Molecular Epidemiology of *Streptococcus agalactiae*: Mobile Elements as Genetic Markers

Shi-Lu Luan

Department of Clinical Microbiology, Biomedical Laboratory Science, Clinical Bacteriology, Umeå University, Umeå, Sweden
2006
Doctoral Dissertation 2006

Copyright © Shi-Lu Luan

Printed in Sweden by Print & Media AB
Umeå, 2006
## CONTENTS

CONTENTS ............................................................................................................................... 3

PAPERS IN THIS THESIS ........................................................................................................ 5

ABBREVIATIONS ................................................................................................................ 6

ABSTRACT ............................................................................................................................ 7

INTRODUCTION .................................................................................................................. 8

*Streptococcus agalactiae* ..................................................................................................... 8

*S. agalactiae* diseases ........................................................................................................ 8

*S. agalactiae* colonization, transmission, risk factors and prevention efforts ................. 9

*S. agalactiae* serotype distribution ................................................................................ 11

*S. agalactiae* genome sequence ..................................................................................... 12

Major virulence factors ..................................................................................................... 13

Horizontal gene transfer .................................................................................................. 18

Mobile genetic elements ................................................................................................. 19

Insertion sequences .......................................................................................................... 19

Group II intron ................................................................................................................... 21

Horizontal gene transfer in *S. agalactiae* ..................................................................... 24

Insertion sequences in *S. agalactiae* ............................................................................. 25

Group II intron identified in *S. agalactiae* ..................................................................... 27

Molecular epidemiology .................................................................................................... 27

Conventional phenotypic typing methods ....................................................................... 29

Genotypic strain typing ................................................................................................... 30

Definitions ........................................................................................................................ 36
AIMS OF THE STUDY ........................................................................................................... 37

RESULTS AND DISCUSSION .............................................................................................. 38

Identification of two additional GBSi1 copies ................................................................. 38
Uneven distribution of GBSi1 in *S. agalactiae* ................................................................ 39
The putative target motif of GBSi1 .................................................................................. 39
A unique DNA fragment with plasmid features is correlated with GBSi1 ...................... 41
Genetic lineages of clinical *S. agalactiae* isolates ........................................................ 42
Clonal relationships among invasive and non-invasive isolates and clone/clonal complex-specific odds ratios ........................................................................................................ 46
Clonal divergence of *S. agalactiae* ................................................................................. 48
Serotype distribution and serotype-specific odds ratio .................................................... 50
Relationship between genotype and capsular serotype .................................................. 51
Distribution of mobile genetic elements ........................................................................ 53
Distribution of the *fbsA* and *spb1* genes among serotype III isolates ...................... 54

CONCLUSIONS ...................................................................................................................... 57

ACKNOWLEDGEMENTS ..................................................................................................... 58

REFERENCES ......................................................................................................................... 61

PAPER I-III
PAPERS IN THIS THESIS

This thesis is based on the following articles and manuscripts, which are referred to in the text by their roman numerals (I-III).

   An inserted DNA fragment with plasmid features is uniquely associated with the presence of the GBSi1 group II intron in \textit{Streptococcus agalactiae}.
   Gene \textit{312}, 305-312.

   Multilocus sequence typing of Swedish invasive group B streptococcus isolates indicates a neonatally associated genetic lineage and capsule switching. \textit{J Clin Microbiol} \textit{43}, 3727-3733.

   Serotype- and Clonal Complex-Specific Differences in Disease Potential of \textit{Streptococcus agalactiae}. (Submitted manuscript).

Published papers are reproduced by permission from Elsevier Science B.V. and the American Society for Microbiology.
ABBREVIATIONS

aa Amino acids
bp Base pair
CC Clonal complex
EOD Early-onset disease
GBS Group B streptococcus
GBSi1 Group II intron in *Streptococcus agalactiae*
HGT Horizontal gene transfer
IEP Intron encoded protein
IS Insertion sequence
kb Kilo base pairs
LOD Late-onset disease
Mb Million base pairs
MGE Mobile genetic elements
MLEE Multilocus enzyme electrophoresis
MLST Multilocus sequence typing
nt Nucleotide
OR Odds ratio
ORF Open reading frame
PCR Polymerase chain reaction
PFGE Pulsed-field gel electrophoresis
REA Restriction endonuclease analysis
RFLP Restriction fragment length polymorphism
SLV Single-locus variant
ST Sequence type
ABSTRACT

Streptococcus agalactiae, also designated group B streptococcus (GBS), is a Gram-positive coccus, and it is an important pathogen that causes invasive disease in neonates, pregnant adults, and non-pregnant adults with predisposing conditions. The group II intron GBSi1 is one of the major mobile genetic elements identified in S. agalactiae. The aim of this thesis was to characterize the GBSi1 distribution pattern, the population structure, and the influence of serotype- and clone-specific properties on the invasive capacity among clinical invasive and non-invasive isolates of S. agalactiae.

Two additional copies of GBSi1 were identified at sites different from the primarily identified scpB-lmb locus. The distribution of GBSi1 was uneven among different serotypes. Three intron copies were only found in isolates of serotype III, and these targeted all the three identified gene loci. In contrast, a single copy of GBSi1 was found in isolates of serotype II and V and only located at the scpB-lmb locus. Furthermore, at the 5' flanking region of the scpB-lmb gene locus, a novel 2.1 kb DNA fragment with plasmid features was identified only in intron carrying isolates. This may suggest that GBSi1 once was brought into the S. agalactiae genome by an integrated plasmid.

Multilocus sequence typing was used to characterize totally 314 invasive and non-invasive S. agalactiae isolates collected in Northern and Western Sweden from the years 1988 to 2004. Five major genetic lineages (clonal complexes) were identified among both invasive and non-invasive isolates, including serotype Ia, Ib, and II to V, indicating a clonal population structure of S. agalactiae isolates. A number of genetically highly related isolates were found to express different capsular types, suggesting that capsule switching occurs rather frequently between isolates. Furthermore, non-invasive isolates belonging to the same clonal complexes displayed more heterogeneity in capsule expression as well as in the distribution patterns of mobile genetic elements than invasive isolates. This indicates that less variability is allowed in a highly selective environment such as the blood. All major clonal complexes and serotypes caused invasive disease, although their ability to do so varied greatly. CC17 was significantly associated with neonatal invasive disease; whereas CC19 was equally common among isolates from adult and neonatal disease, despite that both CC17 and CC19 expressed capsular type III. This striking difference seen between CC17 and CC19 suggests that clonal complex associated properties, in addition to capsular type, play important roles in the virulence of S. agalactiae. CC1, a new emerging clone since early 1990s, has caused substantial amount of disease among adults. In addition, mutually exclusive distribution of mobile elements GBSi1 and IS1548 was seen, and they were shown to constitute genetic markers for serotype III CC17 and CC19 isolates, respectively.

Keywords: Streptococcus agalactiae, group II intron (GBSi1), multilocus sequence typing, capsular serotype, population structure
INTRODUCTION

*Streptococcus agalactiae*

*Streptococcus agalactiae*, also designated group B *strep*tococcus (GBS), is a Gram-positive coccus of the genus Streptococcus. *S. agalactiae* was originally known for causing bovine mastitis, and it was first classified in the 1930s by Lancefield and Hare in the studies on the serological differentiation of streptococci (Lancefield, 1935). *S. agalactiae* emerged as an important human pathogen during the early 1970s, and has remained as the leading cause of mortality and morbidity among neonates (Baker, 2000). *S. agalactiae* is also an important pathogen in pregnant women and adults with underlying diseases.

*S. agalactiae* diseases

**In neonates.** The invasive *S. agalactiae* disease in neonates commonly presents as bacteremia, meningitis, or pneumonia. The incidence of infection among neonates had ranged between 1.0 and 4.7 per 1000 live births in the United States, and incidence rates for most European countries vary between 0.2 and 2 per 1000 live births (Baker, 2000; Trijbels-Smeulders *et al.*, 2004). After the introduction of a *S. agalactiae* screening program in the United States during the 1990s, the incidence of *S. agalactiae* infection in neonates has been reduced to 0.35 per 1000 live births during the last five years (Prevention, 2005).

There are two distinct syndromes of *S. agalactiae* disease, early-onset (EOD) and late-onset disease (LOD), and EOD accounts for 80% of all *S. agalactiae* cases. The EOD develops during the first week of life, usually within the first 24 hours, and is dominated by sepsis and pneumonia (Baker, 1978). EOD is due to an ascending infection from the genital tract or colonization of the infant during delivery. Colonization of the airways of the neonate is rapidly followed by the development of
pneumonia. Breaching of the pulmonary mucosal barrier leads to the entry of *S. agalactiae* into the intravascular space and to the development of severe sepsis. Case-fatality rates for EOD in United States are estimated at 5% (Dermer *et al.*, 2004). Late-onset disease (LOD) usually occurs from one week to three months after delivery, and presents with meningitis as the predominant clinical syndrome (Yagupsky *et al.*, 1991). In this setting the entry of bacteria into the bloodstream is followed by the invasion of the cerebrospinal fluid space. Fatality rate of LOD is reported to range from 2 to 6%. However, a substantial proportion of neonates, especially among meningitis survivors, suffers long-term sequelae such as retarded mind development, deafness and speech delay.

**In adults.** During pregnancy, the primary clinical manifestation of *S. agalactiae* disease is bacteremia, but chorioamnionitis, endometritis and septic abortion also occur (Schrag *et al.*, 2000a; Schrag *et al.*, 2000b). Among non-pregnant adults, *S. agalactiae* can cause urosepsis, pneumonia, as well as skin and soft tissue infections (Farley, 2001). Of note, the incidence of disease among non-pregnant adults, particularly among elderly with underlying conditions, has increased over the past decade. The annual incidence rate was reported to range from 25 to 30 cases per 100,000 adults of age over 65 in 2003 (Edwards and Baker, 2005; Edwards *et al.*, 2005).

*S. agalactiae* colonization, transmission, risk factors and prevention efforts

**Colonization.** *S. agalactiae* exists as normal flora in the genitourinary and gastrointestinal tracts, and 16-40% of healthy adults are asymptptomatically colonized with *S. agalactiae* (Brimil *et al.*, 2006; Hickman *et al.*, 1999; Jones *et al.*, 2006; Manning *et al.*, 2004; Schuchat and Wenger, 1994). The prevalence of colonization among healthy elderly adults (about 25%) is similar to that among women of
childbearing age (Edwards et al., 2005). About half of the neonates born to S. agalactiae carrying mothers become colonized with S. agalactiae. Newborns over 48 hours old are most commonly colonized in the throat, umbilicus, and rectum; and can remain colonized throughout childhood (Hickman et al., 1999).

**Transmission.** Besides being acquired from the mother during labor and delivery as in EOD, S. agalactiae transmission can occur via direct contact with other individuals. This is the proposed way for transmission in LOD and among adults. Transmission is hypothesized to occur via the fecal-oral route or by person-to-person contact (Manning et al., 2000).

**Risk factors.** Maternal risk factors associated with the development of neonatal S. agalactiae disease, such as heavy S. agalactiae colonization, premature delivery, prolonged rupture of membranes, intrapartum fever, S. agalactiae urinary tract infection, history of S. agalactiae infection in a previous child, and African American ethnicity, have been identified (Schuchat et al., 1994). Premature, low birth weight infants are at greatly increased risk of contracting S. agalactiae infection (Schuchat and Wenger, 1994). Among non-pregnant adults, most invasive disease occur in patients with underlying conditions including diabetes, malignancy or high age (Farley, 2001).

**Prevention efforts.** The incidence of neonatal S. agalactiae disease has declined considerably in the United States during the 1990s since the introduction of surveillance programs and intrapartum antibiotic prophylaxis (Moore et al., 2003; Schrag et al., 2000a; Schrag et al., 2002). In Europe there are different prevention strategies, and only a few countries have adopted a screening program for prevention of neonatal S. agalactiae infection (Trijbels-Smeulders et al., 2004). In the USA the incidence of EOD has decreased by 70%, as a result of the prophylactic measures. In contrast, the overall rate of LOD showed little change (Schrag et al., 2000a). On the other hand, widespread use of antibiotic prophylaxis gives rise to concerns about
There is an urge to develop an alternative prevention protocol, a vaccine. There are a number of candidates including capsular conjugate vaccines using traditional carrier proteins such as tetanus toxoid (TT) and mutant diphtheria toxin CRM197, as well as *S. agalactiae* specific proteins such as C5a peptidase. The capsular polysaccharide (CPS)-TT conjugates have advanced most, and have completed clinical phase 1 and phase 2 trials. The conjugate vaccines prepared with CPS from *S. agalactiae* types Ia, Ib, II, III, and V have proved to be safe and highly immunogenic in healthy adults (Paoletti and Kasper, 2003). In particular, the safety and immunogenicity of *S. agalactiae* serotype V-TT conjugate vaccine shown in healthy elderly persons suggest the potential for vaccination to prevent the increase of invasive *S. agalactiae* infections in elderly persons. To date, a number of potential protective antigens have been identified, using classic molecular approaches and multiple genome screening, including the Alp protein family, Sip, and the C5a-peptidase (Cheng et al., 2002a; Cheng et al., 2002b; Lindahl et al., 2005; Maione et al., 2005). However, no definitive protein conjugates have yet been on trial. Recently, *Lactococcus lactis* expressing pili from *S. agalactiae* has been reported to elicit protective immunity and may serve as a live vaccine (Buccato et al., 2006). Although these achievements in vaccine development are promising, substantial regulation and legal challenges remain to be overcome before any commercial *S. agalactiae* vaccine becomes available (Heath and Feldman, 2005).

**S. agalactiae** serotype distribution

Capsular serotyping has been the classic method used in descriptive epidemiology of *S. agalactiae*. According to capsular expression, *S. agalactiae* are subclassified into nine serotypes, Ia, Ib, and II through VIII. There are demographic, geographic, and temporal variations with respect to the predominant serotypes present in the human
population (Blumberg et al., 1996; Hickman et al., 1999; Kieran et al., 1998). Serotype VI and VIII were reported to dominate among pregnant woman in Japan, but were seldom found in other populations (Lachenauer et al., 1999). Nevertheless, investigations from the United States and Sweden have shown that three serotypes (Ia, III, and V) account for 80-96% of EOD in newborns (Harrison et al., 1998; Persson et al., 2004; Shet and Ferrieri, 2004). LOD is dominated by serotype III followed by serotype Ia and V (Harrison et al., 1998). Serotype V emerged during the early 1990s, and has been described as the predominant capsule type recovered from nonpregnant adults with invasive *S. agalactiae* disease in the USA. Additionally, serotype V was the most common serotype (47%) among the colonizing isolates from elderly people. Through 1992-99, among nonpregnant people with invasive disease, serotype Ia, V and III accounted for two third of the isolates in the USA (Edwards et al., 2005).

**S. agalactiae** genome sequence

The year 1995 marked the start of a genomic era with publications on the complete genome sequence of two bacterial pathogens, *Haemophilus influenzae* (Fleischmann et al., 1995) and *Mycoplasma genitalium* (Fraser et al., 1995); since then, more than three hundred bacterial genomes have been fully sequenced. In 2002 two complete genome sequences of *S. agalactiae* were released for a serotype III strain and a serotype V strain (Glaser et al., 2002; Tettelin et al., 2002). The genome sequences of additional six *S. agalactiae* invasive strains were reported in 2005, which include serotype Ia, Ib, II, III, and V (Tettelin et al., 2005).

The *S. agalactiae* genome is around 2 Mb with low G/C content of about 36%, and was predicted to comprise more than 2,100 protein-encoding genes. Of these encoding genes, there have been 14-15 pathogenicity island-like regions identified, which are dispersed around the genome. These islands contain genes encoding known and putative virulence genes, mostly predicted surface proteins, as well as a number of genes related to mobile elements such as prophages, plasmids and conjugative
transposons. The islands may have an important role in acquisition of virulence factors and in genetic diversification of the species (Glaser et al., 2002; Tettelin et al., 2002).

Analysis of the eight *S. agalactiae* genomes revealed a pan-genome consisting of a core genome shared by all isolates, accounting for approximately 80% of the genome, plus a dispensable genome consisting of partially shared and strain-specific genes (Tettelin et al., 2005). Importantly, *S. agalactiae* appears to have a rather large flexible gene pool, and an average of 33 new strain-specific genes is predicted to be identified when a new *S. agalactiae* strain is sequenced. The genetic heterogeneity among GBS isolates is also revealed by whole-genome hybridization using microarrays (Brochet et al., 2006; Tettelin et al., 2002). The information of *S. agalactiae* genomes suggests that gene acquisition, duplication, and reassortment have played an important role in genetic diversity and evolution of *S. agalactiae*.

The genetic relatedness among pathogenic streptococci species was demonstrated by comparison of the *S. agalactiae* genome with that of *Streptococcus pyogenes* (GAS) and *Streptococcus pneumoniae*. Half of the predicted *S. agalactiae* proteins have homologs in either of the two other species, while 683 genes are unique to *S. agalactiae*. Moreover, the chromosomal order of genes is highly conserved between *S. agalactiae* and *S. pyogenes*, stressing the relatedness between these two species, while the gene order shows low conservation between *S. agalactiae* and *S. pneumoniae*, possibly reflecting the importance of transformation and recombination in the evolution of *S. pneumoniae* (Tettelin et al., 2001; Tettelin et al., 2002).

**Major virulence factors**

The broad spectrum of *S. agalactiae* disease manifestations reflects a complex interplay between the host’s innate and adaptive immune systems and bacterial virulence factors. In particular in a newborn there are quantitative and qualititative deficiencies in phagocytes, specific antibodies, and complement components of the
immune system. On the other hand, *S. agalactiae* is part of the normal flora, and thus an opportunistic pathogen. *S. agalactiae* virulence traits mainly include: (a) factors that inhibit immunological clearance; (b) adhesin or invasin that enables penetration of epithelial and endothelial cellular barriers to reach the bloodstream and deeper tissues; (c) toxins that directly injure or disrupt host tissue components.

**Capsule.** The foremost virulence factor of *S. agalactiae* is the surface polysaccharide capsule that inhibits complement-mediated killing. The vast majority of *S. agalactiae* isolates from invasive disease are encapsulated, which provides the basis for the classification of *S. agalactiae* into nine serotypes. As early as in 1976, deficiency in maternal antibody against polysaccharide was found to correlate with susceptibility to neonatal *S. agalactiae* infection (Baker and Kasper, 1976). The role of capsule in virulence was directly addressed by Rubens et al, who found that the nonencapsulated transposon mutant of *S. agalactiae* showed significantly reduced virulence in a neonatal rat model (Rubens *et al.*, 1987). The terminal sialic acid present on the capsule of all serotypes (sialylation of capsule) is crucial for the pathogenic property of CPS, as loss of capsular sialic acid was associated with loss of virulence in the mutant strain (Wessels *et al.*, 1989).

The CPS structure among the serotypes differs in the arrangements of four sugar components into repeating units (Cieslewicz *et al.*, 2005). This structural diversity is associated with genetic diversity of the capsular biosynthesis cluster. The capsular biosynthesis cluster (operon) in serotype III consists of 16 different genes and the gene order in this locus is conserved across all serotypes. The conserved structure of CPS may confer survival advantage, and thus counter balances the antigenic variation provoked by host immunity. Expression of capsule by *S. agalactiae* is regulated by growth rate, and can also be affected by environmental factors such as supplement of human sera in the growth medium (Paoletti *et al.*, 1996; Platt *et al.*, 1994).

**Beta-hemolysin/Cytolysin.** A hallmark of the *S. agalactiae* phenotype is the appearance of beta-hemolysis surrounding colonies growing on a blood agar plate. In
pneumonia caused by *S. agalactiae*, the expression of hemolysin is associated with injury in the lung epithelial cells, which accounts for pulmonary damage, and electron microscopy suggest that the hemolysin acts as a pore-forming cytolysin (Nizet et al., 1996; Nizet et al., 1997a; Nizet et al., 1997b). Hemolysin also induces cytolysis and apoptosis of the phagocytes (Liu et al., 2004). The genetic basis of hemolysin was unraveled by two groups independently by adopting different molecular approaches (Pritzlaff et al., 2001; Spellerberg et al., 1999a; Spellerberg et al., 2000). Spellerberg et al. first discovered the *cyl* operon of 7 kb via analysis of nonhemolytic isogenic mutants, and demonstrated that several genes belonging to the *cyl* operon, namely *acpC* and *cylZABE*, are required for hemolysin production of *S. agalactiae* (Spellerberg et al., 1999a). Subsequent investigation by Pritzlaff et al. utilized a positive approach of expressing a plasmid library of the *S. agalactiae* chromosome in *E. coli*, and indicated that a single ORF, *cylE*, is necessary and sufficient for beta-hemolytic and cytolysin activity (Pritzlaff et al., 2001).

**C5a-peptidase (ScpB).** C5a-peptidase is a surface-localized serine protease, which impairs the leukocyte recruitment to sites of infection through cleaving the chemoattractant C5a (Wexler et al., 1985). Recently it was shown that ScpB is capable of binding to fibronectin (Fn). This property may play a role in promoting cell invasion as a ScpB negative mutant showed a reduced ability to invade human epithelial cells *in vitro* (Cheng et al., 2002b). The *scpB* gene is widely distributed in the genus Streptococcus. It was first identified in *S. pyogenes*, and subsequently also in *S. agalactiae*, group C and G streptococcal isolates. Of note, *scpB* is highly restricted to human isolates including all the serotypes of *S. agalactiae*, and only found in a minority of bovine isolates (Dmitriev et al., 1999; Franken et al., 2001). Importantly, immunization with ScpB or ScpB-type III polysaccharide conjugate vaccines enhances clearance of *S. agalactiae* from lungs of infected mice (Cheng et al., 2002a).

**Lmb.** The Lmb protein is identified as a laminin-binding protein, and may play a role in colonization and invasion of damaged epithelium and translocation of bacteria into
the bloodstream (Spellerberg et al., 1999b). Lmb shows homology to the streptococcal LraI protein family, which have been implicated in adhesion and metal transport. The gene lmb is present in all of the common serotypes of S. agalactiae, and it has also been identified in S. pyogenes, with the sequence being virtually identical in different strains both within and between the two species (Elsner et al., 2002; Terao et al., 2002). However, the ability of the Lmb protein to bind laminin and promote binding to tissues remains controversial, as other studies have failed to detect the binding with recombinant protein and whole bacteria (Elsner et al., 2002; Terao et al., 2002). Evidence is also lacking that Lmb participates in metal transport like other LraI proteins. Our unpublished data shows that Lmb does not elicit protective immunity in mice, probably due to its poor exposure on the surface of S. agalactiae.

**The fibrinogen-binding protein FbsA.** The FbsA is a principal fibrinogen (Fgn) receptor in S. agalactiae, and its importance in virulence has been unfolded by a few recent studies. FbsA has a typical structure of a surface protein, a signal peptide, a repeat region containing various number of 16 aa repeats, and a C-terminal cell wall anchor region with LPXTG motif (Schubert et al., 2002). Even a single repeat unit of FbsA can bind to fibrinogen, and the binding capacity increases with the number of the repeats. Expression of FbsA can be specifically up-regulated by RogB, a protein with significant similarity to members of the RofA-like protein family of transcriptional regulators (Gutekunst et al., 2003). FbsA was first shown to protect S. agalactiae from opsonophagocytosis as growth of a fbsA deletion mutant in human blood was significantly impaired (Schubert et al., 2002). FbsA mediates large-scale aggregation of human plasma fibrinogen, which leads to the formation of a thick layer on the bacterial cell wall itself that becomes an efficient mask against phagocytosis (Pierno et al., 2006). Of note, FbsA also promotes adherence to and invasion of human brain microvascular endothelial cells, indicating that FbsA plays an important role in S. agalactiae caused meningitis (Tenenbaum et al., 2005). Moreover, the presence of an intact copy of fbsA appears to correlate with a neonatally associated virulent lineage of serotype III (Brochet et al., 2006; Luan et al., 2005).
**Alp family surface proteins.** The Alp family proteins are surface associated proteins that contain a number of repetitive sequences and elicit protective immunity in mice. However, the proteins in the Alp family have little or no effect on virulence in the mouse model of invasive infection. Alp proteins have been extensively studied by the research group of G. Lindahl, and characteristics of this family have been fully summarized in a recent review (Lindahl et al., 2005). Four members of the Alp family have so far been identified: the α, Rib, R28, and Alp2 proteins. The Alp proteins are encoded by allelic genes and the expression of a given Alp family protein is strongly correlated with capsular type. All purified Alp proteins elicit protective immunity in mice, and variation in number of repeats appears to allow *S. agalactiae* to escape host immunity. The immunological cross activities are very limited among Alp family members despite their rather high homologies, and cross activity is detected only between the α protein and Alp2 in addition to R28 and Rib.

Except the Alp family protein, other surface proteins may also have essential roles in the protective immunity. Multiple genome screen of *S. agalactiae* isolates including serotype Ia, Ib, II, III and V identified nearly 600 predicted surface proteins, among which 312 proteins were expressed in *E. coli* and tested for induced passive protection in a mouse maternal immunization model (Maione et al., 2005). Four proteins, including Sip from the core genome and three novel proteins (GBS67, GBS80, and GBS104) from the variable gene pool elicits promising protection, and the combination of these four proteins gives rise to high protection against all strains ranging from 59% to 100%.

**Spb1.** The *spb1* gene was identified by subtractive hybridization from a serotype III strain of the putative hypervirulent clone ET1/III-3 (Adderson et al., 2003), and the gene sequence of *spb1* indicates characteristics of a surface protein. The Spb1 protein is mainly involved in bacterial internalization into host cells as a Spb1-negative mutant was significantly reduced in the ability to invade such cells, but showed little difference in adhering to epithelial cells (Adderson et al., 2003).
**Pili.** The proteins GBS80 and GBS104 are protective antigens identified in multiple genome screening, and they can form high molecular weight polymers, which have a pilus-like form revealed by immunogold electron microscopy (Lauer et al., 2005). Similar pilus structure has also been identified in GAS, and as in *S. agalactiae* pili appear to be formed by a sortase-mediated covalent polymerization of proteins containing the LPXTG motif and incorporation of accessory proteins to the pilus backbone (Dramsi et al., 2006; Mora et al., 2005).

**Hyaluronidase.** Hyaluronidase can cleave hyaluronic acid, a major component of the connective tissue, and therefore facilitates bacterial dissemination. However, in a group of clinical isolates recovered from patients with endocarditis an insertion sequence IS1548 was found inserted in the hyaluronidase encoding gene *hylB*, which caused loss of hyaluronidase activity (Granlund et al., 1998). This finding puts a question mark at the role of hyaluronidase in *S. agalactiae* pathogenesis since its activity is not required for all infections.

**Horizontal gene transfer**

Horizontal (lateral) gene transfer (HGT) is a process of acquiring genetic material from one genome to another within or between species, for example the transmission of antibiotic resistance genes among various bacteria, including *S. pneumoniae*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa* (Monroe and Polk, 2000). In prokaryotes DNA can be transferred between cells by three ways, transformation, transduction, and conjugation. The size of the genetic material transferred may range from a few kilobases in natural transformation to several ten-kilobases in phage-mediated transduction, and potentially hundreds of kilobases in conjugation. It is only through recent comparative analysis of multiple prokaryotic genomes that the degree of HGT and the substantial impact of HGT on genetic diversity have become evident. The degree of HGT varies remarkably among species, and provides an important
mechanism for microbes in genetic adaptation to the environment, and has a significant biological impact. Many pathogenic properties are encoded on plasmids, phages or pathogenicity islands and the transfer of such genetic elements is generally thought to be associated with the emergence of pathogenic clones.

Mobile genetic elements

Mobile genetic elements (MGEs) are genomic elements that are capable to translocate themselves within or between genomes. Their size ranges from hundreds of base pairs to more than 100 kbp. The degree of genomic flexibility is dependent on the content of MGEs such as insertion sequences (IS), conjugative transposons, plasmids, and bacteriophages (Kolsto, 1997). Many MGEs are identified in bacterial genome sequences, and they often carry virulence genes required for bacterial pathogenicity and antibiotic resistance. Regardless if MGEs have their own replicating machinery, or become integrated into the host genome, they confer new characters to the host, promote host genome rearrangement (including deletions, inversions and replicon fusions), regulate or affect gene expression, and contribute significantly to environmental adaptation, and genomic diversity in the host bacteria. The following section will focus on two MGEs, namely, insertion sequence (IS) elements and Group II introns, which are prevalent in S. agalactiae and appear to play an important role in evolution of GBS genetic lineages.

Insertion sequences

Insertion sequences (IS) are small DNA sequences, generally less than 2.5 kb in length, encoding functions involved in their own translocation, which can transpose both within and between genomes (Mahillon et al., 1999). They were first detected in 1950 by their ability to generate mutants by insertional inactivation. To date several
hundred ISs have been identified from numerous bacterial and archaeal species, and they fall into over 20 distinct groups (Mahillon et al., 1999). An IS database can be accessed at http://www-is.biotoul.fr/is.html.

Usually, an IS contains a transposase gene, flanked by inverted repeats up to 40 bp. In addition, the IS itself is often flanked also by short, directly repeated (DR) sequences, which are generated upon insertion into the recipient DNA (Mahillon and Chandler, 1998). The organization of a typical IS is illustrated in Figure 1. ISs are the simplest and smallest mobile elements, and they are key components in bigger transposable elements. They are the means by which genomic islands such as pathogenicity islands become transferable, and they often play a role in plasmid integration (Rocha et al., 1999).

**Figure 1.** Organization of a typical insertion sequence. The IS is represented as an black box in which the terminal inverted repeats are shown as dark grey boxes labeled IRL (left inverted repeat) and IRR (right inverted repeat). An open reading frame encoding the transposase (black box) is located in the IS. DR boxes flanking the IS represent short directly repeated sequences (DR) generated in the target DNA as a consequence of insertion. The transposase promoter is localized in IRL. Adapted from Galun (Galun, 2003).

Furthermore, ISs are of particular importance for genome flexibility, as they provide multiple similar sequences for initiating homologous recombination. They are capable of promoting various types of genome rearrangement including deletions, inversions,
and replicon fusions and regulating gene expression (inactivation and antigenic variation by switch on/off), as well as spreading antibiotic resistance genes. Members of this heterogeneous class of mobile genetic elements are capable of promoting various types of genome rearrangement including deletions, inversions, and replicon fusions.

**Group II intron**

Group II introns are genetic elements that have properties of both catalytic RNAs and retroelements. Group II introns were originally discovered in the mitochondrial and chloroplast genomes of lower eukaryotes and higher plants, where they interrupt conserved genes. The first bacterial introns were discovered by PCR screen in 1993, and to date over 80 complete group II introns have been reported. About one quarter of the sequenced bacterial genomes including both Gram-positive and Gram-negative bacteria contains group II introns (Dai and Zimmerly, 2002a; Toro, 2003). The complete group II intron database can be accessed at [http://www.fp.ucalgary.ca/group2introns/](http://www.fp.ucalgary.ca/group2introns/). Mobility of group II introns occur mainly by site-specific insertions (retrohoming), or to a less extent by retrotransposition. Via retrohoming or ‘homing’ group II intron invades specific DNA target sites, typically the unoccupied site in an intronless allele at frequencies approaching 100% (Cousineau *et al.*, 1998; Skelly *et al.*, 1991). In addition, group II introns may ‘retrottranspose’ at low frequencies, $10^{-4}$ to $10^{-5}$, into ectopic sites that resemble the normal homing site. (Dickson *et al.*, 2001; Ichiyanagi *et al.*, 2002).

Group II introns consist of a multifunctional protein encoding gene flanked by ‘RNA domain’ sequence, and the entire size of an intron ranges between a few hundred to a couple of thousand nucleotides. The transcript of a group II intron can fold into a conserved structure with six domains, which has catalytic activity of a ribozyme and carries out RNA splicing and reverse splicing (integration) reactions. This folding requires help from the intron-encoded protein (IEP), which has reverse transcriptase,
RNA maturase, and in some cases DNA endonuclease activities. The mechanism underlying mobility of group II introns have been well described for the yeast mitochondrial introns al1 and al2, and for L1.ltrB in *Lactococcus lactis* (Cousineau *et al*., 1998). Mobility occurs by a remarkable target DNA-primed reverse transcription mechanism in which the excised intron RNA reverse splices directly into a DNA target site and is then reverse transcribed by the IEP. The integration of cDNA is promoted by recombination or repair mechanisms independent of homologous recombination, which may differ among organisms (Cousineau *et al*., 1998; Eskes *et al*., 2000). The endonuclease activity of IEP is not essential, as the group II intron of *Sinorhizobium meliloti*, which lacks the endonuclease domain of the IEP, is mobile *in vivo* (Martinez-Abarca *et al*., 2000; Martinez-Abarca and Toro, 2000).

**Figure 2.** Intron-encoded ORF structure. A typical ORF structure is shown. The largest domain (RT) contains the seven subdomains common to all reverse transcriptases (0–7). Additional domains are domain X, which contributes to a splicing, or maturase, function, and the Zn domain, which has an endonuclease activity utilized in mobility. The Zn domain is absent from many bacterial intron ORFs. The six RNA structural domains surround the ORF and are flanked by exon sequences (E1 and E2). Adapted from Dai *et al* (Dai *et al*., 2003).

An interesting issue concerning bacterial group II introns is whether they behave like retroelements or introns. As group II introns are often inserted outside of genes, after terminal structures, and within IS elements or plasmids, it is proposed that bacterial introns have adapted to function mainly as retroelements (Dai and Zimmerly, 2002a, 2002b). A fundamental question is where group II introns come from. More than two
hundred IEPs of group II introns have been sequenced, and these IEPs can be divided into eight major classes, based on phylogenetic analysis, denoted mitochondrial, chloroplast-like 1 and 2, and bacterial A-E (Toro et al., 2002; Zimmerly et al., 2001). Importantly, the lineage of IEP correlates with RNA structure subclass, implying coevolution of IEP with intron RNA structure (Toor et al., 2001) The phylogeny distribution suggests that group II introns evolved in bacteria and then were transferred to eukaryotes, possibly via endosymbionts that give rise to organelles such as chloroplasts and mitochondria. An origin in archaea is less possible, as the few group II introns in archaea is postulated to come from bacteria via horizontal transfer.

**Figure 3** Phylogeny of the group II intron ORFs and correspondence with RNA structural classes. Phylogenetic relationships of ORFs are summarized based on neighbor-joining analyses. ORFs are divided into eight clades, named mitochondrial, chloroplast-like 1 and 2, and bacterial A–E. Each ORF clade is associated with a distinct RNA structural class (IIA1, IIB1, IIB2, IIC, two other distinct IIB-like, and two distinct IIA/B hybrid classes). The branching pattern is indicated for the *S. agalactiae* intron GBSi1, along with group II introns of the yeast mt introns *coxI*-I1 and -I2, the *L. lactis* L1.LtrB intron, as well as the introns of *E. coli* and the archaea *Methanosarcina acetivorans* (*M.a.*) Adapted from Lambowitz and Zimmerly (Lambowitz and Zimmerly, 2004).
Horizontal gene transfer in *S. agalactiae*

Different types of genes and groups of organisms vary in their propensity for HGT (Gogarten and Townsend, 2005). In *S. agalactiae*, the suggested HGT occurs most frequently in the *scpB-lmb* gene locus and capsule gene cluster. The first evidence of HGT in *S. agalactiae* came from an investigation that showed as high as 98% identity in the *scpB-lmb* locus between *S. pyogenes* and *S. agalactiae* (Franken et al., 2001). The *scpB-lmb* locus was found to be flanked by insertion sequences that constitutes a composite transposon, thus the locus is believed to be acquired via horizontal gene transfer. Interestingly, despite a similar gene organization of this region in *S. pyogenes*, a composite transposon structure was not detected in any isolates of *S. pyogenes*. Furthermore, most *S. agalactiae* isolates of bovine origin lack the *scpB* and *lmb* genes, suggesting that they are acquired or lost together. In contrast, all human isolates were found to have both genes (Dmitriev et al., 1999; Franken et al., 2001). Similarly, human isolates of group C and G streptococci harbor close homologues of the genes, while animal isolates of these bacterial species lack the genes (Franken et al., 2001). Taken together, *lmb* and *scpB* genes seem to have been dispersed via HGT, and they may be of particular importance for the ability of these streptococcal isolates to colonize and/or to cause disease in humans.

The CPS is encoded by the capsular biosynthesis cluster (operon), which consists of 14-18 different genes (*cpsA-cpsM, cpsO*, and *neuA-neuD*) in nine serotypes. The central genes of this locus encode serotype-specific glycosyltransferases and polymerases, and are flanked by genes conserved across diverse capsular serotypes, which are required for sialic acid synthesis (*cpsA-E* and *cpsL*) and capsule transport (*neuBCDA*). Eight of the nine serotypes appear to be closely related both structurally and genetically, whereas serotype VIII is more distantly related and as the only serotype contains rhamnos in the CPS (Cieslewicz et al., 2005). It has been shown that the horizontal transfer of a single gene *cpsH*, the putative CPS polymerase gene, between *S. agalactiae* serotypes Ia and III can lead to seroconversion of the polysaccharide capsule (Chaffin et al., 2000). Comparison of capsule structure and
proteins expressed by the cps operon suggests that the evolution of serotype-specific capsular polysaccharides has occurred through en bloc replacement of specific glycosyltransferase genes, and horizontal gene transfer is probably the mechanism for capsule variation (Cieslewicz et al., 2005).

The complete genome sequences of eight S. agalactiae strains of various serotypes have given a substantial amount of information and indicate that S. agalactiae has acquired virulence associated genes via HGT from other species including S. pyogenes, S. pneumoniae, and Listeria monocytogenes (Tettelin et al., 2005). For example, a strain-specific locus in S. agalactiae type III strain COH1, encoding the preprotein translocase subunits, three glycosyltransferases, and a repetitive surface protein, displays remarkable similarity to a genomic island present in S. pneumoniae TIGR4. A second COH1-specific island encodes surface proteins, which are similar to a fimbrial subunit and to internalin A of Listeria monocytogenes. Another region was found to be shared by strain H36B and S. pneumoniae. This region encodes a α-galactosidase, which may allow H36B to degrade and transport host α-galactosides.

Similar to in S. pyogenes, phage-associated genes account for 10% of all strain-specific genes in S. agalactiae. For example, type Ib isolate H36B contains a 41 kb prophage element that displays strong mosaicism and contains a protein similar to the S. pyogenes phage-associated pyrogenic exotoxin C. Furthermore, an intact copy of the 18-kb conjugative plasmid Tn916 of Enterococcus faecium which encodes tetracycline resistance as well as determinants necessary for its own movement, was detected with more than 95% nucleotide identity in four sequenced S. agalactiae strains, suggesting a recent acquisition of this trait.

**Insertion sequences in S. agalactiae**

Several insertion elements have been identified among S. agalactiae strains (Table 1) (Achard et al., 2005; Franken et al., 2001; Granlund et al., 1998; Rubens et al., 1989;
Spellerberg et al., 2000; Tamura et al., 2000). IS1548, IS861, IS1381, and ISsa4 were found to be present in multiple copies, whereas ISsag have been detected in low copy numbers.

### Table 1. IS elements identified in *S. agalactiae*.

<table>
<thead>
<tr>
<th>IS in <em>S. agalactiae</em></th>
<th>Size (bp)</th>
<th>Locus</th>
<th>Phenotype identified</th>
<th>No. of copies*</th>
<th>IS Family</th>
<th>Presence in GAS</th>
<th>Presence in <em>S. pneumoniae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>IS1548</td>
<td>1316</td>
<td><em>hylB, scpB, cpsD</em></td>
<td>Hyaluronidase Neg.; NT</td>
<td>0-8.</td>
<td>ISAs1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>IS1381</td>
<td>900</td>
<td>chromosome</td>
<td>None</td>
<td>0-5</td>
<td>IS5</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>IS861</td>
<td>1442</td>
<td><em>cpsD</em></td>
<td>NT</td>
<td>0-9</td>
<td>IS3</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>ISsa4</td>
<td>962</td>
<td><em>cylB</em></td>
<td>Non-hemolytic</td>
<td>0-9.</td>
<td>IS982</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ISsag1</td>
<td>1251</td>
<td><em>scpB-lmb</em></td>
<td>None</td>
<td>1</td>
<td>IS3</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>ISsag2</td>
<td>1220</td>
<td><em>scpB-lmb</em></td>
<td>None</td>
<td>2</td>
<td>IS3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>ISsag10</td>
<td>1724</td>
<td><em>cpsD-cpsE</em></td>
<td>Resist. to lincomycin</td>
<td>1</td>
<td>IS1595</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

A. The number of IS copies were estimated according to the references (Achard et al., 2005; Dmitriev et al., 2003; Dmitriev et al., 2004; Franken et al., 2001; Franken et al., 2004; Granlund et al., 1998; Tamura et al., 2000). Abbreviations, Neg., negative; NT, nontypeable by serotyping; resist., resistant.

Some of these elements have been reported to be associated with virulence genes. The most common insertion site of IS1548 is in the loci between the C5a-peptidase gene and the gene encoding the laminin-binding protein Lmb. The other common insertion site of IS1548 is within the *hylB* gene, where it causes inactivation of the hyaluronidase expression (Granlund et al., 1998). IS861 and IS1548 have been found inserted in the capsule gene cluster (Sellin et al., 2000), and ISsa4 was first detected in nonhemolytic strains, with a single copy inserted in *cylB*, which encodes the membrane-spanning domain of the putative hemolysin transporter (Spellerberg et al., 2000).
**Group II intron identified in *S. agalactiae***

A group II intron of 1825 bp, designated GBSi1, was identified in the *scpB-lmb* locus in *S. agalactiae* by Granlund *et al.* in 2001 (Granlund *et al.*, 2001). A sequence highly similar to GBSi1 was found in the capsular polysaccharide biosynthetic locus in *S. pneumoniae*. Both introns belong to intron bacteria class C based on IEP phylogeny analysis. In *S. pyogenes*, a 154 bp fragment of intron was present, showing 85% similarity with GBSi1 sequence. An inverted repeat sequence, with features typical of transcription terminators, was identified immediately upstream of the insertion site for the group II intron both in the *S. agalactiae* and *S. pneumoniae* sequences. This motif is suggested to constitute a target for the *S. agalactiae* intron as well as for rather closely related introns in *Bacillus halodurans, Pseudomonas alcaligenes*, and *Pseudomonas putida*. The IEP of GBSi1 has a RT domain but lacks an endonuclease domain. The self-splicing activity was demonstrated *in vitro* (Granlund *et al.*, 2001); however, mobility of GBSi1 was not tested either *in vitro* nor *in vivo*. Interestingly, a copy of IS1548 can be inserted close by, 9 bp upstream of the putative promoter for *lmb*, whereas the insertion site for GBSi1 is located 88 bp further upstream. PCR screening and sequencing revealed that the presence of group II intron and IS1548 is mutually exclusive among serotype III isolates of *S. agalactiae*.

**Molecular epidemiology**

Molecular epidemiology, the study of the distribution and determinants of disease occurring in human population using molecular technique, is a blend of molecular biology and epidemiology. The development of the PCR and sequencing techniques together with an increasing availability of complete genome sequences from many pathogens has enabled a great advance in molecular typing techniques during the last years. In contrast to the conventional typing methods molecular typing offers advantages of high throughout capacity, reproducibility, high discriminatory power and easy exchange of data across laboratories. These molecular techniques have proven useful in identifying, tracking, and intervention against various infectious
diseases. They have also yielded interesting insights into the biology, virulence mechanisms, and evolution of bacteria (Binnewies et al., 2006; Enright et al., 2002; Maiden, 2006). This new field of molecular epidemiology, bacterial biology, and population genetics has developed in a rapid pace.

Molecular typing techniques target at genetic variation to distinguish between isolates; however, in order to evaluate their utility and interpret the results knowledge about population structures and evolution genetics is required. It has been demonstrated that bacterial genomes have various levels of plasticity, and bacterial population structures vary over a broad range, from highly clonal such as in *Mycobacterium tuberculosis*, to the nonclonal (panmictic) (e.g., *Helicobacter pylori*), with most species being somewhere in between (semiclonal) (Alland et al., 2003; Spratt and Maiden, 1999; Suerbaum et al., 1998). In a strictly clonal organism, genetic variation such as nucleotide substitutions, duplications or deletions only arises by mutation, and these genetic changes can only spread vertically from mother to daughter cells. Therefore in a clonal population, individual lineages are readily recognized by their characteristic patterns of genetic variation, with limited or no sharing of genetic variation between the lineages; moreover, the same history of phylogeny is recalled at all loci (congruence). In contrast, in a totally nonclonal population, due to frequent horizontal gene transfer and recombination, there is no lineage structure and no congruence. It has become apparent that the extremes of fully clonal and nonclonal bacterial species are relatively rare. Many bacteria investigated to date show evidence of horizontal genetic exchange but also contain groups of clonally related isolates, particularly in an epidemiological investigation of short term (years or decades) or during a geographical spread (Maiden, 2006). In addition, natural selection also has an important effect on population structure, and the effect largely depends on the fitness of new genotypes and the size of population (van Belkum et al., 2001). All these factors influence the utility and choice of typing methods in a specific organism.

To be useful in a molecular typing context, the genes chosen for genotyping should show high degrees of intraspecies variability, and should provide unambiguous results.
The genes, used as genetic markers in the typing, range from a single gene, such as in IS-RFLP fingerprinting, to the entire genome, such as in comparative genomics. Thus, these methods detect genetic variations at different scales, and thereby yield different resolution even when they are applied to the same collection of isolates. To obtain sufficient discriminatory power the choice of genetic marker highly depends on the population structure of the species. M. tuberculosis is genetically homogenous by most typing methods, but the IS6110 fingerprinting exploits variability in both the number and genomic position of the highly plastic IS6110. Thus, this method has become the gold standard for epidemiological investigations of M. tuberculosis (van Embden et al., 1993). In other more divergent organisms such as Neisseria meningitidis, multilocus sequence typing (MLST) indexes variation in house keeping genes, which are under stabilizing selection for conservation of metabolic functions, and clusters isolates into distinct lineages despite extensive genetic heterogeneity in the population (Maiden et al., 1998). The increasing availability of complete genomes and comparative genomics will enable rationalizing the choice of genetic markers.

**Conventional phenotypic typing methods**

Conventional typing methods are based on the detection of phenotypes or characteristics expressed by an organism. These typing methods include serotyping (based on antigenicity), phage-typing (lysis by or susceptibility to bacteriophages), biotyping (based on biochemical characteristics according to metabolism), bacteriocin typing (susceptibility to bacteriocins), toxigenicity bioassays (detect toxin production), and typing according to antibiotic resistance patterns. One of the widely used phenotyping methods is serotyping, which uses antibodies to detect the difference in antigenic determinants expressed on the cell surface and subclass organisms into serogroups. For many organisms such as pneumococci, S. agalactiae, and N. meningitidis, serotyping is the primary means of classifying and evaluating isolates. However, some strains do not express surface antigens (nonencapsulated S. agalactiae) or express variant antigens which do not react with the existing antibodies,
such as typing of *S. pyogenes* according to M protein, and are thus nontypeable. Above all, serotyping often fails in differentiating between a large numbers of strains.

Multilocus enzyme electrophoresis (MLEE) has been the typing technique of choice for more than three decades. It detects genetic variations by differences in protein mobility via starch gel electrophoresis. Large-scale MLEE surveys of many pathogens demonstrated that bacterial populations ranges from highly clonal to pandemic (Smith *et al.*, 1993). However, in some cases collections comprised of only invasive isolates have under-estimated the genetic diversity of the species. MLEE is technically demanding and expensive and is therefore difficult to perform outside of research and reference laboratories. With MLEE Musser *et al* identified two distantly related genetic lineages among serotype III isolates recovered from 18 states in the United States, and one of them was designated as highly virulent since it was associated with invasive disease in neonates (Musser *et al.*, 1989). Other investigators, with the exception of Hauge *et al* (Hauge *et al.*, 1996), also have identified distinct groups of serotype III strains with enhanced virulence using this method, and showed that *S. agalactiae* has a clonal population structure (Helmig *et al.*, 1993; Quentin *et al.*, 1995).

**Genotypic strain typing**

**REA/RFLP.** Digestion of whole chromosomal DNA with a frequently cutting restriction endonuclease and separation of DNA fragments by standard (linear) electrophoresis is called restriction endonuclease analysis (REA) or restriction fragment length polymorphism (RFLP). REA can be used directly to type organisms such as *Listeria monocytogenes* with high discriminatory power and reproducibility. To increase resolution, the DNA fragments of RFLP may be hybridized with a labelled probe, which is complementary to a sequence present in multiple copies in the genome. This approach has been applied to subtype a variety of organisms. When ribosomal operon genes are used as probe, the method is called ribotyping. Ribotyping
has applications in the typing and subtyping of various Gram negative bacteria; however, it has been of little use in strain typing for Gram positives, and in most cases can only differentiate to species level.

Blumberg et al first used REA and ribotyping to type invasive *S. agalactiae* isolates, and demonstrated that REA is sufficiently discriminatory for strain differentiation of *S. agalactiae* (Blumberg et al., 1992). REA also has been used to study the population structure of *S. agalactiae*. Takahashi et al subclassified serotype III strains from Tokyo and Salt Lake City into three distinct phylogenetic lineages using REA with two different enzymes (*HindIII* and *Sse83871*). Strains of a single lineage, RDP type III-3, caused the majority (91%) of the invasive disease in neonates, which suggests that this lineage may be more virulent (Takahashi et al., 1998). Virulence related gene clusters, uniquely associated with the RDP type III lineage, were identified using genomic subtractive hybridization, and await further elucidation of their role in *S. agalactiae* pathogenesis (Bohnsack et al., 2002). In addition, RFLP has application in typing isolates nontypeable by serotyping (Sellin et al., 2000). The RFLP analysis of insertion sequences such as IS1381, IS1548, and IS861 also proved to be useful in estimating the relatedness of *S. agalactiae* isolates (Dmitriev et al., 2002; Tamura et al., 2000).

**Pulsed-field gel electrophoresis.** Pulsed-field gel electrophoresis (PFGE) has been used as a golden standard typing method with high resolution and reproducibility for many bacteria. Digestion of whole chromosomal DNA with a rare-cutting enzyme is followed by separation in an agarose gel matrix, in which the orientation of the electronic field across the gel is changed periodically. Although the method detects minor genetic events reflected by the gain or loss of a restriction site, it has high discriminatory power and readily detects large scale genomic variations such as chromosomal rearrangement that occur over a relatively short time scale. PFGE is considered the principle method of choice for epidemiological outbreak investigations, and is currently the method used by a cross nation system, PulseNet, to carry out surveillance on food-borne disease pathogens such as *Salmonella typhimurium*, *E. coli* O157:H7, and *L. monocytogenes* ([http://www.cdc.gov/pulsenet](http://www.cdc.gov/pulsenet)) (Gerner-Smidt et al., 2002).
2006). However, PFGE requires standardized protocol and analysis software, and the results can be ambiguous based on the banding patterns of restricted DNA fragments.

Fasola et al and Gordillo et al first showed that PFGE patterns of *S. agalactiae* were more easily interpreted and sensitive than REA (Fasola et al., 1993; Gordillo et al., 1993). PFGE has been used in epidemiological investigations on transmission, acquisition and dynamics of *S. agalactiae*. Green et al used PFGE to demonstrate that recurring *S. agalactiae* disease in infants is due to mostly the original infecting strain or occasionally a newly acquired strain (Green et al., 1994). Mother/infant isolate pairs usually had identical PFGE patterns, stressing the importance of vertical transmission (Hansen et al., 2004; Melchers et al., 2003). Hansen et al used PFGE to analyse totally 1500 isolates from women during and after pregnancy; and found that all carriers were colonized by a single *S. agalactiae* clone on all occasions spanning up to 2 years (Hansen et al., 2004). As serotype V emerged in the early 1990s, PFGE analysis of serotype V isolates from the United States and France showed that a single clone accounted for more than half of the invasive disease caused by serotype V during the last years (Elliott et al., 1998; Le Thomas-Bories et al., 2001).

**Multilocus sequence typing (MLST).** MLST is based on the principle of MLEE, but genetic variation at housekeeping genes is identified by DNA sequencing. Since MLST can detect both synonymous and nonsynonymous changes, it has higher discriminatory power and fewer loci are needed. In MLST, approximately 500 bp DNA fragments of seven housekeeping genes are amplified and sequenced. The sequences of these genes are compared to known alleles at each locus at the MLST database (http://pubmlst.org and http://mlst.net), and each isolate can be described by a seven-integer allelic profile that defines a sequence type (ST). For example, in *S. agalactiae*, isolates with allelic profile 2-1-1-2-1-1-1 is designated ST17, a predominant clone among serotype III. Since MLST uses sequence data and define isolates as ST with digital numbers, the results are unambiguous, highly reproducible, and can be stored in the internet database, which offers easy comparison between laboratories and over time. Although MLST was conceived as a typing tool for clinical
microbiologist and epidemiologist, sequence data form MLST have been successfully used to infer phylogeny and study population genetics. However, for the latter purposes, data should be interpreted with caution.

MLST usually employs allele fragment of between 400 and 600 bp since this length of nucleotide sequence could be reliably achieved in a single run of the gel-based automated sequencing. However, for population studies, it would be more appropriate to sequence the entire gene. Furthermore, the number of loci examined is the minimum with sufficient resolution in order to reduce the cost and time required for isolate characterization. During the development of the first MLST scheme for the meningococcus, originally 12 loci were tested, and seven were chosen to give reliable identification of invasive meningococci. Moreover, it has proven impossible to identify a set of universal housekeeping genes for a wide range of pathogens, since bacterial genomes are extensively diverse and MLST as a typing scheme measures variability to provide sufficient discriminatory power. In some instances, in order to increase the resolution of MLST, sequencing of genes that are under selection pressure and render greater variation may be used in combination with MLST. MLST has been under rapid development since its first scheme was introduced in 1998 (Maiden et al., 1998). To date MLST schemes have been developed for more than thirty species, and it is currently the main typing method for identifying/tracking highly virulent/antibiotic-resistant clones of several pathogens including *N. meningitidis* (Maiden et al., 1998), *S. pneumoniae* (Brueggemann et al., 2003), and *S. aureus* (Feil et al., 2003).

The analysis of MLST data are based on allelic profile (ST) or more directly the concatenated nucleotide sequences of seven loci. Clustering algorithms such as UPGMA (unweighted pair group method with arithmetic mean) and BURST/eBURST exploits the relatedness of sequence type based on allelic numbers. In particular, eBURST has the capacity to accommodate large collection of isolates, subdivide isolates into clonal complexes (CCs), in which the isolates share high genetic similarity (six or seven alleles), and predicts the common ancestor from which the CC
organizes. eBURST has proven to be useful in tracking clonal diversification. However, for less divergent organisms, it is preferable to analyze nucleotide sequences.

A MLST scheme for *S. agalactiae* was developed by Jones *et al.* in 2003. The following seven loci were selected for the MLST scheme: alcohol dehydrogenase (*adhP*), phenylalanyl tRNA synthetase (*pheS*), amino acid transporter (*atr*), glutamine synthetase (*glnA*), serine dehydratase (*sdhA*), glucose kinase and transketolase (*tkt*). These genes are spread evenly around the chromosome with a distance of 20 kb between two loci.

**Whole genome sequence comparison.** The increasing amount of complete genome sequences available from many pathogens has enabled direct comparison of sequences to identify genetic differences across entire genomes. It has also made it possible to correlate those differences to biological function, and to gain insight into selective evolutionary pressures and patterns of gene transfer or loss, particularly within the context of virulence. Genomic comparison of multiple strains from the same species revealed surprisingly high levels of diversity in a number of organisms, which were underestimated while small genomic regions were analysed using methods such as MLEE and MLST. *E. coli* strains vary in as much as 25% of their genome, and *Salmonella enterica* serovars vary by 10-12%. These variations reflect an acquisition of virulence factors or antibiotic resistance mechanisms, and stress the need for sequencing multiple strains to represent the genetic inventory of a given species.

Analysis of the eight *S. agalactiae* genomes revealed a pan-genome concept and a rather large flexible gene pool (Tettelin *et al.*, 2005). In another study, which reported three complete genome sequences, high sequence conservation was detected, and a remarkable conservation of the order and relative orientation of orthologous genes was observed (Brochet *et al.*, 2006). Although the pathogenicity-like islands were located on putatively flexible genomic regions, they are conserved in the strains tested, which suggest that the islands were present prior to the divergence of *S. agalactiae*. In
addition, significant genetic differences were also identified in the chromosomal backbone, which may contribute to host adaptation.

**Microarray.** Microarray-based comparative genomic hybridization (CGH), in which labeled DNAs from two strains are competitively hybridized to a full-genome microarray, has been described in numerous reports (Fitzgerald and Musser, 2001; Fitzgerald *et al.*, 2001; Gressmann *et al.*, 2005; Taboada *et al.*, 2004). Microarray-based CGH is easy to apply to a large number of strains, e.g., clinical isolates. However, microarray analysis can only identify what is absent from a given strain but cannot identify additional genes an organism may have. The use of a pan-array, consisting of the genomes of several sequenced strains and the use of oligonucleotide array (Affymetrix system), will greatly increase the interrogative capacity and resolution. The development of standardized bioinformatics tools to analyse the microarray data is warranted to understand evolution of bacterial pathogens.

Brochet *et al.* characterized 75 *S. aglactiae* diverse strains using a whole genome DNA microarray based on the NEM316 genome, and identified a similar core genome as described in a previous study (Brochet *et al.*, 2006; Tettelin *et al.*, 2005). Surprisingly, the strain most similar to NEM316 was of bovine origin with only 10 genes lacking, whereas the most divergent strain was a thermosensitive strain isolated from fish, which may adapt to the specific environment through large scale genetic variation. All strains clustered into nine groups, and the clustering was largely independent of serogroups and origin of the isolates with some exceptions (Brochet *et al.*, 2006).
Definitions

Definitions of relevance for this work. Adapted from Riley (Riley, 2004).

**Isolate.** A population of bacterial cells in pure culture derived from a single colony on a primary isolation plate.

**Strain.** An isolate or group of isolates displaying specific genetic or phenotypic properties.

**Genotype.** A specific pattern, or set of marker scores, derived from examination of the DNA of an isolate using a particular typing system.

**Clone.** A group of isolates or strains descending from a common ancestor because of a direct chain of replication.

**Locus.** A chromosomal or genomic location of a gene.

**Allele.** One of several alternative forms of a gene at a given locus.

**Discriminatory power.** The ability to discriminate a set or collection of isolates.

**Genetic diversity.** The diversity of genotypes present in a population of isolates.
AIMS OF THE STUDY

To characterize distribution of GBSi1 among *S. agalactiae* and identify features unique to GBSi1 harboring isolates.

To characterize population structure of *S. agalactiae* invasive and non-invasive isolates.

To investigate whether certain genetic lineages show high virulence potential and preferred tropism in humans.

To identify possible molecular markers for respective genetic lineages.
RESULTS AND DISCUSSION

Identification of two additional GBSi1 copies

In an effort to characterize the distribution pattern of the group II intron (GBSi1) in *S. agalactiae*, we found that there were at least two additional copies of GBSi1 in *S. agalactiae* serotype III reference strain M732 besides the previously identified intron located at the *scpB-lmb* locus (Paper I, Figure 1A, lane 1). Two additional GBSi1 copies, identical to that located at the *scpB* locus, were identified with a reverse PCR scheme, and the insertion sites were referred to as the Y- and Z-site (Figure 4).

![Diagram of gene region flanking GBSi1](Image)

**Figure 4.** Schematic views of the gene region flanking GBSi1 in a serotype III strain M732. The coding regions are shown with grey arrows. *ftsY* encodes a signal recognition particle (Sag0727); *sag0728* encodes the substrate binding protein of an ABC-transporter; *lplA1* encodes a putative lipoate-protein ligase, and *sag0883* encodes a putative cobyric acid synthase. The GBSi1 intron is depicted by an open box with an arrow indicating the orientation of the element.

The Y-site has been previously identified by Bohnsack et al (Bohnsack *et al.*, 2002). At both the Y and Z sites, GBSi1 left all open reading frames intact, even though the intron was inserted directly after the stop codon of the genes located downstream of the intron copies (Figure 4). At the X site, the intron was inserted 88 bp from the
promoter of the \textit{lmb} gene, but no difference in Lmb expression was detected in comparison to intronless strains (Our unpublished results).

**Uneven distribution of GBSi1 in \textit{S. agalactiae}**

To examine the distribution of GBSi1 in various isolates and to determine the copy number of the intron in \textit{S. agalactiae}, Southern blot analysis and PCR screening were performed on totally 109 \textit{S. agalactiae} clinical isolates of serotypes Ia, Ib, and II-VIII. The presence of GBSi1 appeared to be restricted to certain serotypes. The presence of GBSi1 was found only in a subpopulation of 29 isolates belonging to serotype Ia, III, and V but in none of the other serotypes tested. In particular, three copies of GBSi1 were identified in serotype III isolates (15 of 34), while a single copy was found in isolates of serotype II (10 of 17) and V (4 of 13) (Paper I, Figure 1 and Table 1). Remarkably, in serotype III strains harbouring GBSi1, the intron was invariably inserted at all identified targeting sites including X, Y, and Z; in contrast, the single GBSi1 copy present in serotype II and V was located at the \textit{scpB-lmb} locus (X site) (Paper I, Figure 2B). However, Takahashi and colleagues have described serotype Ia isolates with GBSi1 inserted at two loci, X and Y site (Takahashi \textit{et al.}, 2002). The fact that all intron harbouring isolates share the insertion site at the \textit{scpB} locus is intriguing, which indicates that the \textit{scpB} locus is the primary targeting site of the intron, and that the intron may subsequently mobilize to the Y or/and the Z site. The mechanism of mobility, retrohoming or retrotransposition, highly depends on the similarity of the targeting motif between the X site and the other sites.

**The putative target motif of GBSi1**

In order to determine the target motifs of the Y and Z sites, the region flanking the intron insertion sites was analyzed, and a number of conserved motifs were revealed
(Paper I, Figure 3). An inverted repeat sequence followed by a conserved TTTATA intron-binding site 1 (IBS1) was identified upstream of the two additional intron insertion sites, which was consistent with an earlier description of GBSi1 and which is typical of the class C of bacterial group II introns (Granlund et al., 2001). At the site immediately downstream of the intron, a conserved U (δ’ site) followed by UA was identified in the Y and Z sequences. In addition, a common pattern of thymidine/uracil-residues was located at identical positions upstream of the insertion site of all three intron copies. These conserved motifs might be of importance for the target recognition of GBSi1. Comparison of intron insertion sites in the isolates of serotype II, V, and III harbouring GBSi1 revealed identical X and Y sites but remarkable difference at the Z site. The isolates of serotype II and V had an additional 41 nucleotides inserted immediately upstream of the target site Z, which caused a disruption of the GBSi1 target motif at the Z-site.

The distribution pattern of intron may imply that after initial homing to the X site located in the scpB locus, transposition of the intron to the Y and Z sites took place in an ancestor of this lineage. Alternatively, the three sites could be regarded as multiple homing sites for the intron since they all show considerable target similarities. However, it is not apparent if this degree of similarity between the three intron insertion sites is sufficient enough for homing to occur. The fact that all serotype III isolates harbouring GBSi1 was found to contain three copies but never one or two copies argues against that the intron is mobilized to the three sites at equal efficiency. The lack of insertion of GBSi1 in serotype II and V isolates into the Z-site could be explained by disruption of the third target site. However, the fact that no intron was located at the alternative Y site, even though this was identical to that of isolates harbouring three GBSi1 copies, argues against homing into all three sites. An explanation to these seemingly contradictory findings might be that host factors are required for intron mobility, which is not present in all S. agalactiae isolates. Despite the fact that GBSi1 was demonstrated to be able to self-splice in vitro (Granlund et al., 2001), its mobility has not been examined in vivo.
A unique DNA fragment with plasmid features is correlated with GBSi1

The distribution of GBSi1 was uneven among isolates; however, all intron harbouring isolates shared the common insertion site at the scpB locus. It had been recently reported that scpB was located on a putative composite transposon, and this locus was involved in horizontal gene transfer (Franken et al., 2001). Amplification and sequence analysis of the region upstream of scpB in intron harbouring strain M732 revealed an insertion of a 2.1 kb fragment in comparison to that of strain O90R of serotype Ia (Figure 5). Remarkably, all intron-bearing isolates regardless of serotype contained the 2.1 kb fragment. However, the insert was not detected in any intronless isolate.

**Figure 5.** Schematic view of the 2.1 kb fragment present in *S. agalactiae* harbouring GBSi1.

The G/C content of the DNA fragment was 29.7%, notably lower compared to the G/C composition of 35.6-35.7%, reported for the genome of *S. agalactiae* (Glaser et al., 2002; Tettelin et al., 2002). The 2.1 kb fragment seemed to be largely associated with mobility, since an amino-terminal fragment of transposase TnpA and proteins possibly originating from plasmids were encoded by this fragment. It has been suggested that the majority of bacterial group II introns act as retroelements (Dai and Zimmerly, 2002a), and like other mobile genetic elements, such as insertion elements and composite transposons, they may be transferred between bacterial cells via phages, conjugative transposons, or conjugative plasmids by horizontal gene transfer. The fact
that the 2.1 kb DNA fragment encodes remnants of plasmid related proteins and its unique occurrence in intron harbouring isolates suggests that the fragment could have been involved in the uptake process of GBSi1 into *S. agalactiae*. In this scenario, GBSi1 might initially have been located on a plasmid, which served as vehicle for horizontal transfer of the intron. In that case immunity against uptake of the plasmid among a fraction of *S. agalactiae* may account for the uneven distribution of GBSi1.

**Genetic lineages of clinical *S. agalactiae* isolates**

To analyze the genotype distribution, i.e., the population structure, of *S. agalactiae* and to investigate whether there is a correlation between particular genotypes and infection type, MLST was used to characterize *S. agalactiae* clinical isolates. Totally 212 *S. agalactiae* isolates associated with adult and neonatal invasive disease were collected in two geographic areas, approximately 1000 km apart, Northern Sweden (Umeå) during the years 1988-2004 and from Western Sweden (Gothenburg) during the years 1988 to 1997 (Table 2). In addition, 102 non-invasive isolates were collected in the Västerbotten county from wound samples of adults (mean age 56) and vaginal samples women (mean age 32). The total 314 isolates were resolved into 52 unique STs, and these STs were grouped by BURST into five major CCs, two minor CCs and two singletons. In a clonal complex the STs share more than five identical alleles, and it is designated according to its predicted ancestor ST or the representative ST that appeared at highest frequency. CC19 and CC1 were most predominant, together representing half of the isolates; while CC9, CC17, and CC23 accounted for 13-15% of the isolates, respectively (Table 2). The minor CCs, CC7 and CC4, together with two singletons accounted for 7% of the isolates. In addition, the concatenated sequences of seven genes for each ST were used for the phylogeny analysis, and sequences from a *S. pyogenes* M3 strain were added to root the phylogeny. Similar groupings were revealed; however, despite that the majority of CC19 and CC17 isolates express same serotype III capsule, they are relatively distantly related.
Table 2. Distribution of clonal complexes in relation to patient group and geographic area.

<table>
<thead>
<tr>
<th>CCs</th>
<th>No. of isolates (% of total)</th>
<th>Major serotype within CC (%)</th>
<th>Non-invasive (Umeå, 1999-2000)</th>
<th>Invasive</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>V</td>
<td>W</td>
</tr>
<tr>
<td>CC1</td>
<td>75 (24)</td>
<td>V (64)</td>
<td>33</td>
<td>42</td>
</tr>
<tr>
<td>CC9*</td>
<td>46 (15)</td>
<td>Ib (63)</td>
<td>17</td>
<td>14</td>
</tr>
<tr>
<td>CC19</td>
<td>84 (27)</td>
<td>III (87)</td>
<td>17</td>
<td>21</td>
</tr>
<tr>
<td>CC7</td>
<td>8 (3)</td>
<td>V (75)</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>CC17</td>
<td>48 (15)</td>
<td>III (100)</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>CC4</td>
<td>10 (3)</td>
<td>Ia (100)</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>CC23*</td>
<td>40 (13)</td>
<td>Ia (78)</td>
<td>13</td>
<td>17</td>
</tr>
<tr>
<td>Singletons</td>
<td>3 (1)</td>
<td>NA</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Total No.</td>
<td>314</td>
<td></td>
<td>39</td>
<td>63</td>
</tr>
</tbody>
</table>

*Three isolates from child patients belonging to CC9 and one belonging to CC23 are not listed in the table. Abbreviations ‘V’ represents vaginal, ‘W’ wound, ‘A’ adult and ‘N’ neonate. NA not applicable.

Overall, the five predominant genetic lineages (CCs) identified here among *S. agalactiae* isolates recovered in Sweden were similar to that found in a global collection investigated by Jones et al (Jones et al., 2003). This clonal structure has also been shown by a number of other methods, such as PFGE, MLEE, and RAPD (Hauge 43
et al., 1996; Musser et al., 1989; Quentin et al., 1995; Rolland et al., 1999; Sellin et al., 2000; Takahashi et al., 1998). In comparison to MLST analysis of *S. pyogenes*, the diversity of *S. agalactiae* appears to be much lower despite a highly similar genetic backbone shared between these two streptococci (Enright et al., 2001; Glaser et al., 2002). In *S. pyogenes*, the rate of recombination has been reported to be relatively high in the overall population (Kalia et al., 2002). In addition, in accordance with the clonal population structure of *S. agalactiae*, the major STs were present in both the invasive and non-invasive population. In contrast, in a non-clonal organism such as *N. meningitis*, the invasive isolates are just a small fraction of the non-invasive population.

Numerous epidemiological investigations have shown that the serotype distribution of *S. agalactiae* varies both geographically and over time (Harrison et al., 1998; Kalia et al., 2002). In order to determine the geographical difference in the genotype distribution and monitor the population dynamics over the years, the invasive isolates collected from Umeå and Gothenburg during the years 1988-97 were compared (Paper I). CC19 and CC17 were the predominant clonal complexes, together accounting for half of the isolates, whereas CC9 and CC23 were much less common, represented by 10-17% of the isolates in the two regions. A geographical difference in the distribution of clonal complexes was noted for CC1, which was more common in Gothenburg (21.1%) than in Umeå (10.2%). The distribution of CC19, CC17, and CC9 showed stability over time. In contrast, CC23 (mainly serotype Ia isolates) almost disappeared after 1994, whereas CC1 (mainly serotype V isolates) did not appear until the early 1990s but has caused a considerable proportion of invasive diseases ever since. The fluctuation of lineages over time was more apparent in the Western part of Sweden than in the Northern region. A change of genotype in the population might reflect differences in fitness of genotypes, e.g., spread of an antibiotic-resistant clone; or changes in herd immunity. The latter has been noticed in organisms such as pneumococci, where a rise in the frequency of vaccine unrelated genotypes was pronounced in the post-vaccination era (Beall et al., 2006).
Figure 6. Phylogenetic tree determined by the neighbor-joining method. Clonal complexes (CC) are grouped by BURST.
Clonal relationships among invasive and non-invasive isolates and clone/clonal complex-specific odds ratios

Since differences with respect to the genetic lineage distribution of *S. agalactiae* were seen geographically and over time, it is important to take these differences into account and avoid sampling bias in an epidemiological investigation. In order to determine whether correlation exists between a certain CC and disease, the genotype distribution of non-invasive isolates was compared to that of the invasive isolates collected from the same defined population during years 1995-2004 (Paper II). The eBURST analysis of the 39 STs of 185 invasive and non-invasive isolates revealed the same major genetic lineages CC1, CC19, CC9, CC17 and CC23 (Paper III). There was a significant difference in the distribution of the clonal complexes (*P* = 0.001), despite that all the clonal complexes were represented among both invasive and non-invasive isolates. The most prevalent CCs among isolates from wound and vaginal samples were CC1, CC19 and CC23, whereas ST17 in CC17 was represented by only three isolates (Paper III, table 1). In contrast, among isolates from invasive disease, CC19 was the most predominant (35%), followed by CC1 (20%), CC9 (18%), and CC17 (16%). In particular, CC17 accounted for 11% of the adult invasive cases and 22% of the neonatal invasive disease, but was found only among 3% of the non-invasive isolates (chi-square test, *P* = 0.0025). Thus, the OR for causing invasive disease for isolates belonging to CC17 was 6.12. CC19 was the most prevalent clonal complex causing invasive disease in both neonates and adults but was also the second most common among non-invasive isolates (*P* = 0.05). On the other hand, CC1 was more common among non-invasive isolates than among invasive (Table 3).

When the distribution of lineages among adults and neonates with invasive *S. agalactiae* disease was compared; there was a highly significant association between CC17 and neonatal infections (chi-square test, Bonferroni corrected *P* = 0.006, paper II, Table 1 and 2), whereas no correlation between other lineages and a specific patient group could be established (Chi-square test, Bonferroni corrected, *P* > 1).
The existence of a hypervirulent serotype III lineage of *S. agalactiae* (CC17) has been proposed by several research groups (Lin *et al.*, 2006; Musser *et al.*, 1989; Quentin *et al.*, 1995; Takahashi *et al.*, 1998). Musser *et al* and Takahashi *et al* suggested the existence of a highly virulent clone causing neonatal invasive diseases. The basis for the conclusion was that while ET1/RDP III-3 (CC17) was underrepresented among carriage isolates, this clone represented the vast majority of isolates from neonatal invasive diseases, (Musser *et al.*, 1989; Takahashi *et al.*, 1998). Similar studies were also conducted in France and Denmark (Hauge *et al.*, 1996; Quentin *et al.*, 1995). In the French collection of *S. agalactiae*, multiple virulent clones with capacity to invade the central nervous system were identified. However, no significant difference in the lineage distribution among disease associated serotype III isolates, or any difference in pathogenic potential was noted. Recently, in agreement with the conclusions from the Danish study, no highly virulent clone was identified among serotype III isolates collected from North America (Davies *et al.*, 2004). It is important to remember that adult invasive isolates were not examined in any of the aforementioned studies. With our data set it is reported for the first time the significant association of CC17 and neonatal infection, as well as the striking difference between the two major lineages CC19 and CC17. CC19 was commonly found among adults, and caused a significant proportion of both adult and neonatal invasive disease. In contrast, CC17 was predominant in neonatal invasive disease, but it was less frequently found in adult invasive disease and seldom present among the non-invasive isolates from wound and vagina. This finding suggests that there is an age-dependent tropism of CC17, which causes the association seen between CC17 and neonatal invasive disease. It is possible that ST19 represents a highly successful clone, and is composed of isolates with increased capacity to cause diseases both among neonates and adults, whereas CC17 seems to be strongly adapted to neonates.

It should be pointed out that the present study only compared invasive adult and neonatal isolates and non-invasive isolates from adults, and it is possible that the population of neonatal non-invasive/carriage isolates may be different. CC17 was found to be as prevalent as CC19 when colonizing isolates from mothers and neonates
were examined in a national-wide Swedish study (our unpublished data). Whether CC17 is highly efficient in vertical transmission awaits further investigation of carriage isolates from neonates. Furthermore, the clonal properties shown by CC17 and CC19 seem to play an important role in virulence, and future work is warranted to identify genes of potential importance in pathogenesis.

**Clonal divergence of *S. agalactiae***

Each CC has an ancestor ST, which diversify into a number of single locus variants (SLVs) differing at one allele from the ancestor ST. 14 SLVs were identified for ST19, seven for ST1 and ST17, respectively, four for ST9 and two for ST23. Most SLVs were unique for respective geographical area and contained only one or two isolates, except ST107 and ST2, which were represented by a few isolates in both areas (Table 1 and 2). It will be interesting to investigate what features these SLVs have acquired and confer a selection advantage to make them remain persistent in the population.

The SLVs can have arisen from their respective clonal ancestor by point mutation or by recombination, and direct estimate can be made by a method described by Feil et al (Feil et al., 2003). Briefly, putative point mutations are assigned on the basis of two criteria: The SLV allele differs from that of the ancestral ST only at a single nucleotide site, this difference is unique within the database. Those that differ at multiple nucleotide sites, or which differ at a single site but correspond to alleles found elsewhere in the data set, are assigned as having arisen by recombination. The five major clonal complexes provide a total of 34 SLVs for this analysis (Table 2). Twenty-five of these SLVs possess variant alleles that differ at a single nucleotide site from the allele in the putative ancestral ST, and they probably had arisen by point mutation. Nine SLVs had multiple changes in respective allele, and they may have arisen by recombination. The recombination rate is rather high, regarding that *S. agalactiae* has a clonal population structure.
Table 3. Variant alleles within the SLVs of five major clonal complexes

<table>
<thead>
<tr>
<th>ST of clonal ancestor</th>
<th>ST of SLV</th>
<th>Variant locus in SLV</th>
<th>Ancestral allele</th>
<th>SLV allele</th>
<th>No of nt change</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>121</td>
<td>adhP</td>
<td>1</td>
<td>32</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>124</td>
<td>atr</td>
<td>3</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>107</td>
<td>adhP</td>
<td>1</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>28</td>
<td>glnA</td>
<td>2</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>27</td>
<td>glnA</td>
<td>2</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>86</td>
<td>glcK</td>
<td>2</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>175</td>
<td>glnA</td>
<td>2</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>19</td>
<td>164</td>
<td>atr</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>232</td>
<td>ahp</td>
<td>1</td>
<td>45</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>254</td>
<td>adhP</td>
<td>1</td>
<td>9</td>
<td>2</td>
</tr>
<tr>
<td>19</td>
<td>233</td>
<td>adhP</td>
<td>1</td>
<td>46</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>241</td>
<td>pheS</td>
<td>1</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>122</td>
<td>glcK</td>
<td>2</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>123</td>
<td>atr</td>
<td>3</td>
<td>23</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>119</td>
<td>atr</td>
<td>1</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>120</td>
<td>pheS</td>
<td>1</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>125</td>
<td>sdhA</td>
<td>1</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>150</td>
<td>tkt</td>
<td>1</td>
<td>12</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>148</td>
<td>glnA</td>
<td>2</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>244</td>
<td>atr</td>
<td>1</td>
<td>30</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>126</td>
<td>tkt</td>
<td>1</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>23</td>
<td>88</td>
<td>pheS</td>
<td>4</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>23</td>
<td>234</td>
<td>tkt</td>
<td>3</td>
<td>16</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>12</td>
<td>adhP</td>
<td>8</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>adhP</td>
<td>8</td>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>8</td>
<td>adhP</td>
<td>8</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>245</td>
<td>adhP</td>
<td>8</td>
<td>46</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>atr</td>
<td>2</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>151</td>
<td>adhP</td>
<td>1</td>
<td>34</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>139</td>
<td>atr</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>1</td>
<td>153</td>
<td>adhP</td>
<td>1</td>
<td>36</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>183</td>
<td>adhP</td>
<td>1</td>
<td>48</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>231</td>
<td>glcK</td>
<td>2</td>
<td>24</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>235</td>
<td>adhP</td>
<td>1</td>
<td>47</td>
<td>12 (deletion)</td>
</tr>
</tbody>
</table>
Interestingly, the clonal complexes appeared to be more diverged among the wound isolates than among the vaginal or rectal isolates. In the CC1 isolates recovered from wounds, the complex contained in addition to ST1 five single locus variants (SLVs) of ST1 and a double locus variant ST196; whereas among the recto-vaginal isolates CC1 included only ST1, and a single SLV. Clonal divergence was also noted to a less extent for CC19 among isolates from vagina. A possible explanation might be that wound constitutes a less restricted environment compared to the normally acidic milieu in the vagina.

**Serotype distribution and serotype-specific odds ratio**

The surface polysaccharide capsule is considered as the most important virulence factor of *S. agalactiae* since it inhibits complement-mediated killing. To examine the importance of CPS in pathogenesis, the serotype distribution was analyzed and compared among invasive and non-invasive isolates (Paper III). Six serotypes ranging from Ia, Ib and II to V were found to be present among both invasive and non-invasive isolates (Paper III, Table 3). However, a significant difference was found in the distribution of the serotypes (*P = 0.004*). Among invasive isolates, half of the isolates were of serotype III, and the other predominant serotypes were Ib and V (both 14.5%) whereas the other serotypes accounted for less than 10% of the isolates (Paper III, Figure 2). In contrast, among non-invasive isolates, serotype V was most prevalent (29.4%), followed by serotype III, Ib, II and Ia, which accounted for 12-21% of the isolates, respectively. Clearly, serotype III was highly overrepresented among invasive isolates (chi-square test, Bonferroni corrected *P = 0.0008*), and had a high disease potential (OR 3.59). Both serotype Ib and IV had an OR near one, compared to the OR of 0.5 or less, seen in serotype Ia, II, and V isolates. The high prevalence of serotype V among non-invasive isolates was reflected by the fact that 42% of the wound isolates and 33% of the vaginal isolates belonged to CC1 (Table I). *S. agalactiae* of serotype V was present in 12 of 45 invasive isolates from adults (26.7%) compared to in 5 of 37
neonatal isolates (13.5%) (Paper III, Table 2). The serotype IV isolates associated with invasive disease were seen only in adults.

Numerous studies have demonstrated that serotype III was the dominating serotype among invasive isolates, and serotype V has emerged since the early 1990s (Blumberg et al., 1996; Persson et al., 2004). In the present study serotype V was found to dominate among the isolates from wounds, and also caused a substantial amount of invasive disease in adults. Serotype IV isolates has rarely been reported in epidemiological studies. In the present study 12 serotype IV isolates were identified, accounting for 4% of the total isolates. Future studies will reveal whether serotype IV is an emerging invasive serotype among adults. In addition, in the present study no isolates of serotype VI, VII, or VIII were identified, which were reported to dominate among Japanese woman in the nineties (Lachenauer et al., 1999).

Relationship between genotype and capsular serotype

In general, the serotypes appeared to be associated with the overall genotype defined by ST. However, there were a number of STs found to contain isolates of more than one serotype (Paper II and III). CC17 was the only clonal complex that contained a single serotype, since the 48 CC17 S. agalactiae isolates all expressed serotype III capsule. All other clonal complexes were shown to contain isolates of more than one serotype. CC9 and CC1 demonstrated high degree of serotypic variation, and contained isolates of five serotypes (Figure 7).

Moreover, the capsular phenotypes were more heterogeneous among non-invasive isolates than among invasive. Among non-invasive isolates, all clonal complexes except CC17 harboured three serotypes or more. In particular, non-invasive isolates of CC1 expressed all the major serotypes present in the S. agalactiae collection except Ia. In contrast, the invasive CCs harboured fewer serotypes. (Paper III, Fig 2). This
finding suggests less variability is allowed in a highly selective environment such as the blood.

![Chart showing percentage of Clonal Complexes (CC9, CC1, CC23, CC17, CC19, CC7, CC4) with different serotypes (Ia, Ib, II, III, IV, V)](chart)

**Figure 7.** Clonal complex in relation to serotype among invasive and non-invasive *S. agalactiae*.

In addition to the fact that a number of genetically identical isolates expressing different capsular types, the genotypes of isolates belonging to serotype Ia, II, and III were heterogeneous (Paper II, Table 3). The same serotype is seen among isolates that differ at six loci, such as ST17 and ST19 isolates, which share expression of type III capsule, and ST23 and ST4 isolates that encode capsule of type Ia. Taken together these findings indicate that horizontal transfer of capsular genes has occurred in the population. Importantly, the level of capsular switching appears to be high, and exchange of capsule gene clusters between *S. agalactiae* with different genetic backgrounds seems to occur without restriction to certain genetic lineages. Changes at the capsular locus are probably driven by the host immune response, and the new genotype/serotype combination may be lost, or may persist in the population and diversify if increased fitness is acquired. This might explain the observation of an
altered serotype appearing only once or twice, and that the stable clones such as ST19 and 17 of serotype III persists/dominates. Despite that serotype V (CC1) has emerged relatively recently (Blumberg et al., 1996), it appears to be subject to frequent capsule switching. A lesson can be learned from our finding about capsule switching and the fact that genotype shifts has already been noted in the post-vaccination era of S. pneumoniae (Beall et al., 2006). An ideal vaccine for S. agalactiae should include capsule from multiple serotypes and a specific protein to confer herd immunity, and therefore avoid selection pressure on S. agalactiae itself.

**Distribution of mobile genetic elements**

The presence of insertion sequences (IS) and GBSi1 was investigated by PCR screening, followed by sequencing. IS1381 was found to be the most widely distributed element, present in 174 isolates (94%), whereas GBSi1 was only found among 32 isolates (17.3%) (Paper III, Figure 1). IS861 was detected in 98 isolates (53%), and IS1548 in 70 isolates (37.8%). Based on the presence or absence of MGEs, 11 genetic variants were identified among the isolates.

Most isolates harboured either one or three MGEs, 35.1% and 39.5%, respectively; while 45 isolates (24.3%) had two MGEs. CC1, CC17, and CC19 showed different combinations of MGEs, respectively; whereas CC1 and ST23 shared the same combination, where the majority of the isolates harboured only IS1381 (Paper III, Figure 1). Strikingly, among serotype III isolates mutually exclusive distribution was observed only between GBSi1 and IS1548, but not between any of the other MGEs. All CC19 isolates carried IS1548 regardless of source of isolate but none harboured GBSi1, whereas GBSi1 was present in all CC17 isolates. IS1548 was not found among any of the isolates belonging to CC17 (Paper II and III). Interestingly, in the isolates of CC1 and CC9, IS1548 was seldom found among invasive isolates but was common in wound isolates. In contrast, CC19 harbours IS1548 regardless of source of isolates (Paper III).
Importantly, in the present study higher heterogeneity in the MGE patterns, in addition to the serotypic variation, was noted among non-invasive isolates. MGEs are important cause of genome plasticity. Taking into account that there are multiple copies of IS elements in *S. agalactiae* isolates and the rearrangement of the genome they may cause, the snapshot of MGE pattern seen here indicate a substantial genomic diversity, which possibly cause phenotypic differences (Franken et al., 2001; Granlund et al., 1998; Spellerberg et al., 2000). These genetic and possible phenotypic differences may be important for *S. agalactiae* isolates to adapt to different niches.

**Distribution of the *fbsA* and *spb1* genes among serotype III isolates**

To analyze difference in the distribution of virulence factors in the major genetic lineages of serotype III, we examined the presence of *fbsA* and *spb1* by PCR in a number of invasive and noninvasive isolates belonging to CC17 and CC19. The *fbsA* gene was found in 12 out of 13 CC17 isolates tested, and 10 of these isolates harboured the *spb1* genes. In contrast, 40 out of totally 43 isolates of CC19 lacked the *fbsA* and *spb1* gene. Further sequence analysis revealed that CC19 isolates harboured a truncated *fbsA* gene, which resulted in deletion of the signal peptide. Consequently the FbsA protein may not likely be expressed on the surface of CC19 isolates (Our unpublished data). FbsA has been shown to promotes adherence to and invasion of human brain microvascular endothelial cells, indicating that FbsA plays an important role in meningitis caused by *S. agalactiae* (Tenenbaum et al., 2005). Bidet et al analysed 110 serotype III isolates by PFGE, (Bidet et al., 2003; Takahashi et al., 1998), and identified a PFGE group accounted for the vast majority of cerebrospinal fluid (CSF) isolates, and GBSi1 was present in most isolates belonging to this group (Bidet et al., 2003). The majority of this PFGE group likely corresponds to the CC17 lineage, since GBSi1 is a putative genetic marker for CC17 among serotype III isolates. Moreover, the presence of an intact copy of *fbsA* appears to correlate with a neonatally associated virulent lineage of serotype III (Brochet et al., 2006; Luan et al., 2005).
The difference in distribution of CC17 and of CC19 between the neonatal and adult patient groups might reflect a difference in a number of factors, such as adhesins, and other virulence factors expressed. So far it is known that ET1/RDP III-3/ST-17 has hyaluronidase activity, shows higher production of sialic acid, exhibits growth inhibition in chemical defined medium with high concentration of phosphate, and possesses spb1 encoding a protein mediating epithelial cell invasion (Adderson et al., 2003; Musser et al., 1989; Takahashi et al., 1998). In addition to these differences, the present study shows that GBSi1 and IS1548 are genetic markers for CC17 and CC19 of serotype III isolates, respectively, in accordance with previous findings (Granlund et al., 2001; Luan et al., 2003; Takahashi et al., 2002). In the future comparative genomics and identification of genes important for host tropism may provide insight into both pathogenesis and evolution of *S. agalactiae*. 
CONCLUSIONS

- Two additional GBSi1 copies were identified at sites different from the primarily identified scpB-lmb locus. However, insertion of GBSi1 at these three sites was unequal. In isolates harbouring a single copy of GBS1 at the scpB-lmb locus, the third target motif of GBSi1 is divergent whereas the second site is intact. This finding indicates that host factors may be required for transposition into alternative sites and they may not be present in all isolates.

- A unique DNA fragment with plasmid features was identified at the scpB-lmb gene locus, and its presence was correlated with that of GBSi1. This fragment may be a remnant of the vehicle which carried GBSi1 into S. agalactiae, and immunity against uptake of the original vector plasmid might be the reason for the uneven distribution pattern of GBSi1.

- MLST analysis revealed five major genetic lineages in both invasive and non-invasive isolates, indicating a clonal population structure of S. agalactiae. There were a number of genetically highly related isolates expressing different capsular types, which suggests rather frequent genetic exchange between isolates. Furthermore, the non-invasive isolates belonging to the same clonal complexes displayed more heterogeneity in capsule expression as well as in the MGE patterns than among invasive isolates. This indicates that less variability is allowed in a highly selective environment such as the blood.

- All major clonal complexes and serotypes caused invasive disease, although their ability to do so varied greatly. CC17 was significantly associated with neonatal invasive disease; whereas CC19 was equally common among isolates from adult and neonatal disease, despite that both CC17 and CC19 expressed capsular type III. This striking difference seen between CC17 and CC19 suggests that clonal complex associated properties, in addition to capsular type, play important roles in the virulence of S. agalactiae. CC1, a new clone emerging since early 1990s, has caused substantial amount of disease among adults. Mutually exclusive distribution of GBSi1 and IS1548 was seen, and they were shown to constitute genetic markers for serotype III CC17 and CC19 isolates, respectively.
ACKNOWLEDGEMENTS

I am grateful to many people who have helped me in various ways and made a difference in my life over these years.

My supervisor, Prof. Mari Norgren for sharing her vast knowledge, enthusiasm and wisdom about both science and life. Her good leadership, great guidance, and strong support have been very important for me during the PhD study. She is truly a gold mine, and I have always enjoyed and benefited from our stimulating discussions on various topics.

My associate supervisor, Dr. Margareta Granlund, for sharing her extensive knowledge about microbiology and scientific writing. Her intellectual sharpness, down to earth spirit, aptitude for accuracy, and attention to detail have been impressive. She always offered time for fruitful discussions despite her busy schedule, and patiently listened to my problems. Margareta and Mari have been my role models for all the years, and I hope they will keep being my mentors in the future.

Prof. Anders Sjöstedt, for kindly letting me participate in his group’s journal club and showing interest in my research. His leadership creates a good working atmosphere at the department.

Prof. Bengt-Harald Jonsson and his wife Lena Tibell for sharing their vast knowledge about protein chemistry, introducing me to Swedish culture, and inviting me to lovely dinner parties. Their hospitality, charm, and kindness are unforgettable.

My collaborators, Nicola Jones, for friendly hosting my stay at Oxford, teaching me the MLST scheme and emphasizing the quality of research and people; Mats Sellin, for giving advice on scientific writing and having stimulating talks at the lunch room; Patrick Ryden for insights and resolutions of my statistical problems, and for sharing views on a scientific career.
My dear colleagues, who have also become good friends through the daily contacts over these years, for creating good working atmosphere, for making research fun, for delicious desserts and refreshing coffee breaks, for sharing your culture and life experience, and for tons of inspiration, encouragement and support. The names include present and previous members in Mari’s group, Linda, calm and commanding, Elvira, a strong fighter; Carin, well organized and considerate; Anna, beating the odds; Lena, a strong woman with a big heart; and Linda-Marie, bright and enthusiastic; as well as members from Anders’ group, Helena, very talented and tough, whom I really look up to; Henrik, a smart guy with good sense of humor, thanks for helping with my computer and sharing the investment tactics; Lao Lai, Igor, brilliant scientists and wise men, and being my classmates learning Swedish; Carl, inspiring us to adopt a more healthy lifestyle; Blanka, smart and sweet; Constatin, well-trained and cold humor; Svenja, wise, passionate and spoil us with delicious food; Linda B, stylish and a good organizer; Linda N, positive and cheerful; Malin, classic and peaceful; Mattias, an intelligent and kind guy; Marie H, harmonious and bright; Laila, social and enthusiastic. And Siv at CMB, I am grateful for all the high-quality sequences and lovely chats we had. I will miss you all.

The secretaries, Kerstin, Margareta and Anna-Märta for your efficiency and plenty of practical help. Anna-Märta, also for being a gym mate and for a lovely holiday at her cottage.

People at Bakt. Lab, for being friendly and helpful. Eva’s fine leadership creates good working atmosphere. In particular, Stig and his staffs for always preparing media and buffers with high quality; Helin, for typing my isolates; Katrina for teaching me the PFGE technique.

And, Jenny at the international office, for quickly arranging an apartment for me this autumn; Gunnila at Dungen and Sonja at Umeå county for kindly and quickly solving the problem regarding day-care; Jessica and other teachers at Regnbåge, for being
warm-hearted and professional; Kersti at Esclapen and Lars at IKSU for not only treating my physical pain but also inspiring me with their life stories.

My Chinese friends, Jie and Yong-Zhi, Ya-Fang, Sa, Li-Fang, Xiao-Ju, Shu-Yan and Lao lai; as well as my Korean friend, Hyunjung for your everlasting friendship and sharing tips on parenthood. Also thanks to Peng and Ming for kind lifts and interesting chats.

Back in China, Prof. Jun-Bao Li for his kind help and guidance at hard time; Prof. Mei-Ling Jing, for sharing her extensive knowledge on epidemiology and life strategies; and my previous colleagues in Hangzhou for being warm-hearted and supportive.

My family in China, my mother Xiu-Mei, for being strong, open-minded, and always having faith in me; my brother Si-Yao and his wife for being understanding, supportive and taking good care of our mother; and my sister, Shi-Shu for being my best friend. My father, a man with beautiful mind, brilliant humor and wisdom, had fostered my free spirit, and inspired me to challenge myself. I miss you so much.

Last, but most, to my darling Hebin for his unfailing love and support through good and bad times. Our son Ting-Kang, for the joy and happiness he has brought into my life and for all the adventures, small or big, fearful or fun, we have had together and a lot more to come in the future.

This work has been supported by grants from the Swedish Research Council (08675), the Västerbottens Landsting, the Medical Faculty at Umeå University, and Insamlings Stiftelsen at Umeå University to MN, and a grant from the Västerbottens Landsting to MG. SLL is a member of the EuroPathoGenomics Graduate Academy, and have been supported in part by the EU network of Excellence EuroPathoGenomics and grants from the Kempe Foundation.
REFERENCES


