From Department of Microbiology, Tumor and Cell Biology
Karolinska Institutet, Stockholm, Sweden and Swedish Institute for
Infectious Disease Control, Solna, Sweden

GENETIC CONTENT OF CLINICAL PNEUMOCOCCAL ISOLATES AND ITS RELATION TO DISEASE OUTCOME

Jessica Dagerhamn

Karolinska Institutet

SMITTSKYDDSINSTITUTET

Swedish Institute for Infectious Disease Control

Stockholm 2009

All previously published papers were reproduced with permission from the publisher.

Published by Karolinska Institutet.

© Jessica Dagerhamn, 2009
ISBN 978-91-7409-557-9

Printed by
REPROPRINT AB
Stockholm 2009
www.reproprint.se
Gårdsvägen 4, 169 70 Solna

ABSTRACT

Streptococcus pneumoniae is a Gram-positive bacterium that can cause a wide range of diseases. These include otitis media, sinusitis, pneumonia and meningitis. It can, however, also cause asymptomatic carriage thereby behaving almost like a commensal. It is of great interest to try to determine factors playing a role for if the bacterium will cause carriage or whether it will cause an invasive disease (meaning that the bacterium is sampled from the blood or the meninges). The aim of this thesis is to investigate differences in genetic content in invasive and carriage isolates.

In the first study we set out to see how well a molecular typing method, MLST, and genetic content as determined by microarray correlated. We found that isolates differing in up to two MLST alleles and belonging to the same clonal complex clustered together using microarray data. Hence the method of MLST, measuring differences in 7 house-keeping genes, correlates well with genetic content as determined by microarray. We also suggest an alternative typing method, a PCR consisting of 25 accessory genes, which gives the same discrimination as MLST clonal complexes. This method must, however be investigated further.

In paper II we wanted to see how well isolates differing in more than two alleles, and still belonging to the same clonal complex correlated genetically. We found that both the number of alleles differing and if the isolates compared were of the same CC and / or Serotype affected the number of genes differing. For the three larger CCs investigated we saw that for one of them the number of genes increased with the number of alleles differing. For the other two there was a large increase when the serotype of the isolates compared shifted, indicating the possibility that these CC:s in fact come from more than one ancestor ST.

Paper III investigated the genetic content of isolates of different invasive disease potential to see if any particular accessory region was associated with either carriage or invasiveness. No obvious candidate was found. ARs 6 and 34 were found in most of the invasive isolates and were absent in most of the isolates of low invasive disease potential and were tested for invasiveness in a mouse model of infection. We were not able to see any differences between knock-out mutants of these regions and the wild type. This, in addition to the fact that many genes found to be important for virulence in STM screens were absent, suggest that there is a great genetic redundancy affecting the ability of different isolates to cause invasive disease.

In the fourth study we determined the invasive disease potential of isolates from Stockholm, Sweden 1997-2004. Invasiveness for serotype well matched previous studies with serotype 1, 4, 7F having a high invasive disease potential and 6A, 19F and 23 F having a low invasive disease potential. We did, however, find differences between clonal types of the same serotype, for serotypes 14 and 6B. We further investigated 6B by whole genome microarray and whole genome sequencing and found a number of differences. Among these are two different variants of the important protein PspA and the absence of PcpA, important for disease in the lung, from one of our isolates. We were also able to determine that the four serotype 6B isolates differed in presence/ absence of prophages, a factor that may be of importance for virulence.

LIST OF PUBLICATIONS

I. **J Dagerhamn**, C Blomberg, S Browall, K Sjöström, E Morfeldt, B Henriques-Normark
Determination of accessory gene patterns predicts the same relatedness among strains of Streptococcus pneumoniae as sequencing of housekeeping genes does and represents a novel approach in molecular epidemiology. Journal of clinical microbiology, 2008 Mar;46(3):863-8. Epub 2007 Dec 26.

II. C Blomberg*, **J Dagerhamn***, S Browall, B Henriques-Normark
Pneumococcal clonal complexes show varying degree of genetic similarity
Manuscript

III. C Blomberg, **J Dagerhamn**, S Dahlberg, S Browall, J Fernebro, E Morfeldt, S Normark, B Henriques-Normark
Pattern of accessory regions and invasive disease potential in *Streptococcus pneumoniae.* Journal of Infectious Diseases, 2009, 199 (7), 1032-42

IV. **J Dagerhamn***, S Browall*, K Sjöström, C Blomberg, Å Örtqvist, S Normark, B Henriques-Normark
Differences in invasive disease potential relates to inter- and intraclonal genetic variations in pneumococcal isolates
Manuscript

*= contributed equally

CONTENTS

LIST OF ABBREVIATIONS

AOM	Acute otitis media
CAP	Community-acquired pneumonia
CC	Clonal complex
CSP	Competence stimulating peptide
DLV	Double locus variant
IDP	Invasive disease potential
IL	Interleukin
IPD	Invasive pneumococcal disease
IS	Insertion sequence
LTA	Lipoteichoic acid
Mb	Megabases
MLST	Multilocus sequence typing
Mn	Manganese
OR	Odds Ratio
ORFs	Open reading frames
PBP	Penicillin binding protein
PCR	Polymerase chain reaction
PCV-7	7-valent protein conjugated vaccine
PCV-10	10-valent protein conjugated vaccine
PFGE	Pulsed field gel electrophoresis
PI-1	Pilus islet 1
PI-2	Pilus islet 2
PPV-23	23-valent polysaccharide vaccine
SLV	Single locus variant
SRGA	Swedish reference group for antibiotics
ST	Sequence type
STM	Signature tagged mutagenesis
TA	Teichoic acid
TLV	Triple locus variant
TNF	Tumor necrosis factor
Zn	Zink

1 INTRODUCTION

In 1881 Georg Miller Sternberg and Louis Pasteur independently isolated the bacterium later known as *Streptococcus pneumoniae* for the first time (Pasteur 1881; Sternberg 1881). It is a Gram- positive bacterium that shows α-hemolysis on blood agar plates. Its cell wall is composed of several layers of peptidoglycan with bound teichoic acid (TA) and lipoteichoic acid (LTA) anchored in the cell membrane. The pneumococcal cell wall also contains choline bound to the TA. Pneumococci can grow as single bacteria, in pairs or in chains. They show bile solubility and are usually optochin sensitive. All these criteria plus agglutination with antipneumococcal polysaccharide capsule antibodies are taken into consideration when classifying a bacterium as *Streptococcus pneumoniae* (Dowson 2004). The bacteria are surrounded by a polysaccharide capsule and, depending on this capsule, belong to one of currently 91 described serotypes. It was through the transformation of unencapsulated bacteria to encapsulated that Avery was able to show that DNA is the carrier of genetic material (Avery, MacLeod et al. 1979). *S. pneumoniae* cause a wide range of diseases, such as otitis media, pneumonia with or without septicemia and meningitis, as well as asymptomatic carriage. The disease is regarded as invasive if the bacterium is isolated from a normally sterile site such as the blood or the meninges. Risk factors for pneumococcal infections include age (< 2 years or >65), smoking, alcoholism, underlying diseases such as chronic bronchitis, asthma, chronic obstructive pulmonary disease, diabetes, cancer and immunosuppression including HIV (Musher 2004). A prior viral respiratory tract infection, especially caused by influenza, is also a risk factor for acquiring a pneumococcal infection (Musher 2004).

1.1 COLONISATION

Pneumococcal transmission occurs through direct contact with respiratory secretions or inhalation of aerosols from either asymptomatic carriers or from a person with pneumococcal disease (Antao and Hausdorff 2009). When the pneumococcus enters the nasopharynx it faces several challenges. It has to compete with other bacteria already residing. There is also the ciliary beating that works to push the pneumococci out of the nasopharynx. Also, in order to colonize the bacteria must adhere to the surface. Any proteins working to adhere to the nasopharyngeal epithelial cells must stretch through the capsule. It has, however been shown that the capsule is required for pneumococcal colonization even though less amount of it may be beneficial (Magee and Yother 2001). Pneumococci can change the amount of capsule they produce and are found in two phase variants, opaque and transparent. The transparent variant has less capsule (Weiser, Austrian et al. 1994). Nelson and co-workers suggest a model where the opaque variant enters the nasopharynx. The higher amount of capsule allows for escape from mucus. Thereafter the amount of capsule diminishes, giving rise to a transparent variant for which it is easier to adhere (Nelson, Roche et al. 2007). The opaque variant is also found in the invasive infections.

It has been shown that there is a positive correlation between carriage of *S. pneumoniae*, *M. catarrhalis* and *H. influenzae* (Jacoby, Watson et al. 2007). Also, there is a positive correlation with rhinovirus infection and carriage of these pathogens

(Jacoby, Watson et al. 2007). On the other hand, there is a negative correlation between carriage of *S. pneumoniae* and *S. aureus* (Jacoby, Watson et al. 2007).

1.2 DISEASE AND CARRIAGE

The pneumococcus can give rise to a wide variety of diseases ranging from acute otitis media (AOM) to meningitis. It can also, however, give rise to asymptomatic carriage in the nasopharynx, most commonly in pre-school children attending day-care centers. 44 percent of children under 6 years of age has been found to be colonized by *S. pneumoniae* without signs of infection (Zenni, Cheatham et al. 1995) and for children attending day-care centers the figure is as high as 60 % (Boken, Chartrand et al. 1995; Duchin, Breiman et al. 1995; Boken, Chartrand et al. 1996). A study from England showed that by the age of 6 months 54 % of infants have acquired pneumococci at least once. This increases to 97 % by two years of age (Sleeman, Griffiths et al. 2006). Serotypes most often acquired for carriage were serotypes 6B, 19F, 23F, 6A and 14 (Sleeman, Daniels et al. 2005; Sleeman, Griffiths et al. 2006). Serotypes 15C and 7F had the shortest duration of carriage, approximately 6 weeks, and 6B had the longest, approximately 20 weeks (Sleeman, Griffiths et al. 2006). In the same study by Sleeman et al an inverse relationship between duration of carriage and the attack rate of a specific serotype was found (Sleeman, Griffiths et al. 2006). The attack rate is the number of cases of invasive disease per 100 000 acquisitions. There is a larger genetic diversity among isolates causing carriage than among isolates causing invasive disease (Robinson, Edwards et al. 2001; Brueggemann, Griffiths et al. 2003).

1.2.1 Acute otitis media and sinusitis

Pneumococci may be carried into the eustachian tubes where they are normally rapidly removed, mainly due to ciliary action. If the opening of the eustachian tubes is obstructed due to edema caused by e.g. co-existent viral infection or allergy the bacteria may not be cleared. Otitis media is a common disease. It is estimated that by three years of age approximately 80% of US children has had at least one episode (Teele, Klein et al. 1989). *Streptococcus pneumoniae* accounts for approximately 40% of otitis media cases worldwide with *H. influenzae* accounting for another 40 % (Casey and Pichichero 2004). Apart from genetic predisposition (Rovers, Haggard et al. 2002) risk factors include not being breast-fed as an infant, attending a large day-care center and passive smoking (Rovers, Schilder et al. 2004).

Similar to the mechanism for otitis media the osteomeatal complex may be obstructed due to edema. The fluid accumulated in the paranasal cavitites provides a medium for the bacteria to grow in. Again, in this case pneumococci may not be cleared causing sinusitis.

1.2.2 Pneumonia

Pneumonia is a major cause of morbidity and mortality world-wide, especially in developing countries, and it is estimated that in 2002 1.9 million children under 5 years of age died due to acute respiratory infections (Williams, Gouws et al. 2002). Another study showed that 19% of deaths in children less than 5 years of age are estimated to be

due to pneumonia (Bryce, Boschi-Pinto et al. 2005). *S. pneumoniae* is the leading cause of community-acquired pneumonia (CAP). When pneumococci enter the lungs their presence stimulates an inflammatory response.

The most common complications of bacterial pneumonia are pleural empyema and non-purulent plural effusion (Goldbart, Leibovitz et al. 2009) where pleural empyema accounts for 28% of children hospitalized with CAP (Byington, Spencer et al. 2002). The pneumococcus is now the leading etiological agent causing pleural empyema (Byington, Spencer et al. 2002) . The most common serotype causing empyema is serotype 1 (Byington, Spencer et al. 2002; Tan, Mason et al. 2002; Byington, Korgenski et al. 2006; Fletcher, Leeming et al. 2006; Goldbart, Leibovitz et al. 2009).

1.2.3 Sepsis

If the bacteria enter the bloodstream they cause bacteremia. Bacteremia with clinical signs constitutes sepsis. The bacteremia can be primary, i.e. without focus of infection, or secondary, where the bacteremia is the result of an established infection (Musher 2004).

1.2.4 Meningitis

Meningitis is an inflammation of the meninges of the brain (Weber 2004). Incidence in developed countries is approximately 1/100000 per year, as compared to 12/100000 per year in children less than 5 in Africa (Wenger, Hightower et al. 1990; O'Dempsey, McArdle et al. 1996; Berg, Trollfors et al. 2002; Weber 2004). It has been shown that pneumococcal meningitis is associated with higher case fatality rates and a higher number of patients with neurologic sequelae than either *Haemophilus influenzae* or *Neisseria meningitidis* (Baraff, Lee et al. 1993). Mortality rates are up to 40% and sequelae occur in up to 50% of survivors (Bhatt, Lauretano et al. 1993; de Gans and van de Beek 2002; Weber 2004).

1.2.5 Atypical pneumococcal infections

Streptococcus pneumoniae also has the ability to cause other infections such as endocarditits, primary peritonitis, septic arthritis, osteomyelitis, soft tissue infection and conjunctivitits (Musher 2004). Interestingly non-typeable isolates, i.e. isolates without capsule, are prone to cause conjunctivitis (Medeiros, Neme et al. 1998; Porat, Greenberg et al. 2006; Shouval, Greenberg et al. 2006).

1.2.6 Epidemiology

Since pneumococci are so diverse it is important to monitor which subpopulations of bacteria actually cause disease. After vaccine introduction it is also important to monitor changes.

Generally, in young children serogroups 6, 14 and 19 are the most prominent causes of invasive pneumococcal disease worldwide (Hausdorff, Bryant et al. 2000). In older children and adults the diversity is greater and no single serogroup is common all over the world, however, serotype 14 is among the 3 most common serotypes in every region of the world except Asia (Hausdorff, Bryant et al. 2000). This greater diversity is

also shown by the fact that in most regions of the world 4-5 serogroups account for 50 % of invasive pneumococcal disease (IPD) cases in older children and adults, this is in comparison to 2-3 serotypes in younger children (Hausdorff, Bryant et al. 2000).

Sjöström et al (Sjostrom, Spindler et al. 2006) have shown that serotypes such as 11A and 19F act more as opportunistic pathogens while serotypes such as 1 and 7F with a high potential for causing invasive disease act more as primary pathogens. These serotypes primarily infect younger people and give rise to a not so serious invasive disease. Furthermore, they found that certain serotypes caused a high mortality such as serotypes 3, 6A, 11A and 19F. Alanee et al found that serotype played little role in accounting for disease severity, instead host factors such as old age, underlying chronic disease and immunosuppression played the greatest role in determining disease outcome (Alanee, McGee et al. 2007).

Genetically related bacteria are called clones. Pneumococci can be divided into clones using molecular typing methods such as pulsed-field gel electrophoresis (PFGE) and multi-locus sequence typing (MLST). These are further described under methods. When using these methods you can see that not all isolates of the same serotypes are genetically related. It also shows that isolates related by molecular typing not necessarily share the same serotype. MLST is readily comparable between laboratories and defines the sequence type (ST) of an isolate.

1.2.7 Odds Ratio and attack rate for causing invasive disease

In vaccine considerations the rank order of serotypes has been used to guide which serotypes should be included (Hausdorff, Bryant et al. 2000; Hausdorff, Bryant et al. 2000). Another approach is to try to determine how often a serotype will cause invasive disease as compared to carriage. This has most commonly been done by comparing the Odds Ratios (OR) for causing invasive disease. This is calculated by

$$OR=ad/bc$$

where a is the number invasive serotype/clone X, b is the number of non-invasive serotype/clone X, c is the number of invasive serotype/clone non-X and d is the number of non-invasive serotype/clone non-X (Sandgren, Sjostrom et al. 2004) An OR >1 indicates that the serotype/clone has a high invasive disease potential, i.e. when it encounters a host it is more likely to cause disease than carriage. An OR<1 indicates that the serotype/clone has a low invasive disease potential and it is more likely to cause carriage than disease.

Several studies that use this measure have been published previously (Brueggemann, Griffiths et al. 2003; Sandgren, Sjostrom et al. 2004; Hanage, Kaijalainen et al. 2005). However, the population material included in the studies has differed somewhat. In the study by Brueggemann et al (Brueggemann, Griffiths et al. 2003) both the invasive and the carriage cohort were 0-5 years, the isolates were collected from approximately the same time period and all isolates originated from Oxford. In the study by Hanage et al (Hanage, Kaijalainen et al. 2005) the invasive isolates were collected from children 0-24 months from the whole of Finland while the carriage isolates were only collected

from the city of Tampere. In the study by Sandgren et al (Sandgren, Sjostrom et al. 2004) both the invasive and carrier isolates were collected during one year from the same geographic area, Stockholm. However, both adult and children were included among the invasive cases. No serotype studied in these three studies showed consistently an OR significantly above 1 suggesting that there may be differences depending on age or on geographic location. A meta-analysis performed by Brueggemann et al suggested that OR is stable over time and by geographic location (Brueggemann, Peto et al. 2004). This study however calculated OR in a slightly different way,

$$OR=ad/bc$$

where a is the number of invasive serotype X, b is the number of non-invasive serotype X, c= the number of invasive serotype 14 and d is the number of non-invasive serotype 14. This study determined the serotypes with high OR to be 1, 4 and 7F, intermediate OR 6B, 9V and 19A and low OR for causing invasive disease 3, 19F, 23F and 35B (Brueggemann, Peto et al. 2004)

Another way is to study the attack rate. In this case you compare the number of isolates of a serotype that cause invasive disease to how often that serotype is acquired (Sleeman, Griffiths et al. 2006). This was done by Sleeman and co-workers (Sleeman, Griffiths et al. 2006) and they found that OR for invasiveness overall matched the attack rates of individual serotypes.

1.2.8 The human immune response

The human immune systeme is composed of two parts, the innate and the adaptive immune response. The innate immune system is the first line of defense against infections. It is composed of components such as phagocytic cells, proinflammatory cells, antigen-presenting cells, antimicrobial peptides, complement, cytokines and chemokines leading to an inflammatory response. Inflammation brings more antimicrobial cells and proteins to the site of infection. Cytokines activate innate and adaptive defenses and chemokines attract leukocytes. Phagocytic cells ingest and kill bacteria. Complement enhances phagocytosis through opsonisation and induces inflammation. Antigen-presenting cells initiate the adaptive immune responses (Engleberg, DiRita et al. 2007). The adaptive immune system includes components such as antibodies, T cells and B cells. Antibodies are immunoglobulins that specifically bind to a known ligand. The immunoglobulins are synthesized by B cells. T cells recognize and bind antigens (Engleberg, DiRita et al. 2007).

Both the innate and adaptive human immune responses are important in pneumococcal infections. The major symptoms are due to an inflammatory response. Several pneumococcal virulence factors (see below) interact with the immune system. One example is the capsular polysaccharide (see below) important for evasion of the human immune response through the inhibition of complement mediated opsonophagocytosis (Winkelstein 1984).

1.3 VIRULENCE FACTORS

1.3.1 Capsule

The capsule is the major virulence factor of *Streptococcus pneumoniae*. Isolates that are unencapsulated are in principle considered avirulent. However, clinical isolates may be non-encapsulated. Also, one study has shown in a mouse model of infection that factors other than capsular type seem to be of importance (Kelly, Dillard et al. 1994). In this study they switched capsule of three different isolates of capsular types 2, 5, and 6B to a type 3 capsule. They showed that virulence was neither that of the parental strain, nor was it simply as the type 3 capsule donor strain.

The different polysaccharide capsules differ with respect to both their sugar compositions and linkages, however, all capsules perform the same main function of reducing opsonophagocytosis by limiting access of phagocytic receptors to complement bound to the cell wall. Genetically the capsule region of an isolate is flanked by *dexB* and *aliA*. It contains both a serotype-specific region and a common region where genes contained within the type-specific region are unique to a specific serotype or serogroup (Yother 2004). The type-specific region contains glycosyltransferases, polymerases, transporters and enzymes needed for the synthesis of the serotype-specific sugars. The common region encodes genes for modulations of capsule production and, most often, the glycosyltransferase needed to initiate capsule synthesis (Yother 2004).

Pneumococci have been shown to switch their capsular types, and insertion sequence (IS) elements are common in the capsular loci suggesting that transposition has played a role in evolution. Also, capsular switches include DNA replacements from 15 up to more than 25 kb (Coffey, Enright et al. 1998; Coffey, Daniels et al. 1999). The median size of DNA taken up by cells through transformation is 7kb (Yother 2004) suggesting that the large genetic transfer in capsular switches is a rare event.

The amount of capsule that is expressed is believed to vary. It is suggested that it would be important for pneumococci to reduce capsule expression in order to expose adhesins necessary for colonization on their surface. During systemic infection, however the antiphagocytic properties of the capsule would make high-level capsule expression desirable. Pneumococci do in fact exist in two different phase variants, transparent, more common in carriage, and opaque, more common in systemic infection. Although the capsular type is important for virulence, the genetic background in which it is expressed also plays a major role (Yother 2004).

1.3.2 Proteins

Several proteins have been suggested as being important for virulence. Signature-tagged mutagenesis (STM) is one way of detecting these proteins. Three such studies have been performed in *S. pneumoniae* identifying approximately 300 proteins (Polissi, Pontiggia et al. 1998; Lau, Haataja et al. 2001; Hava and Camilli 2002). Several of these are of unknown function but may well be for example adhesins.

1.3.2.1 Pili

Gram-positive pili have only recently been identified. Two types of pili exist in pneumococci. Pilus islet 1 (PI-1) consists of three structural proteins, three sortases and one regulator. PI-1 has been shown to be important for colonization, virulence and the inflammatory response in mice (Barocchi, Ries et al. 2006). It has also been suggested to play a role in the successful transmission of the penicillin non-susceptible clone of ST156 which is present worldwide (Sjostrom, Blomberg et al. 2007). The second pilus islet, pilus islet 2 (PI-2) mediates adhesion to epithelial cells (Bagnoli, Moschioni et al. 2008). Usually these two pilus islet are not found together in the same isolate.

1.3.2.2 PcpA

PcpA is a 79 kDa large protein that is transcribed as a monocistronic transcript of approximately 2.3 kb (Sanchez-Beato, Lopez et al. 1998). It is a choline-binding protein that at its N-terminal has two tandem arrays of five leucin-rich repeats (Sanchez-Beato, Lopez et al. 1998). PcpA is under control of the Mn2+ controlled regulator PsaR (Johnston, Briles et al. 2006) and is only expressed under low manganese conditions such as are found in the lungs and the blood (Glover, Hollingshead et al. 2008). In the human nasopharynx manganese is relatively accessible and PcpA is not expressed there. Glover and co-workers therefore propose that PcpA is a good vaccine candidate that elicits protection against infection and not against carriage. Thereby no ecological niche in the nasopharynx will be left empty for other pathogens to take over (Glover, Hollingshead et al. 2008).

1.3.2.3 LytA

Lyt A is a lytic amidase that is the major autolysin in pneumococci. It cleaves N-acetylmuramoyl-L-alanin leading to cell lysis (Tomasz, Moreillon et al. 1988). Pneumococci undergo autolysis when they e.g. reach stationary phase. A *lytA*- knock-out mutant had significantly lower virulence than wild-type D39 (Berry and Paton 2000).

1.3.2.4 PspA

Purified PspA is able of binding human lactoferrin and competitively inhibit its binding to pneumococci and one proposed function for this protein is iron acquisition at the mucosal surface (Hammerschmidt, Bethe et al. 1999). The N-terminal domain is essential for full pneumococcal virulence (Jedrzejas, Hollingshead et al. 2000). PspA also reduces complement-mediated clearance and phagocytosis of pneumococci (Jedrzejas 2001). It is suggested that the negatively charged pspA would simply repel complement molecules and prevent their interaction with pneumococci (Jedrzejas 2006). Also, increase in bacterial surface charge has been correlated to the decrease in antibacterial phagocytic activity (Jedrzejas 2006). Finally, PspA interact with human IgG antibodies (Jedrzejas 2006). A *pspA*- knockout mutant was attenuated for virulence compared to wild-type D39 after i.p. challenge, both by survival time and survival rate (Berry and Paton 2000). PspA appears to act as a bridging element between choline of TA or LTA and human cytokine-activated cell glycoconjugates (Jedrzejas 2001).

1.3.2.5 PspC

PspC knock-out mutants showed a more than 50 % reduction in its adherence to known pneumococcal ligands in eukaryotic cells (Rosenow, Ryan et al. 1997). The same mutants also showed a 100-fold reduction in carriage in an animal model (Rosenow, Ryan et al. 1997). Transparent pneumococci expressed increased amounts of pspC (Rosenow, Ryan et al. 1997). PspC is capable of binding specifically to human secretory IgA (Hammerschmidt, Talay et al. 1997), which is the most important immunoglobulin in human secretions (Heremans 1974) and represents the major adaptive mucosal defence factor against infectious agents (Underdown and Schiff 1986; Childers, Bruce et al. 1989).

PspC has a mosaic structure, which it has acquired from both inter- and intraspecies genetic transfer (Brooks-Walter, Briles et al. 1999), and there is a close relationship between PspA and PspC (Brooks-Walter, Briles et al. 1999). Recently it was shown that certain alleles of PspC were able to bind the complement inhibitor C4b-binding protein (Dieudonne-Vatran, Krentz et al. 2009). This protein still retains its function and down-regulates the activation of the classical pathway. Hence it is possible for the bacteria to escape eradication by the complement system. It has also been shown that pspC binds C3, the third component of complement (Cheng, Finkel et al. 2000). C3 is of central importance in host defense against bacterial infections as the site of convergence of the classical, alternative and lectin-mediated pathways of complement activation and the key opsonin in the complement cascade (Madsen, Lebenthal et al. 2000). It has been shown that epithelially produced C3 bound to PspC and *pspC* knockout mutants failed to bind C3 (Smith and Hostetter 2000). PspC has also been found to bind factor H (Dave, Brooks-Walter et al. 2001). PspC is important for maintaining pneumococci in the ecological niche for nasal carriage and is necessary for optimal carriage in an adult mouse model (Balachandran, Brooks-Walter et al. 2002). PspC has also been identifed as the bacterial adhesin of the polymeric IgA receptor and it has been shown that binding of PspC to human pIgR and SigA is mediated through SC, the secretory component (Elm, Rohde et al. 2004).

1.3.2.6 Pneumolysin

Pneumolysin is the major pneumococcal cytotoxin. It is conserved among clinical isolates of *S. pneumoniae* although some clones contain a non-hemolytic version (Jefferies, Johnston et al. 2007). It acts by binding to host cell membrane cholesterol, insertion into the targeted membrane and finally formation of pores which leads to host cell lysis (Jedrzejas 2001). Pneumolysin is localized to the cytoplasm and was believed to be dependent on LytA for its release (Jedrzejas 2001).

A pneumolysin-deficient mutant pneumococcal strain showed decreased adherence *in vitro* but not *in vivo* and Rubins and co-workers conclude that the production of pneumolysin does not play a major role in pneumococcal nasopharyngeal colonization (Rubins, Paddock et al. 1998). However, another group was able to show that virulence was attenuated in a pneumolysin-deficient mutant of the laboratory strain D39 (Berry and Paton 2000).

Pneumolysin can also activate the classical complement pathway thereby depleting complement locally (Paton 1996). It can also initiate nitric oxide production from macrophages contributing to tissue damage (Braun, Novak et al. 1999). Pneumolysin is extremely pro-inflammatory and has been shown to affect production of e.g. TNF-α, IL-1β and IL-6 (Mitchell 2004). During the first few hours of bacteremia in a mouse model it is thought that pneumolysin plays a role in preventing inflammation-based immunity, thereby allowing un-checked growth of pneumococci (Benton, Everson et al. 1995).

1.3.2.7 PsaA

PsaA is present and conserved across all serotypes of pneumococci (Romero-Steiner, Pilishvili et al. 2003). It was suggested as a pneumococcal adhesin based on sequence similarities to other streptococcal adhesins (Sampson, O'Connor et al. 1994). However, it has been shown that PsaA is in fact not an adhesin; it has no possibility of protruding outside the cell wall. Its effect on adhesion is probably due to secondary effects modulated by presence or absence of Mn or Zn (Jedrzejas 2001). It is transcribed as part of a polycistronic mRNA including a total of three open reading frames (ORFs) (Sampson, O'Connor et al. 1994).

It has been shown that adherence of a *psaA* knockout mutant to A549 cells was only 9 % of that of the wild type strain, in this case D39 (Berry and Paton 2000). The *psaA* knockout mutant was also less virulent; however, virulence could be restored by reconstitution of *psaA* (Berry and Paton 1996). Berry and Paton also state that the loss of virulence is not due to poorer growth of the mutant (Berry and Paton 1996) a factor that could otherwise contribute (Fernebro, Blomberg et al. 2008).

PsaA has been tested for immunogenicity in a mouse model of infection. Mice were immunised by purified PsaA from a serotype 22 strain. The mice were significantly protected against intravenous challenge by a serotype 3 strain (Talkington, Brown et al. 1996). Immunization of mice with PsaA was only modestly protective against lethal pneumococcal infection (Gor, Ding et al. 2005). It has also been shown that PsaA antibodies can prevent pneumococcal adherence to nasopharyngel epithelial cells (Romero-Steiner, Pilishvili et al. 2003). However, PsaA is not accessible to antibodies in circulation which suggest that it is not ideal as a vaccine candidate (Gor, Ding et al. 2005).

1.3.2.8 PavA

The *pavA* gene is localized to the cell surface even though it lacks traditional secretory or cell surface anchorage signals. It binds fibronectin (Holmes, McNab et al. 2001) and a *pavA*- knockout mutant showed 50 % reduced binding to immobilized human fibronectin. The *pavA* mutant was significantly attenuated in virulence compared to wild-type D39 in a mouse sepsis model (Holmes, McNab et al. 2001). Virulence was also attenuated in a mouse meningitis model (Pracht, Elm et al. 2005) indicating that PavA is important also for pneumococcal survival in the central nervous system. Pracht and co-workers suggest that PavA is important for pneumococcal adherence but not pneumococcal internalisation (Pracht, Elm et al. 2005). They also showed that the effect of PavA on adherence is not due to direct binding to fibronectin (Pracht, Elm et

al. 2005), nor that PavA is directly involved in adherence to a specific receptor but may modulate expression or function of other adherence and virulence determinants (Pracht, Elm et al. 2005).

1.4 TREATMENT AND RESISTANCE

The drug of choice for treating pneumococcal infections in Sweden is penicillin, as it has been for decades. Antibiotic resistance was first reported in South Africa in the 1970s and it is increasing (Witte, Cuny et al. 2008). In Sweden reduced susceptibility to penicillin is 2 % for 2008 (www.rivm.nl/earss) but in for example Romania it is as high as almost 70 % (www.rivm.nl/earss). For erythromycin the resistance rate in Sweden was 5.6 % (www.rivm.nl/earss).

Penicillin and other β-lactams act by binding to the so called Penicillin Binding Proteins (PBPs). Thereby they interfere with the crosslinking of the peptidoglycan in the cell wall (Engleberg, DiRita et al. 2007). There are several PBPs and in order to achieve high level resistance more than one of them needs to be altered in a stepwise process (Hakenbeck, Tarpay et al. 1980; Zighelboim and Tomasz 1980). These alterations usually occur through recombination of PBPs from resistant organisms, yielding mosaic PBP genes (Hakenbeck, Grebe et al. 1999). Resistance to β-lactams may also occur through β-lactamases, this has, however, never been reported in pneumococci (Hakenbeck, Grebe et al. 1999). It has been suggested that one reason for the increase in penicillin resistance is that already resistant clones have a chance to further spread under antibiotic selective pressure (Witte, Cuny et al. 2008). It should, however, be stressed that reduced susceptibility does not always mean treatment failure (Bauer, Ewig et al. 2001; doern, richter et al. 2005) and it has been shown that penicillin resistance is not necessarily a risk factor for a fatal outcome (Henriques, Kalin et al. 2000).

Other drugs used to treat pneumococcal infections are macrolides, commonly erythromycin and clindamycin, thrimethoprim-sulfonamide and fluoroquinolones. The macrolides act by inhibiting protein synthesis (Engleberg, DiRita et al. 2007). Macrolide resistance can occur in one of two ways, either through an efflux pump or by methylation of 23S rRNA (Witte, Cuny et al. 2008). Thrimethoprim-sulfonamide act by inhibiting bacterial dihydrofolate reductase and inhibiting dihydropteroate synthesis (Engleberg, DiRita et al. 2007). Resistance can occur through a mosaic dihydrofolate reductase (Witte, Cuny et al. 2008). Fluoroquinolones act by interfering with DNA replication (Engleberg, DiRita et al. 2007). Resistance occurs through mutations in one or both of the genes *parC*, encoding a topoisomerase and *gyrA*, the A-subunit of DNA gyrase (Witte, Cuny et al. 2008).

Risk factors for acquiring infection with drug resistant pneumococci include >65 years of age, β-lactam therapy in the last 3 months, fluoroquinolone therapy, alcoholism, underlying diseases and immunosuppression (Lode 2007). Macrolide resistance is increasing but the clinical relevance of this resistance has been questioned, just as for β-lactam resistance (Amsden 1999). Newer fluoroquinolones with enhanced activity against pneumococci also show resistance (Chen, McGeer et al. 1999).

1.5 VACCINE

Pneumococcal disease is a major cause of morbidity and mortality worldwide. It is highly desirable to prevent these infections by vaccination. Since pneumococci are extremely diverse it is however difficult to target all pneumococci with one vaccine. Licensed for use as at present are PPV-23, a 23-valent vaccine that gives protection against 23 of the 91 pneumococcal serotypes, PCV-7, a 7-valent vaccine conjugated to a protein as well as PCV10, a 10-valent conjugated vaccine where the protein carrier is from *H. influenzae*. PPV-23, is a polysaccharide based vaccine that has T-cell independent antigens. Therefore it does not yield enough antibody responses in young children who lack the mature B lymphocytes necessary to elicit T cell independent antibody-mediated immunity. The conjugated vaccine, however, yields a T-cell dependent response and therefore gives rise to a better memory for the antigen.

The PPV-23 is a polysaccharide vaccine containing serotypes 1-4, 6B, 7F, 8, 9N, 9V, 10A, 11A, 12F, 14, 15B, 17F, 18C, 19A, 19F, 20, 22F, 23F and 33F. These are chosen because they represent at least 85 to 90% of the serotypes that cause invasive infections (Jedrzejas 2001). PPV-23 has been shown to prevent 45 % of pneumococcal pneumonia, significantly reduce the risk of hospitalization and the risk of death (59%) in elderly >65 years of age in a study from Spain (Vila-Corcoles, Ochoa-Gondar et al. 2006)

PCV-7 is primarily licensed for use in children less than 2 years of age and includes serotypes 4, 6B, 9V, 14, 18C, 19F and 23F. These are the serotypes that are responsible for almost 90 % of invasive pneumococcal disease in children in the USA and Canada (Hausdorff, Bryant et al. 2000). Also, in the same country, they are responsible for approximately 60 % of acute otitis media cases (Hausdorff, Bryant et al. 2000) It has been suggested that PCV-7 could prevent 12000 cases of meningitis and bacteremia, 53000 cases of pneumonia, 116 deaths and 1 million episodes of otitis media per vaccinated birth cohort (Lieu, Ray et al. 2000). The PCV-7 has an efficacy against invasive pneumococcal disease by vaccine serotypes of 89% (Pavia, Bianco et al. 2009). PCV-7 vaccination has decreased rates of pneumococcal infection both in Canada, where incidence in the age group between 1 and 23 months of age decreased from 58.9/100000 to 7.3/100000 (Laupland, Gregson et al. 2009), and in the USA (Poehling, Talbot et al. 2006). In fact, a German study (Herz, Greenhow et al. 2006) showed that an overall decrease of bacteremia cases occurred where there was a decrease in pneumococcal bacteremia after introduction of PCV-7 without a simultaneous increase in bacteremia caused by other organisms. However, there has been reports of increases in pleural empyema (Byington, Korgenski et al. 2006; Gupta, Khaw et al. 2008).

Despite the success in decreasing rates of pneumococcal infections there has been reports regarding increasing rates of pneumococcal infections due to non-vaccine types. The increase is primarily due to increases in serotype 19A (Messina, Katz-Gaynor et al. 2007; Pelton, Huot et al. 2007; Mera, Miller et al. 2008) but increases in serogroups 11, 15, 33 and 35 has also been reported (Mera, Miller et al. 2008). Of concern is that this

increase in serotype 19A often is coupled to antimicrobial resistance (Messina, Katz-Gaynor et al. 2007; Pelton, Huot et al. 2007; Pichichero and Casey 2007)

A10-valent conjugated vaccine has just been licensed including PCV-7 plus serotypes 1, 5 and 7F. A 13-valent protein-conjugated vaccine including also serotypes 3, 6A and 19A is coming soon (Antao and Hausdorff 2009).

1.6 GENETICS

1.6.1 Sequenced laboratory strains

1.6.1.1 TIGR4

The TIGR4 isolate is a clinical isolate from the blood of a 30-year old male patient from Norway (Tettelin, Nelson et al. 2001). It has also been called JNR.7/87, KNR.7/87 and N4. It is a serotype 4 isolate of ST205. This isolate is often used as a model in mouse experiment studies since it is highly virulent in mice.

Its full sequence has been revealed through random shotgun sequencing and it was the first complete genome sequence of *Streptococcus pneumoniae.* It consists of a single circular chromosome almost 2.2 Megabases (Mb) in size. Its G+C content is 39.7 %. It can be found in GenBank under the accession number AE005672. The investigators were able to assign biological roles to 64 % of the predicted proteins. 20 % had no database match. Approximately 5 % of the chromosome is made up of insertion sequences many of which are frame-shifted or degenerate. There are 108 RUP elements. These are thought to act as a non-autonomous insertion sequence. Also in the TIGR4 chromosome are 127 BOX elements. These are modular DNA repeats composed of three subunits. The IS, RUP and BOX elements all contribute to pneumococcal genome plasticity. Several pathways for sugar metabolism are present in the genome as well as a high percentage of ATP-dependent transporters (Tettelin, Nelson et al. 2001).

1.6.1.2 D39

D39 is the progenitor of the isolate that was used by Avery and co-workers to demonstrate that DNA was the carrier of genetic material (Avery, MacLeod et al. 1979). It is a serotype 2 strain first isolated in 1916 (Lanie, Ng et al. 2007) and it is commonly used by researchers worldwide still today. It is deposited in GenBank under the accession number CP000410. Since the sequence of its descendant strain R6 (described below) is known it was sequenced through PCR walking. This means that overlapping PCR products was designed and sequenced based on the R6 genome. By using more than one strain of D39 obtained from different researchers they were able to show that the isolates had remained stable with only 13 mutational differences found between two of the isolates (Lanie, Ng et al. 2007). D39 carries a plasmid, pDP1. This cryptic plasmid contains seven open reading frames but only one of these, a replication protein, has known function.

1.6.1.3 R6

R6 is derived from the highly virulent type 2 strain D39. It was isolated through the steps of R36, which in comparison to D39 has lost its capsule and also lost in virulence.

Next came R36A that in comparison to R36 showed changes in morphology and increased transformability. After this came R6 with yet increased transformability and decreased colony size.

R6 is an unencapsulated isolate of ST128. Its genome consists of just above 2 Mb encoding 2043 predicted proteins and its G+C content is 40 %. Its GenBank accession number is AE007317. It has a 7504 bp deletion in the capsular genes. Over 500 of the genes are predicted to be either located on the surface of the bacterium or secreted. For 23 % of these no function could be predicted. There are systems for utilization of at least 12 carbohydrates. R6 is incapable of synthesizing aspartate, lysine, methionine, threonine, isoleucine, glutamate, arginine, glycine, histidine and leucine. BOX, RUP and IS elements comprise > 3 % of the genome (Hoskins, Alborn et al. 2001).

When comparing R6 to its mother strain D39 synteny is almost 100 %. The vast majority of differences between these isolates are mutational differences. Rearrangements have not occurred between these two isolates (Lanie, Ng et al. 2007).

1.6.2 Transformation

Transformation is the ability to take up DNA and incorporate it into the genome. It was F. Griffith that discovered and named the phenomenon transformation in an experiment where he found that heat-killed encapsulated pneumococci could transfer the ability to produce a capsule to live unencapsulated pneumococci (Griffith 1928).

S. pneumoniae is a naturally transformable bacterium. Genetic transformation also occurs naturally in many other bacterial species, e.g. *Haemophilus influenzae*, *Neisseria gonorrhoeae*, *Neisseria meningitidis* and *Bacillus subtilis*. Bacteria such as *E. coli* can also be transformed, however, this does not occur in nature and the cells must be made competent first. This artificial tranformation is more limited in the amount of DNA the cells can take up, in natural transformation it has, however, been estimated that pneumococci may take up as much as 10 % of its DNA content (Lacks 2004).

The process of transformation is suggested as follows: Double-stranded DNA is bound to the surface of the pneumococcus. Next a single-stranded cleavage occurs at random sites. EndA, a membrane-located nuclease, cleaves the other strand to give a double-strand break. EndA also degrades the complementary strand making it possible for the single-stranded donor strand to enter the cell (Lacks 2004). EndA is essential for uptake of DNA into cells, without it DNA is simply bound to the outside of pneumococcal cells (Lacks, Greenberg et al. 1974).

Pneumococcal cells are not always competent to take up DNA. Competence is controlled in two ways, by early and late genes. There are approximately 20 early genes (Claverys and Havarstein 2007). When extracellular competence-stimulating peptide (CSP) is accumulated it starts a signal transduction system which leads to the up-regulation of CSP. CSP induces competence. This CSP is found in several allelic variants distinct from each other, two of which, CSP-1 and CSP-2 are the most common (Claverys, Martin et al. 2007). The signal transduction system also up-regulates production of comX, the alternative sigma factor, which replaces SigA in

RNA polymerase. This leads to an enhancement of late genes involved in competence (Luo, Li et al. 2003). There are approximately 60 late genes induced by ComX including genes involved in e.g. DNA binding, DNA uptake and processing after the DNA has entered the cell (Claverys and Havarstein 2007).

LytA and other lytic factors are induced in one of the late genes operons, perhaps suggesting a role in lysing of non-competent cells thereby yielding DNA available for uptake. This phenomenon has been termed "pneumococcal fratricide" (Havarstein, Martin et al. 2006). The so called fratricide effectors include CbpD, CibABC, ComM, LytA and LytC all of which have different functions (Claverys, Martin et al. 2007). LytA and LytC are bot lytic enzymes (Claverys, Martin et al. 2007). CbpD is a cell wall degrading toxin; hence it is important for killing neighboring cells (Claverys and Havarstein 2007). CibAB presumably constitute a 2-peptide bacteriocin (Claverys and Havarstein 2007). CibC provides immunity to CibAB and ComM, one of the early genes, yields immunity towards the lytic enzymes (Claverys and Havarstein 2007; Claverys, Martin et al. 2007). Thereby the competent cells do not lyse themselves but rather other cells nearby that are not in the competent state.

1.6.3 Recombination vs mutation

By what means does genetic change occur? Through homologous recombination of pieces of DNA a few kilobases in size or is it through point mutation? The answer is a little bit of both and attempts have been made to determine what influence each of these two events, i.e. recombination and mutation. For the pneumococcus it has been estimated that the rate at which an allele changes is 10:1 in favor of recombination (Feil, Enright et al. 2000). An individual nucleotide is 50-fold more likely to change by recombination than by mutation (Feil, Enright et al. 2000). These estimates have been made by looking at MLST alleles for which a vast amount of data exists. Single locus variants (SLV) were examined and since it is unlikely that more than one point mutation should have occurred when the rest of alleles are unchanged these cases were taken as recombination. It becomes more difficult when alleles only differ in one nucleotide site. This is not necessarily through point mutation since a bacterium could just as easily recombinate an allele with one nucleotide difference as many. Feil and co-workers then made the assumption that a recombining allele should be present elsewhere in the genetic pool and hence it is likely already present in the MLST database. So, an allele that has occurred through point mutation differs in one nucleotide site from its SLV and this allele is a novel allele, not previously present in the MLST database (Feil, Enright et al. 2000; Feil, Holmes et al. 2001; Spratt, Hanage et al. 2001). This implies that drawing phylogenetic trees of bacteria based on sequence of one gene is not valid . Also, in a study comparing 13 streptococcal genomes it was determined that gene gain and gene loss are major determinants of evolution (Marri, Hao et al. 2006).

1.6.4 Genome plasticity

In 2001 the first three genome sequences were made public ((Dopazo, Mendoza et al. 2001; Hoskins, Alborn et al. 2001; Tettelin, Nelson et al. 2001). Since then a vast number has become publicly available. From the study of the first three, one can

conclude that the pneumococcal genome contains over 2000 genes of which approximately 60 % has a predicted function. The G+C content is just below 40 %. IS, BOX and RUP elements, i.e. repetitive elements, were numerous, giving a great potential for genome rearrangements. These kind of repetitive elements have been shown to contribute to genome plasticity in *H. pylori* (Aras, Kang et al. 2003). In comparing the three sequenced strains they varied from each other in about 10 % (Tettelin and Hollingshead 2004). Other studies have revealed that the core genome of a pneumococcus is approximately 70 % (Hakenbeck, Balmelle et al. 2001; Bruckner, Nuhn et al. 2004). The rest is accounted for by the accessory genome, with genes that may or may not be present. In the paper of Tettelin et al (Tettelin, Nelson et al. 2001) a comparative genome hybridisation of 13 additional strains was also carried out. They identified 13 so called regions of diversity. These can be fully or partly present or absent in the genomes studied. Some of these regions may have been originally acquired from other organisms as suggested by their atypical GC content.

Obert and co-workers (Obert, Sublett et al. 2006) performed comparative genome hybridizations using microarray of 72 strains of serotypes 6A, 6B and 14. 42 of these were invasive and 30 were noninvasive. They identified the core genome to be 1553 genes, or 73 % of the TIGR4 genome. In their study they identified 7 regions of diversity to be associated with the noninvasive phenotype and 2 that were associated with the invasive one.

In a study performed by Silva and co-workers (Silva, McCluskey et al. 2006) 13 isolates were investigated also using comparative genome hybridizations by microarray. They represent 7 serotypes, although 6 of the isolates were of serotype 14, and 8 different STs .They identified 3 new regions of diversity for a total of 25 regions of diversity where three or more accessory genes were absent in at least one of the isolates. These regions of diversity range in size from 1.7 to 34.8 kb. No attempt was made to define the pneumococcal core genome. They were able to show that isolates of the same serotype (14) and ST (124) differed in genetic content. These isolates also differed in virulence in a mouse model of infection.

Hiller and co-workers performed a comparative analysis of seventeen fully sequenced strains (Hiller, Janto et al. 2007). They propose that pneumococci possess a supragenome consisting both of the core genome, that is present in all isolates, and an accessory genome that may or may not be present. They found that there are 3170 sequence clusters present in total in the 17 strains. 46 % of these, corresponding to 73 % of coding sequences, are present in all isolates. 36 % are present in two or more isolates and 18 % are unique to a single strain. Of the unique genes 64 % are annotated as hypothetical and approximately 5 % have something to do with phage proteins. The core genes include 30 % hypothetical genes. In an individual isolate 21 to 32 % of the sequence clusters were accessory. They suggest that sequencing of 33 individual isolates will yield 99 % of the supragenome that is present in at least 10 % of the population.

1.6.5 Phages

Another factor that can be of importance for genome plasticity is temperate bacteriophages. In *S.aureus* it has been shown that they may carry accessory virulence factors (Goerke, Pantucek et al. 2009). Previously up to 76 % of clinical isolates have been shown to contain phages (Ramirez, Severina et al. 1999). Another study report 53 % of isolates as carrying at least one prophage (Romero, Garcia et al. 2009). Recently Romero and co-workers analyzed 10 pneumococcal bacteriophages, all of which belong to the Siphoviridae group (Romero, Croucher et al. 2009). They found that the phages can be divided into three groups based on sequence similarities, with 56 % of prophages belonging to group 1, 26 % belonging to group 2 and 11 % belonging to group 3 (Romero, Garcia et al. 2009). However, all phages are organized in five gene clusters, lysogeny, replication, packaging, morphogenesis and lysis (Romero, Croucher et al. 2009). All putative integrases, present in the lysogeny cluster, belong to the integrase family of tyrosine recombinases. Next to the lysogeny cluster the replication cluster is located. This cluster is highly conserved within phage groups but not between them. The packaging cluster is essentially the same in all phages. The morphogeny cluster can differ within phage groups, e.g. most of the sequence differences for phage group 1 are located here. The lysis cluster contains the holin and endolysin genes. The endolysin is a N-acetylmutamoyl-L-alanin amidase. It is 91.5 % similar within phages and 68.3 % identical to the LytA protein. Genes homologous to phage virulence genes, such as *pblB* of *Streptococcus mitis* and *vapE* of *Dichelobacter nodosus*, were identified but many genes had no homologues.

2 AIMS

The general aim with this thesis is to characterize the genetic relationship and genomic content of pneumococcal clinical isolates causing invasive disease or carriage.

2.1 PAPER I

The aim of this paper was to investigate whether molecular typing by (MLST) gives a good estimate of the genetic content of clinical pneumococcal isolates and to find alternative methods for molecular typing.

2.2 PAPER II

The aim of this paper was to examine how well clinical pneumococcal isolates belonging to larger eBURST groups correlate genetically.

2.3 PAPER III

The aim of this paper was to examine whether we can find a special virulence factor or accessory region important for disease outcome.

2.4 PAPER IV

The aim of this paper was to determine the invasive disease potential in children for pneumococcal serotypes and clones and to find associations between bacterial genetic content and disease potential.

3 METHODS

3.1 BACTERIAL ISOLATES

The isolates used in this thesis have been sampled in different ways. R6 and TIGR4 are laboratory strains obtained from Professor Elaine Tuomanen. The PMEN isolates are recognised by the Pneumococcal Molecular Epidemiology Network as being internationally spreading bacterial clones. These were obtained from Professor Alexander Tomasz and Professor Herminia de Lencastre. The invasive isolates have been collected from the clinical microbiology laboratories in Sweden. The carriage isolates come from three different carriage studies performed in 1997, 1999 and in 2004 respectively where children have been sampled at day-care centers in the Stockholm area.

3.2 SEROTYPING AND ANTIBIOTIC SUSCEPTIBILITY TESTING

Serotypin was performed using the gel diffusion method (Halbert, Swick et al. 1955) in combination with the Quellung reaction using serotype-specific sera obtained from Statens Seruminstitut in Copenhagen, Denmark. Drug susceptibility was determined using the disk diffusion method according to the Swedish Reference Group for antibiotics (SRGA) with antibiotic discs from Oxoid.

3.3 PULSED-FIELD GEL ELECTROPHORESIS

The PFGE method was developed in 1984 (Schwartz and Cantor 1984). It is used for molecular epidemiology commonly in outbreak investigations. It is fast and relatively cheap but it is not so easy to compare results between laboratories.

When performing PFGE you imbed the whole genome DNA of your isolate of interest into agarose. Thereafter the DNA is cleaved by a restriction enzyme, apaI in this thesis, and run on a gel. By using electric pulses from various directions it is possible to separate even the larger fragments. The results have then been analysed using the BioNumerics software with the clone inclusion criteria of Tenover et al (Tenover, Arbeit et al. 1995).

3.4 MULTI-LOCUS SEQUENCE TYPING

MLST is another tool for molecular epidemiology developed by Enright and Spratt (Enright and Spratt 1998). It is easily compared between laboratories and is better at determining genetic relationship over time than PFGE. However, the discriminatory power is less than with PFGE.

When performing MLST you sequence approximately 450 bp each of seven house-keeping genes in the pneumococcal genome. For the pneumococcus genes aroE, gdh, gki, recP, spi, xpt and ddl are used. Each locus is given a number depending on the particular sequence. These seven numbers give the allelic profile and also the ST. Isolates differing in one allele are said to be single-locus variants (SLVs) of each other. Isolates differing in two alleles are said to be double-locus variants (DLVs).

3.5 EBURST ANALYSIS

eBURST is an algorithm that tries to determine relatedness between specific STs. The algorithm also attempts to determine which clone is the founder of the clonal complex (CC). In a clonal complex all isolates share the same alleles at least 6 loci with at least one other member of the group. The founder is the ST that has the largest amount of SLVs in that group. If two STs in a group have the same amont of SLVs the ST with the largest number of DLVs is considered the founder (Feil, Li et al. 2004). In this thesis eBURST groups are named after their predicted founder.

3.6 MICROARRAY

The microarray used in this thesis is based on the open reading frames of the two genomes of TIGR4 and R6. A total of 2797 50-oligonucelotides are spotted in triplicates on glass slides.The experiments were carried out using a reference design consisting of an equimolar mix of R6 and TIGR4 as a control. Whole genome DNA is cut into pieces and then labeled with Cy-3 or Cy-5. It is then hybridised to the oligonucelotides. In papers I, II and III 3-4 hybridisations per isolate were carried out including at least one dye-swap experiments. In paper IV two hybridisations including one dye-swap was performed. The microarrays were scanned at 532 and 635 nm.
In the analysis fluorescent spot and background intensities were quantified using the GenePix Pro software. Median pixel intensity was calculated for each spot. Spots showing a reference signal lower than background plus two standard deviations, being saturated or showing obvious abnormalities were removed from the statistical analysis. For statistical analysis the R project for statistical computing was used (www.rproject.org). The log2 fluorescence ratios were normalised to the median. A bayesian linear model and the Holm multiple testing correction test was used to adjust individual p-values (Holm 1979). The isolates were compared to the reference strains as well as to each other. Genes were considered absent if they had an M value of less than -1 and a p-value less than 0.01
This microarray only consists of the full genome of two pneumococcal isolates, TIGR4 and R6. We can never detect genes that are not present in either or these two genomes hence we miss some information in the analysis. Also, the number of genes that are absent in one isolate is not an exact number but more of an approximation. This lies in the nature of the statistical analysis. Some genes may be present but differ enough in sequence over the oligonucleotide to give a false negative result.

4 RESULTS AND DISCUSSION

4.1 PAPER I

The common molecular epidemiological tool, MLST, measures relatedness through sequencing of 7 house-keeping genes. These genes naturally belong to the core genome and it is unclear how well molecular typing with MLST correlates to genetic content of pneumococcal isolates. In paper I we determined genetic content using our whole genome based microarray and correlated the result to sequence type, ST, as determined using MLST. 40 pneumococcal isolates were tested. Some of these belonged to the same clonal complex as determined by eBURST. When performing hierarchical clustering of the microarray results isolates belonging to the same clonal complex clustered together. This suggests that they are closely related. Isolates that did not belong to the same CC did not cluster together suggesting that they are genetically more distant from each other.

All isolates were compared to all others in a pairwise manner to determine the number of genes that differed between the pneumococcal isolates. The isolates in a particular comparison were grouped into one of the following groups: group A clonal complex with the same serotype (24 comparisons), group B clonal complex with different serotypes (25 comparisons), group C not a clonal complex but with the same serotype and group D neither belonging to the same clonal complex nor to the same serotype. Isolates belonging to the same clonal complex were more similar regardless of whether they belonged to the same serotype or not. In contrast, isolates belonging to the same serotype were not more similar to each other than isolates of different capsular types.

The number of differences between isolates belonging to the same clonal complex was investigated in more detail. Isolates belonging to clonal complexes with serotypes of high invasive disease potential, CC306 of serotype 1 and CC191 of serotype 7F, had differences ranging from 0-44 genes. Other clonal complexes had differences ranging from 1-53 genes. Interestingly no differences of the gene content of the isolates of ST306 as well as for the isolates of ST191 could be detected. This suggests that isolates belonging to these invasive STs have few intraclonal differences.

An alternative approach for molecular typing was investigated in the paper. Instead of the expensive and time-consuming method MLST, we found that the absence or presence of genes in the accessory (variable) genome could function as a predictor of relatedness. Presence and absence of 25 genes in 25 accessory regions was determined through PCR. Their presence varied among the 40 isolates tested. Isolates belonging to the same clonal complex did not differ in presence/absence in more than 2 of these 25 genes. In contrast, isolates belonging to different CCs differed in more than 4, hence it seems possible to determine genetic relatedness through PCR alone. This would be a great advantage compared to MLST since it is cheap and fast to use. However it needs to be tested on a larger set of isolates.

4.2 PAPER II

In Paper II the genetic relatedness of clinical isolates belonging to the same clonal complex and differing in up to 4 alleles was investigated using microarray. The two largest clonal complexes in the MLST database, CC156 and CC138 as well as a medium size clonal complex, CC124, were chosen for analysis. Also isolates representing the largest clonal complexes of serotypes 1 and 3 were investigated. There was some variation as to how similar isolates of different clonal complexes were.

For CC138 three serotype 6B isolates and one 19F isolate were analyzed. The 19F isolate was more divergent than the serotype 6B isolates suggesting that either there has been some larger genetic replacement during the serotype switch or that they belong to two branches. One branch including serotype 6B isolates and the other including serotypes 19F and 23F that have come together by chance.

For CC156 the number of genes differing rose with the number of alleles differing. Serotype does not seem to play a great role for genetic diversity in this CC as isolates of even the same ST can have different serotypes.

For CC124 the results are somewhat similar to those for CC138. Here the number of genes differing also increases with number of alleles differing but there is a great leap between 2 alleles differing, with 15 genes differing, and 3 and 4 alleles differing, with 192-225 genes differing. In all comparisons of isolates differing in 3 and 4 alleles the comparisons are between isolates from different branches having different serotypes. Again, either there have been some major genetic rearrangements during serotype switching or the major cluster with serotype 14 and the branch with various serotypes have come together by chance.

For serotype 1 over 100 genes differed between isolates of three of the largest clonal complexes. Only between 51 and 78 genes differed between isolates of CC217 and CC2296, even though isolates compared differed in 5 and 6 alleles. This suggests that perhaps these two clonal complexes represent CCs that have been separated by chance and they may have originated from a common ancestor.

For serotype 3 three clonal complexes were included. The largest CC180 was substantially different from the others. The other two, however, CC260 and CC378 only differ from each other by between 35 and 51 genes. The isolates differ in three alleles. The mean number of genes differing between isolates differing in three alleles is 143 which of course is much higher. This would indicate that these two clonal complexes are in fact one and the same. Interestingly isolates of these clonal complexes belong to the same clone as determined by PFGE.

In general we conclude that differences in number of genes differing between isolates as compared by microarray are determined by the number of alleles differing between isolates, if they belong to the same clonal complex or not and if they are of the same serotype or not. Use of more than one molecular typing method is useful when determining genetic relatedness. Assumptions of genetic similarity within large clonal complexes should be done with caution.

4.3 PAPER III

The invasive disease potential varies between serotypes and perhaps clonal types. The reason for this is unknown. In this paper we investigated presence and absence of accessory regions as determined by microarray in serotypes with a known invasive disease potential.

34 % of the genes in the combined genomes of R6 and TIGR4 were absent in at least one of the isolates tested. Since the microarray method is not set up to detect differences in sequence we cannot conclude that all of these genes are in fact missing. Instead we focus on accessory regions were several consecutive genes were found as absent or present. Of the accessory genes, not belonging to the core genome, 40 % were found in such regions.

In our dataset we found 5 new accessory regions that had not been previously described. These mainly encode hypothetical proteins. Our criteria for these regions were at least 3 genes next to each other in the genome missing in at least one of our isolates.

Through three STM screens approximately 290 genes have been found as being essential for virulence in mouse models of infection. Interestingly about one third of these were absent in our clinical isolates. 32 were found to be located in accessory regions. This suggests that there is a great redundancy in the function of these proteins.

Isolates belonging to CC306 of serotype 1 have been shown to have a high invasive disease potential. The isolates in this CC belongs to two different STs, ST306 (two isolates) and ST228 (one isolate). The two ST306 isolates did not differ in genetic content but the ST306 and ST228 isolates differed in presence/absence of approximately 40 genes. These include AR10, AR19, AR27 and part of AR16. AR10, involved in mannose metabolism and AR19 are present in ST228 which is significantly more virulent than ST306 after intraperitoneal challenge in mice.

Serotype 7F is also regarded as a serotype with high invasive disease potential. Two of the serotype 7F isolates were of the same ST, ST191, and hardly any genetic differences were found among them suggesting that invasive STs are more clonal.

In isolates of the serotypes with high IDP (types 1, 4 and 7F) 26 of the totally 41 ARs were absent in at least one isolate. 17 of these harbor genes identified in STM screens as being essential for mouse virulence. No AR is associated solely with isolates of high IDP, however, AR6 and AR34 stood out as the best candidates. We were however not able to show any difference between knockout mutants of these regions and wild type bacteria in a mouse model of infection. This suggests a high level of redundancy among ARs.

4.4 PAPER IV

In paper IV we determined the invasive disease potential of pneumococcal serotypes and clones by characterizing invasive isolates and carriage isolates from children in the Stockholm area between 1997 and 2004. Serotypes 6B, 14, 7F, 1, 6A, 19A, 18C and 19F were, in rank order, the most common serotypes causing invasive disease, all of which individually accounting for over 5 %. Serotype 6B alone accounted for 24 % of all invasive disease cases. Serotype19F was most common type found among the carriage isolates, followed by 6B, 6A, 23F and 14. Again, all of these serotypes accounted for above 5 % each. The majority of these serotypes are included in the 7-valent protein-conjugated vaccine and all serotypes mentioned above are included in the 13-valent vaccine.

When comparing these results to which serotypes cause disease in adults the picture is somewhat different. The serotypes that individually account for at least 5 % of the invasive disease in adults are in rank order 14, 1, 4, 9V, 23F, 3 and 7F. All these serotypes are included in the 13-valent conjugated polysaccharide vaccine.

Invasive disease potential was calculated for the serotypes found among the children by comparing invasive disease and carriage isolates. Serotypes with an OR >1 for causing invasive disease were 1, 4, 7F, 19A, 18C and 6B. Serotypes with an OR <1 for causing invasive disease were 6A, 23F and 19F. Interestingly, one of the most invasive serotypes in previous studies, serotype 14 was not significantly invasive in our data set. 6 clonal complexes had an OR >1 for causing invasive disease: CC9, where all isolates were invasive, CC482, CC205, CC176, CC306 and CC 191.

CC176 consisted of three PFGE clones. Interestingly, we showed that the invasive disease potential differed with clonal type as determined using PFGE. Four isolates representing the three different PFGE clones with different invasive disease potential, where two were invasive and two non-invasive, were subjected to microarray analysis. Two of these isolates, one invasive and one carriage isolate were of ST176. The other two belonged to ST138. 18 genes differed in presence/absence between these four isolates. The gene pcpA, a choline-binding protein important for lung disease, was absent in one of the carriage isolates. Interestingly the two isolates of ST 138 had different pspA alleles.

Since we can only detect differences in genes present on the chip (ie genes present in TIGR4 and R6) we subjected the four pneumococcal isolates to whole genome sequencing to unravel other differences. Interestingly we found that one of the invasive isolates contained a phage. To study whether we could find a correlation between invasive disease and phages we further studied the presence or absence of phages among our clinical isolates belonging to the PFGE clones of CC176. Phages were present on most of the PFGE clones. Most of these were of phage type 3 but phage type 1 and 2 were present in some of the clones. However, this typing of phages is performed using only a couple of genes and there may be great variation due to recombination between regions of the phages.

5 ACKNOWLEDGEMENTS

I would like to thank my main supervisor, Birgitta Henriques-Normark, for letting me pursuit my ideas and for making it possible for me to go to so many interesting conferences and meetings. Thank you to my second supervisor, Eva Morfeldt. Your knowledge of lab work has been a great help to me. I would also like to thank my "almost" supervisor, Staffan Normark, for always having great ideas when I think I am stuck.

Thank you, Anita for all the help on administrative matters.

A big thank you also to Christel, without your help I would have given up a long time ago! Thank you, Sarah for all the help in the lab.

I would also like to thank the rest of my co-authors, Jenny, Sofia, Karin, Andreas. Of course I also want to thank the rest of the pneumo group, all members past and present.

Thank you Christina for taking care of me in Singapore, I truly learnt a lot!

Mamma, thank you for your tremendous support, for all the help with Linus and thanks for not letting your nervousness show.

Thank you to my sisters, Camilla and Fia, for always believing in me, even though you have no idea of what I am doing.

Thank you, pappa, for at least trying to read my papers.

Thank you, Greger, for being there in times of crisis.

Thank you Tore and Eira, for all the help with Linus but also for letting us stay with you before our house was ready.

Thank you Johan for never thinking anything is a big problem and for supporting me so much at home. Words are not enough to describe what you have done for me. I especially appreciate being able to travel so much.

Finally, Linus, thank you for putting things into perspective. Mamma älskar dig!

6 REFERENCES

Alanee, S. R., L. McGee, et al. (2007). "Association of serotypes of Streptococcus pneumoniae with disease severity and outcome in adults: an international study." Clin Infect Dis **45**(1): 46-51.

Amsden, G. W. (1999). "Pneumococcal macrolide resistance--myth or reality?" J Antimicrob Chemother **44**(1): 1-6.

Antao, V. C. and W. P. Hausdorff (2009). "Global epidemiology of pneumococcal disease--new prospects for vaccine control." Adv Exp Med Biol **634**: 19-29.

Aras, R. A., J. Kang, et al. (2003). "Extensive repetitive DNA facilitates prokaryotic genome plasticity." Proc Natl Acad Sci U S A **100**(23): 13579-84.

Avery, O. T., C. M. MacLeod, et al. (1979). "Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III." J Exp Med **149**(2): 297-326.

Bagnoli, F., M. Moschioni, et al. (2008). "A second pilus type in Streptococcus pneumoniae is prevalent in emerging serotypes and mediates adhesion to host cells." J Bacteriol **190**(15): 5480-92.

Balachandran, P., A. Brooks-Walter, et al. (2002). "Role of pneumococcal surface protein C in nasopharyngeal carriage and pneumonia and its ability to elicit protection against carriage of Streptococcus pneumoniae." Infect Immun **70**(5): 2526-34.

Baraff, L. J., S. I. Lee, et al. (1993). "Outcomes of bacterial meningitis in children: a meta-analysis." Pediatr Infect Dis J **12**(5): 389-94.

Barocchi, M. A., J. Ries, et al. (2006). "A pneumococcal pilus influences virulence and host inflammatory responses." Proc Natl Acad Sci U S A.

Bauer, T., S. Ewig, et al. (2001). "Streptococcus pneumoniae in community-acquired pneumonia. How important is drug resistance?" Med Clin North Am **85**(6): 1367-79.

Benton, K. A., M. P. Everson, et al. (1995). "A pneumolysin-negative mutant of Streptococcus pneumoniae causes chronic bacteremia rather than acute sepsis in mice." Infect Immun **63**(2): 448-55.

Berg, S., B. Trollfors, et al. (2002). "Long-term follow-up of children with bacterial meningitis with emphasis on behavioural characteristics." Eur J Pediatr **161**(6): 330-6.

Berry, A. M. and J. C. Paton (1996). "Sequence heterogeneity of PsaA, a 37-kilodalton putative adhesin essential for virulence of Streptococcus pneumoniae." Infect Immun **64**(12): 5255-62.

Berry, A. M. and J. C. Paton (2000). "Additive attenuation of virulence of Streptococcus pneumoniae by mutation of the genes encoding pneumolysin and other putative pneumococcal virulence proteins." Infect Immun **68**(1): 133-40.

Bhatt, S. M., A. Lauretano, et al. (1993). "Progression of hearing loss in experimental pneumococcal meningitis: correlation with cerebrospinal fluid cytochemistry." J Infect Dis **167**(3): 675-83.

Boken, D. J., S. A. Chartrand, et al. (1995). "Colonization with penicillin-resistant Streptococcus pneumoniae in a child-care center." Pediatr Infect Dis J **14**(10): 879-84.

Boken, D. J., S. A. Chartrand, et al. (1996). "Colonization with penicillin-nonsusceptible Streptococcus pneumoniae in urban and rural child-care centers." Pediatr Infect Dis J **15**(8): 667-72.

Braun, J. S., R. Novak, et al. (1999). "Pneumolysin, a protein toxin of Streptococcus pneumoniae, induces nitric oxide production from macrophages." Infect Immun **67**(8): 3750-6.

Brooks-Walter, A., D. E. Briles, et al. (1999). "The pspC gene of Streptococcus pneumoniae encodes a polymorphic protein, PspC, which elicits cross-reactive

antibodies to PspA and provides immunity to pneumococcal bacteremia." Infect Immun **67**(12): 6533-42.
Bruckner, R., M. Nuhn, et al. (2004). "Mosaic genes and mosaic chromosomes-genomic variation in Streptococcus pneumoniae." Int J Med Microbiol **294**(2-3): 157-68.
Brueggemann, A. B., D. T. Griffiths, et al. (2003). "Clonal relationships between invasive and carriage Streptococcus pneumoniae and serotype- and clone-specific differences in invasive disease potential." J Infect Dis **187**(9): 1424-32.
Brueggemann, A. B., T. E. Peto, et al. (2004). "Temporal and geographic stability of the serogroup-specific invasive disease potential of Streptococcus pneumoniae in children." J Infect Dis **190**(7): 1203-11.
Bryce, J., C. Boschi-Pinto, et al. (2005). "WHO estimates of the causes of death in children." Lancet **365**(9465): 1147-52.
Byington, C. L., K. Korgenski, et al. (2006). "Impact of the pneumococcal conjugate vaccine on pneumococcal parapneumonic empyema." Pediatr Infect Dis J **25**(3): 250-4.
Byington, C. L., L. Y. Spencer, et al. (2002). "An epidemiological investigation of a sustained high rate of pediatric parapneumonic empyema: risk factors and microbiological associations." Clin Infect Dis **34**(4): 434-40.
Casey, J. R. and M. E. Pichichero (2004). "Changes in frequency and pathogens causing acute otitis media in 1995-2003." Pediatr Infect Dis J **23**(9): 824-8.
Chen, D. K., A. McGeer, et al. (1999). "Decreased susceptibility of Streptococcus pneumoniae to fluoroquinolones in Canada. Canadian Bacterial Surveillance Network." N Engl J Med **341**(4): 233-9.
Cheng, Q., D. Finkel, et al. (2000). "Novel purification scheme and functions for a C3-binding protein from Streptococcus pneumoniae." Biochemistry **39**(18): 5450-7.
Childers, N. K., M. G. Bruce, et al. (1989). "Molecular mechanisms of immunoglobulin A defense." Annu Rev Microbiol **43**: 503-36.
Claverys, J. P. and L. S. Havarstein (2007). "Cannibalism and fratricide: mechanisms and raisons d'etre." Nat Rev Microbiol **5**(3): 219-29.
Claverys, J. P., B. Martin, et al. (2007). "Competence-induced fratricide in streptococci." Mol Microbiol **64**(6): 1423-33.
Coffey, T. J., M. Daniels, et al. (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 (Pt 8)**: 2023-31.
Coffey, T. J., M. C. Enright, et al. (1998). "Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of Streptococcus pneumoniae." Mol Microbiol **27**(1): 73-83.
Dave, S., A. Brooks-Walter, et al. (2001). "PspC, a pneumococcal surface protein, binds human factor H." Infect Immun **69**(5): 3435-7.
de Gans, J. and D. van de Beek (2002). "Dexamethasone in adults with bacterial meningitis." N Engl J Med **347**(20): 1549-56.
Dieudonne-Vatran, A., S. Krentz, et al. (2009). "Clinical Isolates of Streptococcus pneumoniae Bind the Complement Inhibitor C4b-Binding Protein in a PspC Allele-Dependent Fashion." J Immunol **182**(12): 7865-77.
doern, g. v., s. s. richter, et al. (2005). "Antimicrobial resistance among Streptococcus pneumoniae in the United States: Have we begun to turn the corner on resistance to certain antimicrobial classes?" clin infect dis **41**: 139 - 148.
Dopazo, J., A. Mendoza, et al. (2001). "Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate." Microb Drug Resist **7**(2): 99-125.
Dowson, C. (2004). What Is a Pneumococcus. The Pneumococcus. E. Tuomanen, T. Mitchell, D. A. Morrison and B. G. Spratt, ASM Press.
Duchin, J. S., R. F. Breiman, et al. (1995). "High prevalence of multidrug-resistant Streptococcus pneumoniae among children in a rural Kentucky community." Pediatr Infect Dis J **14**(9): 745-50.

Elm, C., M. Rohde, et al. (2004). "Characterization of the interaction of the pneumococcal surface protein SpsA with the human polymeric immunoglobulin receptor (hpIgR)." Indian J Med Res **119 Suppl**: 61-5.
Engleberg, N., V. DiRita, et al., Eds. (2007). Schaechter's Mechanisms of Microbial Disease, Lippincott Williams & Wilkins.
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 (Pt 11)**: 3049-60.
Feil, E. J., M. C. Enright, et al. (2000). "Estimating the relative contributions of mutation and recombination to clonal diversification: a comparison between Neisseria meningitidis and Streptococcus pneumoniae." Res Microbiol **151**(6): 465-9.
Feil, E. J., E. C. Holmes, et al. (2001). "Recombination within natural populations of pathogenic bacteria: short-term empirical estimates and long-term phylogenetic consequences." Proc Natl Acad Sci U S A **98**(1): 182-7.
Feil, E. J., B. C. Li, et al. (2004). "eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data." J Bacteriol **186**(5): 1518-30.
Fernebro, J., C. Blomberg, et al. (2008). "The influence of in vitro fitness defects on pneumococcal ability to colonize and to cause invasive disease." BMC Microbiol **8**: 65.
Fletcher, M., J. Leeming, et al. (2006). "Childhood empyema: limited potential impact of 7-valent pneumococcal conjugate vaccine." Pediatr Infect Dis J **25**(6): 559-60.
Glover, D. T., S. K. Hollingshead, et al. (2008). "Streptococcus pneumoniae surface protein PcpA elicits protection against lung infection and fatal sepsis." Infect Immun **76**(6): 2767-76.
Goerke, C., R. Pantucek, et al. (2009). "Diversity of prophages in dominant Staphylococcus aureus clonal lineages." J Bacteriol **191**(11): 3462-8.
Goldbart, A. D., E. Leibovitz, et al. (2009). "Complicated community acquired pneumonia in children prior to the introduction of the pneumococcal conjugated vaccine." Scand J Infect Dis **41**(3): 182-7.
Gor, D. O., X. Ding, et al. (2005). "Relationship between surface accessibility for PpmA, PsaA, and PspA and antibody-mediated immunity to systemic infection by Streptococcus pneumoniae." Infect Immun **73**(3): 1304-12.
Griffith, F. (1928). "The significance of pneumococcal types." J. Hyg. **27**: 113-159.
Gupta, A., F. M. Khaw, et al. (2008). "Outbreak of Streptococcus pneumoniae serotype 1 pneumonia in a United Kingdom school." Bmj **337**: a2964.
Hakenbeck, R., N. Balmelle, et al. (2001). "Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of Streptococcus pneumoniae." Infect Immun **69**(4): 2477-86.
Hakenbeck, R., T. Grebe, et al. (1999). "beta-lactam resistance in Streptococcus pneumoniae: penicillin-binding proteins and non-penicillin-binding proteins." Mol Microbiol **33**(4): 673-8.
Hakenbeck, R., M. Tarpay, et al. (1980). "Multiple changes of penicillin-binding proteins in penicillin-resistant clinical isolates of Streptococcus pneumoniae." Antimicrob Agents Chemother **17**(3): 364-71.
Halbert, S., L. Swick, et al. (1955). "The use of precipitin analysis in agar for the study of human streptococcal infections. II Ouchterlony and Oakley Techniques." J Exp Med **101**: 557-76.
Hammerschmidt, S., G. Bethe, et al. (1999). "Identification of pneumococcal surface protein A as a lactoferrin-binding protein of Streptococcus pneumoniae." Infect Immun **67**(4): 1683-7.
Hammerschmidt, S., S. R. Talay, et al. (1997). "SpsA, a novel pneumococcal surface protein with specific binding to secretory immunoglobulin A and secretory component." Mol Microbiol **25**(6): 1113-24.
Hanage, W. P., T. H. Kaijalainen, et al. (2005). "Invasiveness of serotypes and clones of Streptococcus pneumoniae among children in Finland." Infect Immun **73**(1): 431-5.

Hausdorff, W. P., J. Bryant, et al. (2000). "The contribution of specific pneumococcal serogroups to different disease manifestations: implications for conjugate vaccine formulation and use, part II." Clin Infect Dis **30**(1): 122-40.
Hausdorff, W. P., J. Bryant, et al. (2000). "Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I." Clin Infect Dis **30**(1): 100-21.
Hava, D. L. and A. Camilli (2002). "Large-scale identification of serotype 4 Streptococcus pneumoniae virulence factors." Mol Microbiol **45**(5): 1389-406.
Havarstein, L. S., B. Martin, et al. (2006). "New insights into the pneumococcal fratricide: relationship to clumping and identification of a novel immunity factor." Mol Microbiol **59**(4): 1297-307.
Henriques, B., M. Kalin, et al. (2000). "Molecular epidemiology of Streptococcus pneumoniae causing invasive disease in 5 countries." J Infect Dis **182**(3): 833-9.
Heremans, J. F. (1974). "The IgA system in connection with local and systemic immunity." Adv Exp Med Biol **45**(0): 3-11.
Herz, A. M., T. L. Greenhow, et al. (2006). "Changing epidemiology of outpatient bacteremia in 3- to 36-month-old children after the introduction of the heptavalent-conjugated pneumococcal vaccine." Pediatr Infect Dis J **25**(4): 293-300.
Hiller, N. L., B. Janto, et al. (2007). "Comparative genomic analyses of seventeen Streptococcus pneumoniae strains: insights into the pneumococcal supragenome." J Bacteriol **189**(22): 8186-95.
Holm (1979). "A simple sequentially rejective Bonferroni test procedure." Scandinavian Journal of Statistics **6**: 65-70.
Holmes, A. R., R. McNab, et al. (2001). "The pavA gene of Streptococcus pneumoniae encodes a fibronectin-binding protein that is essential for virulence." Mol Microbiol **41**(6): 1395-408.
Hoskins, J., W. E. Alborn, Jr., et al. (2001). "Genome of the bacterium Streptococcus pneumoniae strain R6." J Bacteriol **183**(19): 5709-17.
Jacoby, P., K. Watson, et al. (2007). "Modelling the co-occurrence of Streptococcus pneumoniae with other bacterial and viral pathogens in the upper respiratory tract." Vaccine **25**(13): 2458-64.
Jedrzejas, M. J. (2001). "Pneumococcal virulence factors: structure and function." Microbiol Mol Biol Rev **65**(2): 187-207 ; first page, table of contents.
Jedrzejas, M. J. (2006). "Unveiling molecular mechanisms of pneumococcal surface protein A interactions with antibodies and lactoferrin." Clin Chim Acta **367**(1-2): 1-10.
Jedrzejas, M. J., S. K. Hollingshead, et al. (2000). "Production and characterization of the functional fragment of pneumococcal surface protein A." Arch Biochem Biophys **373**(1): 116-25.
Jefferies, J. M., C. H. Johnston, et al. (2007). "Presence of nonhemolytic pneumolysin in serotypes of Streptococcus pneumoniae associated with disease outbreaks." J Infect Dis **196**(6): 936-44.
Johnston, J. W., D. E. Briles, et al. (2006). "Mn2+-dependent regulation of multiple genes in Streptococcus pneumoniae through PsaR and the resultant impact on virulence." Infect Immun **74**(2): 1171-80.
Kelly, T., J. P. Dillard, et al. (1994). "Effect of genetic switching of capsular type on virulence of Streptococcus pneumoniae." Infect Immun **62**(5): 1813-9.
Lacks, S. (2004). Transformation. The Pneumococcus. E. Tuomanen, T. Mitchell, D. A. Morrison and B. G. Spratt, ASM Press.
Lacks, S., B. Greenberg, et al. (1974). "Role of a deoxyribonuclease in the genetic transformation of Diplococcus pneumoniae." Proc Natl Acad Sci U S A **71**(6): 2305-9.
Lanie, J. A., W. L. Ng, et al. (2007). "Genome sequence of Avery's virulent serotype 2 strain D39 of Streptococcus pneumoniae and comparison with that of unencapsulated laboratory strain R6." J Bacteriol **189**(1): 38-51.
Lau, G. W., S. Haataja, et al. (2001). "A functional genomic analysis of type 3 Streptococcus pneumoniae virulence." Mol Microbiol **40**(3): 555-71.

Laupland, K. B., D. B. Gregson, et al. (2009). "The changing burden of pediatric bloodstream infections in Calgary, Canada, 2000-2006." Pediatr Infect Dis J **28**(2): 114-7.
Lieu, T. A., G. T. Ray, et al. (2000). "Projected cost-effectiveness of pneumococcal conjugate vaccination of healthy infants and young children." Jama **283**(11): 1460-8.
Lode, H. M. (2007). "Managing community-acquired pneumonia: a European perspective." Respir Med **101**(9): 1864-73.
Luo, P., H. Li, et al. (2003). "ComX is a unique link between multiple quorum sensing outputs and competence in Streptococcus pneumoniae." Mol Microbiol **50**(2): 623-33.
Madsen, M., Y. Lebenthal, et al. (2000). "A pneumococcal protein that elicits interleukin-8 from pulmonary epithelial cells." J Infect Dis **181**(4): 1330-6.
Magee, A. D. and J. Yother (2001). "Requirement for capsule in colonization by Streptococcus pneumoniae." Infect Immun **69**(6): 3755-61.
Marri, P. R., W. Hao, et al. (2006). "Gene gain and gene loss in streptococcus: is it driven by habitat?" Mol Biol Evol **23**(12): 2379-91.
Medeiros, M. I., S. N. Neme, et al. (1998). "Streptococcus pneumoniae and Haemophilus influenzae as etiological agents of conjunctivitis outbreaks in the region of Ribeirao Preto, SP, Brazil." Rev Inst Med Trop Sao Paulo **40**(1): 7-9.
Mera, R., L. A. Miller, et al. (2008). "Serotype replacement and multiple resistance in Streptococcus pneumoniae after the introduction of the conjugate pneumococcal vaccine." Microb Drug Resist **14**(2): 101-7.
Messina, A. F., K. Katz-Gaynor, et al. (2007). "Impact of the pneumococcal conjugate vaccine on serotype distribution and antimicrobial resistance of invasive Streptococcus pneumoniae isolates in Dallas, TX, children from 1999 through 2005." Pediatr Infect Dis J **26**(6): 461-7.
Mitchell, T. (2004). Pneumolysin and Other Virulence Proteins. The Pneumococcus. E. Tuomanen, T. Mitchell, D. A. Morrison and B. G. Spratt, ASM Press.
Musher, D. (2004). A Pathogenic Categorization of Clincial Syndromes Caused by Streptococcus pneumoniae. The Pneumococcus. E. Tuomanen, T. J. Mitchell, D. A. Morrison and B. G. Spratt, ASM Press.
Nelson, A. L., A. M. Roche, et al. (2007). "Capsule enhances pneumococcal colonization by limiting mucus-mediated clearance." Infect Immun **75**(1): 83-90.
O'Dempsey, T. J., T. F. McArdle, et al. (1996). "Pneumococcal disease among children in a rural area of west Africa." Pediatr Infect Dis J **15**(5): 431-7.
Obert, C., J. Sublett, et al. (2006). "Identification of a Candidate Streptococcus pneumoniae Core Genome and Regions of Diversity Correlated with Invasive Pneumococcal Disease." Infect Immun **74**(8): 4766-77.
Pasteur, L. (1881). "Note sur la maladie nouvelle provoquée par la salive d'un enfant mort de la rage." Comptes Rendus **92**: 159-165.
Paton, J. C. (1996). "The contribution of pneumolysin to the pathogenicity of Streptococcus pneumoniae." Trends Microbiol **4**(3): 103-6.
Pavia, M., A. Bianco, et al. (2009). "Efficacy of pneumococcal vaccination in children younger than 24 months: a meta-analysis." Pediatrics **123**(6): e1103-10.
Pelton, S. I., H. Huot, et al. (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**(6): 468-72.
Pichichero, M. E. and J. R. Casey (2007). "Emergence of a multiresistant serotype 19A pneumococcal strain not included in the 7-valent conjugate vaccine as an otopathogen in children." Jama **298**(15): 1772-8.
Poehling, K. A., T. R. Talbot, et al. (2006). "Invasive pneumococcal disease among infants before and after introduction of pneumococcal conjugate vaccine." Jama **295**(14): 1668-74.
Polissi, A., A. Pontiggia, et al. (1998). "Large-scale identification of virulence genes from Streptococcus pneumoniae." Infect Immun **66**(12): 5620-9.

Porat, N., D. Greenberg, et al. (2006). "The important role of nontypable Streptococcus pneumoniae international clones in acute conjunctivitis." J Infect Dis **194**(5): 689-96.
Pracht, D., C. Elm, et al. (2005). "PavA of Streptococcus pneumoniae modulates adherence, invasion, and meningeal inflammation." Infect Immun **73**(5): 2680-9.
Ramirez, M., E. Severina, et al. (1999). "A high incidence of prophage carriage among natural isolates of Streptococcus pneumoniae." J Bacteriol **181**(12): 3618-25.
Robinson, D. A., K. M. Edwards, et al. (2001). "Clones of Streptococcus pneumoniae isolated from nasopharyngeal carriage and invasive disease in young children in central Tennessee." J Infect Dis **183**(10): 1501-7.
Romero-Steiner, S., T. Pilishvili, et al. (2003). "Inhibition of pneumococcal adherence to human nasopharyngeal epithelial cells by anti-PsaA antibodies." Clin Diagn Lab Immunol **10**(2): 246-51.
Romero, P., N. J. Croucher, et al. (2009). "Comparative genomic analysis of ten Streptococcus pneumoniae temperate bacteriophages." J Bacteriol **191**(15): 4854-62.
Romero, P., E. Garcia, et al. (2009). "Development of a prophage typing system and analysis of prophage carriage in Streptococcus pneumoniae." Appl Environ Microbiol **75**(6): 1642-9.
Rosenow, C., P. Ryan, et al. (1997). "Contribution of novel choline-binding proteins to adherence, colonization and immunogenicity of Streptococcus pneumoniae." Mol Microbiol **25**(5): 819-29.
Rovers, M., M. Haggard, et al. (2002). "Heritability of symptom domains in otitis media: a longitudinal study of 1,373 twin pairs." Am J Epidemiol **155**(10): 958-64.
Rovers, M. M., A. G. Schilder, et al. (2004). "Otitis media." Lancet **363**(9407): 465-73.
Rubins, J. B., A. H. Paddock, et al. (1998). "Pneumolysin in pneumococcal adherence and colonization." Microb Pathog **25**(6): 337-42.
Sampson, J. S., S. P. O'Connor, et al. (1994). "Cloning and nucleotide sequence analysis of psaA, the Streptococcus pneumoniae gene encoding a 37-kilodalton protein homologous to previously reported Streptococcus sp. adhesins." Infect Immun **62**(1): 319-24.
Sanchez-Beato, A. R., R. Lopez, et al. (1998). "Molecular characterization of PcpA: a novel choline-binding protein of Streptococcus pneumoniae." FEMS Microbiol Lett **164**(1): 207-14.
Sandgren, A., K. Sjostrom, et al. (2004). "Effect of clonal and serotype-specific properties on the invasive capacity of Streptococcus pneumoniae." J Infect Dis **189**(5): 785-96.
Schwartz, D. C. and C. R. Cantor (1984). "Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis." Cell **37**(1): 67-75.
Shouval, D. S., D. Greenberg, et al. (2006). "Site-specific disease potential of individual Streptococcus pneumoniae serotypes in pediatric invasive disease, acute otitis media and acute conjunctivitis." Pediatr Infect Dis J **25**(7): 602-7.
Silva, N. A., J. McCluskey, et al. (2006). "Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates." Infect Immun **74**(6): 3513-8.
Sjostrom, K., C. Blomberg, et al. (2007). "Clonal success of piliated penicillin nonsusceptible pneumococci." Proc Natl Acad Sci U S A **104**(31): 12907-12.
Sjostrom, K., C. Spindler, et al. (2006). "Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen." Clin Infect Dis **42**(4): 451-9.
Sleeman, K. L., L. Daniels, et al. (2005). "Acquisition of Streptococcus pneumoniae and nonspecific morbidity in infants and their families: a cohort study." Pediatr Infect Dis J **24**(2): 121-7.
Sleeman, K. L., D. Griffiths, et al. (2006). "Capsular Serotype-Specific Attack Rates and Duration of Carriage of Streptococcus pneumoniae in a Population of Children." J Infect Dis **194**(5): 682-8.
Smith, B. L. and M. K. Hostetter (2000). "C3 as substrate for adhesion of Streptococcus pneumoniae." J Infect Dis **182**(2): 497-508.

Spratt, B. G., W. P. Hanage, et al. (2001). "The relative contributions of recombination and point mutation to the diversification of bacterial clones." Curr Opin Microbiol **4**(5): 602-6.
Sternberg, G. M. (1881). "A fatal form of septicaemia in the rabbit, produced by the subcutaneous injection of human saliva." Natl Board of Health Bull **2**: 781-783.
Talkington, D. F., B. G. Brown, et al. (1996). "Protection of mice against fatal pneumococcal challenge by immunization with pneumococcal surface adhesin A (PsaA)." Microb Pathog **21**(1): 17-22.
Tan, T. Q., E. O. Mason, Jr., et al. (2002). "Clinical characteristics of children with complicated pneumonia caused by Streptococcus pneumoniae." Pediatrics **110**(1 Pt 1): 1-6.
Teele, D. W., J. O. Klein, et al. (1989). "Epidemiology of otitis media during the first seven years of life in children in greater Boston: a prospective, cohort study." J Infect Dis **160**(1): 83-94.
Tenover, F. C., R. D. Arbeit, et al. (1995). "Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing." J Clin Microbiol **33**(9): 2233-9.
Tettelin, H. and S. K. Hollingshead (2004). Comparative Genomics of Streptococcus pneumoniae: Intrastrain Diversity and Genome Plasticity. The Pneumococcus. E. Tuomanen, T. Mitchell, D. A. Morrison and B. G. Spratt, ASM Press.
Tettelin, H., K. E. Nelson, et al. (2001). "Complete genome sequence of a virulent isolate of Streptococcus pneumoniae." Science **293**(5529): 498-506.
Tomasz, A., P. Moreillon, et al. (1988). "Insertional inactivation of the major autolysin gene of Streptococcus pneumoniae." J Bacteriol **170**(12): 5931-4.
Underdown, B. J. and J. M. Schiff (1986). "Immunoglobulin A: strategic defense initiative at the mucosal surface." Annu Rev Immunol **4**: 389-417.
Weber, J. (2004). Pathogenesis of Pneumococcal Meningitis. The Pneumococcus. E. Tuomanen, T. Mitchell, D. A. Morrison and B. G. Spratt, ASM Press.
Weiser, J. N., R. Austrian, et al. (1994). "Phase variation in pneumococcal opacity: relationship between colonial morphology and nasopharyngeal colonization." Infect Immun **62**(6): 2582-9.
Wenger, J. D., A. W. Hightower, et al. (1990). "Bacterial meningitis in the United States, 1986: report of a multistate surveillance study. The Bacterial Meningitis Study Group." J Infect Dis **162**(6): 1316-23.
Vila-Corcoles, A., O. Ochoa-Gondar, et al. (2006). "Protective effects of the 23-valent pneumococcal polysaccharide vaccine in the elderly population: the EVAN-65 study." Clin Infect Dis **43**(7): 860-8.
Williams, B. G., E. Gouws, et al. (2002). "Estimates of world-wide distribution of child deaths from acute respiratory infections." Lancet Infect Dis **2**(1): 25-32.
Winkelstein, J. A. (1984). "Complement and the host's defense against the pneumococcus." Crit Rev Microbiol **11**(3): 187-208.
Witte, W., C. Cuny, et al. (2008). "Emergence and spread of antibiotic-resistant Gram-positive bacterial pathogens." Int J Med Microbiol **298**(5-6): 365-77.
www.rivm.nl/earss.
www.rproject.org.
Yother, J. (2004). Capsules. The pneumococcus. E. Tuomanen, T. Mitchell, D. A. Morrison and B. G. Spratt, ASM Press.
Zenni, M. K., S. H. Cheatham, et al. (1995). "Streptococcus pneumoniae colonization in the young child: association with otitis media and resistance to penicillin." J Pediatr **127**(4): 533-7.
Zighelboim, S. and A. Tomasz (1980). "Penicillin-binding proteins of multiply antibiotic-resistant South African strains of Streptococcus pneumoniae." Antimicrob Agents Chemother **17**(3): 434-42.