18	11
29	7B/2 *	2	13	2	4	17	1	1	7180	66	0.023	0.023	35	21	26
30	7B/3	12	19	2	17	6	22	14	230	230	0.500	0.190	104	99	97
31	7C/2	1	8	4	1	9	3	18	7175	43	0.012	0.012	ND	ND	ND
32	7C/3	1	8	4	1	9	3	8	7171	43	0.016	0.016	39	51	51
33	7F/2 *	10	20	14	1	6	274	29	7210	218	0.023	0.012	50	38	30
34	7F/3 *	8	9	2	1	6	1	17	191	191	0.016	0.012	50	38	43
35	7F/4 *	8	335	2	1	6	1	17	7208	191	0.023	0.012	50	38	43
36	7F/5	8	9	2	1	6	1	17	191	191	0.012	0.012	50	38	43
37	7F/6	10	20	14	1	6	274	29	7210	218	0.023	0.012	50	38	30
38	8/2	2	5	1	11	16	3	14	53	62	0.016	0.023	ND	ND	ND
39	8/3	2	5	1	11	16	3	14	53	62	0.012	0.012	42	4	55
40	8/4	7	157	15	11	93	1	70	7203	1480	0.016	0.016	42	5	39

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
41	8/5	7	9	15	11	42	1	70	404	1480	0.016	0.016	42	4	40
42	9A/1 *	7	11	10	1	6	1	14	312	162	0.012	0.006	16	58	21
43	9L/2 *	7	16	1	8	14	11	14	5979	124	0.023	0.012	38	25	51
44	9L/3	7	334	8	8	6	28	14	7204	292	0.023	0.012	37	14	56
45	9L/4	7	16	8	8	6	49	14	7240	2174	0.012	0.016	38	14	56
46	9N/2 *	8	5	7	12	6	16	6	7205	None3782	0.023	0.023	34	58	51
47	9N/3	8	5	7	12	15	16	6	3983	None3782	0.023	0.023	ND	ND	ND
48	9N/4	2	13	2	4	6	1	1	71	66	0.023	0.023	34	58	27
49	9N/5	2	13	2	4	6	1	1	71	66	0.012	0.016	34	58	26
50	9N/6 *	2	13	2	4	6	1	1	71	66	0.016	0.012	34	58	26
51	9V/2	7	11	10	5	6	20	6	7200	Singleton7200	0.012	0.008	ND	ND	ND
52	9V/4 *	7	2	40	1	10	1	45	123	113	0.008	0.012	58	58	45
53	9V/5 *	7	11	10	1	6	8	14	162	162	0.016	0.012	101	64	21
54	9V/6 *	7	11	10	1	6	8	14	162	162	0.016	0.012	101	64	21
55	10A/1	5	7	4	2	10	1	27	97	460	0.016	0.012	ND	ND	ND
56	10B/2	1	5	9	16	6	4	197	5505	Singleton5505	0.016	0.012	ND	ND	ND
57	10C/2	7	5	4	1	6	28	18	2862	Singleton2862	0.012	0.012	ND	ND	ND
58	10F/2 *	2	53	2	1	6	19	14	7186	490	0.016	0.016	45	49	53
59	11A/2	13	5	53	1	15	445	14	7214	Singleton7214	0.012	0.016	ND	ND	ND
60	11B/2	1	5	53	1	17	17	45	7165	Singleton7165	0.023	0.023	66	14	31
61	11C/1 *	7	16	1	16	14	11	14	7201	124	0.023	0.012	49	45	51
62	11D/1	2	5	29	12	16	3	14	62	62	0.023	0.012	ND	ND	ND

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
63	11F/2	12	281	4	1	15	1	1	7212	None5995	0.012	0.006	21	40	25
64	12A/2	1	2	7	18	17	14	17	5977	3122	0.012	0.012	33	13	3
65	12A/3	1	2	7	18	17	14	17	5977	3122	0.012	0.016	33	53	3
66	12A/5	1	2	7	18	17	14	17	5977	3122	0.016	0.012	33	13	3
67	12A/6	1	2	7	18	17	14	17	5977	3122	0.016	0.016	33	13	3
68	12A/7	1	2	7	18	17	14	17	5977	3122	0.012	0.008	33	13	3
69	12A/8	1	2	7	18	17	14	17	5977	3122	0.012	0.012	33	13	3
70	12A/9	1	2	7	18	17	14	17	5977	3122	0.008	0.012	33	13	3
71	12A/10	1	2	7	18	17	14	17	5977	3122	0.012	0.008	33	13	3
72	12A/11	1	2	7	18	17	14	17	5977	3122	0.012	0.012	33	13	3
73	12A/12	1	2	7	18	17	14	17	5977	3122	0.012	0.012	33	13	3
74	12A/13	1	2	7	18	17	14	17	5977	3122	0.012	0.012	33	13	3
75	12A/14	1	2	7	18	17	14	17	5977	3122	0.012	0.012	33	11	3
76	12B/1	2	153	111	1	15	20	15	7187	None3745	0.064	0.023	ND	ND	ND
77	12F/2 *	7	2	4	10	197	444	15	7194	Singleton7194	0.012	0.016	38	14	16
78	12F/3 *	18	5	4	1	310	1	6	7228	4399	0.012	0.012	21	14	12
79	12F/4	7	2	4	10	197	444	15	7194	Singleton7194	0.012	0.008	38	14	16
80	12F/5 *	10	20	14	1	6	1	29	218	218	0.016	0.016	33	14	30
81	12F/6 *	10	20	14	1	6	1	29	218	218	0.012	0.008	33	14	30
82	13/2	1	8	1	15	31	441	8	7173	Singleton7173	0.016	0.012	27	46	18
83	14/2 *	7	5	1	8	14	11	14	124	124	0.016	0.012	52	42	51
84	14/3 *	8	8	4	15	17	12	31	875	1106	0.012	0.008	70	20	46

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
85	14/4 *	7	5	1	8	14	1	14	134	124	0.012	0.016	52	43	51
86	14/5 *	1	5	4	5	5	3	8	15	15	0.016	0.008	7	52	37
87	14/6	2	13	2	8	17	79	288	7183	5401	0.008	0.012	17	23	26
88	14/7 *	7	5	1	8	14	25	14	7198	124	0.016	0.012	52	43	49
89	14/8 *	8	8	4	15	39	12	498	7206	554	0.75	0.38	1	1	4
90	14/9 *	7	5	1	8	14	11	14	124	124	0.008	0.008	52	43	49
91	15A/2	1	5	12	5	9	15	5	7166	Singleton7166	0.016	0.012	ND	ND	ND
92	15B/2	1	1	1	1	1	1	1	1	1	0.012	0.012	ND	ND	ND
93	15B/3	1	1	1	1	9	1	1	2008	1	0.012	0.012	ND	ND	ND
94	15C/2	1	8	1	1	1	1	1	7172	1	0.023	0.012	30	64	23
95	15F/2	13	8	4	4	10	28	6	7217	None5041	0.012	0.012	53	55	28
96	15F/3	2	5	4	66	10	1	11	7177	1232	0.016	0.016	53	56	3
97	16A/1	5	300	1	16	17	79	14	6543	Singleton6543	0.012	0.016	ND	ND	ND
98	16F/2	1	5	4	4	6	58	14	7167	None659	0.016	0.006	21	15	20
99	17A/2	18	12	2	16	6	19	245	2599	2599	0.012	0.016	43	25	8
100	17F/1 *	2	1	1	1	6	31	14	574	574	0.012	0.012	26	9	55
101	17F/2 *	7	2	40	1	10	1	45	123	113	0.012	0.016	58	9	45
102	17F/3 *	7	5	1	1	6	31	14	392	392	0.012	0.008	26	9	55
103	17F/4 *	7	5	1	1	6	31	14	392	392	0.012	0.012	26	9	55
104	18A/2	25	31	4	16	32	28	44	241	241	0.016	0.008	44	39	59
105	18B/2 *	7	2	1	1	10	1	6	4706	113	0.012	0.008	58	35	51
106	18C/1	7	2	1	1	10	1	6	4706	113	0.012	0.006	58	35	51

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
107	18C/2 *	7	2	1	1	10	1	6	4706	113	0.012	0.016	58	35	51
108	18C/3 *	7	2	1	2	10	1	6	7195	113	0.008	0.008	58	35	51
109	18F/1 *	2	13	1	1	6	31	14	7181	490	0.023	0.012	61	27	54
110	18F/2	1	5	9	1	6	11	14	4863	Singleton4863	0.023	0.016	28	13	17
111	19A/2	18	5	9	1	9	1	14	7226	Singleton7226	0.016	0.012	ND	ND	ND
112	19A/3	18	5	9	1	9	1	14	7226	Singleton7226	0.012	0.012	19	47	46
113	19A/5	18	5	9	1	9	1	14	7226	Singleton7226	0.012	0.008	19	65	46
114	19B/2	8	13	14	4	17	4	14	199	199	0.032	0.016	35	23	35
115	19C/2	2	13	2	12	17	1	14	3785	Singleton3785	0.023	0.012	ND	ND	ND
116	19F/2	1	1	1	1	1	1	1	1	1	0.012	0.016	47	7	23
117	19F/3	1	5	1	5	1	1	8	425	425	0.010	0.006	32	51	51
118	19F/4	18	5	4	1	15	1	6	7229	4399	0.012	0.012	21	51	24
119	19F/5 *	18	5	4	1	15	1	6	7229	4399	0.012	0.012	21	51	24
120	19F/6	1	1	1	1	1	1	1	1	1	0.012	0.008	35	7	22
121	19F/7 *	18	5	13	1	27	1	6	7230	Singleton7230	0.016	0.016	21	51	12
122	19F/8 *	18	5	4	1	15	1	6	7229	4399	0.012	0.012	21	51	24
123	19F/10 *	18	5	13	1	27	1	6	7230	Singleton7230	0.012	0.012	21	51	12
124	19F/11 *	2	13	2	4	6	1	1	71	66	0.016	0.016	34	58	26
125	19F/12 *	5	5	7	7	8	5	4	87	87/88	0.190	0.125	68	66	61
126	19F/14	4	4	2	4	4	1	1	81	81	0.380	0.380	67	72	62
127	20/2	15	8	8	1	15	1	31	7221	235	0.016	0.016	ND	ND	ND
128	20/3	15	8	8	18	15	1	31	235	235	0.016	0.012	ND	ND	ND

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
129	20/4	1	2	7	18	17	14	17	5977	3122	0.012	0.016	33	11	49
130	21/2	8	10	9	8	1	10	14	7207	Singleton7207	0.016	0.023	17	7	33
131	21/3	8	10	2	16	1	26	1	193	193	0.023	0.012	17	24	36
132	22A/2 *	2	13	4	1	6	31	14	7182	490	0.023	0.016	69	53	48
133	22A/3	2	8	4	16	17	1	14	7178	Singleton7178	0.023	0.016	ND	ND	ND
134	22F/2	13	161	34	1	6	1	6	7219	Singleton7219	0.012	0.006	36	59	12
135	23A/2 *	1	8	9	2	6	4	6	439	439	0.016	0.012	29	13	2
136	23B/1	1	5	4	4	15	1	8	7168	Singleton7168	0.016	0.012	12	22	1
137	23B/2	18	13	8	6	3	6	8	1349	1349	0.125	0.094	U	100	93
138	23F/2	1	5	69	1	6	1	14	7170	Singleton7170	0.016	0.012	65	8	44
139	23F/4 *	2	22	2	4	308	1	1	7184	Singleton7184	0.125	0.094	2	2	5
140	23F/5 *	1	8	9	2	6	4	6	439	439	0.012	0.012	29	13	2
141	23F/6	4	4	2	4	4	1	1	81	81	1.000	0.750	67	72	62
142	23F/10 *	16	8	9	1	6	4	72	515	439	0.023	0.008	29	13	63
143	24A/2	2	9	9	1	6	17	9	5980	Singleton5980	0.012	0.008	21	13	36
144	24B/2	13	5	40	134	1	1	18	7215	Singleton7215	0.016	0.012	33	44	1
145	24F/2	1	8	4	16	307	20	9	7174	Singleton7174	0.016	0.008	17	25	36
146	24F/3	2	8	2	1	17	1	1	7179	None263	0.023	0.012	58	69	32
147	25A/1	10	5	87	18	10	49	6	7209	Singleton7209	0.023	0.012	ND	ND	ND
148	25A/2	142	10	4	5	17	33	277	5407	5407	0.023	0.016	ND	ND	ND
149	25F/2	1	15	87	134	10	1	6	5981	Singleton5981	0.023	0.012	5	50	7
150	25F/3	5	15	4	1	6	1	6	105	5604	0.016	0.012	13	50	7

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
151	25F/4	5	15	4	1	6	1	6	105	5604	0.016	0.012	13	50	7
152	27/2	2	17	107	24	9	105	14	1475	1475	0.032	0.016	26	70	38
153	28A/2	16	44	14	16	9	226	17	7225	Singleton7225	0.012	0.006	51	45	19
154	28F/2	18	5	4	1	6	77	14	7227	None2661	0.008	0.006	29	25	41
155	29/2	2	12	2	8	6	1	14	5982	Singleton5982	0.023	0.012	11	25	50
156	31/2	1	2	29	18	6	1	18	7163	568	0.008	0.012	33	13	51
157	31/3	1	2	29	18	6	1	18	7163	568	0.016	0.012	33	13	51
158	31/4	1	2	29	18	6	1	18	7163	568	0.012	0.012	33	13	51
159	32A/2	5	8	1	62	6	77	412	5983	Singleton5983	0.023	0.012	33	22	9
160	32F/2	5	8	1	62	6	77	412	5983	Singleton5983	0.016	0.012	33	22	9
161	33A/2	2	5	29	18	42	3	18	1012	1012	0.016	0.008	60	38	1
162	33A/3	2	5	29	18	42	3	18	1012	1012	0.012	0.016	ND	ND	ND
163	33B/2	16	5	166	1	13	14	18	2864	Singleton2864	0.016	0.006	6	4	30
164	33C/2	2	169	1	1	6	226	14	2865	Singleton2865	0.016	0.008	8	9	55
165	33D/2	2	8	4	10	17	1	9	2863	Singleton2863	0.023	0.016	15	53	58
166	33D/3	2	8	4	10	17	1	9	2863	Singleton2863	0.012	0.016	15	53	58
167	33F/2	5	35	29	12	9	45	18	7192	100	0.012	0.008	ND	ND	ND
168	34/2	13	8	1	1	13	1	18	7218	Singleton7218	Unknown	Unknown	ND	ND	ND
169	34/3	7	16	19	16	6	19	14	7202	Singleton7202	0.012	0.016	40	48	6
170	34/4	45	15	6	12	9	14	14	7235	1884	0.008	0.008	35	23	19
171	34/5	13	5	1	12	9	1	18	7216	None1041	0.008	0.008	54	57	1
172	35A/3	29	13	2	16	43	1	79	7242	Singleton7242	0.016	0.012	ND	ND	ND

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
173	35B/2	18	12	1	1	14	77	14	7234	Singleton7234	0.016	0.012	ND	ND	ND
174	35B/3	54	5	1	5	6	12	9	7236	Singleton7236	0.012	0.012	25	53	36
175	35B/4	8	13	4	8	6	22	34	198	198	0.012	0.008	64	31	3
176	35C/2 *	7	2	9	1	10	1	45	7196	113	0.012	0.012	58	35	45
177	35C/3 *	7	2	1	1	10	1	45	5989	113	0.023	0.012	58	35	45
178	35F/2	1	2	4	62	6	1	18	7164	Singleton7164	0.016	0.008	ND	ND	ND
179	36/2	109	12	1	1	6	1	18	7237	Singleton7237	0.012	0.012	ND	ND	ND
180	36/3	7	5	1	1	10	7	15	595	None128	0.008	0.008	53	26	3
181	36/4	18	5	2	18	6	20	9	7231	Singleton7231	0.016	0.016	ND	ND	ND
182	36/5	7	12	53	4	6	1	18	4031	None4031	0.064	0.019	ND	ND	ND
183	37/2	29	33	19	1	37	22	31	1279	447	0.016	0.012	14	34	47
184	37/3	29	33	19	1	36	22	31	447	447	0.016	0.008	14	34	47
185	37/5	29	33	19	1	36	22	31	447	447	0.012	0.008	14	34	47
186	37/6	192	34	19	1	36	22	445	7243	Singleton7243	0.012	0.012	14	34	51
187	37/7	29	33	19	1	37	22	31	1279	447	0.016	0.012	14	34	47
188	37/8	29	33	19	1	36	22	31	447	447	0.012	0.006	14	34	47
189	37/9	29	33	19	1	36	22	31	447	447	0.008	0.008	ND	ND	ND
190	38/2	10	43	41	18	13	49	6	393	393	0.012	0.012	ND	ND	ND
191	39/2	5	7	329	1	13	1	27	7189	Singleton7189	0.012	0.016	16	28	48
192	40/2	2	31	1	1	236	335	6	7185	Singleton7185	0.012	0.012	10	34	52
193	41A/2	226	5	1	1	10	28	120	7238	Singleton7238	0.016	0.016	62	56	51
194	41A/3	226	5	1	1	10	28	120	7238	Singleton7238	0.012	0.012	62	56	51

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
195	41F/2	13	2	69	1	9	1	14	7213	Singleton7213	0.012	0.012	55	32	44
196	42/2	1	1	4	1	18	58	17	433	433	0.016	0.012	53	19	19
197	43/3	7	5	1	10	10	16	6	7197	Singleton7197	0.016	0.012	22	37	3
198	43/4	5	61	95	16	218	443	14	7193	None5990	0.016	0.012	22	18	45
199	44/3	1	194	274	1	14	77	1	7176	Singleton7176	0.016	0.016	57	43	26
200	45/2	5	10	7	190	6	19	1	7191	Singleton7191	0.016	0.012	26	13	26
201	45/3	227	8	4	5	306	19	5	7239	Singleton7239	0.012	0.008	23	29	45
202	45/4	5	10	7	190	6	19	1	7191	Singleton7191	0.012	0.012	26	13	26
203	45/5	5	8	8	1	6	19	497	7190	Singleton7190	0.012	0.016	26	12	55
204	46/2	7	5	4	10	9	3	8	7162	Singleton7162	0.008	0.008	ND	ND	ND
205	47A/1	15	8	8	5	15	1	31	7220	235	0.012	0.008	48	3	51
206	47F/1	13	280	53	17	14	1	31	5991	None5991	0.016	0.012	ND	ND	ND
207	48/2	18	9	1	32	6	1	413	7233	7232	0.016	0.012	41	10	3
208	48/3	18	9	4	32	6	1	413	7232	7232	0.012	0.012	41	10	3
209	48/4	18	9	4	32	6	1	413	7232	7232	0.016	0.023	9	51	3
210	48/5	18	9	4	32	6	1	413	7232	7232	0.016	0.012	41	10	3
211	48/6	18	9	4	32	6	1	413	7232	7232	0.016	0.012	41	10	3
212	Ala243 *	1	5	4	5	5	27	8	13	15	1.000	0.750	92	81	73
213	Ala263 *	7	5	1	8	14	11	14	124	124	0.023	0.016	83	43	74
214	Ala283	7	5	1	8	14	11	14	124	124	0.008	0.008	52	43	49
215	Ala287	1	5	4	5	5	27	8	13	15	2.000	1.000	92	81	73
216	Ala289 *	7	5	1	8	14	11	14	124	124	0.016	0.012	83	43	49

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
217	Ala291	1	5	4	5	5	27	8	13	15	3.000	1.000	92	81	73
218	Ala292 *	7	5	1	8	14	11	14	124	124	0.012	0.012	52	43	49
219	Ala317 *	1	5	4	5	5	27	8	13	15	0.750	0.250	92	81	73
220	GE1	2	22	4	18	27	4	8	367	367/4247	0.250	0.064	124	22	93
221	GE2	5	5	4	1	9	1	9	369	2557	0.120	0.015	125	100	115
222	GE3	7	13	8	6	6	12	8	277	277	0.125	0.047	140	115	93
223	GE4	7	13	8	6	6	12	8	277	277	0.500	0.060	140	107	93
224	GE6	8	29	4	15	17	12	31	1106	1106	0.500	0.250	141	110	111
225	GE7	5	6	1	2	6	3	4	90	90	0.750	0.500	67	79	68
226	GE9	7	11	10	1	86	8	1	1043	162	4.000	0.500	75	77	62
227	GE10	7	11	10	1	86	8	1	1043	162	0.500	0.250	75	77	62
228	ICE7 *	8	9	2	1	6	1	17	191	191	0.023	0.016	50	38	43
229	ICE10	10	20	14	1	6	1	29	218	218	0.016	0.016	33	14	30
230	ICE11 *	2	13	9	1	6	19	14	490	490	0.012	0.008	81	43	55
231	ICE13 *	1	5	4	5	5	1	8	9	15	0.016	0.008	7	52	37
232	ICE22 *	8	9	2	1	6	1	17	191	191	0.012	0.012	50	38	43
233	ICE23 *	10	20	14	1	6	1	29	218	218	0.012	0.012	50	38	30
234	ICE27 *	18	3	4	1	15	1	14	655	422/476	0.008	0.012	21	80	48
235	ICE46 *	7	5	1	8	14	11	14	124	124	0.012	0.012	52	43	49
236	ICE50 *	1	5	4	5	5	27	8	13	15	1.500	1.000	92	81	73
237	ICE57 *	2	8	2	4	6	1	1	66	66	0.016	0.016	34	58	26
238	ICE186 *	7	2	1	1	10	1	21	113	113	0.008	0.006	58	35	51

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
239	ICE211 *	4	4	2	4	4	1	1	81	81	0.750	0.380	67	72	62
240	ICE269 *	7	2	1	1	10	1	21	113	113	0.006	0.003	58	35	51
241	ICE501 *	7	2	1	1	10	1	21	113	113	0.006	0.004	58	35	51
242	ICE570 *	1	5	4	5	5	1	8	9	15	0.016	0.012	7	52	37
243	ICE587	7	5	1	8	14	11	14	124	124	0.012	0.012	52	43	49
244	ICE594 *	7	5	1	8	14	11	14	124	124	0.008	0.008	52	43	49
245	PMEN1 †	4	4	2	4	4	1	1	81	81	1.000	1.500	67	72	62
246	PMEN2 *	5	6	1	2	6	3	4	90	90	0.500	0.750	80	79	72
247	PMEN3 *†	7	11	10	1	6	8	1	156	162	1.500	0.750	67	72	62
248	PMEN4 *	1	8	6	2	6	4	6	37	439	0.125	32.000	87	81	2
249	PMEN5 *	1	5	4	11	9	3	16	18	18	1.500	1.000	75	77	70
250	PMEN6	7	13	42	6	10	6	56	268	176	1.000	0.750	84	89	75
251	PMEN7	2	13	8	25	25	6	8	75	Singleton75	0.190	0.094	91	84	93
252	PMEN8	7	22	1	2	5	1	14	185	185	0.190	0.125	90	83	88
253	PMEN9 *	1	5	4	5	5	1	8	9	15	0.016	0.016	7	52	37
254	PMEN10 *	1	5	4	1	5	3	3	20	15	8.000	1.000	74	U	69
255	PMEN11	7	5	4	26	10	3	55	175	Singleton175	6.000	0.500	74	U	71
256	PMEN12 *	5	6	1	2	6	1	28	238	90	4.000	0.750	88	86	94
257	PMEN13 *	1	8	9	5	11	1	12	41	None41	1.500	0.500	85	82	90
258	PMEN14 *	15	16	19	15	6	20	26	236	271/320	2.000	0.750	86	87	86
259	PMEN15 *	15	29	4	21	30	1	14	242	242	0.750	0.750	89	88	81
260	PMEN16	7	13	8	1	10	6	36	173	176	8.000	4.000	93	90	89

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
261	PMEN17 *	7	6	9	2	6	1	67	384	384	1.500	1.000	99	101	73
262	PMEN18 *	2	8	7	4	6	1	1	67	66	4.000	12.000	102	103	87
263	PMEN19	16	12	9	1	41	33	33	289	289	0.003	0.016	82	43	66
264	PMEN20	20	28	1	1	15	14	14	315	315	0.120	0.032	106	85	91
265	PMEN21	7	14	4	12	1	1	14	177	177	0.032	0.032	94	43	21
266	PMEN22 *	5	6	1	2	6	1	14	273	90	0.023	0.047	14	97	55
267	PMEN23	6	11	1	1	15	72	77	376	2090	1.000	0.750	77	78	76
268	PMEN24	18	12	4	44	14	77	93	377	558	1.000	0.750	107	91	96
269	PMEN25	2	5	36	12	17	21	14	63	63	0.120	0.120	108	92	82
270	PMEN26	7	13	8	6	1	6	8	338	176	0.120	0.094	31	13	82
271	PMEN27	10	18	4	1	7	19	9	217	217	0.006	0.016	26	76	67
272	PMEN28	12	8	13	5	16	4	20	306	306	0.006	0.016	42	59	17
273	PMEN29	10	31	4	1	6	4	94	615	615	0.012	0.016	26	18	34
274	PMEN30	8	10	2	16	1	26	1	193	193	0.016	0.094	98	24	26
275	PMEN31	7	15	2	10	6	1	22	180	180	0.023	0.023	71	38	65
276	PMEN32	12	19	2	17	6	22	14	230	230	1.000	0.750	100	99	85
277	PMEN33	2	5	1	11	16	3	14	53	62	0.016	0.023	42	4	84
278	PMEN34 *	10	20	14	1	6	1	29	218	218	0.016	0.016	33	14	30
279	PMEN35 *	7	5	1	8	14	11	14	124	124	0.016	0.023	52	43	49
280	PMEN36 *	7	2	1	1	10	1	21	113	113	0.023	0.023	58	35	51
281	PMEN37	8	13	14	4	17	4	14	199	199	0.012	0.023	35	23	35
282	PMEN38	10	5	4	5	13	10	18	205	205	0.012	0.016	21	18	1

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
283	PMEN39 *	8	9	2	1	6	1	17	191	191	0.012	0.016	50	38	43
284	PMEN40	13	8	13	5	17	4	8	304	304	0.012	0.012	100	59	51
285	PMEN41	1	5	7	12	10	1	14	327	395	0.012	0.023	105	98	31
286	PMEN42	8	37	9	29	2	12	53	344	344	0.094	0.125	103	U	95
287	PMEN43	8	5	2	27	2	11	71	448	448	0.094	0.094	95	U	83
288	PN1	7	13	8	6	6	6	494	7151	277	1.000	0.060	144	117	114
289	PN2	1	13	4	5	6	214	495	7152	Singleton7152	1.000	0.060	24	U	117
290	SA1	1	8	9	5	11	1	12	41	None41	4.000	1.000	85	82	99
291	SA2	1	8	9	5	11	1	12	41	None41	2.000	1.000	85	82	103
292	SA3	1	8	9	5	11	1	12	41	None41	4.000	2.000	85	82	99
293	SA4	1	8	9	5	11	1	12	41	None41	4.000	0.500	85	82	99
294	SA5	1	8	9	5	11	1	12	41	None41	2.000	0.500	85	82	99
295	SA6	1	8	9	5	11	1	12	41	None41	2.000	0.500	85	82	90
296	SA7	6	62	5	30	88	8	6	1094	1094	>4.000	2.000	85	82	100
297	SA8	1	8	9	5	11	1	165	1605	None41	4.000	2.000	85	82	100
298	SA9	1	8	9	5	11	1	165	1605	None41	4.000	0.500	85	82	99
299	SA10	2	8	9	5	11	1	13	1656	None68	4.000	2.000	115	82	110
300	SA11	6	62	5	30	88	8	6	1094	1094	4.000	2.000	109	82	103
301	SA12	6	62	5	30	88	8	6	1094	1094	4.000	2.000	109	82	103
302	SA13	6	62	5	30	88	8	6	1094	1094	2.000	2.000	109	82	103
303	SA14	6	62	5	30	88	8	6	1094	1094	4.000	2.000	109	82	103
304	SA15	6	62	5	30	88	8	6	1094	1094	2.000	1.000	110	104	102

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
305	SA16	6	62	5	30	88	8	6	1094	1094	4.000	2.000	109	82	103
306	SA17	6	62	5	30	88	8	6	1094	1094	4.000	1.000	110	104	102
307	SA18	6	62	5	30	88	8	6	1094	1094	2.000	1.000	110	104	102
308	SA19	6	62	5	30	88	8	6	1094	1094	4.000	2.000	109	82	103
309	SA20	6	62	5	30	88	8	6	1094	1094	4.000	2.000	109	82	103
310	SA21	6	62	5	30	88	149	6	1607	1094	4.000	1.000	U	82	101
311	SA22	7	5	1	8	14	11	14	124	124	0.250	0.250	112	43	98
312	SA23	7	13	8	6	25	6	8	172	176	0.500	0.125	111	84	93
313	SA24	10	11	4	1	6	149	5	1608	None1608	0.250	0.250	114	116	107
314	SA25	10	11	4	1	6	149	5	1608	None1608	0.500	0.125	113	106	102
315	SA26	10	99	4	1	6	149	5	1610	None1608	0.500	0.250	113	106	102
316	SA27	13	5	4	5	5	3	8	5174	15	0.023	0.016	7	52	37
317	SA28	13	5	4	5	5	3	8	5174	15	0.012	0.012	7	52	37
318	SA31	6	62	5	30	88	8	6	1094	1094	2.000	1.500	110	104	102
319	SA32	6	62	5	30	88	8	6	1094	1094	4.000	0.750	109	82	103
320	SA34	18	5	2	5	27	19	9	5178	None5177	0.015	0.015	127	111	36
321	SA35	18	5	2	5	27	19	9	5178	None5177	0.030	0.080	128	111	36
322	SA36	18	5	2	5	27	19	9	5178	None5177	0.250	0.047	127	111	36
323	SA37	12	62	9	10	3	20	6	5179	Singleton5179	0.120	0.030	123	113	113
324	SA38	6	62	5	30	88	8	6	1094	1094	16.000	0.750	129	82	103
325	SA39	1	8	9	5	11	1	12	41	None41	0.750	0.250	85	82	90
326	SA40	1	8	9	5	11	1	12	41	None41	1.000	0.250	85	106	102

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
327	SA41	6	62	5	30	88	8	6	1094	1094	0.500	0.120	143	106	102
328	SA42	15	16	19	5	6	20	26	763	271/320	0.008	0.008	126	112	34
329	SA43	1	8	9	5	11	1	12	41	None41	0.250	2.000	85	82	90
330	SA44	6	62	5	30	88	8	6	1094	1094	2.000	2.000	109	82	103
331	SA45	6	62	5	30	88	8	6	1094	1094	4.000	5.000	109	82	103
332	SA47	1	8	9	5	11	1	165	1605	None41	2.000	0.500	85	82	100
333	SA48	2	8	9	5	11	1	13	1656	None68	4.000	1.000	115	U	110
334	SA49	58	13	54	1	241	1	18	7255	Singleton7255	0.023	0.008	133	U	3
335	SA51	7	22	1	2	5	1	14	185	185	0.125	0.030	90	U	88
336	SA52	2	62	1	25	6	20	5	2421	2421/2909	0.250	0.125	131	U	127
337	SA53	1	6	29	10	6	79	18	7247	1010	0.250	0.250	138	U	125
338	SA54	2	5	1	2	292	220	27	7084	None5431	0.125	0.125	139	97	129
339	SA55	2	13	8	25	25	6	8	75	Singleton75	0.250	0.125	91	84	93
340	SA56	6	62	5	30	88	8	6	1094	1094	2.000	1.000	109	82	103
341	SA57	7	22	1	2	5	1	14	185	185	0.125	0.060	136	97	123
342	SA58	2	8	111	5	17	1	6	7249	Singleton7249	0.125	0.060	113	53	120
343	SA59	6	62	1	30	88	8	6	7251	1094	0.250	0.060	134	106	102
344	SA60	7	13	8	6	25	6	8	172	176	0.125	0.060	111	84	93
345	SA61	7	5	8	6	25	6	8	5183	176	0.125	0.030	111	84	U
346	SA62	6	62	5	30	88	8	6	1094	1094	1.000	0.500	142	82	130
347	SA63	7	5	1	8	14	11	14	124	124	0.125	0.060	112	43	98
348	SA64	54	8	4	5	9	1	18	7254	Singleton7254	0.125	0.030	137	118	119

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
349	SA65	8	13	331	10	2	4	496	7252	Singleton7252	0.250	0.125	135	U	128
350	SA66	7	22	1	2	5	1	14	185	185	0.125	0.030	90	83	88
351	SA67	2	62	1	2	6	14	5	7250	Singleton7250	0.125	0.030	138	U	118
352	SA68	228	62	5	30	88	8	6	7257	1094	0.500	0.250	109	82	103
353	SA69	1	8	9	5	11	149	6	7248	Singleton7248	0.500	0.250	85	U	100
354	SA70	7	5	8	6	25	6	8	5183	176	0.250	0.060	111	U	93
355	SA71	4	4	2	4	4	8	1	6215	81	0.250	0.250	67	72	62
356	SA72	6	62	5	30	88	8	6	1094	1094	1.000	0.250	109	82	121
357	SA73	1	8	9	5	11	1	12	41	None41	4.000	1.000	U	82	90
358	SA74	4	4	2	4	4	8	1	6215	81	1.000	0.250	67	72	62
359	SA75	4	4	2	4	4	8	1	6215	81	0.500	0.250	67	72	62
360	SA76	1	5	4	30	27	20	1	7246	Singleton7246	0.125	0.030	U	U	124
361	SA77	4	4	2	4	4	8	1	6215	81	1.000	0.500	67	72	62
362	SA78	11	5	4	8	14	11	499	7253	Singleton7253	0.250	0.125	132	82	126
363	SA79	7	13	8	6	25	6	8	172	176	0.125	0.060	111	84	93
364	SA80	4	4	2	4	4	1	1	81	81	0.500	0.250	67	72	62
365	SA81	4	4	2	4	4	8	1	6215	81	0.500	0.250	67	72	62
366	SA82	7	5	4	8	14	11	U	U	124	0.250	0.030	113	U	122
367	SA83	2	5	1	11	16	3	14	53	62	0.047	0.023	42	4	55
368	SA84	13	5	4	5	5	3	8	5174	15	0.060	0.030	7	52	37
369	SN1	5	6	1	2	6	3	4	90	90	0.250	2.000	117	79	68
370	SN2	7	11	10	1	6	8	1	156	162	1.500	0.250	67	72	62

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
371	SN3	7	11	10	1	6	8	1	156	162	0.750	0.250	67	72	62
372	SN4	7	11	10	1	6	8	1	156	162	2.000	2.000	67	77	62
373	SN5	7	11	10	1	6	8	1	156	162	2.000	0.500	67	72	62
374	SN6	5	6	9	2	6	3	4	3999	90	1.500	1.000	80	79	106
375	SN7	5	6	1	2	6	442	4	7161	90	1.000	0.250	121	79	68
376	SN8	7	11	10	1	6	8	1	156	162	2.000	0.380	67	72	62
377	SN9	7	11	10	1	6	8	1	156	162	1.000	0.060	67	72	62
378	USA1	7	2	1	1	10	1	21	113	113	<=0.03	<=0.06	58	35	92
379	USA2 *	7	2	1	1	10	1	21	113	113	<=0.03	<=0.06	58	35	80
380	USA3	7	2	1	1	10	1	21	113	113	<=0.03	<=0.06	58	35	92
381	USA4	7	5	1	8	14	11	14	124	124	<=0.03	<=0.06	83	43	49
382	USA5	7	5	1	8	14	11	14	124	124	<=0.03	<=0.06	52	43	49
383	USA6 *	7	5	1	8	14	11	14	124	124	<=0.03	0.06	96	43	49
384	USA7	7	5	1	8	14	11	14	124	124	<=0.03	<=0.06	83	43	49
385	USA8 *	2	8	2	4	6	492	1	8119	66	<=0.03	<=0.06	34	94	26
386	USA9 *	2	8	2	4	6	1	1	66	66	0.094	0.094	72	75	78
387	USA10	2	8	2	4	6	1	1	66	66	0.125	0.125	72	75	78
388	USA11 *	2	8	2	4	6	1	1	66	66	0.064	0.064	97	95	26
389	USA12 *	2	8	2	4	6	1	1	66	66	<=0.03	<=0.06	34	58	26
390	USA13 *	2	8	2	4	6	1	1	66	66	<=0.03	<=0.06	34	58	26
391	USA14	8	9	2	1	6	1	17	191	191	<=0.03	<=0.06	50	38	43
392	USA15	8	9	2	1	6	1	17	191	191	<=0.03	<=0.06	50	38	43

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
393	USA16 *	8	9	2	1	6	1	17	191	191	0.004	0.008	50	38	43
394	USA17	8	9	2	1	6	1	17	191	191	0.016	0.016	50	38	43
395	USA18 *	10	20	14	1	6	1	29	218	218	<=0.03	<=0.06	76	14	30
396	USA19	10	20	14	1	6	1	29	218	218	<=0.06	<=0.06	50	38	30
397	USA20 *	10	20	14	1	6	1	29	218	218	<=0.03	<=0.06	50	38	30
398	USA21	2	13	9	1	6	19	14	490	490	<=0.03	<=0.06	81	96	55
399	USA22 *	2	13	9	1	6	19	14	490	490	0.500	0.500	77	93	79
400	USA23	1	8	1	2	6	4	6	33	439	2.000	2.000	67	105	105
401	USA24	4	4	2	4	4	1	1	81	81	4.000	2.000	67	79	62
402	USA25	1	5	4	11	9	43	1	7245	None22	2.000	1.000	119	77	105
403	USA26	4	4	2	4	4	1	1	81	81	4.000	2.000	67	77	62
404	USA27	1	5	9	2	6	4	14	1906	439	<=0.03	<=0.06	29	13	49
405	USA28	4	4	2	4	4	1	1	81	81	2.000	0.500	67	77	62
406	USA29	1	5	4	11	9	3	16	18	18	2.000	1.000	75	77	70
407	USA30	4	4	2	4	4	1	1	81	81	2.000	2.000	67	77	62
408	USA31	61	53	122	16	6	38	14	7256	Singleton7256	1.000	2.000	118	100	108
409	USA32	4	4	2	4	4	1	1	81	81	2.000	1.000	67	72	62
410	USA33	5	6	1	2	6	1	28	238	90	1.000	0.250	120	86	109
411	USA34	1	8	9	5	11	1	12	41	None41	8.000	1.000	85	82	90
412	USA35	4	4	2	4	4	1	1	81	81	4.000	2.000	67	109	62
413	USA36	4	4	2	4	4	1	1	81	81	2.000	2.000	67	77	62
414	USA37	4	4	2	4	4	1	1	81	81	4.000	2.000	67	72	62

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
415	USA38	4	4	2	4	4	1	1	81	81	4.000	1.000	67	72	70
416	USA39	5	6	1	2	6	3	4	90	90	2.000	2.000	80	79	68
417	USA40	4	4	2	4	4	1	1	81	81	2.000	2.000	67	72	62
418	USA41	4	4	2	4	4	1	1	81	81	4.000	2.000	67	72	62
419	USA42	4	4	2	4	4	1	1	81	81	4.000	2.000	67	72	62
420	USA43	4	4	2	4	4	1	1	81	81	4.000	4.000	122	72	104
421	USA44	4	4	2	4	4	1	1	81	81	2.000	2.000	67	72	62
422	USA45	4	4	2	4	4	1	1	81	81	2.000	1.000	67	72	62
423	USA46	4	4	2	4	4	1	1	81	81	4.000	2.000	67	72	62
424	USA47	4	4	2	4	4	1	1	81	81	2.000	1.000	67	72	62
425	USA48	1	8	6	2	6	4	6	37	439	1.000	8.000	116	81	2
426	USA49	4	4	2	4	4	1	1	81	81	4.000	1.000	67	72	62
427	670 †	5	6	1	2	6	3	4	90	90	Unknown	Unknown	80	79	68
428	70585 †	16	12	9	1	41	33	33	289	289	Unknown	Unknown	12	43	66
429	CDC0288-04 †	10	20	14	1	9	1	29	220	218	Unknown	Unknown	76	14	30
430	CDC1087-00 †	8	9	2	1	6	1	17	191	191	0.030	0.060	50	38	43
431	CDC1873-00 †	6	11	1	1	15	72	77	376	2090	4.000	1.000	77	78	76
432	CDC3059-06 †	8	13	14	4	17	4	14	199	199	0.120	0.060	78	22	77
433	CGSP14 †	1	5	4	5	5	3	8	15	15	2.000	Unknown	67	72	62
434	D39 †	7	5	1	1	10	7	15	595	None128	Unknown	Unknown	73	26	3
435	JJA †	2	8	2	4	6	1	1	66	66	I	Unknown	72	75	78
436	MLV-016 †	2	5	29	12	16	3	14	62	62	Unknownn	Unknown	79	38	55

#	Isolate	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*	ST	CC	Penicillin MIC	Cefotaxime MIC	*pbp2x*	*pbp1a*	*pbp2b*
437	P1031 †	10	5	4	1	7	19	9	303	217	S	Unknown	26	76	67
438	R6 †	7	5	1	1	10	7	15	595	None128	Unknown	Unknown	53	74	3
439	SPnINV104 †	12	5	13	5	17	4	20	227	306	S	S	4	59	64
440	SPnINV200 †	1	5	4	5	5	1	8	9	15	S	S	7	52	37
441	SPnOXC141 †	7	15	2	10	6	1	22	180	180	0.012	S	71	38	65
442	TIGR4 †	10	5	4	5	13	10	18	205	205	Unknown	Unknown	21	18	1

* Isolate selected for whole-genome sequencing. † Genome retrieved from Genbank. ND = Not Done. U = Unobtainable. MIC = Minimum inhibitory concentration (µg/ml). MICs for some isolates for which genome sequences where retrieved from Genbank were unknown. In some cases penicillin/cefotaxime susceptibility was indicated as S = susceptible, I = intermediate-resistant. ST = Sequence Type as defined by multilocus sequence typing. CC = Clonal Complex.

APPENDIX 2. Unobtainable *pbp* sequences.

As described in Chapter 3, some *pbp* nucleotide sequences could not be obtained even after repeated PCR and sequencing attempts using multiple primer combinations and/or PCR conditions. Each of these "difficult" *pbp* sequences is discussed in further detail below. For reference, Figures A3.1 – A3.3 show the relative binding site positions for all *pbp* primers. Nucleotide sequences for primers which were not listed in Table 2.3, are shown in Table A3.1.

Table A3.1 Primers for attempted PCR and sequencing of difficult *pbp* sequences.

Gene	Primer Name	Primer Sequence (Forward Primers (F) 5'-3', Reverse Primers (R) 3'-5')
pbp2x	Sp55F	GGT CAA ACC TGC CAG GAA CT
	Sp34F	ATC TTG GAG CGG GCT TCA GC
	*pbp2x*F3	AGA GGC AGC AAG CAC AAC TCG
	*pbp2x*F6	ATG ATC CAA ATG ACC AGT CTG TCC G
	*pbp2x*F7	AGC CCT CAA GTC TGG TAC GGC
	*pbp2x*R2	ATT GTT GTC AAT AGA AGA AGC TAA CG
	*pbp2x*R3	ACA TTA CTG GAG AGA GCG AAA CC
pbp1a	Sp41F	AGC CCT ATC TGC TCT CTT TGT AG
	Sp42F	CGG TCA TCA TAT AGG CTG TCG
	Sp42R	CGA CAG CCT ATA TGA TGA CCG
	Sp44R	CGT ATC CTG GGA GCT TTC T
	pbp1aF3	AGT CGA TTC CGA GAC CAT TTA GG
	pbp1aF4	ATA CCA CCA TTA GAA AAG GCG GC
	pbp1aR1	ATG TCG GAA CTC GTG CCA TGA AGG
	pbp1aR4	ACM GGA CGT GGA GCC TAC C
pbp2b	*pbp2b*_R1	ATT TGG CCT TCC AAG ATA GCG TGG

pbp2x for isolate SA21

PCR for the Sp36F+Sp33R and *pbp2x*F1+Sp33R primer pairs failed. PCR was successful for the Sp35F+Sp33R and Sp55F+Sp34R primer pairs. Sequences were obtained using primers Sp55F, Sp35F, Sp34R and Sp33R. The Sp55F sequence could not be

successfully aligned with those from Sp35F, Sp34R and Sp33R. The resulting Sp35F-Sp34R-Sp33R contig was ~511 bp short of the required allele length. Based on the available sequence, two new primers were designed in an attempt to allow sequencing of the missing region. Sequencing reactions using these primers (*pbp2x*R2 and *pbp2x*R3) were attempted using PCR product obtained from the Sp55F+Sp34R primer pair. These reactions were not successful, meaning that the SA21 *pbp2x* sequence was too short and could not be included in the *pbp2x* analyses.

pbp2x for isolate SA73

Repeated PCR attempts with the Sp36F+Sp33R primer pair yielded no, or only weak product from which sequencing reactions subsequently failed. Electrophoresis of the PCR product obtained using the *pbp2x*F1+Sp33R primer pair indicated that mis-priming had occurred. PCRs with the Sp37F+Sp33R and Sp55F+Sp34R primer pairs were successful; however, only primer Sp37F produced a successful sequencing reaction. Analysis of this sequence showed diversity in both the Sp36F and *pbp2x*F1 primer binding regions, explaining the reasons for the failures of these primers. A new primer, *pbp2x*F5, was designed as an alternative to Sp36F and *pbp2x*F1. Sequencing reactions using the *pbp2x*F5, Sp35F, Sp34R and Sp33R primers were attempted using the Sp37F+Sp33R PCR product. Unfortunately, only the sequencing reaction for the Sp33R primer was successful.

pbp2x for isolate SA76

PCR failed for both the Sp36F+Sp33R and *pbp2x*F1+Sp33R primer pairs. PCR product was obtained using both the Sp36F+Sp52R and *pbp2x*F1+Sp52R primer pairs.

Subsequently, sequences were obtained using primers *pbp2x*F1, Sp36F, Sp35F, Sp34F, Sp34R and Sp52R. Alignment of these sequences produced two contigs separated by ~100 bp around the Sp34R primer binding region. The nucleotide sequences of these contigs were used to design two new primers (*pbp2x*F6 and *pbp2x*F7) in an attempt to close the gap. Unfortunately, neither of these primers produced successful sequencing reactions, meaning that the SA76 *pbp2x* sequence was incomplete.

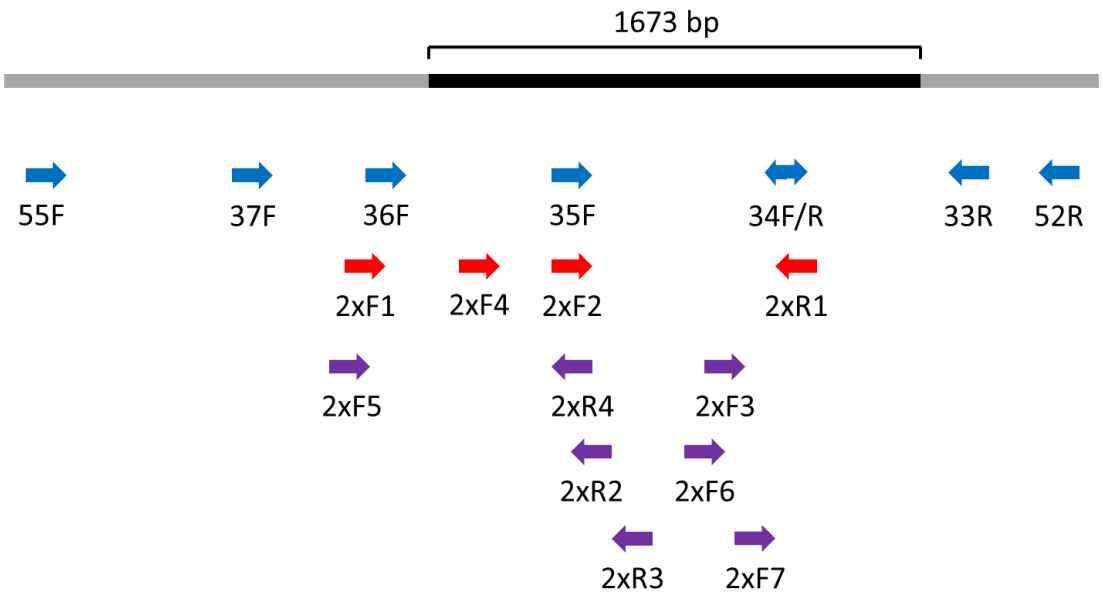


Figure A3.1. *pbp2x* primer binding sites.
Binding site positions for *pbp2x* primers are indicated by arrows relative to the chromosomal region captured within *pbp2x* alleles (black block). Blue arrows represent primers originally designed by ABB. Red arrows represent alternative primers, designed by KLW, which were used for PCR/sequencing when the original primers had failed. Purple arrows represent primers designed by KLW for use with specific isolates when the original and alternative primers had failed. "Sp" and "*pbp*" prefixes for primer names are not shown.

***pbp1a* for isolates PMEN10, PMEN11, PMEN42 and PMEN43**

No PCR product was obtained using any of the following primer combinations: Sp47F+Sp43R, *pbp1a*F1+Sp43R, Sp41F+Sp44R, Sp41F+Sp43R, *pbp1a*F+Sp43R, Sp47F+Sp44R. The PMEN10 isolate was later selected for whole-genome sequencing, from which the *pbp1a* sequence was obtained. This sequence is discussed further here

but, for reasons of consistency was excluded from the previous analyses described in this thesis.

Examination of the PMEN10 genomic regions flanking the *pbp1a* locus showed that the upstream region was highly divergent from that of the R6 reference strain. This diversity affected the Sp41F, Sp47F and *pbp1a*F1 primer binding regions, and likely prevented these primers from binding during PCR. The PMEN10 genome sequence was also highly divergent from that of R6 throughout the *pbp1a* locus. Interestingly, the resultant peptide sequence was predicted to be 10 amino acids longer than that of R6, with the addition of a 'QNPQVQQPQQ' motif at the C terminus. (Note that the *pbp1a* reading frame is in the reverse orientation.) It is not clear to what extent this change affected the PMEN10 clone, but given that it was a widely disseminated, high-level penicillin-resistant cause of disease, it is unlikely that the effect was detrimental.

pbp1a for isolate PN2

Sp47F+Sp43R and *pbp1a*F1+Sp43R primer PCR reactions failed. Sp41F+*pbp1a*R2 and Sp41F+*pbp1a*R3 primer PCR reactions yielded no, or only weak product from which no sequences could be obtained. A weak PCR product was obtained for the Sp41F+Sp43R primer pair, but Sp41F and Sp43R sequencing reactions repeatedly failed. The Sp41F+Sp43R primer PCR was repeated with a reduced annealing temperature (52°C) in an attempt to increase the product yield. The product itself was subsequently used as the template for an additional PCR, to further increase yield. Gel electrophoresis of the final product suggested that these changes may have resulted in mis-priming. The

product at the correct gel band size was purified in preparation for sequencing. Unfortunately, sequencing reactions for the Sp41F and Sp43R primers failed.

pbp1a for isolate SA48

PCR for each of the primer pairs Sp47F+Sp43R, *pbp1a*F1+Sp43R, *pbp1a*F1+*pbp1a*R2 and *pbp1a*F1+*pbp1a*R3 yielded no, or only weak product from which no sequences could be obtained. PCR was repeated for the *pbp1a*F1+Sp43R primer pair with a reduced primer annealing temperature (52°C). PCR product was gel-purified but sequencing reactions with the *pbp1a*F1 and Sp43R primers failed.

pbp1a for isolate SA65

PCR for the primer pairs Sp47F+Sp43R, *pbp1a*F1+Sp43R and *pbp1a*F1+*pbp1a*R2 failed or yielded only weak product. PCR for primer pair *pbp1a*F1+*pbp1a*R3 yielded a strong product; however, no successful sequences could be obtained. PCR for primer pairs Sp41F+Sp43R and *pbp1a*F+Sp44R yielded weak products which were subsequently boosted by repeat PCR with reduced primer annealing temperature (52°C). Products were gel-purified before sequencing was attempted using primers Sp41F/*pbp1a*F1 and Sp43R/Sp44R, respectively. Unfortunately, all sequencing reactions failed.

pbp1a for isolates SA49, SA51, SA52, SA53, SA69, SA70 and SA76

Strong PCR products were obtained using multiple primer combinations; however, sequencing reactions for the Sp43R, Sp44R, *pbp1a*R2, *pbp1a*R3 primers, covering the 3' region of the sequence, repeatedly failed. Sequencing reactions also failed for two

additional primers (*pbp1a*F3 and *pbp1a*F4) designed using the available 5' sequences. Thus the 3' region of these sequences could not be obtained.

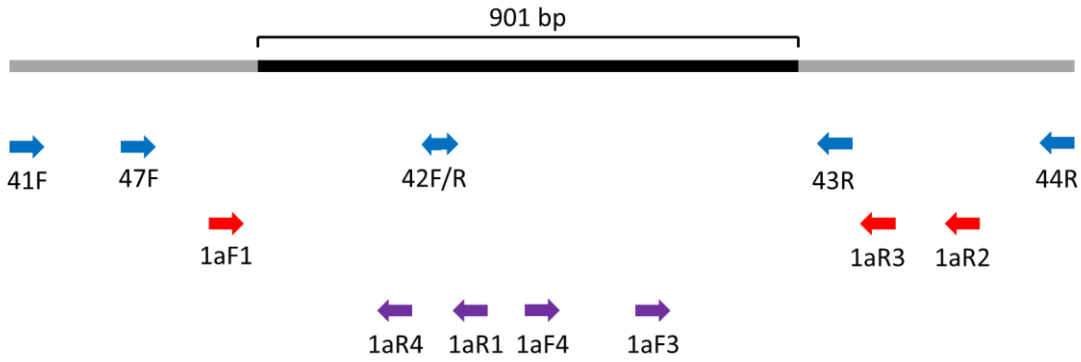


Figure A3.2. *pbp1a* primer binding sites.
Binding site positions for all *pbp1a* primers are indicated by arrows relative to the chromosomal region captured within *pbp1a* alleles (black block). Blue arrows represent primers originally designed by ABB. Red arrows represent alternative primers, designed by KLW, which were used for PCR/sequencing when the original primers had failed. Purple arrows represent primers designed by KLW for use with specific isolates when the original and alternative primers had failed. "Sp" and "*pbp*" prefixes for primer names are not shown.

pbp2b for isolate SA61

PCR for primer pair *pbp2b*_2F+*pbp2b*_4R yielded a strong product from which sequences for primers *pbp2b*_3F, *pbp2b*_3R and *pbp2b*_4R were obtained. Repeated sequencing reaction attempts for primer *pbp2b*_2F yielded a sequence with a "messy" 5' region which was rejected for quality control reasons. The resulting *pbp2b* contig was 256 bp short of the allele length. This sequence was used to design the primer *pbp2b*R1. Sequence reactions using *pbp2b*R1 yielded sequence which, over the same region, also failed to meet quality standards. Thus the 5' region of the SA61 *pbp2b* sequence could not be obtained.

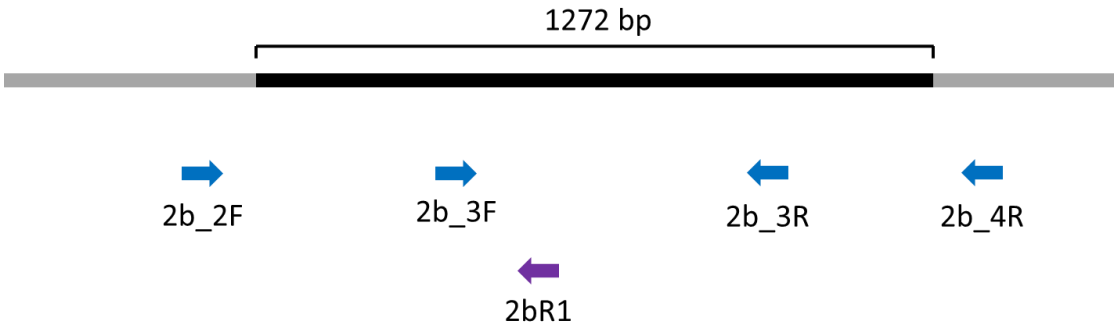


Figure A3.2. *pbp2b* primer binding sites.
Binding site positions for all *pbp2b* primers are indicated by arrows relative to the chromosomal region captured within *pbp2b* alleles (black block). Blue arrows represent primers originally designed by ABB. The purple arrow represents the primer designed by KLW, for use with isolate SA61 following sequencing failure of the *pbp2b*_2F primer. *"pbp"* prefixes for primer names are not shown.

APPENDIX 3. Conference abstracts describing works in this thesis and related works.

Prevalence and sequence diversity of bacteriocin genes *blpMN* in *Streptococcus pneumoniae*

K. LINNEA EK[1], **KELLY L. WYRES**[1], KARL G. KRISTINSSON[2], AND ANGELA B. BRUEGGEMANN[1]

[1]University of Oxford, Oxford, United Kingdom; [2]Landspitali University Hospital and University of Iceland, Reykjavik, Iceland.

Presented at the *7th International Symposium of Pneumococci and Pneumococcal Diseases*, Tel Aviv, Israel, 2010.

Background & Aims: Bacteriocins are small antimicrobial peptides found in many bacterial lineages. The *blp* gene cluster in pneumococci contains genes encoding a number of bacteriocin-like compounds. Previous work suggested that two genes, *blpM* and *blpN*, are necessary for inhibitory activity. The current aims were to: (i) detect the prevalence of *blpMN* within a diverse collection of pneumococci*;* (ii) determine the prevalence of *blpMN* among carriage and invasive strains; (iii) identify any serotype or genotype association with *blpMN;* and (iv) resolve the sequence diversity of *blpMN*.

Methods: Pneumococci recovered from Icelandic children <7 yrs of age (invasive disease and healthy children attending day care centres) from 1993-2006 had previously been characterised by serotype and MLST. 161 strains were selected to obtain a collection of 16 serotypes and 35 genotypes. A PCR assay was designed to amplify *blpMN*; amplicons were visualised by electrophoresis and sequenced.

Results: 134 (83%) strains tested positive for *blpMN*: 84/97 (87%) from carriage and 50/64 (78%) from invasive disease. For all genotypes and serotypes tested, between 50-100% of strains were positive, apart from serotype 12F (0/1 strain tested) and ST218 (0/7 strains tested). A phylogenetic tree was constructed from the *blpMN* sequences, which showed three large families that could further be divided into 8 clusters. Some strains with the same serotypes and/or genotypes had identical *blpMN* sequences, but some possessed *blpMN* genes from different families.

Conclusions: A high proportion of invasive and carriage strains were positive for *blpMN*. Strains may be sharing *blpMNs* through horizontal gene transfer.

Nasopharyngeal carriage of *Streptococcus pneumoniae* in urban and rural Nepal, using silica desiccant packages for storage and transport

J. METZ[1], S. HANIEH[1], R. PRADHAN[2], A. JOSHI[2], D. SHAKYA[2], L. SHRESTHA[2], A. SHRESTHA[2], J. HABENS[2], S. KELLY[1], B.D. MAHRAJAN[2], T. JOHN[1], E. BOHLER[3], **K. WYRES**[4], A. BRUEGGEMANN[4], D.F. KELLY[1], A.M. WERNO[5], S. THORSON[2], N. ADHIKARI[2], D. MURDOCH[5], A.J. POLLARD[1] AND THE OXFORD VACCINE GROUP, UNIVERSITY OF OXFORD

[1]Oxford Vaccine Group, University of Oxford, Oxford, UK; [2]Department of Paediatrics, Patan Hospital, Kathmandu, Nepal; [3]Okhaldunga Community Hospital, Okhaldunga, Nepal; [4]Department of Zoology, University of Oxford, Oxford, UK; [5]Department of Microbiology, University of Otago, Christchurch, New Zealand.

Presented at the *7th International Symposium of Pneumococci and Pneumococcal Diseases*, Tel Aviv, Israel, 2010.

Background & Aims: We have recently demonstrated a significant burden of childhood pneumococcal disease in Kathmandu, the majority of clinical isolates being non-PCV7 serotypes. Most Nepali children live in rural settings where there are limited microbiological culture facilities. We compared nasopharyngeal carriage prevalence and serotype distribution of *Streptococcus pneumoniae* between urban and rural locations using a silica desiccant package (SDP) for transport and storage of pneumococcal isolates and compared this with standard methods.

Methods: Nasopharyngeal swabs were collected from healthy Nepali children aged 6 weeks to 2 years in urban Kathmandu and rural Okhaldhunga (N=1101). Dual swabs were collected in Kathmandu, one swab transported in skim-milk-tryptone-glucose-glycerin (STGG) media and cultured within 8 hours, and the other, stored 14 days in SDPs prior to culture. Single swabs were collected from Okhaldhunga, stored 14 days in SDPs and cultured.

Results: Pneumococci were found in 708 of the 1101 children. In the urban site, pneumococci were isolated from 41% (95% CI 36.7-44.9) of swabs stored in SDP compared to 59% (95% CI 54.4 - 62.6) of those transported in STGG (N=574). In rural areas using the silica dessicant method 69% (95% CI 64.9 - 73) of children were found to be pneumococcal carriers. Serotyping of these isolates is currently underway.

Conclusions: The prevalence of pneumococcal carriage was higher in the rural area than the urban site. Although isolation rates are not identical to those using standard methods high isolation rates were obtained with the SDPs. Serotype distribution data will allow further comparison between these techniques.

Pneumococci through the looking glass: evolutionary history since 1937

K. WYRES[1], L. LAMBERTSEN[2], L. MCGEE[3], L. DE GOUVEIA[4], A. VON GOTTBERG[4], J. LIÑARES[5], M. JACOBS[6], K. KRISTINSSON[7], B. BEALL[3], K. KLUGMAN[8], R. HAKENBECK [9] AND A. BRUEGGEMANN[1]

[1]Univ of Oxford, Oxford, United Kingdom; [2]SSI, Copenhagen, Denmark; [3]CDC, Atlanta, GA; [4]NICD, Gauteng, South Africa; [5]Hosp Univ Bellvitge, Barcelona, Spain; [6]Case Western Reserve Univ, Cleveland, OH; [7]Landspitali Hosp, Reykjavik, Iceland; [8]Emory Univ, Atlanta, GA; [9]Univ Kaiserslautern, Kaiserslautern, Germany.

Presented at the *51st Interscience conference on Antimicrobial Agents and Chemotherapy*, Chicago, USA, 2011.

Background & Aims: Pneumococci face selective pressures from antimicrobial and vaccine use. To better understand how this bacterium will respond we must understand its evolutionary history. Using a unique set of isolates from 1937 - 2007 we studied evolution with respect to serotype switching, clonal stability and horizontal gene transfer of penicillin (pen)-binding proteins (*pbps*).

Methods: Serotype and multilocus sequence typing (MLST) data were collected for 384 isolates and 43 Pneumococcal Molecular Epidemiology Network strains. Isolates were assigned to clonal complexes (CCs) by goeBURST. Regions of the *pbp2x, pbp1a* and *pbp2b* genes were sequenced. Data from 16 public genomes were used in analyses. 29 countries/5 continents were represented.

Results: 1) 25/163 CCs were represented by isolates of >=2 serotypes, suggestive of >=40 independent serotype switching events, >=36 of which preceded pneumococcal conjugate vaccine introduction. 2) The variability of serotypes and *pbps* within a CC was inconsistent: some were stable for decades; some changed frequently. CC191^{7F} (n = 12), the most globally distributed, invasive, serotype 7F CC, was highly stable with no *pbp* changes from 1937 - 2005. Conversely, CC66 (n = 14) was highly variable, with evidence of at least 4 serotype switching events (7B/9N/14/19F/23F) from 1952 - 2005 and several *pbp* changes via mutation (n = 4) and recombination (n = 8). 3) 26 unique *pbp2x* ± serotype ± *pbp1a* combinations were found among >1 CC suggesting possible co-transfer of these regions.

Conclusions: This collection of historical and modern isolates has revealed key aspects of pneumococcal evolution over the past 7 decades, data that influence our understanding of the problem of pen resistance and vaccine-induced changes in the population.

Serotype distribution of nasopharyngeal isolates of *Streptococcus pneumoniae* in rural and urban children in Nepal

M. HAMALUBA[1], S. HANIEH[1], J.A. METZ[1], R. PRADHAN[2], D. SHAKYA[2], L. SHRESTHA[2], A. SHRESTHA[2,] A. JOSHI[2], J. HABENS[2], B. MAHARJAN[2],T. JOHN[1], S. KELLY[1], L. YU[4], **K.L. WYRES**[5], R. FISHER[5], E. BOHLER[6], D. KELLY[1], S. THORSON[2], A.M. WERNO[3,7], A.B. BRUEGGEMANN[5], N. ADHIKARI[2], D.R. MURDOCH[3,7] AND A.J. POLLARD[1]

[1]Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford, United Kingdom; [2]Patan Academy Paediatric Research Unit, Patan Hospital, Kathmandu, Nepal; [3]University of Otago, Christchurch, New Zealand; [4]Centre for Statistics in Medicine, Oxford, UK; [5]Department of Zoology, University of Oxford, Oxford, UK; [6]Okhaldunga Community Hospital, Okhaldunga, Nepal; [7]Canterbury Health Laboratories, Christchurch, New Zealand.

Presented at the *8th International Symposium of Pneumococci and Pneumococcal Diseases*, Iguaça Falls, Brazil, 2012.

Background & Aims: Most developed countries have vaccination programmes based on infant pneumococcal conjugate vaccination, no such vaccine is licensed in Nepal. Here, we characterize the nature of circulating *S. pneumoniae* in Nepal by examining the serotype and genotype distribution of nasopharyngeal isolates.

Methods: Nasopharyngeal swabs were obtained from children aged 6 weeks to 24 months, resident in urban and rural Nepal between January and December 2009. Swabs were transported using two different methods, cultured and genotyped by multilocus sequence typing. Serotyping was by PCR using primers that allowed detection of 69 distinct serotypes. *cpsA* negative samples were considered to be nontypeable *S. pneumoniae.*

Results: 1101 children were enrolled of whom 574 were from an urban setting. *S. pneumoniae* was isolated from 320 of the rural samples (N = 527) and 334 of the urban. Though a small proportion of these isolates are yet to be characterized the following picture is emerging. The serogroups/types most commonly identified, in order of decreasing frequency, were: 6, NT, 15B/C, 14, 34, 23F, 11A, 19F, 35B, 17F, 18, 9V and 23. 33% of the isolates from both settings (216/654) were types contained in PCV10 and 36% (236/654) were PCV13 vaccine-types. Most serogroups were common to both settings except serogroup 18 and serotype 9V (largely found in the rural area) and serotypes 17F, 21,23A in the urban. 51% of the isolates had novel sequence types.

Conclusions: Most *S. pneumoniae* isolates circulating in Nepali children appear to comprise serotypes not included in PCV10 or PCV13.

Molecular epidemiology of pneumococci recovered from the nasopharynx of young children in Nepal

KELLY L. WYRES[1], ROBERTA FISHER[1], MAINGA HAMALUBA[2], SARAH HANIEH[2,3], ANIP JOSHI[3], JOCELYN HABENS[3], RAHUL PRADHAN[3], SARAH KELLY[2], ERIK BOHLER[4], DAVID R. MURDOCH[5,6], DOMINIC KELLY[2], ANDREW J. POLLARD[2], STEPHEN THORSON[3] AND ANGELA B. BRUEGGEMANN[1]

[1]Department of Zoology, University of Oxford, Oxford, UK; [2]Department of Paediatrics, Oxford Vaccine Group, University of Oxford, Oxford, UK; [3]Patan Academy Paediatric Research Unit, Patan Hospital, Kathmandu, Nepal; [4]Okhaldhunga Community Hospital, Okhaldhunga, Nepal; [5]University of Otago, Christchurch, New Zealand; [6]Canterbury Health Laboratories, Christchurch, New Zealand.

Presented at the *30th Annual Meeting of the European Society of Paediatric Infectious Diseases*, Thessaloniki, Greece, 2012.

Background & Aims: Although rates of pneumococcal carriage and disease are highest in developing countries, the pneumococcal molecular epidemiology within these regions is poorly understood. Nasopharyngeal isolates from urban (Kathmandu) and rural (Okhaldhunga) sites were genotyped by multilocus sequence typing (MLST). The main aims were to characterise the overall population structure and identify major genotypic differences between locations.

Methods: 599 pneumococcal isolates, 299 from Kathmandu and 300 from Okhaldhunga, were recovered from healthy children <2 yrs of age and genotyped. Alleles and sequence types (STs) were assigned; 576 isolates with complete STs were analysed. goeBURST was used to define clonal complexes (CCs).

Results: 302 unique STs were detected; 187 STs were found in Kathmandu and 160 in Okhaldhunga. 229 STs were newly recognised and described 330 isolates (57% of the collection). There were 11 major CCs with ≥10 isolates (n; predominant serogroup/type(s)): CC63 (28; 14, 15BC); CC4209 (27; 15BC); CC193 (22; 21, 17F); CC6025 (22; 11A); CC5613 (20; 6, 15A); CC7695 (17; 35B, 18); CC4217 (12; 6); CC5080 (12; 23A); CC1439 (12; 34); CC4881 (10, 9V); and CC176 (10; 23F, 19F). Some CCs predominated in Kathmandu (CC193, 5080, 1439) and others in Okhaldhunga (CC6025, 7695, 4881); the rest were roughly equally distributed between locations.

Conclusions: The majority of genotypes in this study were novel and thus provide important information about the molecular epidemiology of carriage pneumococci in Nepal. These data also provide a baseline from which to measure genetic changes after a pneumococcal conjugate vaccine is implemented.