GENETIC DIVERSITY AND VIRULENCE GENE CHARACTERIZATION OF GROUP B STREPTOCOCCUS

By

Amber Cody Springman

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Comparative Medicine and Integrative Biology

2012

ABSTRACT

GENETIC DIVERSITY AND VIRULENCE GENE CHARACTERIZATION OF GROUP B STREPTOCOCCUS

By

Amber Cody Springman

Group B Streptococcus (GBS, *Streptococcus agalactiae*) is a Gram-positive, β-hemolytic bacterium that was first described as a causative agent in bovine mastitis. GBS has since become a leading cause of pneumonia, sepsis, and meningitis newborns and has also emerged as an opportunistic pathogen in the elderly and adults with underlying medical conditions. The accessibility to a collection of GBS strains from colonized women, newborns, adults, elderly, and bovines has provided an excellent basis for conducting molecular epidemiological studies. The use of molecular subtyping methods including multilocus sequence typing (MLST), PCR-based RFLP, and mulitiplex PCR has allowed for inferences into the phylogenetic relationships among strains and the identification of specific virulence determinants associated with different clonal lineages. During the first study of this dissertation, the extent of genetic diversity and allelic variation within key virulence genes were analyzed in a panel of 94 GBS strains from diverse sources (e.g. pregnant women, newborns, elderly adults, and bovines). In the second study, a more inclusive set of 295 human-derived and bovine-derived strains representing 73 sequence types were evaluated for the presence of and variation in horizontally acquired pilus islands (PIs). Extensive genetic analyses conducted in these studies revealed that

recombination and horizontal gene transfer have played a major role in the diversification of GBS clonal lineages. Furthermore, evolutionary pressures have favored the selection of specific allelic determinants that has led to the emergence of highly specialized lineages that have become successful at causing disease in humans. Overall, this research has strong implications for vaccine development since many of the virulence genes examined herein have been proposed as vaccine candidates. The high degree of sequence variability observed in many of the potential vaccine targets underscores the difficulty of developing a broadly protective, universal vaccine against GBS. Continued efforts to assess the genetic diversity of GBS populations and the mechanisms involved in pathogenesis will prove critical for establishing new prevention measures to combat GBS-associated disease.

ACKNOWLEDGMENTS

I would first and foremost like to thank my major advisor, Dr. Shannon Manning. She has provided me with unconditional support and been by my side through this entire process. I appreciate all her patience and would have not made it through this degree without her. Secondly, my co-advisor Dr. Dele Davies has been encouraging me since I was an undergraduate researcher, then a technician, and all the way through graduate school. He has been an incredible mentor to me and I am thankful for everything he has taught me through the years. My other committee members – Dr. Linda Mansfield and Dr. Terry Marsh have also provided me great advice and guidance during this time. Dr. Mansfield was there for our lab during a difficult period and it has always been comforting to know I can count on her. And, Dr. Marsh who has known me since I was a stressed out undergraduate to a stressed out graduate student, has always motivated me to keep moving forward and "*Carpe diem."*

I would also like to extend my deepest appreciation to Drs. Vilma Yuzbasiyan-Gurkan and Vicky Hoelzer-Maddox. I could not have managed my day-to-day life as a graduate student without their continuous help and guidance. I am so proud to say I represent the Comparative Medicine and Integrative Biology Program and I commend them for building such an excellent program.

And, I cannot forget my lab mates, undergraduates, fellow graduate students, and closest friends: Lindsey Ouellette, Rhia LeVeque, Barbie Gadsden,

J.P. Jerome, Jessica St. Charles, all of those in the CMIB program, David Lacher, Teresa Bergholtz, Seth Walk, Adam Nelson, Galeb Abu-Ali, Lukas Wick, Weihong, Maggi Lewis, Katie Schaffer, Jacob Sinkoff, Amber Million, and so many others. And a very special thanks to Hans Steinsland who helped me with everything I ever needed, from computer problems to giving my cat a bath in the lab. All of you have made this an unforgettable experience – I know I will be always be able to count on you as close colleagues and, more importantly, as friends.

I especially want to thank Beth Whittam. She has been my #1 fan and done so much for me over the years. Also, my Dad has been so important to me and I am so appreciative of everything he did (and put up with) in raising me and for always believing in me. In addition, I want to thank my Mom, brother Shane, sister Heidi, sister-in-law Inga, all the kids (Jake, Sasha, Jaden, Lexie, Cole, Corbin, McKenna, and Macy), step Mom Peg and step Dad Jim, and last but not least, my Great Aunt Dorty. Of course, I want to thank all of my cats: Jewel, Dayday, Hines, and my beloved Beyonce.

Lastly, I would like to dedicate my dissertation to Dr. Tomas S. Whittam. Not only did he introduce me to the field of Microbial Evolution, but he also opened the doors for me to pursue a career in an area that I truly love. For all of the life lessons and great times – Tom, I thank you.

TABLES OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

GROUP B STREPTOCCUS:

DISEASE, GENETIC VARIABILITY, AND VIRULENCE

INTRODUCTION

Historical View

The earliest documentation of the Gram positive bacterium, *Streptococcus agalactiae*, was in 1887 by the French veterinarian and microbiologist, Edmond Nocard who described the organism as an animal pathogen causing bovine mastitis [1]. In 1933, Rebecca Lancefield introduced a serological classification system for grouping *Streptococcus* from humans, animals, and dairy products [2]. The Lancefield system classified β-hemolytic *S. agalactiae* into group B *Streptococcus,* a group then comprised of strains from bovines with mastitis and milk [2]. The first report of group B *Streptococcus* (GBS) causing fatal newborn infections came from Fry in *The Lancet,* 1938 [3]. By the 1970's, the occurrence of GBS causing newborn disease and meningitis increased dramatically [4-7]. Although measures for preventing infections have been introduced since that time, GBS remains a leading cause of bacterial disease in newborns [8].

Current View of Disease

Today, GBS is mostly thought of as a commensal organism that resides in the genitourinary and lower gastrointestinal tract of healthy adults. Approximately 20% to 40% of women are colonized with GBS during pregnancy [9]. While maternal colonization is usually asymptomatic, it places the newborn at immediate risk for vertical transmission. Up to 50% of newborns born to colonized mothers will also become colonized with GBS. Among colonized newborns, 1-2% will develop early onset

disease (EOD) within the first 6 days of life [10,11]. Another 1-2% of newborns will develop late onset disease (LOD) between 7 and 29 days of life [10,11]. Both classes of disease can manifest as pneumonia, sepsis, and meningitis. While EOD is a result of vertical transmission from mother to newborn, LOD is likely a result of horizontal acquisition from colonized individuals [10].

Based on the Centers for Disease Control and Prevention (CDC) guidelines [12], pregnant women are universally screened for GBS at 35-37 weeks gestation, and if colonized, are administered intrapartum antibiotic prophylaxis (IAP) to reduce the risk of transmission. Since the CDC guidelines were implemented in 1996 [13] and recently updated in 2010 [12], the incidence of EOD has decreased by 33%, however, [14,15] the incidence of LOD has not been affected [11,16-18]. Consequently, GBS remains the leading cause of bacterial disease in newborns with an overall incidence rate of ~0.3 infections per 1,000 live births [8]. Evidence that GBS is becoming resistant to second-line antibiotics [17,19,20] and is emerging as a pathogen among adult [21] and elderly populations [22,23] reinforces the need to seek alternative prevention measures for preventing disease in both newborns and the elderly.

MOLECULAR EPIDEMIOLOGY

GBS can be further subgrouped into nine capsular serotypes, Ia, Ib, II-VIII, based on the antigenic variation of the polysaccharide capsule [24]. In addition to conventional serotyping methods using latex agglutination [25], molecular methods that target DNA polymorphisms in the capsular gene cluster (i.e. PCR-based RFLP) are often used to

characterize strains [26-31]. Epidemiological studies have shown that the prominent disease-causing capsular types in the United States and other developed countries are Ia, Ib, II, III [20,32,33], and more recently, capsular type V [21,23,32]. Capsular type III contributes to the majority of newborn invasive disease followed by types Ia, V, II, and Ib, respectively [18,20,33]. In non-pregnant adults, particularly adults with underlying conditions and the elderly, type V is the most frequently associated with disease [33,34]. Disease in adults and the elderly varies from that in newborns and manifests mainly as skin and soft tissue infections and bacteremia. Urinary tract infections and pneumonia are additional complications seen among the elderly with GBS infections [34]. Estimated case fatality rates for young infants and elderly adults with invasive GBS disease are ~4% and 12%, respectively [16]. As a result, both newborn and elderly populations are considered high-risk and have been the primary focus for vaccine development [32,34].

In additional to capsular genotyping, other molecular tools have been introduced to study the epidemiology of GBS [35-37]. For instance, a multilocus sequence typing (MLST) scheme for GBS [37] has proven highly useful for evaluating epidemiological associations and for understanding phylogenetic relationships [27,28,38]. The MLST procedure involves sequencing of seven conserved housekeeping genes whereby each sequence variant is assigned an allele and the combination of alleles at all 7 loci is used to designate a sequence type (ST). Epidemiological analyses have shown that the capsular type III strains frequently associated with invasive newborn disease primarily belong to ST17 and ST19 [37,39]. Moreover, ST17 accounts for the majority of meningitis and LOD [28,39] cases and has been coined as a “highly virulent” GBS clone

[35,37,40]. Phylogenetic analysis based on MLST data from human and bovine GBS has revealed a close relationship of ST17 strains to some bovine-derived strains, which led to the suggestions that the highly virulent ST17 clone was derived from a bovine ancestor [41,42]. However, recent whole genome analysis of a representative bovine-derived strain revealed a much greater evolutionary distance from the other human-derived strains [43], highlighting a limitation in the MLST method and its ability to adequately portray the evolutionary history of GBS. The question remains as to what factors contributed to the emergence of GBS in humans and the specific characteristics that distinguish human-derived and bovine-derived GBS strains will be explored in more detail throughout this dissertation.

EVOLUTION AND GENETIC DIVERSITY

The increasing number of genome sequences available for GBS has allowed for detailed investigations of the evolutionary history and genetic diversity of GBS. Currently there are nine complete or draft genome sequences available for GBS including, NEM316 (III, ST23) [44], 2603V/R (V, ST110) [45], A909 (Ia, ST7) [46], COH1 (III, ST17) [46], 515 (Ia, ST23) [46], CJB111 (V, ST1) [46], H36B (Ib, ST6) [46], 18RS21 (II, ST19) [46], and FSL S3-026 (III, ST67) [43]. With the exception of FSL S3-026, a bovine strain responsible for clinical mastitis [43], the other genomes are human strains representing different serotypes and phylogenetic lineages or STs [46].

Comparative genome analysis of the eight human GBS strains (average size 2.2 Mb) revealed a pan-genome [47] consisting of a core genome of 1806 genes shared

among all strains (i.e. ~80% of any single genome) and a dispensable genome of 907 genes that are strain-specific or partially shared among strains [46]. The pan-genome model for GBS indicates an 'open' genome, meaning that with each additional genome sequence it can be predicted that an average of 33 new strain-specific genes will be identified [46,48]. Interestingly, many of the strain-specific genes, including known virulence genes, are present within the dispensable genome and carried on genomic islands [45,46]. Although these genomic islands do not fit the classical definition of pathogenicity islands, they are often flanked by mobile elements and contain atypical nucleotide compositions (i.e. GC content), revealing their acquisition through horizontal gene transfer (HGT) [44-46]. The dispensable genome also contains a large number of insertion sequences which are present in a relatively high frequency in the bovine genome strain (i.e. 4.3% of the total genome) [43,46], suggesting that HGT has played an important role in host specificity and diversification of GBS strains.

Evolutionary analyses have provided more insight into the events that have shaped the genome complexity of GBS. A marked diversity observed among GBS strains having the same capsular type and ST [46,49,50], for example, demonstrates that genetic recombination has played an important role in GBS evolution. Indeed, it has been shown experimentally that large chromosomal exchanges up to 334kb can occur between strains through conjugation [51]. Furthermore, the distribution of SNPs across multiple genomes indicates that large chromosomal rearrangements have likely occurred in natural populations [51]. Gaining a better understanding of the extent of genetic variation in GBS populations and how this variation impacts virulence will provide clues as to how GBS emerged as a human pathogen.

NEWBORN PATHOGENESIS AND VIRULENCE DETERMINANTS

The primary prerequisite for newborn disease is maternal genitourinary colonization. The vaginal tract is a highly acidic environment (pH 3.8-4.5) that favors GBS adherence to epithelial cells, in turn, placing newborns at direct risk for infection [52]. Vertical transmission from colonized mothers can occur *in utero* following infection of the placental membrane, or via aspiration of contaminated vaginal fluids during labor and delivery [53]. Entry into the lungs, an environment of near-neutral pH, is known to trigger the expression of specific virulence factors that activate an inflammatory response and cause respiratory distress and pneumonia [53]. The ability of GBS to invade the respiratory epithelium permits access to the bloodstream where infection can rapidly progress to septicemia [53]. During a systemic infection, GBS can disseminate to multiple organs and ultimately infect the central nervous system through penetration of the blood-brain barrier [53]. At each stage of pathogenesis, GBS expresses specific virulence factors that facilitate bacterial interactions with the host. Several GBS virulence factors are discussed below while highlighting the role they play in disease.

Extracellular matrix binding proteins

Binding to eukaryotic extracellular matrix (ECM) components is a primary step in bacterial colonization and invasion. The ECM is a macromolecular structure composed of proteins such as laminin, fibronectin, and fibrinogen that provide support to epithelial and endothelial cells. GBS encodes several surface proteins, namely Lmb, ScpB, and FbsA, which bind specifically to ECM components. Together these proteins mediate the

attachment of GBS to various tissues (i.e. vaginal epithelium, respiratory epithelium, and brain endothelium) and promote dissemination during infections [54].

In particular, the Lmb (laminin binding) protein binds to human laminin [55] and the ScpB (Streptococcal C5a peptidase group B) protein can function as an adhesin and bind to fibronectin [56]. Interestingly, the genes encoding ScpB and Lmb are located adjacent to one another on a 16kb composite transposon [57]. Insertion sequences that flank the *scpB-lmb* region have facilitated the horizontal exchange of this element, which shares 98% identity to the respective genes in *Streptococcus pyogenes* [57]. As opposed to bovine GBS strains, all humans GBS strains harbor the *scpB-lmb* element [50,57]. There are naturally occurring mutations within ScpB that result from a 4 amino-acid deletion [50,58]. While this deletion does eliminate ScpB peptidase activity, it does not affect its affinity for fibronectin [58]. Functionally, the role of both ScpB and Lmb has proven critical for the adherence of GBS to human cells [55,59]. This is a clear example of how the acquisition of virulence-associated genes through horizontal gene transfer has contributed to the ability of GBS to cause human disease.

The surface localized protein, FbsA (fibrinogen-binding protein from *S. a**galactiae***), has a specific role in binding fibrinogen [60]. Fibrinogen, a soluble glycoprotein, is found in high concentrations in blood plasma and is important for blood coagulation [61]. FbsA is a highly repetitive protein composed almost entirely of 16-amino-acid repeats that are responsible for its fibrinogen-binding activity [60]. While nearly all GBS strains contain the gene encoding FbsA, they can differ in the number of repeat regions they contain and can vary in the composition of their repeating units [60].

In contrast to other repetitive genes (i.e. Alpha-C protein genes, discussed in more detail below), the variation in *fsbA* is not a result of recombination between single repeats, but a result of selective pressures to maintain binding activity [60]. Mechanistically, the number of repeats present in FbsA affects the capacity of GBS to bind fibrinogen whereby the greater number of repeats either increases the affinity for fibrinogen and/or the amount of fibrinogen bound by FbsA [60]. The binding to fibrinogen contributes to GBS virulence by inhibiting the deposition of complement factor C3b on the bacterial surface which in turn prevents opsonophagocytic killing by the immune system [53]. As discussed next, this is one of several mechanisms used by GBS to evade immune recognition and increase its survival during the course of infection.

Proteins involved in immune evasion

In addition to FbsA, other proteins including ScpB, CspA (cell surface associated protein), BibA (GBS immunogenic bacterial adhesion), and most importantly, CPS (capsular polysaccharide) are involved in GBS immune evasion. As mentioned, the primary function of ScpB is as a serine protease which functions to cleave C5a, a major chemoattractant generated by the activation of complement in blood [62]. The inactivation of C5a through ScpB cleavage reduces the recruitment of neutrophils to the sites of infection and decreases immunological clearance. Closely related to ScpB is CspA, another serine protease that cleaves fibrinogen into fibrin-like fragments [63]. The adhesive fragments generated by CspA coat the bacterial surface and interfere with immunological recognition [63]. BibA is an immunogenic bacterial adhesion that binds

to human C4-binding protein, a regulator of the classical complement pathway [64]. The function of BibA confers resistance to neutrophilic killing and is required for GBS survival in the blood [64]. In addition to its anti-phagocytic activity, BibA induces the production of IgG-specific antibodies and is being explored as a novel target for therapeutic intervention [65].

Central to GBS immune evasion is the capsular polysaccharide, CPS, which is also the most well studied virulence factor in GBS. The CPS protein contains a terminal sialic acid (Neu5Ac) bound to galactose in a α2→3 linkage, an epitope widely expressed on the surface of human cells [66]. This is a classic mechanism of molecular mimicry that allows GBS to resemble host 'self' and avoid immune recognition. Similar to FbsA, the sialyated capsule interferes with the deposition of C3b and protects GBS from phagocytic killing [67]. Together these strategies used by GBS to circumvent host clearance mechanisms and are why individuals with undeveloped or compromised immune systems (i.e. newborns and the elderly) are particularly prone to GBS-associated disease.

Proteins involved in tissue adherence and cellular destruction

GBS has become a successful human pathogen due, in part, to its ability to adhere to various tissues including the vaginal epithelium, respiratory epithelium, and blood-brain barrier endothelium. A family of surface adhesions, known as Alpha-C proteins, mediates GBS binding to human cervical epithelial cells and interacts with host cell glycosaminoglycan (GAG) to promote bacterial internalization [68]. This group is defined by the Alpha-C and Rib proteins but includes a number of antigenic variants

known as Alp2, Alp3, Alp4, and Epsilon [69]. The genes that encode these proteins contain large, identical, tandem repeating units that vary in size due to events of recombination [70]. In addition to the binding activity of Alpha-C proteins, it has been shown that the number of repeats in these genes can undergo internal deletions during the course of infection as a means of avoiding the immune system [71].

Recently, GBS was found to encode pilus structures, long filamentous appendages on the cell surface that have been mostly studied in Gram negative bacteria. These structures were identified during a bioinformatics analysis to identify antigenic surface-associated proteins that could be tested as vaccine candidates [72,73]. The multigenome investigation revealed three distinct, horizontally acquired pilus islands (PI's): PI-1, PI-2a, and PI-2b [74]. Each PI encodes all the structural components and catalytic enzymes necessary for assembling pilus structures on the cell surface. The PI's include three structural proteins all containing a LPXTG motif – the backbone protein (BP) and two ancillary proteins (AP-1, AP-2) – and two dedicated sortase enzymes belonging to the sortase C superfamily (SrtC) [74]. In contrast to Gram-negative organisms that depend on the aid of chaperones for pilus assembly, polymerization of Gram-positive pili relies on the Sec-dependent secretion system for transport of the LPXTG proteins to the cell surface [75]. Although the PI-encoded SrtC is responsible for polymerization of pilus structures, the conserved housekeeping sortase, SrtA, is required for the covalent attachment of pili to the cell wall [76].

Since pilus structures were identified in GBS, their role in pathogenesis has been under intense investigation. GBS infection models have demonstrated that AP1 of PI-1 and PI-2a (AP1-1 and AP1-2a) confer the adhesive capacity of GBS to human tissues

[77,78] while the BP of PI-1 and PI-2a (BP-1 and BP-2a) mediate its intracellular and paracellular translocation across cells [79-81]. In addition, PI-2a pili have a specific involvement in GBS biofilm formation [78,82] and the Spb1/SAN1518 protein of PI-2b is suggested to enhance intracellular survival of GBS in macrophages [83]. GBS pilus components are also highly immunogenic, making them appealing vaccine candidates. Recently, a combinatorial pilus-based vaccine containing BP-1, BP-2b, and AP1-2a revealed good correlates of protection in mice and is a promising strategy for preventing GBS infections [84].

In addition to the adhesive proteins involved in GBS invasion, other factors contribute to this process by causing direct cellular injury and damaging host protective barriers. The β-haemolysin/cytolysin (β-H/C), encoded by the *cylE* gene, is a pore-forming toxin that damages lung and brain tissues, impairs cardiac function, and induces host inflammatory responses [85-87]. Another pore-forming toxin, known as the CAMP factor, is a secreted protein that produces discrete transmembrane pores and cell lysis [88]. The CAMP factor, encoded by the *cfb* gene, has been exploited for decades to diagnostically identify GBS. When grown next to colonies of *Staphlococcus aureus* on a blood agar plate, GBS will produce a distinct zone of hemolysis that distinguishes it from other organisms [89]. The activity of both β-H/C and CAMP are responsible for the widespread tissue damage and the systemic spread seen during GBS infections.

REGULATION OF VIRULENCE FACTORS

The propensity of GBS to exist as both a commensal organism and as a highly invasive pathogen reflects its ability to tightly regulate virulence gene expression under different host conditions [90]. Bacterial gene regulation is primarily accomplished through two-component signaling systems (TCS). A typical TCS is comprised of a membrane-associated sensor histidine kinase that responds to an environmental stimulus and cognate response regulator that mediates a cellular response, mostly through altering the expression of target genes [91]. There are at least 20 known TCS in GBS of which the CovRS (Cov, control of virulence) system is most recognized for its role in pathogenesis [90,92-94].

In the CovRS (or CsrRS) system, CovS serves as the sensor histidine kinase and responds to changes in environmental pH, while its cognate response regulator, CovR, acts as a repressor and binds directly to the regulatory regions of several genes [93,95]. Under neutral pH conditions, CovS is active and autophosphorylates at a conserved histidine residue, and subsequently phosphorylates CovR at a conserved aspartate residue in its receiver domain [96]. Phosphorylation of CovR at the receiver domain induces promoter binding via its effector domain and increases the repression of target genes [96]. In contrast, CovS is inactive under acidic conditions, which results in the dephosphorylation of CovS and CovR and the decreased repression of target genes by CovR [96].

The response of the CovRS system to acidic stress has a global impact on gene expression in GBS. In addition to genes involved in transport, metabolism, and stress response, many of the CovRS-regulated genes are known virulence factors [95]. Specifically, *bibA*, *cylE*, *fbsA*, *scpB* and pilus components are all under the control of the CovRS system and are over-expressed at a high pH [95]. The pH-dependent regulation of CovRS conforms to the lifestyle of GBS, which takes on a colonizing phenotype while inhabiting the vaginal tract then transitions into an invasive phenotype while infecting the newborn [96]. During asymptomatic colonization of the vaginal tract, an acidic environment, virulence genes are down-regulated and genes required for colonization (i.e. transporters and adherence factors) are up-regulated [95]. When newborn transmission occurs, GBS exposure to the lungs results in a shift from an acidic to near-neutral environment, which leads to the inactivation of CovRS and the up-regulation of virulence genes [95].

The modulation of gene expression such as that seen under the CovRS system is a reflection of how GBS adequately adapts to changing host environments. The appropriate expression of genes in response to environmental conditions can provide GBS a survival advantage and impact its virulence in the host [90]. The increased expression of virulence factors is not only important for GBS pathogenesis but also carries implications for vaccine development [95]. The protective efficacy of a vaccine will depend on the surface exposure of targeted antigens. In this sense, the increased expression of virulence targets would increase the opsonophagocytic killing of GBS by immune serum and enhance a vaccine's effectiveness at preventing GBS-associated disease [95].

EFFORTS FOR VACCINCE DEVELOPMENT

Given that GBS continues to cause life-threatening infections in vulnerable populations and current prevention measures have been suboptimal at preventing disease, there are vigorous efforts to identify vaccine targets and/or alternative therapeutic interventions that will alleviate the threat of disease [97]. For decades, the capsular polysaccharide has been the primary vaccine target, however, logistical issues have hindered its progress and recent efforts have shifted toward identifying novel vaccine candidates.

CPS gained its first appeal as a promising vaccine candidate from the early observation that the risk of newborn disease is inversely correlated to the levels of maternal CPS specific antibodies in serum at the time of delivery [97]. The fact that newborns can acquire maternal opsonizing antibodies through the placenta provides a rationale for targeting the vaccination of women at childbearing age as a means of protecting newborns against GBS infections [32]. Although purified CPS was initially determined to be poorly immunogenic, the coupling of CPS to a tetanus toxoid (TT) carrier in the form of a conjugated vaccine greatly enhances its protective role [32]. Currently, the goal is to clinically evaluate a multivalent conjugate vaccine containing the five most prevalent capsular types: Ia, Ib, II, III, and IV [32]. While indications are that these preparations are highly effective and safe, the limitation of a pentavalent CPS-TT vaccine only providing coverage to a subset serotypes has created considerable apprehension and delayed its progress toward advanced clinical trials.

More recently, the advent of whole-genome sequencing has introduced a new, more streamlined approach to vaccine development. The availability of genomic information allows for multi-strain, *in silico* analyses that can predict antigens that are likely to be good vaccine candidates [98]. GBS was one of the first bacterial pathogens that this approach, referred to as reverse vaccinology, was exploited [72]. In an attempt to identify a universal GBS vaccine, the eight human genomes were interrogated for putative surface-associated and secreted proteins that could be used as vaccines [72]. Among the eligible proteins selected, four of them conferred protection in mice [72]. Interestingly, three of these protective antigens (the fourth being the surface immunogenic protein, Sip) were determined to be components of GBS pilus structures [73]. The basis of this work not only provoked attention regarding the presence of pili in GBS but also led to an investigation as to whether a vaccine exclusively containing pilus components could be used to prevent disease [84]. A combinatorial pilus-based vaccine containing components from each pilus island (i.e. BP of PI-1 and PI-2a, and AP1 of PI-2b) was subsequently shown to elicit broad protection in mice [84]. It has also been postulated that because pili play in role in promoting bacterial adhesion, a pilus-based vaccine might also elicit antibodies that block GBS colonization [84].

As efforts continue to optimize and evaluate GBS vaccines, further investigations into the molecular epidemiology of GBS will prove critical for the development process. It is with hopes that the work put forth for this dissertation will enhance our understanding of the genetic diversity within GBS populations and contribute to the exploration into developing new and improved strategies for preventing disease caused by GBS.

REFERENCES

REFERENCES

1. Nocard MaRM (1887) Sur une mammite contagieuse des vaches laitieres. Ann Inst Pasteur 1:109.

2. Lancefield RC (1933) A Serological Differentiation of Human and Other Groups of Hemolytic Streptococci. J Exp Med 57: 571-595.

3. Fry RM (1938) Fatal infections by haemolytic streptococcus group B. The Lancet 1:199-201.

4. Hood M, Janney A, Dameron G (1961) Beta hemolytic streptococcus group B associated with problems of the perinatal period. Am J Obstet Gynecol 82: 809-818.

5. Eickhoff TC, Klein JO, Daly AK, Ingall D, Finland M (1964) Neonatal Sepsis and Other Infections Due to Group B Beta-Hemolytic Streptococci. N Engl J Med 271: 1221-1228.

6. Butter MNW DC (1967) Streptococcus agalactiae as a cause of meningitis in the newborn, and of bacteremia in adults. Antonie van Leeuwenhoek 33:439-450.

7. Becroft DM, Farmer K, Mason GH, Morris MC, Stewart JH (1976) Perinatal infections by group B beta-haemolytic streptococci. Br J Obstet Gynaecol 83: 960-966.

8. Prevention CfDCa (2012) Active Bacterial Core Surveillance Report, Emerging Infections Program Network, Group B Streptococcus, 2010.

9. Hansen SM, Uldbjerg N, Kilian M, Sorensen UB (2004) Dynamics of Streptococcus agalactiae colonization in women during and after pregnancy and in their infants. J Clin Microbiol 42: 83-89.

10. Schuchat A, Wenger JD (1994) Epidemiology of group B streptococcal disease. Risk factors, prevention strategies, and vaccine development. Epidemiol Rev 16: 374-402.

11. (2005) Early-onset and late-onset neonatal group B streptococcal disease--United States, 1996-2004. MMWR Morb Mortal Wkly Rep 54: 1205-1208.

12. Verani JR, McGee L, Schrag SJ (2010) Prevention of perinatal group B streptococcal disease--revised guidelines from CDC, 2010. MMWR Recomm Rep 59: 1-36.

13. (1996) Prevention of perinatal group B streptococcal disease: a public health perspective. Centers for Disease Control and Prevention. MMWR Recomm Rep 45: 1-24.

14. Schrag SJ, Zell ER, Lynfield R, Roome A, Arnold KE, et al. (2002) A population-based comparison of strategies to prevent early-onset group B streptococcal disease in neonates. N Engl J Med 347: 233-239.

15. Phares CR, Lynfield R, Farley MM, Mohle-Boetani J, Harrison LH, et al. (2008) Epidemiology of invasive group B streptococcal disease in the United States, 1999-2005. JAMA 299: 2056-2065.

16. Schrag SJ, Zywicki S, Farley MM, Reingold AL, Harrison LH, et al. (2000) Group B streptococcal disease in the era of intrapartum antibiotic prophylaxis. N Engl J Med 342: 15-20.

17. (2007) Perinatal group B streptococcal disease after universal screening recommendations--United States, 2003-2005. MMWR Morb Mortal Wkly Rep 56: 701-705.

18. Jordan HT, Farley MM, Craig A, Mohle-Boetani J, Harrison LH, et al. (2008) Revisiting the need for vaccine prevention of late-onset neonatal group B streptococcal disease: a multistate, population-based analysis. Pediatr Infect Dis J 27: 1057-1064.

19. Manning SD, Foxman B, Pierson CL, Tallman P, Baker CJ, et al. (2003) Correlates of antibiotic-resistant group B streptococcus isolated from pregnant women. Obstet Gynecol 101: 74-79.

20. Imperi M, Gherardi G, Berardi A, Baldassarri L, Pataracchia M, et al. (2011) Invasive neonatal GBS infections from an area-based surveillance study in Italy. Clin Microbiol Infect 17: 1834-1839.

21. Harrison LH, Dwyer DM, Johnson JA (1995) Emergence of serotype V group B streptococcal infection among infants and adults. J Infect Dis 171: 513.

22. Farley MM, Harvey RC, Stull T, Smith JD, Schuchat A, et al. (1993) A population-based assessment of invasive disease due to group B Streptococcus in nonpregnant adults. N Engl J Med 328: 1807-1811.

23. Blumberg HM, Stephens DS, Modansky M, Erwin M, Elliot J, et al. (1996) Invasive group B streptococcal disease: the emergence of serotype V. J Infect Dis 173: 365-373.

24. Henrichsen J (1985) Nomenclature of GBS antigens. Antibiot Chemother 35: 303-304.

25. Slotved HC, Elliott J, Thompson T, Konradsen HB (2003) Latex assay for serotyping of group B Streptococcus isolates. J Clin Microbiol 41: 4445-4447.

26. Manning SD, Lacher DW, Davies HD, Foxman B, Whittam TS (2005) DNA polymorphism and molecular subtyping of the capsular gene cluster of group B streptococcus. J Clin Microbiol 43: 6113-6116.

27. Manning SD, Lewis MA, Springman AC, Lehotzky E, Whittam TS, et al. (2008) Genotypic diversity and serotype distribution of group B streptococcus isolated from women before and after delivery. Clin Infect Dis 46: 1829-1837.

28. Manning SD, Springman AC, Lehotzky E, Lewis MA, Whittam TS, et al. (2009) Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada. J Clin Microbiol 47: 1143-1148.

29. Imperi M, Pataracchia M, Alfarone G, Baldassarri L, Orefici G, et al. (2010) A multiplex PCR assay for the direct identification of the capsular type (Ia to IX) of Streptococcus agalactiae. J Microbiol Methods 80: 212-214.

30. Madzivhandila M, Adrian PV, Cutland CL, Kuwanda L, Schrag SJ, et al. (2011) Serotype distribution and invasive potential of group B streptococcus isolates causing disease in infants and colonizing maternal-newborn dyads. PLoS One 6: e17861.

31. Ippolito DL, James WA, Tinnemore D, Huang RR, Dehart MJ, et al. (2010) Group B streptococcus serotype prevalence in reproductive-age women at a tertiary care military medical center relative to global serotype distribution. BMC Infect Dis 10: 336.

32. Baker CJ, Edwards MS (2003) Group B streptococcal conjugate vaccines. Arch Dis Child 88: 375-378.

33. Harrison LH, Elliott JA, Dwyer DM, Libonati JP, Ferrieri P, et al. (1998) Serotype distribution of invasive group B streptococcal isolates in Maryland: implications for vaccine formulation. Maryland Emerging Infections Program. J Infect Dis 177: 998-1002.

34. Edwards MS, Baker CJ (2005) Group B streptococcal infections in elderly adults. Clin Infect Dis 41: 839-847.

35. Musser JM, Mattingly SJ, Quentin R, Goudeau A, Selander RK (1989) Identification of a high-virulence clone of type III Streptococcus agalactiae (group B Streptococcus) causing invasive neonatal disease. Proc Natl Acad Sci U S A 86: 4731-4735.

36. Nagano Y, Nagano N, Takahashi S, Murono K, Fujita K, et al. (1991) Restriction endonuclease digest patterns of chromosomal DNA from group B beta-haemolytic streptococci. J Med Microbiol 35: 297-303.

37. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, et al. (2003) Multilocus sequence typing system for group B streptococcus. J Clin Microbiol 41: 2530-2536.

38. Evans JJ, Bohnsack JF, Klesius PH, Whiting AA, Garcia JC, et al. (2008) Phylogenetic relationships among Streptococcus agalactiae isolated from piscine, dolphin, bovine and human sources: a dolphin and piscine lineage associated with a fish epidemic in Kuwait is also associated with human neonatal infections in Japan. J Med Microbiol 57: 1369-1376.

39. Davies HD, Jones N, Whittam TS, Elsayed S, Bisharat N, et al. (2004) Multilocus sequence typing of serotype III group B streptococcus and correlation with pathogenic potential. J Infect Dis 189: 1097-1102.

40. Jones N, Oliver KA, Barry J, Harding RM, Bisharat N, et al. (2006) Enhanced invasiveness of bovine-derived neonatal sequence type 17 group B streptococcus is independent of capsular serotype. Clin Infect Dis 42: 915-924.

41. Bisharat N, Crook DW, Leigh J, Harding RM, Ward PN, et al. (2004) Hyperinvasive neonatal group B streptococcus has arisen from a bovine ancestor. J Clin Microbiol 42: 2161-2167.

42. Bohnsack JF, Whiting AA, Martinez G, Jones N, Adderson EE, et al. (2004) Serotype III Streptococcus agalactiae from bovine milk and human neonatal infections. Emerg Infect Dis 10: 1412-1419.

43. Richards VP, Lang P, Bitar PD, Lefebure T, Schukken YH, et al. (2011) Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae. Infect Genet Evol 11: 1263-1275.

44. Glaser P, Rusniok C, Buchrieser C, Chevalier F, Frangeul L, et al. (2002) Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease. Mol Microbiol 45: 1499-1513.

45. Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, et al. (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proc Natl Acad Sci U S A 99: 12391-12396.

46. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, et al. (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A 102: 13950-13955.

47. Medini D, Donati C, Tettelin H, Masignani V, Rappuoli R (2005) The microbial pan-genome. Curr Opin Genet Dev 15: 589-594.

48. Mira A, Martin-Cuadrado AB, D'Auria G, Rodriguez-Valera F (2010) The bacterial pan-genome:a new paradigm in microbiology. Int Microbiol 13: 45-57.

49. Brochet M, Couve E, Zouine M, Vallaeys T, Rusniok C, et al. (2006) Genomic diversity and evolution within the species Streptococcus agalactiae. Microbes Infect 8: 1227-1243.

50. Springman AC, Lacher DW, Wu G, Milton N, Whittam TS, et al. (2009) Selection, recombination, and virulence gene diversity among group B streptococcal genotypes. J Bacteriol 191: 5419-5427.

51. Brochet M, Rusniok C, Couve E, Dramsi S, Poyart C, et al. (2008) Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of Streptococcus agalactiae. Proc Natl Acad Sci U S A 105: 15961-15966.

52. Tamura GS, Kuypers JM, Smith S, Raff H, Rubens CE (1994) Adherence of group B streptococci to cultured epithelial cells: roles of environmental factors and bacterial surface components. Infect Immun 62: 2450-2458.

53. Doran KS, Nizet V (2004) Molecular pathogenesis of neonatal group B streptococcal infection: no longer in its infancy. Mol Microbiol 54: 23-31.

54. Lindahl G, Stalhammar-Carlemalm M, Areschoug T (2005) Surface proteins of Streptococcus agalactiae and related proteins in other bacterial pathogens. Clin Microbiol Rev 18: 102-127.

55. Spellerberg B, Rozdzinski E, Martin S, Weber-Heynemann J, Schnitzler N, et al. (1999) Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of Streptococcus agalactiae to human laminin. Infect Immun 67: 871-878.

56. Beckmann C, Waggoner JD, Harris TO, Tamura GS, Rubens CE (2002) Identification of novel adhesins from Group B streptococci by use of phage display reveals that C5a peptidase mediates fibronectin binding. Infect Immun 70: 2869-2876.

57. Franken C, Haase G, Brandt C, Weber-Heynemann J, Martin S, et al. (2001) Horizontal gene transfer and host specificity of beta-haemolytic streptococci: the role of a putative composite transposon containing scpB and lmb. Mol Microbiol 41: 925-935.

58. Tamura GS, Hull JR, Oberg MD, Castner DG (2006) High-affinity interaction between fibronectin and the group B streptococcal C5a peptidase is unaffected by a naturally occurring four-amino-acid deletion that eliminates peptidase activity. Infect Immun 74: 5739-5746.

59. Cheng Q, Stafslien D, Purushothaman SS, Cleary P (2002) The group B streptococcal C5a peptidase is both a specific protease and an invasin. Infect Immun 70: 2408-2413.

60. Schubert A, Zakikhany K, Schreiner M, Frank R, Spellerberg B, et al. (2002) A fibrinogen receptor from group B Streptococcus interacts with fibrinogen by repetitive units with novel ligand binding sites. Mol Microbiol 46: 557-569.

61. Fuss C, Palmaz JC, Sprague EA (2001) Fibrinogen: structure, function, and surface interactions. J Vasc Interv Radiol 12: 677-682.

62. Chmouryguina I, Suvorov A, Ferrieri P, Cleary PP (1996) Conservation of the C5a peptidase genes in group A and B streptococci. Infect Immun 64: 2387-2390.

63. Harris TO, Shelver DW, Bohnsack JF, Rubens CE (2003) A novel streptococcal surface protease promotes virulence, resistance to opsonophagocytosis, and cleavage of human fibrinogen. J Clin Invest 111: 61-70.

64. Santi I, Scarselli M, Mariani M, Pezzicoli A, Masignani V, et al. (2007) BibA: a novel immunogenic bacterial adhesin contributing to group B Streptococcus survival in human blood. Mol Microbiol 63: 754-767.

65. Santi I, Maione D, Galeotti CL, Grandi G, Telford JL, et al. (2009) BibA induces opsonizing antibodies conferring in vivo protection against group B Streptococcus. J Infect Dis 200: 564-570.

66. Angata T, Varki A (2002) Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. Chem Rev 102: 439-469.

67. Marques MB, Kasper DL, Pangburn MK, Wessels MR (1992) Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci. Infect Immun 60: 3986-3993.

68. Bolduc GR, Baron MJ, Gravekamp C, Lachenauer CS, Madoff LC (2002) The alpha C protein mediates internalization of group B Streptococcus within human cervical epithelial cells. Cell Microbiol 4: 751-758.

69. Creti R, Fabretti F, Orefici G, von Hunolstein C (2004) Multiplex PCR assay for direct identification of group B streptococcal alpha-protein-like protein genes. J Clin Microbiol 42: 1326-1329.

70. Lachenauer CS, Creti R, Michel JL, Madoff LC (2000) Mosaicism in the alpha-like protein genes of group B streptococci. Proc Natl Acad Sci U S A 97: 9630-9635.

71. Madoff LC, Michel JL, Gong EW, Kling DE, Kasper DL (1996) Group B streptococci escape host immunity by deletion of tandem repeat elements of the alpha C protein. Proc Natl Acad Sci U S A 93: 4131-4136.

72. Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, et al. (2005) Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science 309: 148-150.

73. Lauer P, Rinaudo CD, Soriani M, Margarit I, Maione D, et al. (2005) Genome analysis reveals pili in Group B Streptococcus. Science 309: 105.

74. Rosini R, Rinaudo CD, Soriani M, Lauer P, Mora M, et al. (2006) Identification of novel genomic islands coding for antigenic pilus-like structures in Streptococcus agalactiae. Mol Microbiol 61: 126-141.

75. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in gram-positive pathogens. Nat Rev Microbiol 4: 509-519.

76. Nobbs AH, Rosini R, Rinaudo CD, Maione D, Grandi G, et al. (2008) Sortase A utilizes an ancillary protein anchor for efficient cell wall anchoring of pili in Streptococcus agalactiae. Infect Immun 76: 3550-3560.

77. Dramsi S, Caliot E, Bonne I, Guadagnini S, Prevost MC, et al. (2006) Assembly and role of pili in group B streptococci. Mol Microbiol 60: 1401-1413.

78. Konto-Ghiorghi Y, Mairey E, Mallet A, Dumenil G, Caliot E, et al. (2009) Dual role for pilus in adherence to epithelial cells and biofilm formation in Streptococcus agalactiae. PLoS Pathog 5: e1000422.

79. Pezzicoli A, Santi I, Lauer P, Rosini R, Rinaudo D, et al. (2008) Pilus backbone contributes to group B Streptococcus paracellular translocation through epithelial cells. J Infect Dis 198: 890-898.

80. Maisey HC, Hensler M, Nizet V, Doran KS (2007) Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells. J Bacteriol 189: 1464-1467.

81. Krishnan V, Gaspar AH, Ye N, Mandlik A, Ton-That H, et al. (2007) An IgG-like domain in the minor pilin GBS52 of Streptococcus agalactiae mediates lung epithelial cell adhesion. Structure 15: 893-903.

82. Rinaudo CD, Rosini R, Galeotti CL, Berti F, Necchi F, et al. (2010) Specific involvement of pilus type 2a in biofilm formation in group B Streptococcus. PLoS One 5: e9216.

83. Chattopadhyay D, Carey AJ, Caliot E, Webb RI, Layton JR, et al. (2011) Phylogenetic lineage and pilus protein Spb1/SAN1518 affect opsonin-independent phagocytosis and intracellular survival of Group B Streptococcus. Microbes Infect 13: 369-382.

84. Margarit I, Rinaudo CD, Galeotti CL, Maione D, Ghezzo C, et al. (2009) Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm. J Infect Dis 199: 108-115.

85. Doran KS, Chang JC, Benoit VM, Eckmann L, Nizet V (2002) Group B streptococcal beta-hemolysin/cytolysin promotes invasion of human lung epithelial cells and the release of interleukin-8. J Infect Dis 185: 196-203.

86. Doran KS, Liu GY, Nizet V (2003) Group B streptococcal beta-hemolysin/cytolysin activates neutrophil signaling pathways in brain endothelium and contributes to development of meningitis. J Clin Invest 112: 736-744.

87. Hensler ME, Miyamoto S, Nizet V (2008) Group B streptococcal beta-hemolysin/cytolysin directly impairs cardiomyocyte viability and function. PLoS One 3: e2446.

88. Lang S, Palmer M (2003) Characterization of Streptococcus agalactiae CAMP factor as a pore-forming toxin. J Biol Chem 278: 38167-38173.

89. Darling CL (1975) Standardization and evaluation of the CAMP reaction for the prompt, presumptive identification of Streptococcus agalactiae (Lancefield group B) in clinical material. J Clin Microbiol 1: 171-174.

90. Rajagopal L (2009) Understanding the regulation of Group B Streptococcal virulence factors. Future Microbiol 4: 201-221.

91. Beier D, Gross R (2006) Regulation of bacterial virulence by two-component systems. Curr Opin Microbiol 9: 143-152.

92. Jiang SM, Cieslewicz MJ, Kasper DL, Wessels MR (2005) Regulation of virulence by a two-component system in group B streptococcus. J Bacteriol 187: 1105-1113.

93. Lamy MC, Zouine M, Fert J, Vergassola M, Couve E, et al. (2004) CovS/CovR of group B streptococcus: a two-component global regulatory system involved in virulence. Mol Microbiol 54: 1250-1268.

94. Jiang SM, Ishmael N, Dunning Hotopp J, Puliti M, Tissi L, et al. (2008) Variation in the group B Streptococcus CsrRS regulon and effects on pathogenicity. J Bacteriol 190: 1956-1965.

95. Santi I, Grifantini R, Jiang SM, Brettoni C, Grandi G, et al. (2009) CsrRS regulates group B Streptococcus virulence gene expression in response to environmental pH: a new perspective on vaccine development. J Bacteriol 191: 5387-5397.

96. Lembo A, Gurney MA, Burnside K, Banerjee A, de los Reyes M, et al. (2010) Regulation of CovR expression in Group B Streptococcus impacts blood-brain barrier penetration. Mol Microbiol 77: 431-443.

97. Baker CJ, Kasper DL (1976) Correlation of maternal antibody deficiency with susceptibility to neonatal group B streptococcal infection. N Engl J Med 294: 753-756.

98. Rappuoli R (2001) Reverse vaccinology, a genome-based approach to vaccine development. Vaccine 19: 2688-2691.

CHAPTER 2

SELECTION, RECOMBINATION, AND VIRULENCE GENE DIVERSITY AMONG GROUP B STREPTOCOCCAL GENOTYPES

A. Cody Springman[1,2], David W. Lacher[1], Guangxi Wu[1,2],

Nicole Milton[1,2], Thomas S. Whittam[1], H. Dele Davies[2], and Shannon D. Manning[1,2]

Microbial Evolution Laboratory, National Food Safety & Toxicology Center[1], and Department of Pediatrics & Human Development[2], Michigan State University, E. Lansing, Michigan 48824

ABSTRACT

Transmission of group B *Streptococcus* (GBS) from mothers to neonates during childbirth is a leading cause of neonatal sepsis and meningitis. Although subtyping tools have identified specific GBS phylogenetic lineages to be more important in neonatal disease, little is known about the genetic diversity of these lineages or the role that recombination and selection play in the generation of emergent genotypes. Here we examined genetic variation, selection and recombination in 7 multilocus sequence typing (MLST) loci from 38 GBS sequence types (STs) representing 7 clonal complexes (CCs), and performed DNA sequencing and PCR-based RFLP analysis of several putative virulence genes to identify gene content differences between genotypes. Despite the low level of diversity in the MLST loci, a neighbor-net analysis revealed a variable range of genetic exchange among GBS CCs, suggesting that recombination is partly responsible for the diversity observed between genotypes. Recombination also is important for several virulence genes, as some gene alleles had evidence for lateral gene exchange across divergent genotypes. The CC-17 lineage, which is associated with neonatal disease, is relatively homogeneous and therefore, appears to have diverged independently with an exclusive set of virulence characteristics. These data suggest that different GBS genetic backgrounds have distinct virulence gene profiles that may be important for disease pathogenesis. Such profiles could be used as markers for the rapid detection of strains with an increased propensity to cause neonatal disease, and may be considered useful vaccine targets.

INTRODUCTION

Group B *Streptococcus* (GBS) is a leading cause of neonatal sepsis, pneumonia and meningitis [1], and causes infections in pregnant women, nonpregnant adults, and the elderly with underlying medical conditions. Maternal GBS colonization is a main risk factor for neonatal disease and roughly 20-40% of pregnant women are colonized [2,3]. Colonization rates of up to 31% and 34% have been documented in young men [4] and nonpregnant women [4,5], respectively, whereas a rate of 22% has been observed in individuals over 65 years [6]. GBS also has been identified as the cause of bovine mastitis in up to 45% of symptomatic bovines [7]. Nine distinct polysaccharide capsule types (serotypes) are known and the serotype distribution varies by population.

The genetic diversity of GBS populations has been studied using a variety of different methods including restriction fragment length polymorphism (RFLP) [8], ribotyping [9,10], pulsed-field gel electrophoresis [11], multilocus enzyme electrophoresis (MLEE) [12], random amplification of polymorphic DNA [13], restriction digestion pattern (RDP) typing [14] and multilocus sequence typing (MLST) [15]. By utilizing methods that focus on conserved genetic changes within GBS strains, virulent GBS clones that have diversified genetically can be identified. Both MLEE and MLST can distinguish the major GBS serotype III clones associated with neonatal invasive disease as clonal sequence type (ST)-17 in the MLST system [15-17] or electrophoretic type (ET)-1 in the MLEE system [12]. This clone is also evident in the RDP system as RDP-III [14].

A recent study of 75 GBS strains representing different sources and STs

reported that the ST-17 lineage is relatively homogeneous and contains a unique set of surface proteins [18]. Homogeneity within a GBS lineage that is significantly associated with neonatal disease is likely important for disease pathogenesis, though few studies have been conducted to identify specific differences in virulence characteristics between lineages. Similarly, the role of selection and recombination in the generation of STs as well as known virulence genes has not been fully explored and requires investigation. Here, we assess the genomic diversity of GBS strains representing a variety of clonal genotypes, examine evidence for selection and recombination, and evaluate the extent of DNA polymorphism and allelic variation in several putative virulence genes.

MATERIALS AND METHODS

***Bacterial strains*.** We selected 94 GBS strains for analysis that represented different multilocus sequence types (STs), origins, and capsular serotypes (Table 1); the most common STs observed in other MLST studies [16,17,19] were overrepresented. Thirty-one strains were well characterized controls obtained from the University of Michigan (n=11), ATCC (n=1; ATCC 12403 [20]), GlaxoSmithKlein (n=6) [21], or the Channing Laboratory (n=11) [22]. The bovine strains were isolated from milk expressed by bovines with mastitis in the United States (n=3) [23] or the United Kingdom (n=3) [24]. Invasive strains from neonates (n=16), adults (n=5) and elderly patients (n=3) were previously collected via population-based surveillance in Alberta, Canada [25], while the

35 colonizing strains were recovered from a cohort of pregnant women in Calgary [2].

Multilocus sequence typing (MLST) and capsule genotyping. MLST was applied to all 94 GBS strains by amplifying and sequencing PCR fragments for seven housekeeping genes *(adhP, atr, glcK, glnA, pheS, sdhA, tkt)* as described [26]. Consensus sequences were trimmed in SeqMan (DNASTAR; Madison, WI) and the GBS database (http://pubmlst.org/sagalactiae/) [27] was used for allele and ST assignments. A previously described [28] PCR-based RFLP assay determined the polysaccharide capsule (*cps*) genotype.

Uncovering allelic variation in cspA, gbs2018, scpB and sip. Published sequences available for *sip* (surface immunogenic protein), *scpB* (C5a peptidase), *cspA* (serine protease), and *gbs2018* (surface protein) were downloaded from GenBank, aligned using the ClustalW algorithm in MegAlign (DNASTAR), and examined for variable regions that could be targeted in RFLP and florescent RFLP (fRFLP) [29] assays. MEGA4 [30] was used to construct neighbor-joining trees [31] and classify the different alleles for each gene. Alleles were labeled numerically and minor variants, which represented alleles that clustered at >98% bootstrap support with a major allele, were labeled with the major allele number followed by a number. After determining how many alleles were present for each gene among the published sequences, an *in silico* restriction digest analysis was conducted to map the expected number and size of DNA fragments generated following digestion with a specific enzyme. A variable region ranging between 2,000 and 4,000 bp was examined in *gbs2018*, while a 1,920 bp PCR

product was examined for *sip*, which utilized primers located in the two flanking genes (*gbs0030* and *gbs0032* [20]) (Table 2). For *scpB*, all 1,411 bp were analyzed, but because GenBank strain AF189002 [32] contained a 12 bp deletion that could not be detected via gel electrophoresis, a conventional RFLP assay was not developed. Instead, an enzyme that produced a 5' overhang, which allows for incorporation of a fluorescently labeled nucleotide, was chosen for use in fRFLP [29]. For *cspA*, there was insufficient sequence variation among the known alleles to distinguish them by RFLP, so DNA sequencing was used instead.

PCR-based RFLP analyses. Primers were developed for each gene that targeted conserved sequences flanking variable regions identified in the published sequence analysis (Table 2). Digestion of PCR products specific for *gbs2018* and *sip* were performed in a 30µl reaction using the following restriction enzymes and conditions: 5U *Dde*I at 37°C overnight for *gbs2018*; and 5U *Taq*$^{\alpha}$I at 65°C overnight as well as 5U *Ban*I, *Ban*II, *Hpa*I at 37°C overnight for *sip*. Conventional RFLP was used to examine the digested PCR products and band sizes were estimated using Pro-Score/RFLP software, version 2.39 [DNA ProScan, Inc.; Nashville, TN]. Rather than RFLP, a portion of the *cspA* gene was sequenced in all of the strains included in the study (Table 2) and an established multiplex PCR protocol [33,34] was used to identify the different alleles of the Alp protein family (*bca*, *alp2*, *alp3*/*alp4*, *epsilon*, and *rib*). The fRFLP protocol developed for *scpB* was used to capture the 12 bp deletion identified in the published sequence analysis. First, 15 µl of PCR product was purified with 6 µL of ExoSAP-IT

[USB; Cleveland, OH] to remove unincorporated dNTPs and primers. The reaction was incubated at 37°C for 15 min and 80°C for 15 min for enzyme inactivation and digested overnight with 5U of *Dde*I at 37°C. Restriction fragments were fluorescently labeled using the GenomeLab Methods Development Kit (Beckman Coulter) by combining 2 µl restriction digest product, 1.5 µl 10x reaction buffer, 0.1 µl ddUTP, 0.1 µl *Taq* DNA polymerase and ddH_2O to 15 µl, and incubating at 60°C for 1 hour. The labeled samples were purified with Sephadex G-50 Fine columns, dried using vacuum centrifugation, and resuspended in 10 µl deionized formamide. Finally, after mixing the samples with CEQ DNA Size Standard 600 (Beckman Coulter), fRFLP was conducted using previously described conditions: a capillary temperature of 50°C, denaturation for 2 min at 90°C, injection at 2 kV for 30 sec, and separation at 4.8 kV for 1 hour [29]. The DNA fragment sizes were visualized using CEQ2000XL software, ver. 4.3.9 (Beckman Coulter; Fullerton, CA).

Identifying additional virulence gene alleles. Because the published sequence data did not represent the extent of variation that could be uncovered in the four virulence genes, DNA sequencing was completed on all GBS strains that yielded unexpected banding patterns in the RFLP analyses. Furthermore, a subset of additional GBS strains were selected for sequencing based on their placement within a neighbor-joining tree [31] derived from the MLST data. A CEQ2000XL DNA sequencer (Beckman Coulter) was used to sequence the gene-specific PCR products, though the number of

strains, DNA fragment sizes, and conditions differed by gene (Table 2). Consensus sequences for *cspA*, *gbs2018*, and *sip* were aligned to published sequences with MegAlign using the ClustalW algorithm (DNASTAR).

Phylogenetic analysis. The neighbor-joining [31] and neighbor-net algorithms [35], untransformed distances (*p* distance) and bootstrap confidence values based on 1000 replications were applied to concatenated MLST and virulence gene sequences using MEGA4 [30] and SplitsTree 4 [36]. To test for the role of past recombination in generating allelic variation, the pairwise homoplasy index (Phi) [37] was calculated in SplitsTree 4 [36].

The action of natural selection on molecular variation was inferred from the number of synonymous substitutions per synonymous site (d_S) and the number of nonsynonymous substitutions per nonsynonymous site (d_N) estimated by the modified Nei-Gojobori method using MEGA4 [30]. In addition to the Phi test, the genetic algorithm recombination detection (GARD) approach with a β-Γ rate distribution and 3 rate classes was used to further evaluate recombination and identify recombination breakpoints for each gene (www.datamonkey.org/GARD) [38]. If breakpoints were identified, then the gene was broken up into segments defined by the breakpoints for further analysis. Testing for positive or negative selection on individual codons within each segment was examined using the single likelihood ancestor counting (SLAC) method (www.datamonkey.org) [39], which also was used to calculate the ratio of d_N to d_S and estimate the 95% confidence interval. Finally, the F84 distance substitution

model from DAMBE (http://dambe.bio.uottawa.ca/dambe.asp) [40] examined the level of substitution saturation by plotting the estimated number of transitions and transversions against the genetic distance for all pairwise comparisons of the alleles for the locus in question. Saturation was observed when the number of transversions exceeded the number of transitions and the transition frequency reached a plateau.

RESULTS

Genetic diversity of MLST loci. A total of 38 STs were represented in this study of 94 GBS strains (Table 1). The 38 STs had 72 variable nucleotide sites among all seven concatenated housekeeping gene sequences (3,546 bp), and were defined by 6-10 alleles per gene (Table 3). The synonymous rate of substitution (d_S) ranged from a low of 0.77 for *tkt* to a high of 1.88 for *atr* with an average of 1.46 synonymous substitutions per 100 synonymous sites. The nonsynonymous rate (d_N) per 100 nonsynonymous sites was lower ranging between 0.04 and 0.46 for *glnA* and *tkt* (Table 3). Tests for selection operating on the allele sequences at each MLST locus based on the SLAC method found no individual sites (codons) to be under significant negative or positive selection, indicating that the MLST loci are evolving neutrally. Furthermore, phylogenetic network analysis revealed no evidence for recombination in any locus except *pheS*, which was not confirmed using the GARD approach. Significant recombination, however, was detected for all 38 STs combined (Table 3).

Population structure and recombination in MLST loci. Among the 94 GBS strains, 31 of the 38 STs clustered into 1 of 7 clonal complexes (CCs); 7 singleton STs were identified that were not part of a cluster (Fig. 1). The largest cluster, CC-23, contained eight STs that grouped together with 98% bootstrap support, whereas the smallest clusters were comprised of only three STs each that grouped together with >87% bootstrap support (Fig. 1). The bovine strains represented three singleton STs (67, 72, and a novel ST (new1)) and one cluster (CC-61) that was not closely related to the CCs containing the human strains (Fig. 1). Of all CCs, however, CC-61 was most closely related to CC-17, though the bootstrap value at the node separating the two clusters was only 42%. The STs within each CC differed by 1.3 (CC-1) to 4.0 (CC-12) nucleotides, and the overall difference between all 7 CCs was only 19.4 nucleotides.

Phylogenetic network analysis clearly separated the CCs identified in the neighbor-joining phylogeny (Fig. 2). CC-23 was located most distantly from the other CCs, while the CC-17 lineage was adjacent to the bovine-specific CC-61 lineage in both the network and neighbor-joining analyses (Figs. 1,2). The neighbor-net network also revealed that a significant degree of recombination has played an important role in the generation of GBS STs, particularly in the sample of singleton STs evaluated. There were too few parsimonious-informative sites in the MLST genes from each CC to detect recombination within CCs, and no recombination was detected ($P = 0.29$) between CCs 17, 23 and 61 (Fig. 2), which represent the predominant clusters at the bottom of the neighbor-joining tree (Fig. 1). After grouping the two singleton STs (ST-22 and ST-32) with CCs 17, 23, and 61, however, there was evidence for recombination ($P = 4.2 \times 10^{-}$

^{4}). Similarly, recombination also was not detected (P = 0.16) between the predominant clusters at the top of the neighbor-joining tree (Fig. 1), though after the singleton STs 36, 103 and new1 were included, recombination was detected ($P = 9.0 \times 10^{-4}$).

Virulence gene diversity. Several major alleles were identified for cspA (n=12) gbs2018 (n=6), sip (n=8), and scpB (n=4) in this population of 38 GBS STs using a combination of PCR-based RFLP assays and sequencing (Table 4). Many of these alleles have not been previously described including *scpB* allele 4, which contains a 128 bp deletion. The most polymorphic gene was *gbs2018*. Considerably more variability was observed in the 2000-4000 bp *gbs2018* PCR fragment than was expected based on the *in silico* analysis. Therefore, major allele assignments for *gbs2018* were made based on the published sequence data from 20 strains representing 6 previously identified alleles [18]; four minor variants were identified in these sequences as well (Supplemental Table 1). Sequencing *gbs2018* in an additional set of 19 strains did not reveal any new alleles, however, a range of 0 to 8 variants were found per allele by performing RFLP of the larger PCR fragment in the remaining 55 strains. In all, 12 *gbs2018* variants were found (Table 4) and 5 strains with *gbs2018* allele 5 contained a 4 bp insertion. The average number of nucleotide differences between the 6 *gbs2018* alleles in the first 600 bp of sequence is 44.9 with a total of 495 (83%) variable nucleotide sites (Table 4). The phylogenetic network constructed from the *gbs2018* sequences revealed several parallel paths of mutation, which is indicative of phylogenetic incompatibilities in the early divergence of alleles providing significant

evidence for recombination (Table 4). Evidence for recombination also was detected in the GARD analysis, as numerous recombination breakpoints were detected. Despite this, saturation analysis revealed that the observed substitutions within *gbs2018* are saturated, thereby rendering all other phylogenetic inferences invalid, including both the SplitsTree and GARD analyses. This saturation level is likely due to the high degree of variation and large number of internal repeats previously identified within *gbs2018* [41], which makes sequence alignments difficult.

By contrast, both *sip* and *cspA* were similar to the MLST loci with fewer variable nucleotide sites and low d_S and d_N rates that are within the range and slightly greater than the average for the seven MLST genes (Table 4). The complete sequence analysis of 41 strains identified 8 major *sip* alleles. Four of the 8 major alleles were identified from 14 previously published sequences (Table 2); one minor variant (*sip*3.2), which contained a 4 bp deletion, was observed in the *sip* allele 3. Sequencing *sip* in an additional 27 strains revealed four novel alleles (*sip*5, *sip*6, *sip*7, *sip*8), and six minor variants (*sip*1.2, *sip*1.3, *sip*1.4, *sip*3.3, *sip*3.4, *sip*8.2) that contain SNPs and/or deletions ranging in size from 60 to 120 bp. In addition, the *sip*5 allele contained a 6 bp insertion at nucleotide 733. For *cspA*, 12 major alleles were identified among the 88 strains sequenced during the course of the study; only 3 of the 12 alleles were present among the 8 genome strains with published sequence data available [20]. Despite the large number of variants found, the 12 *cspA* alleles contain only 18 variable sites, while the 8 *sip* alleles contain 43 variable sites (Table 4). The phylogenetic networks for both genes revealed parallel paths, though only *sip* had significant evidence for recombination (Table 4); the GARD analysis identified two recombination breakpoints in

sip. By contrast, no breakpoints were identified for *cspA* indicating a lack of evidence for recombination, a finding that is consistent with the neighbor-net analysis. Because of the recombination detected in *sip* as well as the lack of evidence for saturation, the d_S and d_N rates were calculated independently for each gene segment using SLAC analysis (Table 4).

Clonal distribution of virulence gene alleles. In general, some CCs are more homogeneous than others with regard to the distribution of specific gene alleles, though this varies depending on the gene of interest (Fig. 3, Supplemental Table 2). For example, only CC-1 contains more than one *gbs2018* allele, as the other CCs are comprised of one predominant allele type. The distribution of *gbs2018* allele 1 suggests that lateral gene exchange is likely responsible for the presence of allele 1 in more divergent genetic backgrounds (Fig. 3C). Additional evidence for lateral gene exchange was identified for *sip* allele 1, *scpB* alleles 1 and 2, the epsilon alpha-like protein gene, and *cspA* allele 1, as each allele is located in multiple independent lineages (Fig. 3). By contrast, evidence for vertical inheritance was observed for the alpha C (*bca*) and *rib* alpha-like protein genes, *cspA* allele 2, *gbs2018* allele 2, and *sip* allele 4, as these alleles are present in closely related CCs and are consistent with the neighbor-joining phylogeny (Fig. 3). Some alleles also have a limited distribution across different CCs, with the most notable lineage (CC-17) containing three unique alleles for *sip*, *cspA*, and *gbs2018* (Fig. 3). CC-61 has unique alleles for *cspA* and *gbs2018*, while the PCR assays for *scpB* were negative (Fig. 3), which is presumably due to a deletion within the PCR primer site or the absence of *scpB* in some strain types.

DISCUSSION

Little is known about the genetic diversity and virulence characteristics of common GBS STs that have been found in multiple populations worldwide. This study of the genetic diversity of GBS strains representing a variety of clonal genotypes confirms the finding that different genetic backgrounds have distinct virulence gene profiles that may be important for disease pathogenesis [18]. This finding is corroborated by comparative genomic data indicating that GBS has an "open pan-genome" when compared to other pathogens [42], which contributes to the overall diversity. Here we examined the population structure of 38 distinct GBS STs and found that the MLST genes appear to be evolving neutrally since no individual codons are under negative or positive selection as inferred by d_N/d_S ratios. The pairwise homoplasy index also indicated that the individual MLST loci lack evidence for recombination, providing additional support [15] that these housekeeping genes are useful for detailed phylogenetic analyses.

The neighbor-joining phylogeny grouped the 38 STs into seven distinct CCs with >80% bootstrap support (Fig. 1); prior studies [19,43] have identified similar clusters. The overall level of diversity is low, though this sample of STs does not represent the complete diversity within the GBS population, as over 440 STs have been described to date [27]. Despite the low diversity level observed in the MLST loci, the extent of networking at the base of each CC in the neighbor-net analysis reveals a range of genetic exchange from low to high as marked by the MLST alleles. For example, CC-23 has a low rate of exchange, CC-7 and CC-12 have medium rates, and CC-17 and

CC-61 have high recombination rates (Fig. 2). These findings suggest the hypothesis that genotypic diversity is attributable to recombination between GBS genotypes in nature, and that some genotypes are likely recombining at a higher frequency than others.

Most population prevalence studies of GBS disease have found that several common GBS STs (e.g., STs 1, 12, 19, 17, 23) occur most frequently, though these studies also report additional singleton STs or variants [16,17,19,43]. It is possible that the more prevalent STs represent the oldest GBS lineages, or that there has been one or more selective sweeps that have removed genotypes with "unfavorable" characteristics and promoted those with "favorable" ones. Furthermore, recombination may be responsible for the generation of closely related STs within a given CC as well as the singletons located on distinct branches of the phylogenetic tree. Because shifts in the type of GBS strains important for disease have been observed [44,45], the high level of recombination among different genotypes combined with the divergence of successful clones could potentially contribute to additional shifts, thereby influencing disease prevention protocols, particularly vaccination. Several GBS vaccine candidates have been investigated including the polysaccharide capsule [46], alpha [47] and beta [48] C proteins, and Sip [49], which are promising alternatives to antibiotic chemoprophylaxis of women during childbirth. It is particularly important, however, for such vaccines to be multivalent and have an ability to recognize different genetic variants in order to effectively combat GBS disease.

In this study, the high level of recombination observed among all STs (Fig. 2) may be important for the generation of novel virulence gene alleles, and correlates with the

degree of lateral gene exchange observed for some genes (e.g., alpha like protein genes, *scpB*) (Fig. 3). Indeed, it is possible that lateral gene exchange is responsible for the distribution of some virulence gene alleles across divergent genotypes, while in other cases, vertical inheritance contributes to the distribution of specific alleles among closely related genotypes. In some genes, however, there is evidence for both lateral and vertical gene transmission, as certain alleles are found in closely related genotypes (e.g., *gbs2018*-2, *sip*2), while others are found in divergent genotypes (e.g., *gbs2018* allele 1, *sip*1) (Fig. 3).

The CC-17 lineage, which contains several unique virulence gene alleles (e.g., *cspA*-3, *gbs2018*-3, *sip*-2) and is relatively homogeneous [18], appears to have diverged independently with an exclusive set of virulence characteristics. The same is true for CC-61, the lineage comprised only of bovine strains. Such virulence characteristics were likely to be initially acquired through point mutation or recombination, as was demonstrated by the high level of gene exchange for MSLT loci representing both lineages (Fig. 2). It is therefore possible that the success of both clones in separate hosts can be attributed to subsequent and independent divergence. This hypothesis is supported by evidence demonstrating that GBS strains from humans and bovines represent distinct populations [18,23,24,50] and that ST-17, a member of CC-17 associated with neonatal disease [16,17,19,51,52] and meningitis [52], evolved from a bovine ancestor [24].

Although recombination appears to be important in the generation of new virulence gene alleles (e.g., *sip*), there was no statistical evidence for positive selection using the SLAC method within *sip* or *cspA*. This is an unexpected finding for two

surface proteins anticipated to be under immunogenic pressure, as such surface proteins are often highly variable with mosaic allele structures and evidence of positive selection [53,54]. For *sip*, several large in-frame deletions were identified in certain alleles, and therefore it is possible that the generation of deletions is a way of creating protein structural variants that may allow GBS to evade host immune system responses. Previous work on Sip has demonstrated that it elicits protective host immune responses [21], and is one of four antigens identified in a multiple genome scan that increases the survival rate of mice challenged with GBS [49]. Despite this, differences in surface accessibility [49] and Sip antibody responses [21] have been observed, and such differences could be attributable to allelic variation within *sip*. Most notably, we found that strains belonging to the neonatal disease-associated CC-17 lineage only possessed the *sip*2 allele. The same is true for *cspA* and *gbs2018*, as the CC-17 lineage was comprised of strains with only *cspA* allele 3 and *gbs2018* allele 3 (Fig. 3); ST-17 strains were shown to have a single allele of *gbs2018* in prior studies as well [18,41]. Together these findings suggest that the independent divergence of the CC-17 lineage has likely led to the dissemination of strains with virulence gene characteristics important for neonatal disease development. Such findings, however, require further investigation and future studies should focus on examining the genetic diversity in a larger subset of GBS genotypes from different sources. In the meantime, specific virulence gene alleles (e.g., *sip*2, *gbs2018*-3, *cspA*3) can be used as markers to rapidly identify GBS infection with a strain belonging to the highly virulent CC-17 lineage.

ACKNOWLEDGEMENTS

We thank Dr. Claire Fraser at The Institute of Genomic Research (TIGR), Drs. Michael Cieslewicz and Lawrence Madoff of the Channing Laboratory at Harvard University, Dr. Betsy Foxman of the University of Michigan for supplying GBS control strains; and Drs. Martin Wiedmann and Nicola Jones for providing bovine strains. We also thank Dr. Hans Steinsland, Katherine Schaeffer, and Maggi Lewis for technical assistance.

This study was supported by Public Health Service grant AI066081 from the National Institutes of Health.

REFERENCES

REFERENCES

1. Schrag SJ, Zywicki S, Farley MM, Reingold AL, Harrison LH, et al. (2000) Group B streptococcal disease in the era of intrapartum antibiotic prophylaxis. N Engl J Med 342: 15-20.

2. Davies HD, Adair C, McGeer A, Ma D, Robertson S, et al. (2001) Antibodies to capsular polysaccharides of group B Streptococcus in pregnant Canadian women: relationship to colonization status and infection in the neonate. J Infect Dis 184: 285-291.

3. Hansen SM, Uldbjerg N, Kilian M, Sorensen UB (2004) Dynamics of Streptococcus agalactiae colonization in women during and after pregnancy and in their infants. J Clin Microbiol 42: 83-89.

4. Bliss SJ, Manning SD, Tallman P, Baker CJ, Pearlman MD, et al. (2002) Group B Streptococcus colonization in male and nonpregnant female university students: a cross-sectional prevalence study. Clin Infect Dis 34: 184-190.

5. Manning SD, Neighbors K, Tallman PA, Gillespie B, Marrs CF, et al. (2004) Prevalence of group B streptococcus colonization and potential for transmission by casual contact in healthy young men and women. Clin Infect Dis 39: 380-388.

6. Edwards MS, Rench MA, Palazzi DL, Baker CJ (2005) Group B streptococcal colonization and serotype-specific immunity in healthy elderly persons. Clin Infect Dis 40: 352-357.

7. Keefe GP (1997) Streptococcus agalactiae mastitis: a review. Can Vet J 38: 429-437.

8. Hauge M, Jespersgaard C, Poulsen K, Kilian M (1996) Population structure of Streptococcus agalactiae reveals an association between specific evolutionary lineages and putative virulence factors but not disease. Infect Immun 64: 919-925.

9. Blumberg HM, Stephens DS, Licitra C, Pigott N, Facklam R, et al. (1992) Molecular epidemiology of group B streptococcal infections: use of restriction endonuclease analysis of chromosomal DNA and DNA restriction fragment length polymorphisms of ribosomal RNA genes (ribotyping). J Infect Dis 166: 574-579.

10. Huet H, Martin C, Geslin P, Grimont F, Quentin R (1993) Ribotyping of Streptococcus agalactiae strains isolated from vaginas of asymptomatic women. Res Microbiol 144: 457-465.

11. Rolland K, Marois C, Siquier V, Cattier B, Quentin R (1999) Genetic features of Streptococcus agalactiae strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and hylB gene analysis. J Clin Microbiol 37: 1892-1898.

12. Musser JM, Mattingly SJ, Quentin R, Goudeau A, Selander RK (1989) Identification of a high-virulence clone of type III Streptococcus agalactiae (group B Streptococcus) causing invasive neonatal disease. Proc Natl Acad Sci U S A 86: 4731-4735.

13. Limansky AS, Sutich EG, Guardati MC, Toresani IE, Viale AM (1998) Genomic diversity among Streptococcus agalactiae isolates detected by a degenerate oligonucleotide-primed amplification assay. J Infect Dis 177: 1308-1313.

14. Takahashi S, Adderson EE, Nagano Y, Nagano N, Briesacher MR, et al. (1998) Identification of a highly encapsulated, genetically related group of invasive type III group B streptococci. J Infect Dis 177: 1116-1119.

15. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, et al. (2003) Multilocus sequence typing system for group B streptococcus. J Clin Microbiol 41: 2530-2536.

16. Jones N, Oliver KA, Barry J, Harding RM, Bisharat N, et al. (2006) Enhanced invasiveness of bovine-derived neonatal sequence type 17 group B streptococcus is independent of capsular serotype. Clin Infect Dis 42: 915-924.

17. Luan SL, Granlund M, Sellin M, Lagergard T, Spratt BG, et al. (2005) Multilocus sequence typing of Swedish invasive group B streptococcus isolates indicates a neonatally associated genetic lineage and capsule switching. J Clin Microbiol 43: 3727-3733.

18. Brochet M, Couve E, Zouine M, Vallaeys T, Rusniok C, et al. (2006) Genomic diversity and evolution within the species Streptococcus agalactiae. Microbes Infect 8: 1227-1243.

19. Bisharat N, Jones N, Marchaim D, Block C, Harding RM, et al. (2005) Population structure of group B streptococcus from a low-incidence region for invasive neonatal disease. Microbiology 151: 1875-1881.

20. Glaser P, Rusniok C, Buchrieser C, Chevalier F, Frangeul L, et al. (2002) Genome sequence of Streptococcus agalactiae, a pathogen causing invasive neonatal disease. Mol Microbiol 45: 1499-1513.

21. Brodeur BR, Boyer M, Charlebois I, Hamel J, Couture F, et al. (2000) Identification of group B streptococcal Sip protein, which elicits cross-protective immunity. Infect Immun 68: 5610-5618.

22. Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, Peterson S, et al. (2002) Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V Streptococcus agalactiae. Proc Natl Acad Sci U S A 99: 12391-12396.

23. Sukhnanand S, Dogan B, Ayodele MO, Zadoks RN, Craver MP, et al. (2005) Molecular subtyping and characterization of bovine and human Streptococcus agalactiae isolates. J Clin Microbiol 43: 1177-1186.

24. Bisharat N, Crook DW, Leigh J, Harding RM, Ward PN, et al. (2004) Hyperinvasive neonatal group B streptococcus has arisen from a bovine ancestor. J Clin Microbiol 42: 2161-2167.

25. Davies HD, Raj S, Adair C, Robinson J, McGeer A (2001) Population-based active surveillance for neonatal group B streptococcal infections in Alberta, Canada: implications for vaccine formulation. Pediatr Infect Dis J 20: 879-884.

26. Manning SD, Lewis MA, Springman AC, Lehotzky E, Whittam TS, et al. (2008) Genotypic diversity and serotype distribution of group B *Streptococcus* isolated from women before and after delivery. Clin Infect Dis 46: 1829-1837.

27. Jolley KA, Chan M-S, Maiden MC (2004) mlstdbNet-distributed multi-locus sequence typing (MLST) database. BMC Bioinformatics 5: 1-8.

28. Manning SD, Lacher DW, Davies HD, Foxman B, Whittam TS (2005) DNA polymorphism and molecular subtyping of the capsular gene cluster of group B Streptococcus. J Clin Microbiol 43: 6113-6116.

29. Lacher DW, Steinsland H, Whittam TS (2006) Allelic subtyping of the intimin locus (eae) of pathogenic Escherichia coli by fluorescent RFLP. FEMS Microbiol Lett 261: 80-87.

30. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.

31. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

32. Bohnsack JF, Takahashi S, Hammitt L, Miller DV, Aly AA, et al. (2000) Genetic polymorphisms of group B streptococcus scpB alter functional activity of a cell-associated peptidase that inactivates C5a. Infect Immun 68: 5018-5025.

33. Creti R, Fabretti F, Orefici G, von Hunolstein C (2004) Multiplex PCR assay for direct identification of group B streptococcal alpha-protein-like protein genes. J Clin Microbiol 42: 1326-1329.

34. Gherardi G, Imperi M, Baldassarri L, Pataracchia M, Alfarone G, et al. (2007) Molecular epidemiology and distribution of serotypes, surface proteins, and antibiotic resistance among group B streptococci in Italy. J Clin Microbiol 45: 2909-2916.

35. Bryant D, Moulton V (2004) Neighbor-net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21: 255-265.

36. Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23: 254-267.

37. Bruen TC, Philippe H, Bryant D (2006) A simple and robust statistical test for detecting the presence of recombination. Genetics 172: 2665-2681.

38. Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SD (2006) Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23: 1891-1901.

39. Kosakovsky Pond SL, Frost SD (2005) Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21: 2531-2533.

40. Xia X, Xie Z (2001) DAMBE: software package for data analysis in molecular biology and evolution. J Hered 92: 371-373.

41. Lamy MC, Dramsi S, Billoet A, Reglier-Poupet H, Tazi A, et al. (2006) Rapid detection of the "highly virulent" group B streptococcus ST-17 clone. Microbes Infect 8: 1714-1722.

42. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, et al. (2005) Genome analysis of multiple pathogenic isolates of Streptococcus agalactiae: implications for the microbial "pan-genome". Proc Natl Acad Sci U S A 102: 13950-13955.

43. Bohnsack JF, Whiting A, Gottschalk M, Dunn DM, Weiss R, et al. (2008) Population structure of invasive and colonizing strains of Streptococcus agalactiae from neonates of six U.S. Academic Centers from 1995 to 1999. J Clin Microbiol 46: 1285-1291.

44. Blumberg HM, Stephens DS, Modansky M, Erwin M, Elliot J, et al. (1996) Invasive group B streptococcal disease: the emergence of serotype V. J Infect Dis 173: 365-373.

45. Rench MA, Baker CJ (1993) Neonatal sepsis caused by a new group B streptococcal serotype. J Pediatr 122: 638-640.

46. Baker CJ, Edwards MS (2003) Group B streptococcal conjugate vaccines. Arch Dis Child 88: 375-378.

47. Pannaraj PS, Kelly JK, Rench MA, Madoff LC, Edwards MS, et al. (2008) Alpha C protein-specific immunity in humans with group B streptococcal colonization and invasive disease. Vaccine 26: 502-508.

48. Pannaraj PS, Kelly JK, Madoff LC, Rench MA, Lachenauer CS, et al. (2007) Group B Streptococcus bacteremia elicits beta C protein-specific IgMand IgG in humans. J Infect Dis 195: 353-356.

49. Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, et al. (2005) Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science 309: 148-150.

50. Finch LA, Martin DR (1984) Human and bovine group B streptococci: two distinct populations. J Appl Bacteriol 57: 273-278.

51. Lin FY, Whiting A, Adderson E, Takahashi S, Dunn DM, et al. (2006) Phylogenetic lineages of invasive and colonizing strains of serotype III group B Streptococci from

neonates: a multicenter prospective study. J Clin Microbiol 44: 1257-1261.

52. Manning SD, Springman AC, Lehotzky E, Lewis MA, Whittam TS, et al. (In press) Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada. J Clin Microbiol.

53. McGraw EA, Li J, Selander RK, Whittam TS (1999) Molecular evolution and mosaic structure of alpha, beta, and gamma intimins of pathogenic Escherichia coli. Mol Biol Evol 16: 12-22.

54. Lachenauer CS, Creti R, Michel JL, Madoff LC (2000) Mosaicism in the alpha-like protein genes of group B streptococci. Proc Natl Acad Sci U S A 97: 9630-9635.

Table 1. Number of group B streptococcal strains representing 38 multilocus sequence types (STs) by polysaccharide capsule (*cps*) genotype, date of isolation, and strain origin.

ST	Inv	Col	Bov	Unk	*cps*	Date	Origin
ST-1 (n=20)	8	10		2	5, 5v, 6, 7, 8	1992-2002	US, Canada, Japan
ST-2 (n=6)	5	1			2, 4, 4v	1970's, 1996-2001	Israel, US, Canada
ST-6 (n=1)	1				1b	1930's	Unknown
ST-7 (n=1)	1				1a	1930's	Unknown
ST-8 (n=5)	2	2		1	1b, 5	1996-1999	US, Canada
ST-10 (n=1)		1			1b	2000	Canada
ST-12 (n=6)	1	5			1b, 2	1999-2002	Canada

Table 1 (cont'd).

ST-14 (n=1)				1	6	Unknown	Unknown
ST-17 (n=6)	5			1	3	1973-1980's, 1995-1996	US, Canada
ST-19 (n=7)	2	4		1	2, 3	1930's, 1998-2000	US, Canada
ST-22 (n=3)	1	2			2	1999, unknown	US, Canada
ST-23 (n=9)	6	3			1a, 2, 3, NT	1930's, 1995-2000	US, Canada
ST-24 (n=1)				1	1a	Unknown	US
ST-25 (n=1)	1				1a	1930's	US
ST-26 (n=2)				2	5	1970's, unknown	US
ST-28 (n=2)	1	1			2	1999-2000	Canada
ST-31 (n=1)	1				3	1999	Canada

Table 1 (cont'd).

ST-32 (n=1)		1			3	2000	Canada
ST-33 (n=1)	1				3	1999	Canada
ST-34 (n=1)	1				3	1995	Canada
ST-36 (n=1)		1			3	1999	Canada
ST-49 (n=1)		1			1a	1999	Canada
ST-61 (n=1)			1		3	1991-1995	US
ST-67 (n=1)			1		2	Unknown	United Kingdom
ST-72 (n=1)			1		2	Unknown	United Kingdom
ST-76 (n=1)			1		2	Unknown	United Kingdom
ST-88 (n=1)		1			1a	2000	Canada

Table 1 (cont'd).

ST-91 (n=1)			1		3	1991-1995	US
ST-103 (n=1)		1			1a	1999	Canada
ST-106 (n=1)		1			3	1999	Canada
ST-110 (n=1)				1	5	Unknown	Unknown
ST-130 (n=1)		1			8	1999	Canada
ST-148 (n=1)	1				3	2001	Canada
ST-411 (n=1)	1				2	1997	Canada
ST-412 (n=1)	1				1a	1996	Canada
ST-414 (n=1)		1			4	1999	Canada
ST-444 (n=1)		1			1a	2000	Canada

Table 1 (cont'd).

ST-new1 (n=1)			1		2	1991-1995	US

Inv=Invasive; Col= Colonizing; Bov=Bovine; Unk=Unknown. Strains with missing information (i.e., "unknown") are typically well characterized control strains

Table 2. Primers for PCR-based RFLP analysis and DNA sequencing, fragment sizes and assay conditions for four genes. Primers (except *gbs2018*) are labeled by the gene or locus name, F or R for forward and reverse, and location within the gene in bp from the start codon, and were developed based on the NEM316 genome sequence (AL732656 [20]).

Table 2 (cont'd).

Gene	Method	Primers	Size (bp)	Annealing temp (°C) (time)	Elongation temp (°C) (time)
cspA	Seq	F2391 (5'-CGAAGTTCCTGGTTCAGAAGATT-3') R2964 (5'-TACTGCAGGACGAGCTTTGAAG-3')	574	53 (30 s)	60 (2 min)
gbs2018	PCR[a]	O1a (5'-AAAATAAACGTGGTCCTATCCTAATAAA-3') O2a (5'-AAAGGCAAAGTTCTGATGAGGTT-3')	2,000 to 4,000	54 (1 min)	72 (4 min)
	Seq	Same forward as PCR (O1a)	600	54 (30 s)	60 (2 min)
scpB	PCR	F284 (5'-CAGCAACCTCAAAAGCGACTATTA-3') R1994 (5'-ACGGTGACTTGTTTGCTGCTATT-3')	1,411	54 (1 min)	72 (1 min)
	Seq[b]	F907 (5'-TGATAGTAGCTTTGGGGGCAAG-3') R1373 (5'-CTTGCCCCCAAAGCTACTATCA-3')	1,411	56 (20 s)	60 (2 min)
sip	PCR	F48 (5'-GCTATTATCAGTCGCAAGTGTTCA-3') R1265 (5'-GCAGTAA CGCCACCACGAT-3')	1,218	54 (1 min)	72 (1 min)
	Seq[c]	F464 (5′-TTGTTTCGCCAATGAAGACATA-3′) R485 (5′-TATGTCTTCATTGGCGAAACAA-3′) F1133 (5′-ACTCTACACAAAATATGGCAGCAA-3′) R1156 (5′-TTGCTGCCATATTTTGTGTAGAGT-3′) F541 (5′-AGTCAAGCAGCAGCTAATGAACAG-3′) R564 (5′-CTGTTCATTAGCTGCTGCTTGACT-3′)	1,920	54 (30 s)	60 (2 min)

Table 2 (cont'd).

[a] Slightly modified primers developed in a prior report [41] were utilized to amplify a *gbs2018* fragment between 2000 and 4000 bp depending on the number of repeats. Sequence analysis was conducted on the first 600 bp (5'end), however, for strains sequenced in this study, only single stranded DNA was examined.

[b] Sequencing of *scpB* utilized three primer sets: the original PCR primers, the F907 and R1373, and the reverse complement of the F907 and R1373. The one strain chosen for *scpB* sequencing yielded a PCR fragment smaller than the expected size using the original primer set; the sequenced fragment is between F284 and R1994, as there is a 128 bp deletion within the original primer site.

[c] *sip* sequencing used three primer sets; *sip* allele 3d required the use of a different primer set (F541 and R564) because of a deletion within F464 and R485 binding sites.

Table 3. Sequence variation in seven multilocus sequence typing genes from 38 sequence types (STs). The fragment size, number (No.) of alleles and variable nucleotide are highlighted, as are the number of synonymous (d_S) and nonsynonymous (d_N) substitutions per synonymous site, the ratio of d_N/d_S and its 95% confidence interval (CI), and the pairwise homoplasy index (Phi) *P*-value.

Table 3 (cont'd).

MLST locus	Size (bp)	No. major alleles	No. variable sites	d_S × 100 (mean ± SE)	d_N × 100 (mean ± SE)	d_N / d_S [95% CI]	Phi (*P*)
adhP	498	10	11	1.61 ± 0.66	0.23 ± 0.11	0.20 [0.06, 0.47]	1.0
atr	501	8	12	2.33 ± 0.74	0.26 ± 0.15	0.09 [0.02, 0.23]	1.0
glcK	459	8	10	1.91 ± 0.75	0.23 ± 0.13	0.13 [0.03, 0.33]	1.0
glnA	498	8	9	1.36 ± 0.55	0.21 ± 0.12	0.11 [0.03, 0.29]	1.0
pheS	501	7	8	1.46 ± 0.69	0.32 ± 0.19	0.13 [0.03, 0.33]	0.04[a]
sdhA	519	6	12	2.58 ± 0.91	0.47 ± 0.23	0.18 [0.07, 0.39]	0.16
tkt	480	8	10	0.89 ± 0.54	0.68 ± 0.30	0.67 [0.31, 1.24]	0.16
STs[b]	3,546	38	72	1.48 ± 0.26	0.20 ± 0.05	0.13 [0.10, 0.18]	6.5×10^{-10}

[a] Despite evidence for recombination using the Phi test [36], a more thorough analysis using the genetic algorithm recombination detection (GARD) [38] failed to identify any recombination breakpoints in *pheS*.

[b] Represents concatenated gene data for all 38 STs.

Table 4. Sequence variation in three putative virulence genes among 94 group B streptococcal strains. The fragment size, number (No.) of major alleles, variants and variable nucleotide sites are noted, as are the pairwise homoplasy index (Phi) *P*-value, mean ratio of d_N/d_S and its 95% confidence interval (CI), and sites showing evidence for negative selection. The genetic algorithm recombination detection (GARD) approach found evidence for two recombination breakpoints in *sip* at 653 bp and 987 bp, which were used to define the three *sip* gene segments used in the single likelihood ancestor counting (SLAC) method analysis. Recombination breakpoints were not observed for *cspA* and the positive evidence for saturation in the 600 bp region of *gbs2018* precluded further GARD and SLAC analyses.

Table 4 (cont'd).

Gene	Size (bp)	No. major/ minor alleles	No. vari-able sites	Phi (*P*)	Full gene d_N / d_S [95% CI]	Seg. 1 d_N / d_S [95% CI]	Seg. 2 d_N / d_S [95% CI]	Seg. 3 d_N / d_S [95% CI]	No. negatively selected sites (location)
cspA	574[a]	12/2	18	0.57	0.15 [0.06, 0.29]	--	--	--	1
sip	1,311	8/7	43	7.9×10^{-4}	0.22 [0.14, 0.32]	0.42 [0.18, 0.82]	0.41 [0.20, 0.74]	0.05 [0.01, 0.13]	1 (Seg. 3)
gbs2018	600	6[b]/11	495	0.00	--	--	--	--	--

[a] A larger fragment (1,490 bp) of *cspA* also was evaluated in 26 strains. Only the data for the shorter fragment (574 bp) are highlighted, as the results for both fragments were similar in each category.

[b] *gbs2018* allele 6 was not observed in GBS strains analyzed in this study, but was identified previously (AM051294) [18].

Figure 1. Phylogenetic relationships of 38 group B streptococcal multilocus sequence types (STs). The numbers at the different nodes represent bootstrap values and only those relationships with > 80% bootstrap support are indicated. The colored circles show the different clusters, or clonal complexes (CCs), that contained STs grouping together with >87% bootstrap support. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

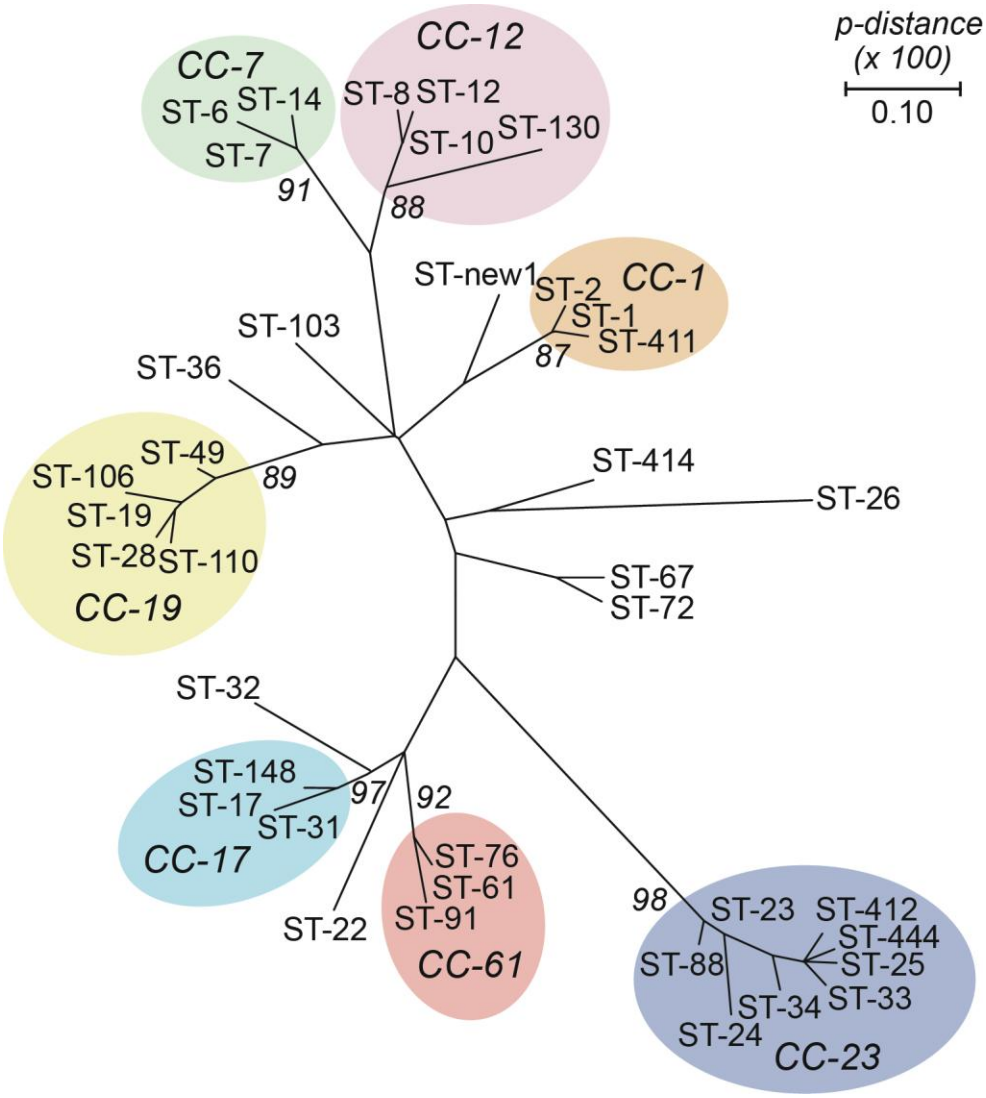


Figure 2. Neighbor-net analysis revealed a complex network with extensive recombination (illustrated as parallelograms) between 38 group B streptococcal multilocus sequence types (STs). Circles highlight the clonal complexes (CCs) identified in the neighbor-joining phylogeny, and the red STs are the six bovine strains.

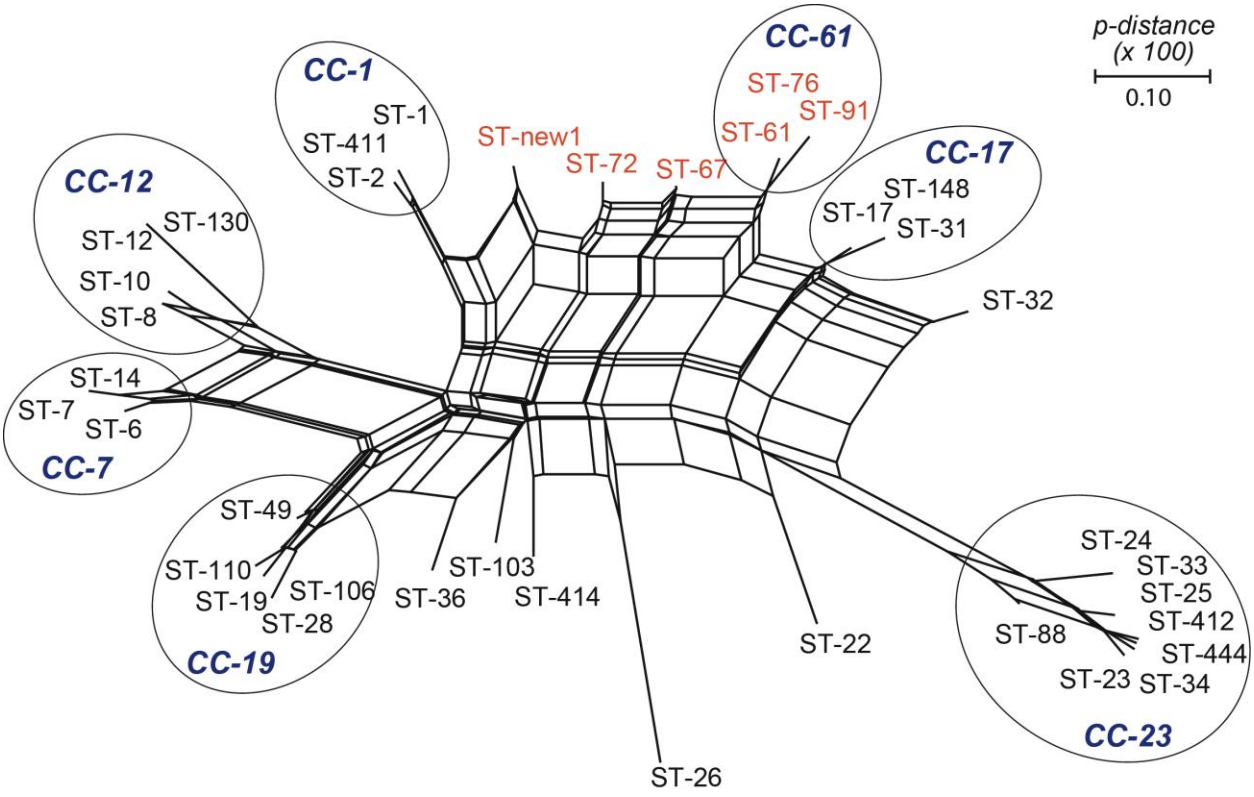


Figure 3. The frequency and distribution of major alleles representing five GBS virulence genes by clonal complex (CC) ranked in order of the neighbor-joining phylogenetic tree (Fig. 1). The frequency of alleles specific for alpha C protein genes (A), *cspA* (B), *gbs2018* (C), *scpB* (D), and *sip* (E)) are indicated by different colors. Only STs that were part of a CC were plotted, as singleton STs located near each CC typically shared the same allele profile and thus, should not be grouped together as one category.

Figure 3 (cont'd).

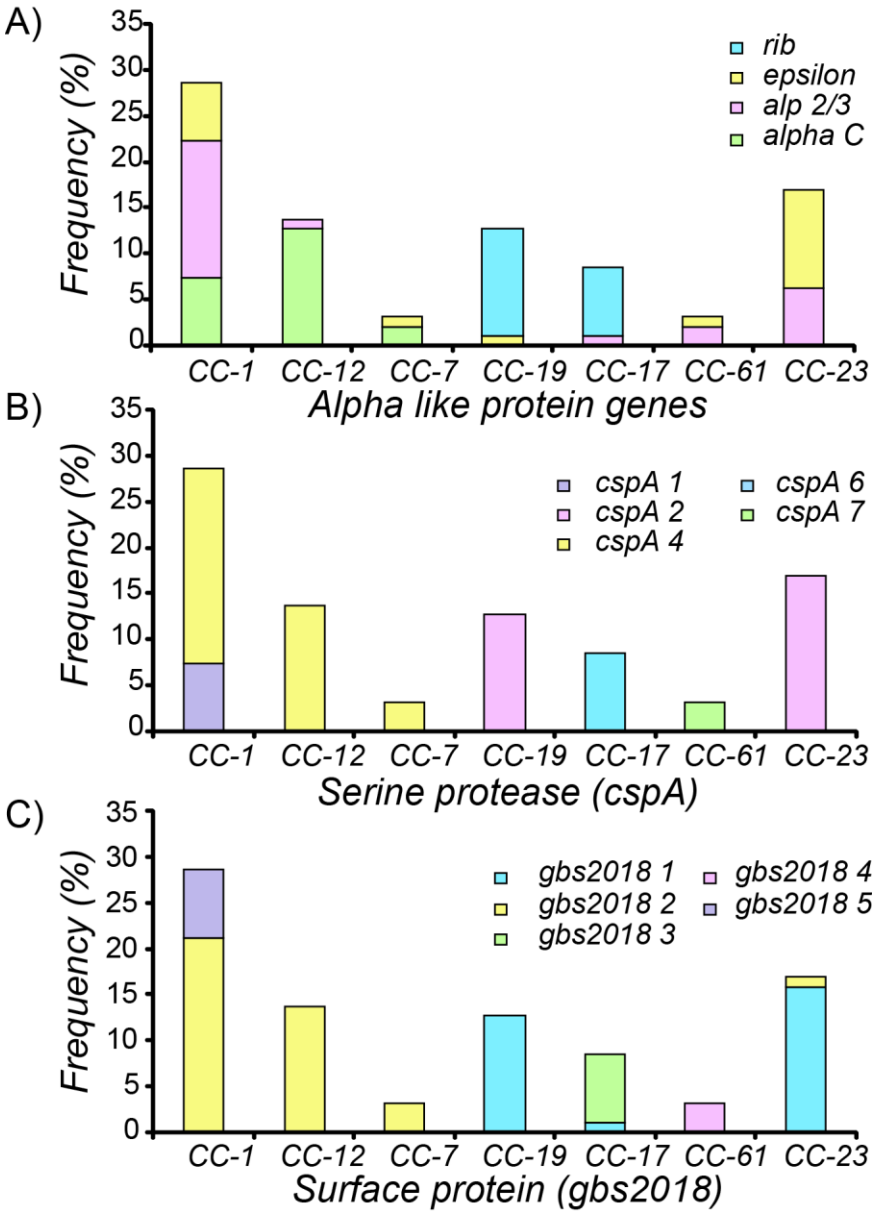
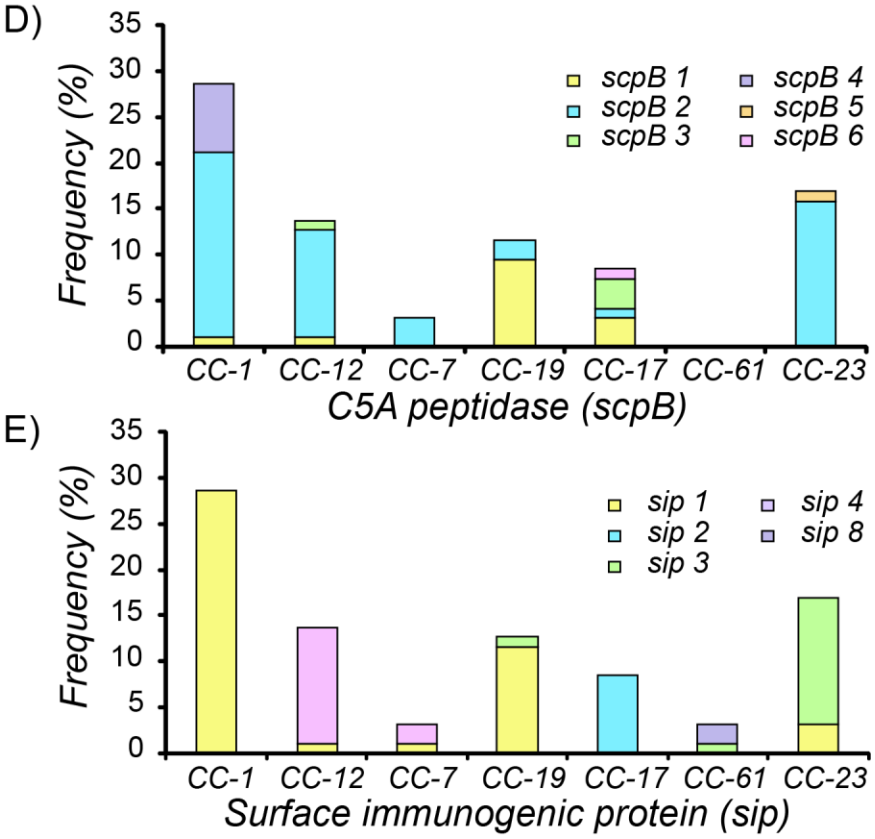


Figure 3 (cont'd).



CHAPTER 3

PILUS DISTRIBUTION AMONG CLONAL LINEAGES OF GROUP B STREPTOCOCCUS (GBS): AN EVOLUTIONARY AND CLINICAL PERSPECTIVE

A. Cody Springman[1,2], David W. Lacher[3], Emily A. Waymire[2], Samantha L. Wengert[1], H. Dele Davies[2], and Shannon D. Manning[1]

Department of Microbiology and Molecular Genetics[1], Department of Pediatrics & Human Development[2], Michigan State University, East Lansing, MI, U.S.A.

Food and Drug Administration[3], Laurel, MD, U.S.A.

ABSTRACT

Group B *Streptococcus* (GBS) is an important opportunistic pathogen in humans and bovines. Certain GBS lineages (or clonal complexes (CCs)) are associated with invasive newborn disease (CC-17, CC-19), asymptomatic colonization (CC-1, CC-23), and bovine mastitis (CC-61, CC-67). GBS pilus structures play an integral role in bacterial colonization and invasion of host tissues and in biofilm formation. Here, we developed a multiplex PCR to detect the presence of three horizontally acquired pilus islands (PIs): PI-1, PI-2a and PI-2b, and examined their distribution among 295 GBS strains representing 73 sequence types (STs) belonging to eight clonal complexes (CCs). In addition, PCR-based RFLP was used to evaluate sequence variability in the pilus backbone protein (BP) genes of PI-2a and PI-2b. Among all strains examined, one of the PI-2 variants was present, with the vast majority of human-derived strains containing PI-1 in addition to PI-2a or PI-2b. In contrast, PI-1 was entirely lacking among the bovine-specific lineages, CC-61 and CC-67, and other bovine-derived strains indicating a role of PI-1 in human infections. Sequence and RFLP analyses revealed six highly divergent alleles in the BP gene of PI-2a and three closely related alleles in the BP gene of PI-2b. Strains within the CC-17 and CC-19 lineages uniformly contained PI-1 and PI-2b (allele 2) or PI-1 and PI-2a (allele 1), respectively. In contrast, CC-1 and CC-23 lineages were more heterogeneous and contained multiple pilus profiles and allelic determinants. The bovine-specific lineages, CC-61 and CC-67, and all other bovine-derived strains, contained a unique allelic determinant (allele 3) in the PI-2b BP gene that was not found in human-derived strains. Biofilm assays conducted

on a subset of 136 strains representing the most common pilus profiles revealed a high degree of strain variation. Among the highest biofilm formers were strains from CC-1 and CC-23, whereas strains from CC-17 and CC-19 were relatively insufficient at forming biofilms. In addition, environmental acidification was found to be a key factor in driving GBS biofilm production, a phenotype likely exploited by GBS during colonization of the vaginal tract. This evaluation of pilus types and biofilm production across diverse strain populations will be useful for future development of vaccines that target pilus structures in GBS.

INTRODUCTION

While Group B *Streptococcus* (GBS, *Streptococcus agalactiae*) was originally described as a causative agent in bovine mastitis [1], it has emerged as an important opportunistic pathogen in humans. Most often GBS is found as commensal organism in the urogenital and lower gastrointestinal tracts of healthy adults. However, women colonized with GBS during pregnancy carry the risk of transmitting the bacterium to their newborn [2]. Newborns infected with GBS can develop life threatening infections including pneumonia, sepsis, and meningitis [3]. In addition, GBS causes serious complications among the elderly and adults with underlying medical conditions where skin and soft tissue infections, urinary tract infections, and bacteremia can result [4].

Molecular epidemiological studies utilizing mulitlocus sequence typing (MLST) have shown that certain GBS lineages, or clonal complexes (CC's), predominate among human and bovine populations [5-7]. Strains belonging to CC-17 and CC-19 are

strongly associated with newborn invasive disease [5,6], with CC-17 strains considered as “hyperinvasive” since they have accounted for the highest frequency of newborn disease [8] and meningitis [7]. Pregnant women are significantly more likely to be colonized with strains belonging to the CC-1 and CC-23 lineages [6], both of which also cause disease in susceptible adults and the elderly [9] [Springman et al., unpublished data]. By contrast, CC-61 and CC-67 strains have been found exclusively within bovine populations and are considered to be highly adapted to the bovine host [10] [Springman et al., unpublished data].

Phylogenetic analysis based on MLST data indicates a close evolutionary relationship between CC-17 strains and certain bovine-derived strains (e.g. CC-61 and CC-67 strains), which has led to the suggestion that the hyperinvasive newborn lineage emerged from a bovine ancestor [8]. Comparative genome analyses have shown, however, that human-derived and bovine-derived GBS strains contain distinct genetic characteristics that confer fitness toward a particular host [10]. We have shown previously, that the overall genetic diversity of GBS has been strongly influenced by recombination between strains [11]. In addition, the acquisition of genes through horizontal gene transfer (HGT) has also influenced GBS evolution and the divergence of clonal lineages [11-13].

Recently GBS was shown to harbor three separate pilus islands (PIs) known as PI-1, PI-2a, and PI-2b [14,15]. Each PI displays a similar operon structure that encodes three structural proteins (i.e., the backbone protein (BP) and two ancillary proteins (AP1 and AP2)) and two pilus-specific sortase enzymes (belonging to class C) [16]. Pilus assembly relies on the sortase enzymes that recognize the conserved LPXTG amino

acid motif on the structural proteins, followed by the covalent attachment of these subunits to each other and the cell wall peptidoglycan [17,18]. A distinct difference between PI-1 and the PI-2 variants are their target sites for integration. PI-1 is a 16kb element that integrates between the genes 0633 and 0652, whereas the PI-2 variants are found at a separate locus, between the genes 1410 and 1403 [16]. In turn, strains may contain PI-1 in addition to PI-2a or PI-2b but cannot contain the PI-2 variants together since they are mutually exclusive. PI-1 is also distinct in that it is flanked by direct repeats signifying this element was acquired horizontally and may maintain the ability to excise once integrated into the bacterial chromosome.

As in other pathogens, pilus structures in GBS mediate bacterial interactions with the host. *In vitro* models of GBS infection have shown that the PI-1 and PI-2a ancillary proteins 1 (AP1-1 and AP1-2a) initiate adherence to various tissues, whereas the backbone proteins (BP-1 and BP-2a) facilitate the invasion and paracellular translocation of host cells [19-21]. Furthermore, PI-2a pili contribute specifically to biofilm formation [22,23] and the presence of the PI-2b protein, Spb1/SAN1518, is thought increase intracellular survival in macrophages [24]. *In vivo*, GBS pilus components are highly immunogenic and a combinatorial pilus-vaccine containing the BP of PI-1 and PI-2b and the AP of PI-2a has been shown to elicit opsonophagocytic antibodies that confer protection in mice [25].

Given the role of pili in GBS colonization and invasion of tissues, the type of pilus present is likely to impact GBS virulence and its ability to infect different hosts (i.e. bovines and humans). Here, we determined the presence/absence of PI-1, PI-2a and PI-2b in 295 strains from humans and bovines and evaluated the distribution of each PI

element across the main phylogenetic lineages. In addition, an analysis of sequence variability was conducted on the main structural genes of PI-2 pili (i.e. BP gene of PI-2a and PI-2b) to determine if specific alleles correspond to phylogenetic lineage, host specificity, and virulence. Lastly, the degree of biofilm formation was tested in a subset of strains to determine if the presence of specific pilus types and if environmental acidification affects the production of biofilms in GBS. This comprehensive analysis of pili and biofilm formation across diverse strain populations will be important for guiding current efforts to develop a pilus-based vaccine against GBS.

MATERIALS AND METHODS

Bacterial strains and DNA extraction. A total of 295 GBS strains with available epidemiological data were included in this study. These strains represented 73 STs and included 120 invasive strains from blood or cerebral spinal fluid of newborns [7], 99 colonizing strains from vaginal/rectal swabs of pregnant and non-pregnant women [5], 51 bovine strains from bulk milk samples in cases of subclinical mastitis, and 25 reference strains. GBS cultures inoculated from freezer stocks were grown overnight in Todd-Hewitt (TH) broth (Acumedia) at 37°C with 5% CO_2 and genomic DNA was isolated using the MO BIO UltraClean® Microbial Isolation Kit (MO BIO Laboratories, Inc.).

Multilocus sequence typing (MLST) and phylogenetic analyses. A well-established MLST scheme for GBS was used to sequence type all GBS strains utilized in this study

[6,26]. Briefly, seven conserved genes (*adhP, pheS, atr, glnA, sdhA, glcK, tkt*) were PCR amplified and sequenced, and then sequence types (STs) were assigned using the PubMed MLST database (http://pubmlst.org/sagalactiae). A neighbor-joining (NJ) phylogeny [27] based on 1000 bootstrap replications was constructed on the concatenated MLST sequences using MEGA4 [28] and displayed in a radial format. Clonal complexes (CCs) were assigned to STs that either clustered together with significant bootstrap support (i.e. ≥ 70%) or have been defined in prior studies. All other STs were classified separately as singletons. The eBURSTv3 algorithm [29] was applied to the MLST allelic profiles from all 295 strains and provided a model for predicting the possible gain or loss of PI-1. Pilus profiles were plotted manually on the NJ and eBURST diagrams using Adobe Illustrator CS5 (Adobe Systems Incorporated).

Multiplex PCR for detection of GBS pilus islands (PIs). Multiplex primers were developed to target and amplify a conserved sortase gene (*srtC*) within each of the three pilus islands: *sag647* (PI-1), *sag1406* (PI-2a) and *san1517* (PI-2b). As an internal control, primers for a conserved MLST gene (*adhP*) were included in each multiplex PCR reaction (Table 5). Each 25 µl reaction contained: 10X buffer II (Applied Biosystems), 2.5 µl 2 mM dNTP, 2.0 µl 25 mM $MgCl_2$, 0.5 µl 10 mM each primer, 1.5 U AmpliTaq Gold (Applied Biosystems), 1 µl 15 ng/µl genomic DNA template, and 15.7 µl ddH20. Thermocycling conditions utilized an initial soak of 94°C for 10 min., followed by 35 cycles of: 92°C for 1 min., 53°C for 1 min, and 72°C for 30 sec.; and a final step of 72°C for 5 min. Amplified PCR products were visualized on ethidium bromide

stained 1.5% agarose gels and the variable size products were used to determine the presence/absence of a particular pilus type. Any strains found lacking PI-1 where further screened for the presence of an intact PI-1 integration site as described by Martins et al. [30]. Briefly, this PCR utilized primers targeting the genes upstream and downstream of the PI-1 integration site. Amplification of a 684 bp fragment indicated the integration site was intact whereas no visible fragment indicated the integration site was occupied by a genetic element other than PI-1.

PCR-based RFLP for PI-2a and PI-2b backbone protein (BP) genes. Based on *in silico* analysis of available GenBank sequences, restriction enzymes were chosen to differentiate between allelic variants in the PI-2a and PI-2b BP genes. Due to the high degree of sequence conservation in the PI-1 BP gene (i.e. 19 variable sites out of 9594 bp in 6 strains), PI-1 was not included in the variability analysis. Gene fragments were amplified using primers flanking the PI-2a BP gene (*gbs59*) and within the PI-2b BP gene (*san1519).* PCR for *gbs59* was performed in a 25 ul reaction and utilized 10 mM of primers and LA *Taq* (Takara Bio, Inc.) with the following thermal cycling conditions: initial soak at 94°C for 1 min., followed by 30 cycles of 94°C for 30 sec., 54°C for 30 sec., and 68°C for 3 min., then a final soak at 72°C for 10 min. PCR for *san1519* was performed under the same conditions as the multiplex PCR but with an annealing temperature of 55°C and 1.5 min. extension. PCR products were enzymatically digested with *Pvu*II (New England BioLabs, Inc.) for PI-2a and with *Ssp*I (New England BioLabs, Inc.) for PI-2b (Table 6). RFLP patterns were visualized on a standard 1.5%

agarose gel and alleles were assigned based on patterns predicted from *in silico* analysis.

Biofilm Assay. A subset of 136 strains that represented the most common CCs and pilus profiles were tested for biofilm formation. Overnight cultures inoculated from freezer stocks were grown in TH broth, and then diluted 1:20 in fresh TH supplemented with 1% glucose (THG). Into a 96-well plate, 100 ul of diluted culture was added, with 4 technical replicates included per strain. Plates were grown under static conditions at 37°C with 5% CO_2 for 20 hrs. Following incubation, unattached bacteria were removed through two washes with of water (200ul), and attached bacteria were stained with 100 ul crystal violet (CV) for 10 min. After staining, unbound CV was removed through 2-3 washing with water, and bound CV was solubilized with 100 ul of 95% ethanol. Biofilm formation was quantified by measuring the absorbance ($O.D._{595}$) using a plate reader (Beckman Coulter, Inc.) and measurements were calculated as the sample value minus the media (blank) control. GBS strain 515 (PI-2a) was included on each plate as a positive control. All assays were repeated 2-3 biological replicates and the average across technical and biological replicates is shown.

To test the affect of pH on GBS biofilm formation, six genome or reference strains were chosen according to their pilus profile, lineage, and source: NEM316 (PI-1,2a, CC-23, human-derived, invasive), 515 (PI-2a, CC-23, human-derived, invasive), GB00402 (PI-1,2a, CC-19, human-derived, invasive), GB00418 (PI-1,2b, CC-17, human-derived, invasive), GB01447 (PI-2b, CC-17, bovine-derived, invasive), and FSL

03-26 (PI-2b, CC-67, bovine-derived, invasive). Biofilm assays were conducted as above with each strain tested under the following conditions: (1) THG adjusted to pH 7.0 with NaOH, (2) THG buffered using 100 mM Na_2HPO_4/NaH_2PO_4 at pH 7.0, (3) THG adjusted to pH 5.0 with HCL, (4) TH adjusted to pH 5.0 with HCL.

RESULTS

Distribution of pilus islands among GBS clonal complexes. A total of 295 GBS strains, including invasive strains from newborns, elderly adults, and bovines and colonizing strains from pregnant and non-pregnant women, were represented in this study. Based on MLST, the 295 strains were defined by 73 STs: 51 STs were human-derived, 18 STs were bovine-derived, and 4 STs were shared among humans and bovines. Phylogenetic analysis of the MLST sequence data grouped the 73 STs into eight recognized clonal complexes (CCs): CC-17, CC-61, CC-67, CC-1, CC-7, CC-12, CC-19, and CC-23 (listed in clockwise order as depicted by the radial phylogeny) (Fig. 4). Overlay of the epidemiological data shows the CC-17 and CC-19 lineages were mostly comprised of invasive newborn strains, the CC-1 and CC-23 lineages were comprised of both colonizing strains from women and invasive strains from elderly adults, and the CC-61 and CC-67 lineages were comprised exclusively of bovine strains.

Plotting the type PI element present along the tree phylogeny reveals GBS pilus type was strongly correlated with strain lineage (Fig. 4). As can also be seen from the histogram (Fig. 5), at least one of the PI-2 variants was present among all strains,

whereas PI-1 was variably present/absent. The majority (73%) of the STs from humans contained PI-1 and 95% of STs isolated from a patient with invasive disease had PI-1. Among the 51 strains from bovines PI-1 was found in only seven (14%), which consisted of those CCs shared between humans and bovines (i.e. CC-1, CC-19, and CC-23). With the exception of one bovine strain representing CC-61, a ST-91 strain, all other bovine-derived strains from CC-61 and CC-67 lacked PI-1. When analysing those strains lacking PI-1, the PI-1 integration site was occupied by another genetic element in 16 of the 51 (31%) bovine strains. In contrast, all human-derived strains without PI-1 contained an intact integration site, indicating they may be eligible recipients for the PI-1 element, or may have lost this element due to selective pressures encountered in the host.

The type of PI-2 variant present was equally distributed on either side of the MLST phylogeny, though there were exceptions within some CCs (Fig. 4). For example, on the right side of the phylogeny CC-17, CC-61, and CC-67 all contained PI-2b, while CC-1 was more diverse with three of 10 STs having PI-2b and multiple pilus profiles seen in ST-1 (PI-1,2a and PI-1,2b) and ST-2 (PI-1,2a, PI-1,2b, and PI-2a). In contrast, the left side of the phylogeny (e.g., CC-7, CC-12, CC-19, and CC-23) was comprised of strains containing PI-2a, though the ST-103 singleton and two of four STs within CC-7 contained P1-2b.

Allelic variation in PI-2a and PI-2b backbone protein (BP) genes. To evaluate the genetic variability within PI-2a and PI-2b, a PCR-based RFLP was developed to identify alleles in the BP gene of PI-2a and PI-2b. Out of 295 GBS strains, 170 strains (58%)

contained PI-2a while the other 125 strains (42%) contained PI-2b. A high degree of variability was observed in the BP of PI-2a (*gbs59*), which was defined by six major allelic determinants (Fig. 6A). The allelic relationships based on nucleotide alignments revealed two distinct groups: group 1 containing alleles 1-3 and group 2 containing alleles 4-6. The BP gene of PI-2b (*san1519*) was relatively conserved and was defined by three allelic determinants, designated alleles 1-3 (Fig. 7A).

The distribution of alleles across CCs was distinct for both the PI-2a and PI-2b BP genes. Those CCs containing PI-2a were mostly represented by multiple allelic determinants with allele 6 predominating across lineages (Fig. 6B). The CC-1 and CC-23 linages associated with colonization in women and elderly infections revealed the highest heterogeneity with five and three allelic determinants represented, respectively. Interestingly, the CC-19 lineage associated with newborn infections was almost entirely represented by allele 1, an allelic determinant rarely seen in other lineages. Among those CCs with PI-2b, the type of allele present was more conserved across lineages (Fig. 7B). The hyperinvasive newborn lineage, CC-17, was entirely represented by allele 2, with the exception of one bovine-derived strain that contained allele 3. Moreover, the allele 3 variant of PI-2b BP gene was unique to bovine-derived strains irrespective of clonal lineage and the bovine-specific lineages, CC-61 and CC-67, were comprised entirely of this allele.

Hypothetical model for gain or loss of pilus island 1 (PI-1). The variable presence/absence of PI-1 among closely related genotypes within the same CCs suggests genetic instability of this element. In contrast to the PI-2 variants, the PI-1

element is flanked by direct repeats indicating it has been acquired through HGT and may maintain the ability to be excised following integration. To model the possible gain/loss of PI-1, an eBURST analysis was conducted on the MLST profiles of all 295 strains. eBURST is a simplistic way of depicting the divergence of closely related genotypes from a founding, or ancestral, genotype. The founding genotype is assigned to the ST that varies from the largest number of STs at a single locus, and all these related STs are then used to define a group. Analysis of the STs from all strains included in this study identified three mutually exclusive groups comprising five of the main CCs: CC-1 and CC-19 (group 1), CC-67 and CC-17 (group 2) and CC-23 (group 3) (Fig. 8). In each group, the founding genotype/ST was used to hypothesize the direction of PI-1 gain/loss during diversification of STs. In figure 8, these events of gain/loss are depicted as solid arrows. For instance, ST-1 was the predicted founder of group 1 and ST-19 was the predicted co-founder, and each of them contributed to the divergence of multiple STs that together represent their respective CC. In the case of ST-1, predicted founder of CC-1, multiple pilus profiles were present (e.g., PI-1,2a and PI-1,2b) and the loss of PI-1 could have occurred in four STs during diversification from ST-1. In contrast, the diversification of ST-19, and CC-19 as a whole, likely involved the gain of PI-1 from the intermediary ST-44 that links the two complexes. Within group 2, the bovine-specific ST-67 was the predicted founder and the closely related hyperinvasive newborn clone, ST-17, was the predicted co-founder. In this case, the divergence of the newborn invasive lineage, ST-17, from the ancestral bovine lineage would have involved a gain of PI-1. Lastly within group 3, the predicted founder, ST-23, contained multiple pilus profiles (e.g., PI-1,2a, PI-1,2b, and PI-2a) and it was not clear

as to which genotype profile contributed to diversification within the complex, therefore, events of possible PI-1 loss are indicated by a dashed arrow. In one instance, there was indication that PI-1 loss occurred during the diversification ST-435 from the intermediary ST-88.

Biofilm formation and the role of environmental acidification. Recently, PI-2a-encoded pili were shown to have a specific involvement in GBS biofilm formation [22]. A subset of 136 strains representing different pilus profiles and allelic variants were analyzed for a differential ability to form biofilms (Fig. 9A). Although strains with PI-1,2a containing allele 2 produced relatively more biofilms, the degree of biofilm formation was highly strain-dependent in which closely related strains within the same CC exhibited a large degree of variability. Overall, the highest degree of biofilm formation was seen in strains belonging to CC-1 and CC-23, the two lineages frequently associated with asymptomatic colonization and elderly infections. In contrast, strains belonging to the newborn invasive lineages, CC-17 and CC-19, exhibited relatively low biofilm formation, as did the bovine-specific lineages, CC-61 and CC-67.

Previous work conducted by Rinaudo et al. demonstrated that biofilm formation by GBS was only evident in media supplemented with glucose at concentrations equal to or above 0.4% [22]. This led us to hypothesize that a decrease in pH associated with metabolism of the excess sugar may be the main driving force in GBS biofilm formation. To test this notion, we measured biofilm formation under various conditions: TH with glucose at pH 7.0, TH with glucose buffered at pH 7.0, (3) TH with glucose at pH 5.0 (4) TH (no supplemented glucose) at pH 5.0. As expected, under buffered conditions at pH

7.0 with adequate glucose, GBS strains failed to produce biofilms (Fig. 9B). In contrast, in acidic conditions at pH 5.0, regardless glucose supplementation, biofilm formation was restored to optimum levels. From these results, we confirmed that environmental acidification is important factor for promoting GBS biofilm formation.

DISCUSSION

As was demonstrated here and in prior studies, GBS strains derived from bovines and humans contain distinctive genetic characteristics that reflect the independent divergence of these two populations [11,31] [Springman et al., unpublished data]. Although it is unclear as to what factors contributed to emergence of GBS in humans, it appears as though acquisition of the PI-1 element through horizontal gene transfer may have facilitated this process. In addition, the epidemiology and source of a given strain is correlated with phylogenetic lineage and pilus type.

The difference in the distribution frequencies of PI elements between bovines and humans strains indicates that these structures may have impacted host specialization and virulence of GBS. The vast majority of human strains (73%) contained PI-1 with 95% of those strains from invasive cases of disease having the PI-1 element. Conversely, among the 51 bovine strains examined, only those strains belonging CCs also found in humans contained PI-1. Together these findings suggest that horizontal acquisition of PI-1 occurred when human strains diverged from bovine stains and this event likely increased the fitness of certain strains and their ability to

colonize humans. In this sense, PI-1 could interchangeably be referred to as a GBS "fitness island" that confers an advantage in the human host.

Evidence that certain pilus alleles are associated with strain virulence was observed in the newborn invasive lineages. Both CC-17 and CC-19 were almost entirely comprised of strains having PI-1,2b (allele 2) and PI-1,2a (allele 1), respectively. The high conservation of allelic type seen among these strains reflects strong selective pressures have limited the variability in these lineages and led to their homogeneity. As such, allele 2 in PI-2b and allele 1 in PI-2a may have specific functions *in vivo* that enhance the ability of CC-17 and CC-19 strains to cause newborn infections. In contrast, the CC-1 and CC-23 lineages associated with asymptomatic colonization and elderly infections exhibited much higher diversity overall with multiple pilus profiles and allelic types represented among these strains. Given strains from these lineages are also found in bovine populations, the variability in the pilus BP of CC-1 and CC-23 may contribute to their greater versatility and ability to colonize different hosts.

Undergoing antigenic variation in surface proteins is a common mechanism utilized by GBS to evade immunological recognition in the host. The seven highly variable alleles in the PI-2a BP gene likely arose as means of providing strains with a functional advantage as they were forced to adapt to changing environments.
The presence of identical alleles among unrelated strains also indicates this locus is a "hot spot" for genetic recombination. Conversely, the low sequence variability seen in the PI-2b BP gene is evidence of more constrained evolutionary history. Interestingly, strains belonging to the hyperinvasive newborn lineage, CC-17, and the closely related bovine-derived lineages, CC-61 and CC-67, all contained PI-2b, however, the type of

allelic variant was distinct. This difference in PI-2b BP allele, in addition to the common lack of PI-1 in bovine-derived strains, reflects the independent divergence of these two strain populations and their distinctive features that permit specificity toward either the bovine or human host.

The dynamics and instability of horizontally acquired elements, in particular PI-1, has not been fully explored in GBS. Following events of HGT, certain elements may maintain the ability to be excised and can be lost as a result of evolution *in vivo*. Such elements typically encode integrases or are flanked by direct repeats, as seen in the case of the PI-1 element. To demonstrate the possible gain or loss of PI-1, a hypothetical model based on eBURST analysis was constructed. The direction of PI-1 gain or loss was predicted based on the ancestral (or founding) genotype, and in many cases, revealed a loss of PI-1 during the diversification of closely related genotypes. Since pilus structures are immunological targets, the loss of this highly exposed surface antigen could provide an advantage in evading the host immune system during infection. This model also revealed that acquisition of PI-1 in the CC-17 newborn invasive lineage was an event that may have promoted the diversification of this lineage from the closely related bovine strains, in turn, allowing these strains to colonize and infect humans. Moreover, the presence of a separate genetic element occupying the PI-1 integration site in many bovine-derived strains would prevent them from acquiring this element and restrict their transmission into human populations.

Although GBS biofilm formation does not appear to be a trait of virulence, the requirement for environmental acidification suggests this phenotype is important for colonization of the vaginal tract. The enhanced ability of strains in the CC-1 and CC-23

lineages to form biofilms is consistent with this notion since these strains have been found to be associated with persistent colonization in women. In addition to aiding in vaginal colonization, GBS biofilm production may also increase its resistance to antibiotics thereby negatively impacting the effect of intrapartum antibiotic prophylaxis at preventing GBS transmission to the newborn.

The distribution patterns of PI elements and the conservation of these structures across phylogenetic lineages will be critical for focusing efforts in the development of a pilus-based vaccine. With respect to PI-1, the variable presence of this element among human-derived strains would limit the protection across all strains. Additional concerns may arise if there is a propensity for the PI-1 element to be excised during the course of infection, in which case coverage would also be reduced. As we gain a better understanding of how GBS pili are involved with disease pathogenesis, new strategies that block the assembly and/or function of pili may be devised that are sufficient for preventing GBS associated disease.

REFERENCES

REFERENCES

1. Nocard MaRM (1887) Sur une mammite contagieuse des vaches laitieres. Ann Inst Pasteur 1:109.

2. Hansen SM, Uldbjerg N, Kilian M, Sorensen UB (2004) Dynamics of Streptococcus agalactiae colonization in women during and after pregnancy and in their infants. J Clin Microbiol 42: 83-89.

3. Prevention CDCa (2012) Active Bacterial Core Surveillance Report, Emerging Infections Program Network, Group B Streptococcus, 2010.

4. Edwards MS, Baker CJ (2005) Group B streptococcal infections in elderly adults. Clin Infect Dis 41: 839-847.

5. Manning SD, Schaeffer KE, Springman AC, Lehotzky E, Lewis MA, et al. (2008) Genetic diversity and antimicrobial resistance in group B streptococcus colonizing young, nonpregnant women. Clin Infect Dis 47: 388-390.

6. Manning SD, Lewis MA, Springman AC, Lehotzky E, Whittam TS, et al. (2008) Genotypic diversity and serotype distribution of group B streptococcus isolated from women before and after delivery. Clin Infect Dis 46: 1829-1837.

7. Manning SD, Springman AC, Lehotzky E, Lewis MA, Whittam TS, et al. (2009) Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada. J Clin Microbiol 47: 1143-1148.

8. Bisharat N, Crook DW, Leigh J, Harding RM, Ward PN, et al. (2004) Hyperinvasive neonatal group B streptococcus has arisen from a bovine ancestor. J Clin Microbiol 42: 2161-2167.

9. Salloum M, van der Mee-Marquet N, Valentin-Domelier AS, Quentin R (2011) Diversity of prophage DNA regions of Streptococcus agalactiae clonal lineages from adults and neonates with invasive infectious disease. PLoS One 6: e20256.

10. Richards VP, Lang P, Bitar PD, Lefebure T, Schukken YH, et al. (2011) Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae. Infect Genet Evol 11: 1263-1275.

11. Springman AC, Lacher DW, Wu G, Milton N, Whittam TS, et al. (2009) Selection, recombination, and virulence gene diversity among group B streptococcal genotypes. J Bacteriol 191: 5419-5427.

12. Brochet M, Couve E, Zouine M, Vallaeys T, Rusniok C, et al. (2006) Genomic diversity and evolution within the species Streptococcus agalactiae. Microbes Infect 8: 1227-1243.

13. Brochet M, Rusniok C, Couve E, Dramsi S, Poyart C, et al. (2008) Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of Streptococcus agalactiae. Proc Natl Acad Sci U S A 105: 15961-15966.

14. Maione D, Margarit I, Rinaudo CD, Masignani V, Mora M, et al. (2005) Identification of a universal Group B streptococcus vaccine by multiple genome screen. Science 309: 148-150.

15. Lauer P, Rinaudo CD, Soriani M, Margarit I, Maione D, et al. (2005) Genome analysis reveals pili in Group B Streptococcus. Science 309: 105.

16. Rosini R, Rinaudo CD, Soriani M, Lauer P, Mora M, et al. (2006) Identification of novel genomic islands coding for antigenic pilus-like structures in Streptococcus agalactiae. Mol Microbiol 61: 126-141.

17. Dramsi S, Caliot E, Bonne I, Guadagnini S, Prevost MC, et al. (2006) Assembly and role of pili in group B streptococci. Mol Microbiol 60: 1401-1413.

18. Telford JL, Barocchi MA, Margarit I, Rappuoli R, Grandi G (2006) Pili in gram-positive pathogens. Nat Rev Microbiol 4: 509-519.

19. Pezzicoli A, Santi I, Lauer P, Rosini R, Rinaudo D, et al. (2008) Pilus backbone contributes to group B Streptococcus paracellular translocation through epithelial cells. J Infect Dis 198: 890-898.

20. Maisey HC, Hensler M, Nizet V, Doran KS (2007) Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells. J Bacteriol 189: 1464-1467.

21. Krishnan V, Gaspar AH, Ye N, Mandlik A, Ton-That H, et al. (2007) An IgG-like domain in the minor pilin GBS52 of Streptococcus agalactiae mediates lung epithelial cell adhesion. Structure 15: 893-903.

22. Rinaudo CD, Rosini R, Galeotti CL, Berti F, Necchi F, et al. (2010) Specific involvement of pilus type 2a in biofilm formation in group B Streptococcus. PLoS One 5: e9216.

23. Konto-Ghiorghi Y, Mairey E, Mallet A, Dumenil G, Caliot E, et al. (2009) Dual role for pilus in adherence to epithelial cells and biofilm formation in Streptococcus agalactiae. PLoS Pathog 5: e1000422.

24. Chattopadhyay D, Carey AJ, Caliot E, Webb RI, Layton JR, et al. (2011) Phylogenetic lineage and pilus protein Spb1/SAN1518 affect opsonin-independent phagocytosis and intracellular survival of Group B Streptococcus. Microbes Infect 13: 369-382.

25. Margarit I, Rinaudo CD, Galeotti CL, Maione D, Ghezzo C, et al. (2009) Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm. J Infect Dis 199: 108-115.

26. Jones N, Bohnsack JF, Takahashi S, Oliver KA, Chan MS, et al. (2003) Multilocus sequence typing system for group B streptococcus. J Clin Microbiol 41: 2530-2536.

27. Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4: 406-425.

28. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24: 1596-1599.

29. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG (2004) eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186: 1518-1530.

30. Martins ER, Melo-Cristino J, Ramirez M (2010) Evidence for rare capsular switching in Streptococcus agalactiae. J Bacteriol 192: 1361-1369.

31. Bohnsack JF, Whiting AA, Martinez G, Jones N, Adderson EE, et al. (2004) Serotype III Streptococcus agalactiae from bovine milk and human neonatal infections. Emerg Infect Dis 10: 1412-1419.

TABLES

Table 5. Pilus Island (PI) multiplex PCR indicating gene targets, primer sequences, and expected size fragments.

PI	Gene target	Sequence	Size (bp)
Control	adhP_F162	ACGCATTTTGGGTCACGA	783
Control	adhP_R944	GTATCCACAGGCACTTTTTCAAC	
PI-1	SAG647_F496	CTACCAACGGCCAAGCTATTTACC	394
PI-1	SAG647_R889	TAGCCGCTTTTTCATTCTTTCTCC	
PI-2a	SAG1406_F356	AACTCCCTATATTTGCAGGTTCAA	243
PI-2a	SAG1406_R598	CGGGTGTAACGACTTTTATCTGAT	
PI-2b	SAN1517_F57	GGGGGTAGGCTTAATGGCTTAT	519
PI-2b	SAN1517_R575	TCCGGTTTAACTGTTCTGATTTGAT	

Table 6. PCR-based RFLP for backbone protein (BP) genes of PI-2a and PI-2b.

(A) PCR targets, primer sequences, and the average size fragments expected.

PI	Gene target	Sequence	Avg. Size (kb)
PI-2a	GBS59_Fup	CAACACGCATATTCCACCAAAAG	2.8
PI-2a	GBS59_Rdn	TCTAACATACGGGCGTACTCTGCT	
PI-2b	SAN1519_F1485	GCCCGCACAAACAACCTACAC	1.7
PI-2b	SAN1519_R3167	AAATGGGCGTCAATATCAATGGA	

(B) PI-2a BP gene digestion with *Pvu*I yielded six major allelic variants; GenBank reference sequences for each variant are listed.

Major Allele	Reference Sequence	Fragment sizes					
1	18RS21			39	261	983	1546
1	2603V-R			39	261	983	1546
2	EU929876	33	294	303	513	725	955
3	H36B			336	336	664	1443
4	EU929869					733	2024
5	515					1264	1472
5	EU929888					1264	1508
6	CJB111				609	949	1168
6	NEM316				612	949	1165
6	EU929963				609	949	1393

(C) PI-2b (*SAN1519*) digested with *SspI* yielded 3 allelic variants; GenBank reference sequences for each variant are listed.

Allele	Reference Sequence	Fragment length				
1	A909		81	251	599	749
2	COH1	171	251	311	351	599
3	FSL03-26		171	251	599	662

FIGURES

Figure 4. Radial neighbor-joining (NJ) phylogeny showing 73 multilocus sequence types (STs) and their pilus profiles. The number at the end of each branch represent each STs represented among the 295 GBS strains. STs shown with grey shading contained invasive strains, STs shown in red contained of bovine-associated strains, while all others contained colonizing or reference strains. Clonal Complexes (CC's) are assigned to those ST's clustering together with high bootstrap support, with bootstrap values indicated at the node of each complex. Pilus profiles for each ST are plotted as colored circles: PI-1 (teal), PI-2a (aqua), and PI-2b (yellow). Among strains/STs lacking PI-1, those containing a PI-1 integration site occupied by another genetic element are indicated with an open (grey) circle.

Figure 4 (cont'd).

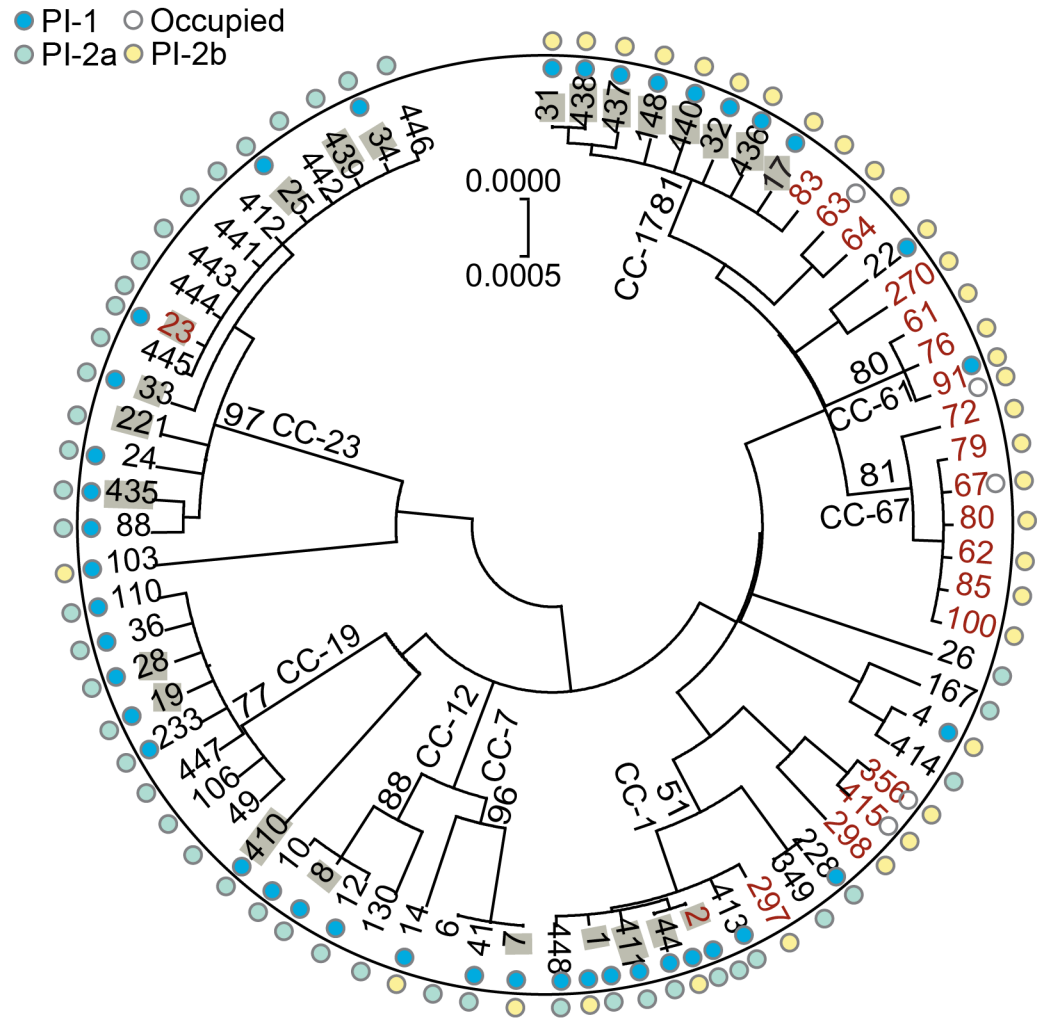


Figure 5. Frequency of pilus islands (PI) PI-1, PI-2a, and PI-2b among GBS clonal complexes (CC's). A total of 295 stains were screened for the presence of PI-1, PI-2a, and PI-2b using multiplex PCR. Strains were grouped into CC's based multilocus sequence typing analysis, and the frequency of each PI element within each CC is shown. Stains with STs not belonging to a major CC were combined into a group of singletons. The CC-17 and CC-19 lineages associated with newborn invasive disease uniformly contained the pilus profiles PI-1,2b and PI-1,2a, respectively. The CC-1 and CC-23 lineages associated with asymptomatic colonization, and the CC-61 and CC-67 lineages are bovine-specific.

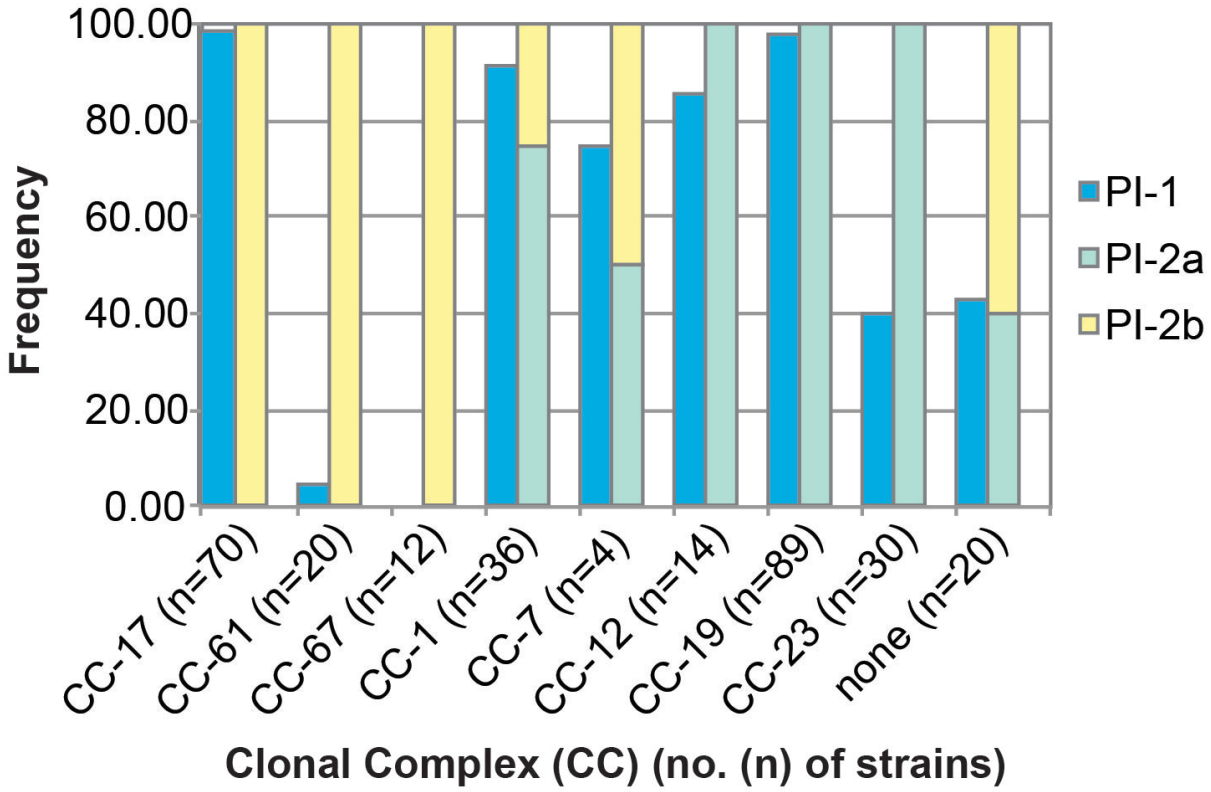


Figure 6. Variability analysis of the PI-2a backbone protein (BP) gene.

(A) Phylogenetic analysis based on the nucleotide alignments of PI-2a BP gene from available GenBank sequences indicated six major allelic types (A1-6) that were highly divergent (bootstrap values based on 1000 replications are indicated at the nodes). The six major alleles are delineated into 2 main groups (group 1: A1-A3 and group 2: A4-A6).

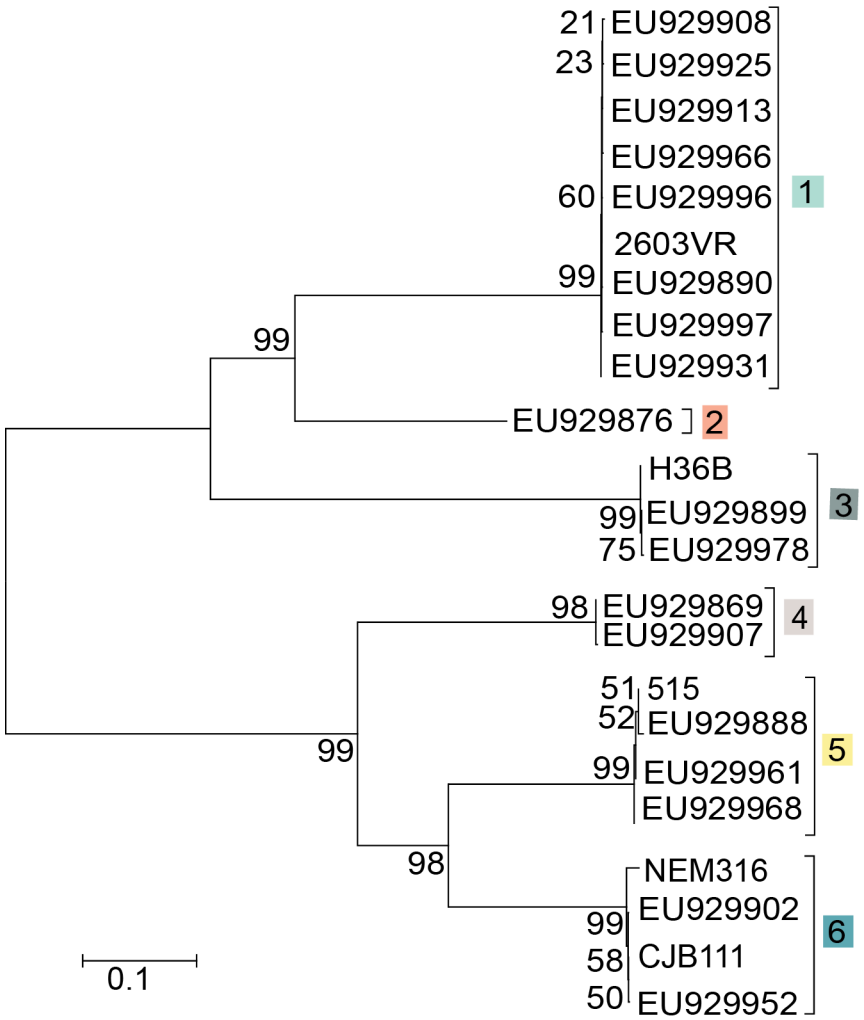


Figure 6 (cont'd).

(B) Frequency of six alleles in PI-2a BP gene across clonal complexes (CCs). A total of 170 out of 295 strains contained PI-2a. The bovine-specific lineages, CC-61 and CC-67, which entirely lacked PI-2a, are indicated in red.

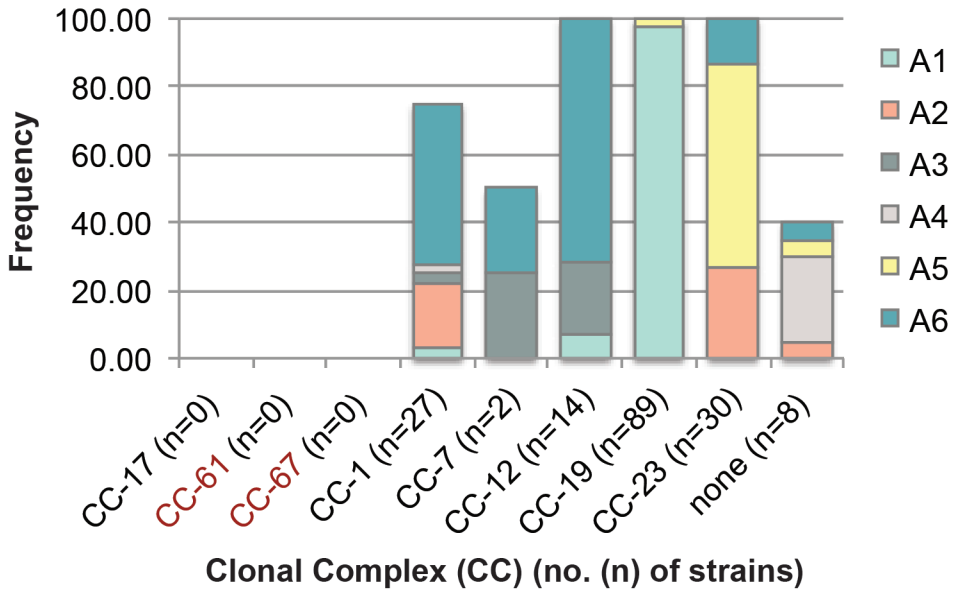


Figure 7. Variability analysis of the PI-2b backbone protein (BP) gene.

(A) Phylogenetic analysis based on the nucleotide alignments of the PI-2b BP gene from available GenBank sequences revealed three allelic types (A1-3). The bovine genome strain, FSL 03-26, is indicated in red.

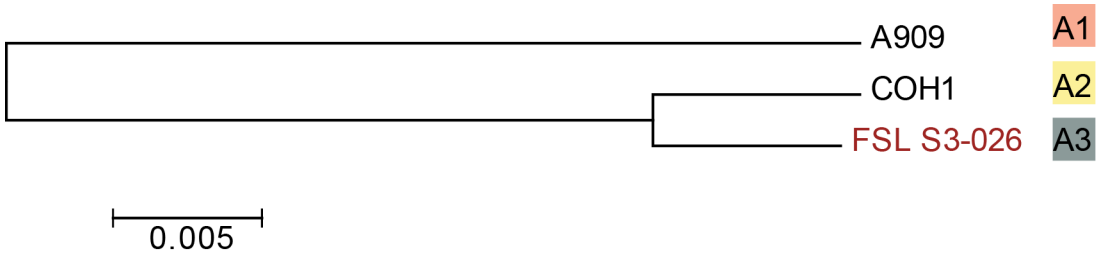


Figure 7 (cont'd.)

(B) Frequency of three alleles in PI-2b BP gene among clonal complexes (CCs). A total of 125 out of 295 strains contained PI-2b. Bovine-specific lineages, CC-61 and CC-67, are indicated in red.

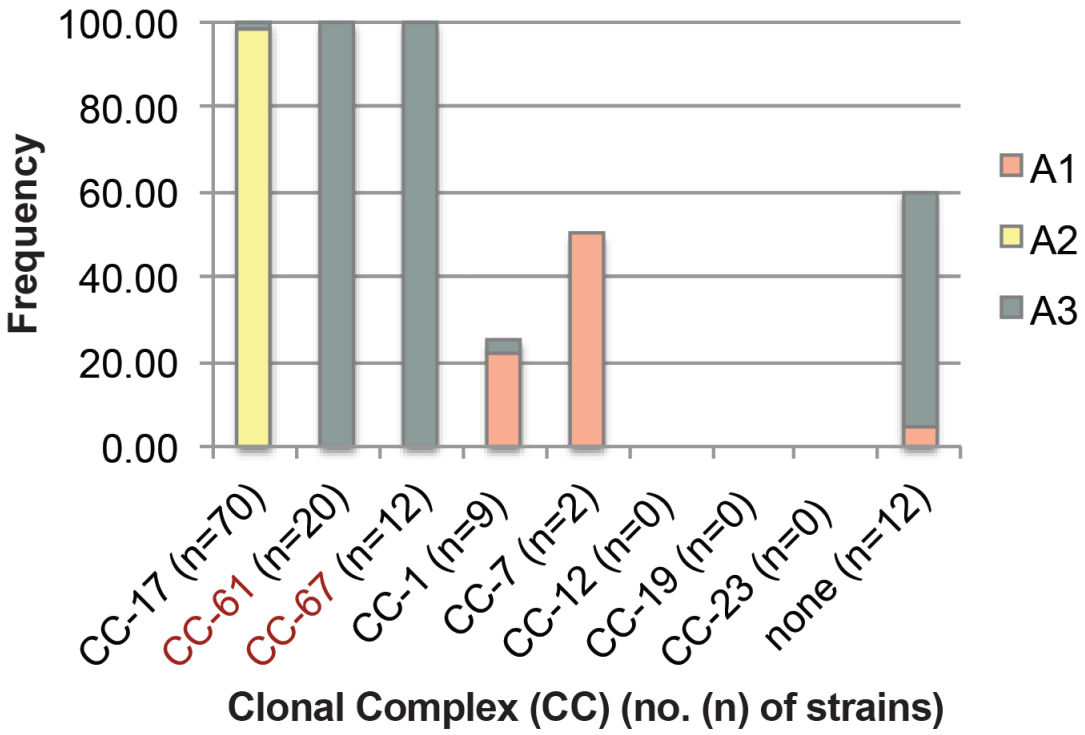


Figure 8. Hypothetical model for the gain/loss of pilus island 1 (PI-1) within GBS clonal complexes (CCs). eBURST analysis was conducted on the MLST allelic profiles of all 295 strains. STs are color-coded based on clinical source: invasive strains (grey shading), bovine strains (red), and colonizing or reference strains (black). The diameter of circles relates to the number of strains for each ST with pilus profiles listed by each ST in grey. The predicted founding (blue) and co-founding (yellow) genotypes were used to predict possible events of PI-1 gain or loss, as indicated by the grey arrows. Instances where the founding genotype was represented by multiple pilus profiles (e.g., ST-23, the founding genotype of CC-23, contained strains with PI-1,2a and strains with PI-2a only), the ancestral genotype was uncertain and a dashed arrow was used to indicate the possible loss of PI-1.

Figure 8 (cont'd.)

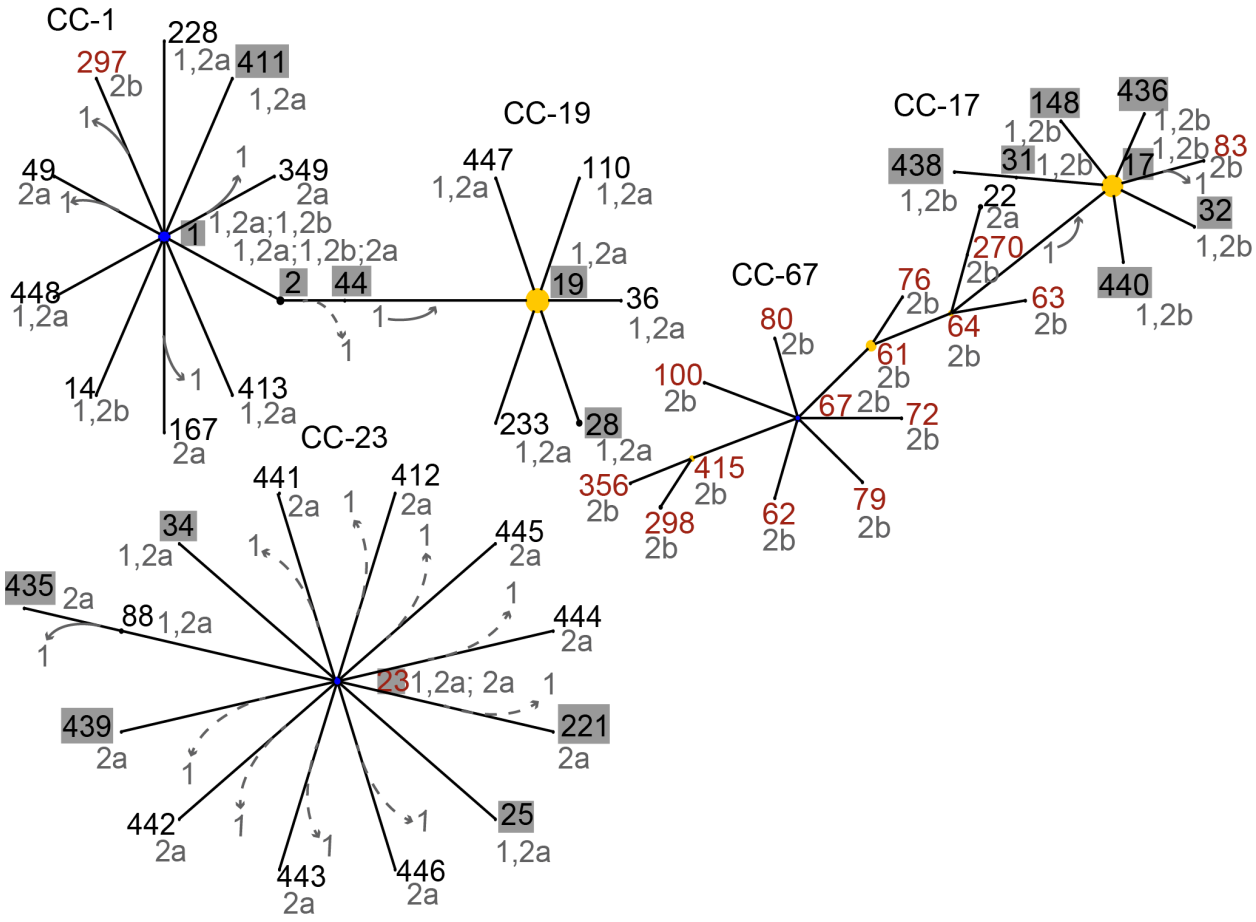


Figure 9. GBS biofilm formation and the role of environmental acidification.

(A) Biofilm formation was measured in 136 strains representing the most common clonal complexes (CCs) and pilus profiles. Each bar represents the average across strains within the same CC having the same pilus profile and PI-2a or PI-2b allelic variant (error bars represent between strain variation). Bovine-specific lineages CC-61 and CC-67 are shown in red.

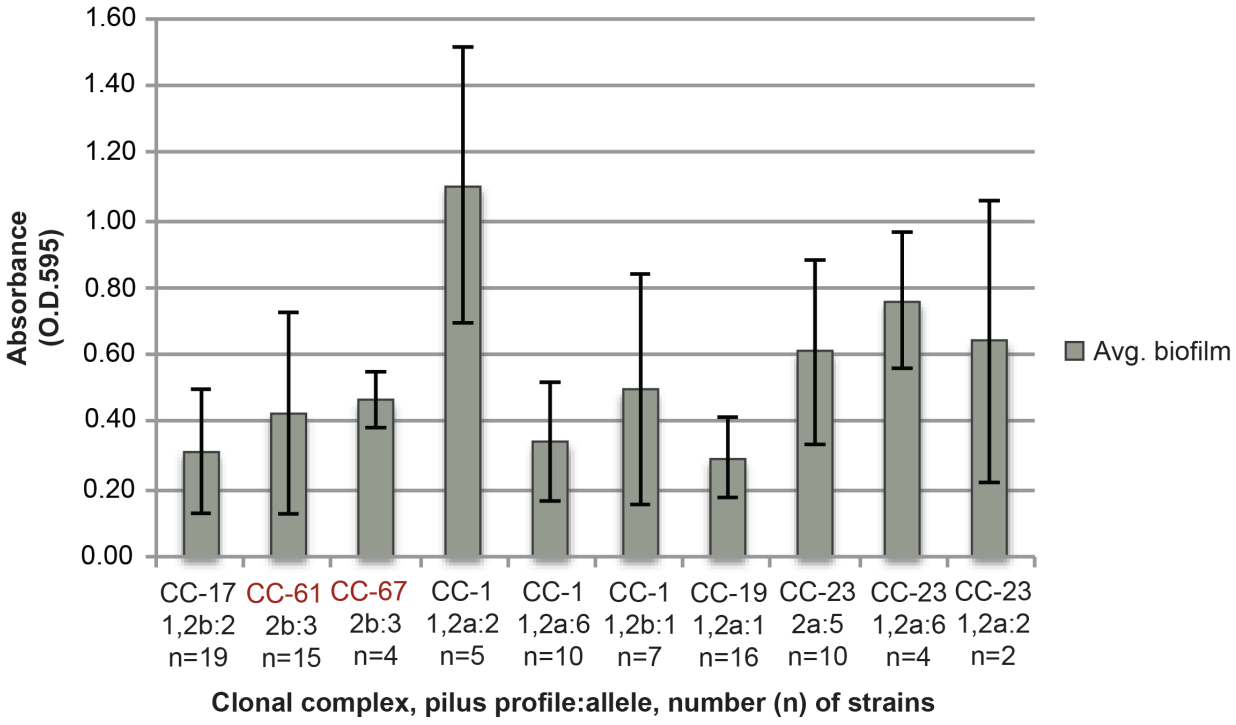
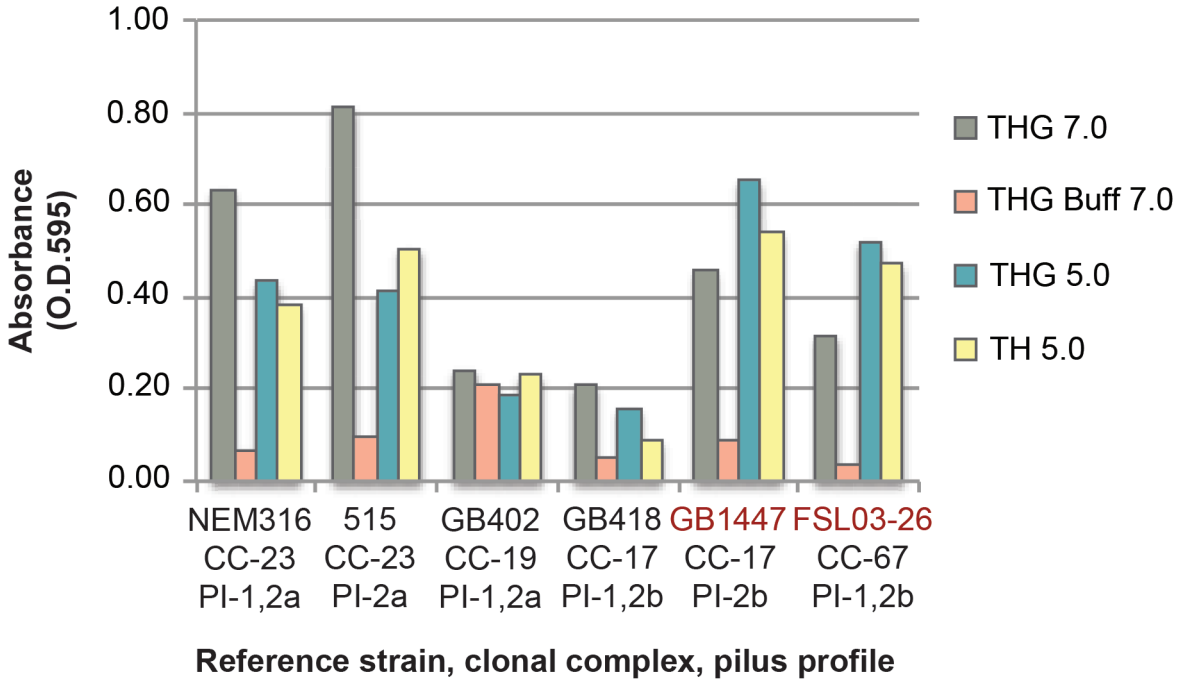


Figure 9 (cont'd.)

(B) Selected genome or reference strains were tested for their ability to form biofilms under the following conditions: Todd-Hewitt (TH) at pH 7.0, TH with 1% glucose (THG) buffered at pH 7.0, and THG at pH 5.0.



CHAPTER 4

CONCLUSIONS AND CHAPTER SUMMARIES

CONCLUSIONS

Bacterial disease caused by GBS remains a serious problem in newborns, susceptible adults, and the elderly. Molecular epidemiological studies using multilocus sequence typing (MLST) have shown that specific GBS clonal lineages, or clonal complexes (CC's), predominate among different populations. For instance, CC-1 and CC-23 are common colonizing lineages that also cause disease in adults and the elderly; CC-17 and CC-19 are considered newborn invasive lineages with the CC-17 lineage causing the most severe disease [1]. CC-61 and CC-67, however, are bovine-specific lineages that have become well adapted to the bovine host [2]. An extensive strain collection of GBS strains from colonized women, newborns, adults, the elderly, and bovines has allowed for in-depth analysis into the molecular epidemiology of GBS. The development of molecular subtyping methods such as multilocus sequence typing (MLST), PCR-based RFLP, and mulitiplex PCR have provided excellent tools for characterizing strains. The work conducted within this dissertation utilized molecular tools and evolutionary analyses along with epidemiological data to evaluate the genetic diversity of GBS populations and characterize specific virulence genes required for human disease. The common goal of the studies was to identify specific bacterial factors that contribute to differences in virulence and host specificity observed among GBS clonal lineages. It was hypothesized that those lineages frequently associated colonization and/or invasion of humans will contain certain characteristics that distinguish them from bovine-derived GBS strains.

CHAPTER SUMMARIES

Chapter 1 - Group B Streptococcus: Disease, Genetic Variability, and Virulence

GBS was first identified as a causative agent in bovine mastitis and has since emerged as an important pathogen in humans. Although healthy adults are commonly colonized with GBS, newborns are particularly vulnerable to GBS infection due to their immature immune system. Disease caused by GBS is multifactorial and requires an array of virulence factors that function at each stage of the infection process: adherence to host tissues, invasion of cellular barriers, and evasion of the immune system. The virulence factors, Lmb, ScpB, and FbsA, allow for binding extracellular matrix components fibrinogen, fibronectin, and laminin, respectively [3-5]. Surface adhesions such as pili and Alpha-C like proteins promote colonization and invasion of host cells [6,7]. Other factors, CylE and CAMP, cause direct cellular injury facilitate GBS invasion [8,9]. The ability of GBS to effectively avoid immunological clearance is mediated by BibA, FbsA, ScpB, and Cps, all of which interfere with complement-mediated opsonization [10-12]. Many of these virulence factors have been proposed as potential vaccine targets, therefore, it is important to evaluate their distribution and genetic variability across diverse GBS strain populations.

Chapter 2 - Selection, Recombination, and Virulence Gene Diversity Among Group B Streptococcal Genotypes

While several studies have identified putative GBS virulence factors and their mechanisms in pathogenesis, little has been shown regarding the extent of genetic variability and the role of selection and recombination among diverse GBS strain populations. In this study, an extensive molecular and evolutionary analysis of 94 GBS strains representing different multilocus sequence types, origins, and capsular genotypes was conducted. PCR-based typing methods (i.e. RFLP, fRFLP, or direct sequencing) were used to uncover allelic variation in a number of virulence genes and then analyzed for evidence of recombination and natural selection. Phylogenetic analyses were used to investigate the clonal relationships among diverse STs and to determine evolutionary pressures that have shaped the divergence of different lineages.

Among the 94 GBS strains, most of the 38 STs that were included clustered into one of seven major clonal complexes (CCs). Recombinational analysis revealed varying degrees of recombination leading to the diversification of lineages, although, the MLST genes appeared to be evolving neutrally without the influence of any strong positive or negative selection pressures. The extent of virulence gene diversity also differed between CCs with some CCs revealing more homogeneity than others. For instance, the CC-17 hyperinvasive newborn lineage contained specific allelic determinants in *sip*, *cspA*, and *bibA* (*gbs2018*) reflecting the independent divergence of this lineage. The closely related bovine-derived strains contained unique alleles in several genes and lacked the *scpB* element that may restrict their ability to colonize and infect humans. In contrast, the CC-1 and CC-23 lineages that commonly colonize humans often contained

multiple allelic determinants suggesting that exposure to different environmental conditions (i.e. the vaginal tract) have favored greater diversity in these lineages. In all, the results of this study have shown that human-derived and bovine-derived GBS strains represent genetically distinct populations in which selective pressures have likely favored the emergence specific genotypes that have become highly successful at causing human disease.

Chapter 3 - Pilus Distribution among Clonal Lineages of Group B Streptococcus (GBS): An evolutionary and clinical perspective

As shown from the previous study, evolutionary pressures such as recombination and selection have played a major role in the diversification of clonal lineages. Recently, three genetic elements encoding pilus structures were identified in GBS. These pilus islands (PIs), PI-1, PI-2a, and PI-2b, represent distinct operons that encode all the structural and enzymatic proteins required for pilus biosynthesis [13]. Each pilus structure is made of a backbone protein (BP) and two ancillary proteins (APs) that are polymerized and attached to the cell surface by sortase enzymes. Functional analyses have shown pili are important for GBS attachment and invasion of host cells and for biofilm formation [14]. In addition, pilus components are highly immunogenic and represent promising vaccine candidates [15].

The presence/absence of each PI was determined in 295 strains from humans and bovines using a novel multiplex PCR targeting conserved genes within each operon. In addition, a PCR-based RFLP was used to evaluate sequence variability in

the BP gene of PI-2a and PI-2b. In total, these strains represented 73 STs and the distribution of each PI was evaluated based on the phylogenetic framework of the MLST phylogeny. The type of PI element was strongly correlated with clonal lineage and the epidemiology of a strain. The vast majority of human-derived GBS strains contained PI-1 in addition to PI-2a or PI-2b. Among strains responsible for invasive disease, PI-1 was present 95% of the time. In contrast, bovine-derived strains belonging to the CC-61 and CC-67 lineages all lacked PI-1. Together these findings suggest that the acquisition of PI-1 has enhanced the ability of certain strains to cause human disease. RFLP analysis revealed seven major alleles in PI-2a and 3 alleles for PI-2b. The newborn invasive lineages, CC-17 and CC-19, both represented homogeneous groups and contained PI-1,2b (allele 2) and PI-1,2a (allele 1), respectively, which has likely contributed to the increased virulence of these strains. Although, the hyperinvasive CC-17 lineage and the closely related CC-61 and CC-67 lineages both contained PI-2b, the bovine-derived strains contained a unique allelic variant that was not seen in human-derived strains indicating specific allelic determinants may contribute to GBS host specificity. Lastly, the lineages associated with colonization and elderly infections, CC-1 and CC-23, were more diverse and contained multiple pilus profiles and allelic types which may reflect different evolutionary pressures acting on these strains.

Overall, GBS biofilm formation was highly variable among strains within and across lineages. Certain strains CC-1 and CC-23 with PI-1,2a (allele 2) revealed high biofilm formation while strains from the newborn invasive lineages, CC-17 and CC-19, were generally poor at forming biofilms. Interestingly, GBS biofilm formation appears to be driven by environmental acidification, which may enhance colonization of GBS in the

vaginal tract. This work suggests that the presence of specific pilus types can impact GBS host specificity and virulence potential. The variation of GBS pilus types across different lineages presents a challenge in developing a broadly protective vaccine that targets pilus components. Future studies focused on understanding how pilus structures are involved in GBS pathogenesis will advance efforts in finding new approaches to prevent infections caused by GBS.

CONCLUDING REMARKS

The recommended use of intrapartum antibiotic prophylaxis has not been fully effective at eliminating GBS transmission to newborns. The increasing incidence of disease in elderly populations and the eminent threat antibiotic resistance in GBS poses a serious threat for the future. Based on molecular epidemiology studies is it known that GBS clonal lineages contribute disproportionately to human disease; however, the specific bacterial factors that contribute to these differences are not fully understood. It is believed that GBS has emerged as a successful human pathogen due in part to its ability to readily exchange genetic material and acquire virulence characteristics that have allowed it persist in human populations. Although vaccination remains an attractive approach to preventing disease, the inherent genetic variability of circulating GBS strains has made it challenging to identify universal vaccine targets. Continued efforts to identify novel antigenic proteins and to understand the mechanisms involved in pathogenesis will prove critical for developing new strategies for preventing GBS-associated disease.

REFERNCES

REFERNCES

1. Manning SD, Springman AC, Lehotzky E, Lewis MA, Whittam TS, et al. (2009) Multilocus sequence types associated with neonatal group B streptococcal sepsis and meningitis in Canada. J Clin Microbiol 47: 1143-1148.

2. Richards VP, Lang P, Bitar PD, Lefebure T, Schukken YH, et al. (2011) Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted Streptococcus agalactiae. Infect Genet Evol 11: 1263-1275.

3. Cheng Q, Stafslien D, Purushothaman SS, Cleary P (2002) The group B streptococcal C5a peptidase is both a specific protease and an invasin. Infect Immun 70: 2408-2413.

4. Spellerberg B, Rozdzinski E, Martin S, Weber-Heynemann J, Schnitzler N, et al. (1999) Lmb, a protein with similarities to the LraI adhesin family, mediates attachment of Streptococcus agalactiae to human laminin. Infect Immun 67: 871-878.

5. Schubert A, Zakikhany K, Schreiner M, Frank R, Spellerberg B, et al. (2002) A fibrinogen receptor from group B Streptococcus interacts with fibrinogen by repetitive units with novel ligand binding sites. Mol Microbiol 46: 557-569.

6. Bolduc GR, Baron MJ, Gravekamp C, Lachenauer CS, Madoff LC (2002) The alpha C protein mediates internalization of group B Streptococcus within human cervical epithelial cells. Cell Microbiol 4: 751-758.

7. Dramsi S, Caliot E, Bonne I, Guadagnini S, Prevost MC, et al. (2006) Assembly and role of pili in group B streptococci. Mol Microbiol 60: 1401-1413.

8. Doran KS, Chang JC, Benoit VM, Eckmann L, Nizet V (2002) Group B streptococcal beta-hemolysin/cytolysin promotes invasion of human lung epithelial cells and the release of interleukin-8. J Infect Dis 185: 196-203.

9. Lang S, Palmer M (2003) Characterization of Streptococcus agalactiae CAMP factor as a pore-forming toxin. J Biol Chem 278: 38167-38173.

10. Harris TO, Shelver DW, Bohnsack JF, Rubens CE (2003) A novel streptococcal surface protease promotes virulence, resistance to opsonophagocytosis, and cleavage of human fibrinogen. J Clin Invest 111: 61-70.

11. Santi I, Scarselli M, Mariani M, Pezzicoli A, Masignani V, et al. (2007) BibA: a novel immunogenic bacterial adhesin contributing to group B Streptococcus survival in human blood. Mol Microbiol 63: 754-767.

12. Marques MB, Kasper DL, Pangburn MK, Wessels MR (1992) Prevention of C3 deposition by capsular polysaccharide is a virulence mechanism of type III group B streptococci. Infect Immun 60: 3986-3993.

13. Rosini R, Rinaudo CD, Soriani M, Lauer P, Mora M, et al. (2006) Identification of novel genomic islands coding for antigenic pilus-like structures in Streptococcus agalactiae. Mol Microbiol 61: 126-141.

14. Konto-Ghiorghi Y, Mairey E, Mallet A, Dumenil G, Caliot E, et al. (2009) Dual role for pilus in adherence to epithelial cells and biofilm formation in Streptococcus agalactiae. PLoS Pathog 5: e1000422.

15. Margarit I, Rinaudo CD, Galeotti CL, Maione D, Ghezzo C, et al. (2009) Preventing bacterial infections with pilus-based vaccines: the group B streptococcus paradigm. J Infect Dis 199: 108-115.