GENERAL INTRODUCTION

"Pathogenesis and symbiosis are relatively rare among bacterial species." (Ochman & Moran, 2001).

"Although we will never conquer infectious diseases, we can certainly learn to live in greater harmony with them. This may perhaps be our most effective intervention strategy. Discovering how to do so will be the great challenge for molecular epidemiologists of the future." (Foxman & Riley, 2001).

A. BRIEF INTRODUCTION

In microbiology, we are in the era of and witness to many paradigm shifts (Bull, ***et al.***, 2000). Group B streptococcus (GBS, ***Streptococcus agalactiae***) and human ureaplasmas (***Ureaplasma parvum*** and ***U. urealyticum***) – are clinically and phylogenetically related potential urogenital tract and perinatal pathogens (Razin ***et al***., 1998; Robertson ***et al***., 2002; Schuchat, 1998). They were selected as examples to demonstrate the paradigm shifts – from traditional to molecular microbiology, then to genomic and bioinformatics-based microbiology (Glass ***et al***., 2000; Glaser ***et al.***, 2002; Tettelin ***et al.***, 2002). In the post-genomic era, comparative, combinatorial and integrated studies, and bioinformatics have become popular strategies, because they can solve problems and generate new knowledge more efficiently than the other stratgies (Dobrindt & Hacker, 2001). When we look back at the history of microbiology, it is not difficult to find that significant progress in one defined specific area (for example, study of one bacterial species or even one test method, etc.) can be the basis for studies in related areas or even the whole microbiological field. For example, ***S. pneumoniae*** research has contributed significantly to studies of GBS serotyping, treatment, and prevention (Obaro & Adegbola, 2002). The aim of this thesis is to determine whether an integrated study

of GBS and human ureaplasma will reveal aspects of both that may not be apparent from studying them separately.

B. WHY GBS AND UREAPLASMAS?

B1. GBS and ureaplasmas.

GBS and ureaplasmas are among the most frequent bacteria associated with urogenital tract infection in women (Horner *et al.*, 2003) and perinatal diseases (Goncalves *et al.*, 2002; Razin *et al.*, 1998; Schuchat, 1998). Both belong to the low G+C group of bacteria – 35.6% for *S. agalactiae* and 25.5% for human ureaplasmas (Glass *et al.*, 2000; Glaser *et al.*, 2002; Tettelin *et al.*, 2002).

S. agalactiae is the only species of group B streptococcus (GBS). GBS is a facultative Gram-positive diplococcus and the leading cause of serious bacterial infections (bacteremia, pneumonia, and meningitis) in newborns (Chaffin *et al.*, 2000; Manning, 2003). An indispensable GBS virulence determinant is its serotype-specific capsular polysaccharide (CPS), which prevents the deposition of host complement factor C3b and inhibits opsonophagocytosis (Chaffin *et al.*, 2000).

Human ureaplasmas include two species: *U. parvum* and *U. urealyticum* and belong to the class *Mollicutes*. Mollicutes are the smallest free-living organisms, which lack cell walls, pass through "bacteria-retaining" filters, require cholesterol for membrane function and growth and use UGA codon for tryptophan. Their feature of without cell wall may be largely responsible for their biologic properties including lack of a Gram stain reaction and nonsusceptibility to many commonly prescribed antimicrobial agents, including beta lactams (Razin *et al.*, 1998). Ureaplasmas have a requirement, unique among mollicutes, for urea. ATP is generated through an electrochemical gradient produced by ammonia liberated during the intracellular hydrolysis of urea by urease (Blanchard & Barile, 1989).

B2. Clinical aspects.

B2.1. GBS.

GBS was originally known as cause of bovine mastitis and was not shown to be a human pathogen until 1938 (Manning, 2003). Currently, GBS remains a common cause of neonatal disease despite prevention efforts (Schuchat, 1998). Clinical syndromes of GBS disease in newborns include sepsis, meningitis, pneumonia, cellulitis, osteomyelitis, and septic arthritis (Schuchat, 1998). In addition to acute illness, which is itself costly, GBS infections in newborns can result in death and disability like cerebral palsy (Schuchat, 1998; Russell & Buttery, 2003).

B2.1.1. Early-onset disease (EOD).

Approximately 80% of infant GBS infections occur in the first few days (1-6) of life, so-called EOD. Newborns with EOD acquire the organism, before or during delivery, from their mothers who are colonized with GBS in the genital tract. Most EOD results from ascending spread of the organism into the amniotic fluid, where aspiration of contaminated amniotic fluid leads to invasive disease in some infants (Schuchat, 1998). Clinically, infants with EOD are often premature and prone to develop sepsis, respiratory distress and leukopenia (Ho *et al*., 1999).

B2.1.2. Late-onset disease (LOD).

GBS LOD occurs in infants between 1 week and 2 to 3 months of age. The pathogenesis of LOD is less well understood, although some cases probably reflect acquisition of the organism during passage through the birth canal (Schuchat, 1998). Clinically, LOD often presents with meningitis and sepsis (Ho *et al*., 1999).

B2.1.3. GBS infections among pregnant women.

Among pregnant women, GBS can cause clinical illness ranging from mild urinary tract infection to life-threatening sepsis and meningitis; chorioamnionitis, endometritis and septic abortion also occur (Manning, 2003). Most invasive

maternal infections are bloodstream infections, and osteomyelitis; endocarditis, and meningitis have also been described (Schuchat, 1998).

B2.1.4. GBS infections among elderly adults.

Recent studies have shown that a substantial burden of GBS-related illness occurs outside the high-risk periods – pregnancy and early infancy. The clinical presentations of invasive GBS diseases among non-pregnant adults (especially in the elderly people, and with underlying medical conditions) most often take the form of primary bacteremia, and skin, soft tissue or bone infections (Schuchat, 1998).

B2.1.5. GBS infection incidence trends.

The introduction of intrapartum antibiotic prophylaxis (IAP) has lowered the incidence of early-onset GBS sepsis by 50-80%, and the incidence among pregnant women also decreased (Moore *et al*., 2003; Schuchat, 1998). In contrast, the number of GBS-attributable conditions among non-pregnant adults, particularly the elderly with underlying conditions, is increasing, ranging from 4.1 to 7.2 cases per 100,000 adults over the past decade (Manning, 2003; Schuchat, 1998).

B2.1.6. GBS colonization.

Many individuals are asymptomatically colonized with GBS. Depending on the characteristics of the study population and the detection methods utilized, 10 to 40% of individuals are colonized. Newborns over 48 hours old are most commonly colonized in the throat and rectum and can remain colonized throughout childhood. At the onset of sexual activity, colonization generally shifts to the genitourinary tract (Manning, 2003). Only a small proportion of people colonized with GBS suffer any ill effects (Feikin *et al*., 2001; McDonald & Chambers, 2000).

B2.2. Ureaplasmas.

Like GBS, human ureaplasmas are also associated with a variety of urogenital tract diseases, although they cause adverse effects in an even smaller proportion of those colonized (Razin *et al*., 1998; Horner *et al*., 2003).

B2.2.1. Adverse pregnancy outcomes.

Adverse pregnancy outcomes, including maternal complications, chorioamnionitis, preterm birth and respiratory complications of the preterm neonate have been attributed to ureaplasmas (Abele-Horn *et al*., 2000). Ureaplasmas have been isolated from the blood of women with postpartum or postabortal fever but not from afebrile women following abortion or from healthy pregnant women (Taylor-Robinson & Furr, 1997; Razin *et al*., 1998). Ureaplasmas have been isolated from maternal blood, umbilical cord blood, and the blood and cerebrospinal fluid (CSF) of neonates (Taylor-Robinson & Furr, 1997; Razin *et al*., 1998). But some others study showed that (1) infection of the cerebrospinal fluid by ureaplasma is infrequent, (2) ureaplasma organisms are frequently present in tracheal aspirate specimens but do not appear to be related to the presence or the subsequent development of respiratory disease, and (3) initiation of erythromycin treatment at 1 to 3 weeks of age does not alter the clinical course (Heggie *et al*., 1994).

B2.2.2. Infants ureaplasma infection.

Colonization of infants by ureaplasmas may occur by ascent of the microorganisms from the lower genital tract of the mother at the time of delivery or by direct invasion of the fetus in utero. The relationship between acute and chronic lung disease (CLD) of prematurity and ureaplasmas is still controversial (Abele-Horn *et al*., 1998; Castro-Alcaraz *et al*., 2002; Hannaford *et al*., 1999; Heggie *et al*., 2001). Colonization of amniotic fluid with ureaplasmas may be associated with subsequent meningitis and other central nervous system (CNS) complications (Taylor-Robinson

& Furr, 1997; Razin *et al*., 1998).

B2.2.3. Ureaplasma related other infections.

Ureaplasmas usually reside extracellularly in the respiratory and urogenital tracts and rarely penetrate the submucosa, except in association with immunosuppression or instrumentation, when they may invade the bloodstream and disseminate to numerous organs and tissues (Razin *et al*., 1998). Ureaplasmas are the most common "typical" bacteria that cause "culture-negative" infectious arthritis in persons with hypogammaglobulinemia (Taylor-Robinson & Furr, 1997; Razin *et al*., 1998). Ureaplasmas sometimes can be detected by culture or PCR in the synovial fluid of persons with rheumatoid arthritis but their precise contribution to this disease is uncertain (Taylor-Robinson & Furr, 1997).

B2.2.4. Ureaplasma morbidity and mortality.

Assessing morbidity and mortality of diseases specifically caused by ureaplasmas is difficult because few studies systematically evaluate them, and some conditions in which they are implicated are polymicrobial (Taylor-Robinson & Furr, 1997; Razin *et al*., 1998). Deaths occurr in some neonates with bloodstream invasion and meningitis due to ureaplasmas; but, in others, the organisms spontaneously disappear from cerebrospinal fluid (CSF) without treatment (Taylor-Robinson & Furr, 1997; Razin *et al*., 1998).

B2.2.5. Ureaplasma colonization.

Human ureaplasmas can establish colonization in utero or at birth; they have been found on the skin and in the oropharynx, lung, and lower genital tract of many newborns. The persistence of ureaplasma in the oropharynx and upper respiratory tract and in the male genital tract usually declines after birth, but as many as 20% of prepubertal girls harbor these organisms in the lower genital tract (Taylor-Robinson

& Furr, 1997; Razin *et al.*, 1998). Human ureaplasmas are the most frequent bacteria isolated from the lower and upper genital tracts of women, especially pregnant women (Knox *et al.*, 1997; Razin *et al.*, 1998). Ureaplasmas are found in the cervix or vagina of 40-80% of asymptomatic, sexually active women (Cunningham *et al.*, 1996) and in the urethras of 5-20% of asymptomatic, sexually active men (Razin *et al.*, 1998; Taylor-Robinson, 1986).

The gastrointestinal tract may also be a potential reservoir: 9-18% of infants had rectal colonization with ureaplasmas (Sanchez & Regan, 1987; Sanchez & Regan, 1990); 57% of women attending a venereal disease clinic had positive anal or rectal swabs for ureaplasmas (Munday *et al.*, 1981).

B2.3. Similarities and differences between GBS and ureaplasmas.

B2.3.1. Similarities.

- Both GBS and human ureaplasmas are common perinatal, urogenital, and opportunistic pathogens;
- Both GBS and human ureaplasmas have high "colonization" but low "infection" rates.

B2.3.2. Differences.

- GBS is more virulent than ureaplasmas – GBS causes more serious morbidity and a higher mortality; its pathogenic role is much more clearly defined than that of ureaplasma. Studies of ureaplasma pathogenesis are often controversial.
- Both organisms have significant age/time-patterns. The incidence of ureaplasma neonatal colonization and infection is inversely related to gestational age or chronological age and is rare in full term infants. GBS infections are more common in preterm infants but also occur at term and in the first few months of life. GBS also causes significant infections in elderly adult; whereas

ureaplasmas do not.

- Ureaplasma colonization rates (20-80%) are higher than GBS (15-40%); ureaplasma colonization rates in pregnant women (40-80%) are higher than in non-pregnant women (~20%) because of the possible effects of oestrogens (Reid *et al.*, 1993). GBS colonization rates are similar between pregnant and non-pregnant women (Hoshina *et al.*, 1991; Manning, 2003);

B2.4. Conditional pathogens.

According to traditional thinking, a pathogen was a microbe that was genetically endowed with a factor that, when expressed, caused disease. This postulate became central to the concept of the monomicrobial aetiology of infectious diseases (Larsen & Monif, 2001). However, the mere presence of potentially pathogenic species does not necessarily constitute disease when defined in terms of symptoms (Larsen & Monif, 2001).

"Conditional" pathogens could mean that only some strains are pathogenic or that virulence factors are only sometimes expressed. It also could mean that host susceptibility plays a role in disease/infection. We are confronted with having to explain why apparently commensal bacteria can cause disease (Larsen & Monif, 2001).

For GBS and ureaplasmas, their potential for causing disease does not guarantee that disease will occur with their presence (Larsen & Monif, 2001). In fact, they rarely do cause disease – most colonized people are healthy. Although GBS is a leading cause of perinatal and maternal postpartum septicaemia, the incidence of disease is grossly disproportional to that of colonization (Larsen & Monif, 2001). The high colonization rates and low morbidity/mortality rates of GBS and especially ureaplasmas suggest that they are not "pure" pathogens but conditional pathogens (Goncalves *et al.*, 2002; Razin *et al.*, 1998; Schuchat, 1998).

B2.4.1. Quantity and diseases.

B2.4.1.1. GBS.

There is evidence of an association between maternal vaginal GBS concentration and neonatal infections after birth, especially in women with premature ruptures of membranes (Itakura *et al*., 1996). Heavy GBS colonization at 23 to 26 weeks' gestation is associated with an increased risk of delivering a preterm, low-birth-weight infant but is not a reliable predictor of neonatal GBS sepsis (Regan *et al*., 1996).

B2.4.1.2. Ureaplasmas.

The degree of colonization with ureaplasma also possible correlates with adverse pregnancy outcomes, but two independent studies got controversial results (Abele-Horn *et al*., 2000; Heggie *et al*., 2001). But for the "conditional" nature of ureaplasmas, this point needs to be further studied before a definite conclusion.

B3. Virulence and pathogenesis.

B3.1. Difference in GBS and ureaplasma pathogenicity.

From the viewpoint of pathogenicity, GBS is more virulent than ureaplasmas; they can cause serious, invasive and potentially fatal infection (not only "morbidity" but also "mortality") (McDonald & Chambers, 2000; Schuchat, 1998). Of the two ureaplasma species, there are differences between *U. parvum* and *U. urealyticum* in distribution and even some possible pathogenicity in different populations (Deguchi *et al*., 2004; Domingues *et al*., 2002; Knox *et al*., 2003). For example, in young males with non-gonococcal urethritis (NGU) *U. urealyticum* is more common than *U. parvum* (Povlsen *et al*., 2002; Deguchi *et al*., 2004). Usually ureaplasma infection-related clinical presentations are more "benign" than GBS infection; and they are rarely directly implicated in patient death (only "morbidity" but no or very

rare "mortality") (Razin *et al.*, 1998). Future research is required to determine why some women develop ascending intrauterine infections with either GBS or ureaplasmas and others do not; further, what interventions may reduce the deleterious effect of systemic fetal inflammation (Finlay & Falkow, 1997; Goncalves *et al.*, 2002).

B3.2. Common themes of GBS and ureaplasma pathogenesis.

Despite these differences in pathogenicity, GBS and ureaplasmas share many common themes in their pathogenesis.

B3.2.1. Adhesins.

There are only a few common types of adhesins – the first key step in causing disease (Baseman *et al.*, 1996) – used by pathogens to adhere to host substrates. These include:

- **Major surface protein antigens:** The GBS surface protein family is defined by C alpha and Rib (Stalhammar-Carlemalm *et al.*, 1999; Lachenauer *et al.*, 2000). Ureaplasma multiple banded antigen (MBA) may also be related to adhesins (Smith *et al.*, 1994; Zheng *et al.*, 1995).
- **Other proteins:** GBS laminin binding protein (Lmb) mediates GBS attachment to human laminin (Spellerberg *et al.*, 1999).

B3.2.2. Other virulence factors.

Several other common themes that are employed to avoid host immune defences:

- **Resistance to opsonophagocytosis –** GBS capsular polysaccharide and proteins (Harris *et al.*, 2003),
- **Cleavage of human fibrinogen or host immune components –** GBS cell-surface-associated protein [*cspA*]) (Harris *et al.*, 2003) and ureaplasma IgA1 protease (Robertson *et al.*, 1984).

- **Antigenic variation –** both GBS and human ureaplasma major surface antigens (Zheng *et al.*, 1994; Lachenauer *et al.*, 2000),
- **Camouflage –** GBS by binding host molecules (for example, GBS IgA binding protein, C beta) (Heden *et al.*, 1991) and ureaplasma MBA by possible mimicry (Ang *et al.*, 2002; Jacobs *et al.*, 1995).

B3.2.3. Mobile genetic elements (mge).

Virulence factors may be associated with mge or located within pathogenicity islands (PIs), which ensures that new strains of pathogens evolve constantly. Two GBS genomic studies have supported this theory (Glaser *et al.*, 2002; Tettelin *et al.*, 2002). Analysis of ureaplasma *tetM* and Tn*1545*-like (Tn*916*) transposon resistance determinants also support this point (de Barbeyrac *et al.*, 1996; Taraskina *et al.*, 2002).

B3.3. Capsular polysaccharide.

GBS capsular polysaccharide (CPS) is recognized as a key virulence factor of GBS (Schuchat, 1998), and its synthesis is determined by the capsular polysaccharide synthesis (*cps*) gene cluster (Chaffin *et al.*, 2000). Ureaplasmas also have capsule-like polysaccharide structures that contain glucosyl-like resides (Whitescarver & Furness, 1975; Robertson & Smook, 1976).

B3.4. Proteins.

B3.4.1. GBS:

a) **GBS surface variable protein antigen family:** Bca, Rib, Alp2, Alp3 (Gravekamp *et al.*, 1998; Lachenauer *et al.*, 2000) and R1-R4 (Moyo *et al.*, 2001a, b; Moyo & Maeland, 2003) are related to GBS virulence and are targets for serosubtyping.

b) **GBS C beta protein (Bac):** is a human IgA binding protein (Berner *et al*., 2002).

c) **Other proteins**. Many other proteins can contribute to GBS pathogenesis, which was described in two recent reviews (Manning, 2003; Spellerberg, 2000). For example, a cell surface protein designated Fbs, a R-like protein named BPS (group B protective surface protein) (Erdogan *et al*., 2002), a glutamine synthetase, the heat shock protein (Hsp70), and surface immunogenic protein (Sip) (Brodeur *et al*., 2000), hyaluronidase/hyaluronate lysase (gene *hylB*) (Granlund *et al*., 1998; Bohnsack *et al*., 2001), C5a-peptidase (ScpB) (Hauge *et al*., 1996; Adderson *et al*., 2000), GBS CAMP-factor (encoded by *cfb*) (Jurgens *et al*., 1987; Podbielski *et al*., 1994) and haemolysin/cytolysin (probable structural gene, *cylE*) (Pritzlaff *et al*., 2001) are all identified as probable virulence factors.

B3.4.2. Ureaplasmas.

B3.4.2.1. Multiple banded antigen (MBA).

MBA exposed on the cell surface are the major antigenic determinants in ureaplasmas. Antisera containing antibodies to these components inhibit growth and metabolism of ureaplasmas and, in the presence of complement, cause lysis of the organisms. These properties are used in various serologic tests that differentiate between ureaplasma species and serovars and detect antibodies to ureaplasmas in sera of patients (Razin *et al*., 1998).

B3.4.2.1. Human immuoglobulin A1 protease.

The 14 recognised human ureaplasma serovars express human immunoglobulin A1 (IgA1) protease activity (Robertson *et al*., 1984). Extensive cleavage of IgA1 could result in a local functional IgA1 deficiency that may facilitate colonization (Kilian *et al*., 1996).

B4. The role of antibody.

B4.1. GBS – protection.

The fact that serotype-specific anticapsular antibody opsonizes homologous GBS strains suggested the protective role of anticapsular antibody (Baker & Kasper, 1976; Schuchat, 1998). Seroype-specific antibody to capsular polysaccharide is protective, but such antibodies are rare (Hoshina, 1997; Campbell *et al.*, 2000; Mikamo *et al.*, 2000). A study of 1207 pregnant women in Canada showed that low level of protective serotype-specific capsular polysaccharide IgG antibody is one major risk factor for disease – in particular serotype III GBS is more invasive than other serotypes in this population may be due, at least in part, to poor maternal serotype III CPS-specific antibody response (Basham *et al.*, 1996; Davies *et al.*, 2001). GBS surface protein antibodies also can protect the host from infection (Hoshina, 1997; Schuchat, 1998). In one study, elevated levels of urinary antibodies were associated with an increased risk of preterm delivery, which suggests that a mucosal inflammatory response to urogenital infection may stimulate the onset of preterm labour. The results suggest that, unlike circulating serum antibody, mucosal GBS antibody may serve as adverse markers for preterm labour (McKenzie *et al.*, 1994).

B4.2. Ureaplasmas – adverse markers.

For ureaplasmas, one hypothesis is that mucosal immunity mediates resistance to infection and systemic immunity contributes substantially to the adverse effects (Cunningham *et al.*, 1996; Razin *et al.*, 1998), which is different from the role of GBS antibody (Davies *et al.*, 2001; McKenzie *et al.*, 1994). The presence of antibody may be an adverse marker in ureaplasma infection, suggesting possible autoimmune or immunopathogenic effects (Horowitz *et al.*, 1995; Abu-Shakra *et*

al., 1999). In such cases, molecular mimicry may play some role (Baseman *et al*., 1996).

B5. Prevention efforts.

B5.1. GBS.

In order to decrease the morbidity and mortality associated with GBS disease in newborns, the Centres for Disease Control and Prevention (CDC) and other organizations developed guidelines to provide women at risk of delivering a GBS-diseased infant with intrapartum prophylaxis. Penicillin and ampicillin are the primary agents used. Clindamycin and erythromycin are recommended for use in penicillin-allergic women. The consensus prophylaxis strategies are associated with frequent antibiotic use – at least 24% of women receive antibiotics during labour and delivery (Schuchat, 1998). Since the development of the GBS prevention program, the incidence of GBS neonatal sepsis has decreased significantly (Manning, 2003).

Because antibiotic resistance is a major public health concern, and antimicrobial prophylaxis cannot prevent most LOD, GBS-related stillbirths, or prematurity and does not address GBS disease in nonpregnant adults (Schuchat, 1998), alternative prevention protocols must be evaluated and implemented. For GBS, an alternative to intrapartum prophylaxis is vaccine, which should significantly reduce the number of women receiving antibiotics (Manning, 2003). Multiple GBS serotypes can cause perinatal and adult diseases, and multivalent GBS vaccines are being developed. Excellent immune responses to polysaccharides conjugated to tetanus toxoid were reported. Though the initial target of GBS vaccine development was the use of vaccines in pregnant women, during the second trimester, GBS vaccines could potentially be given to all women of reproductive age before pregnancy occurs, since conjugate vaccines are likely to lead to long-term protection in adults (Schuchat, 1998).

B5.2. Ureaplasmas.

It is not clear that any prevention strategy is justified for ureaplasma infection, considering the much lower pathogenicity compared with GBS (Gelfand, 1993). Moreover, any prevention strategy would be a challenge, since both intrapartum prophylaxis antibiotics and vaccines are associated with problems. The lack of ureaplasma cell wall makes them nonsusceptibe to many commonly prescribed antimicrobial agents, for example penicillin and ampicillin (Razin *et al.*, 1998). The possible adverse effects of circulating antibody (Horowitz *et al.*, 1995) because of molecular mimicry (Baseman *et al.*, 1996) makes vaccine also a poor option.

C. WHY GENOTYPING?

C1. Several concepts or glossary.

C1.1. Genotyping.

Genotyping of microorganisms is the means to discriminate between and catalogue microbial nucleic acid molecules (van Belkum *et al.*, 2001). It is widely used in basic and applied microbiological research.

- In basic microbiological research, genotyping is used in the fields of taxonomy, phylogenetic relationships, evolutionary dynamics, population genetics, and epidemiology (van Belkum *et al.*, 2001).
- In applied microbiological research, genotyping clearly affects several areas: the epidemiological investigation of outbreaks of infectious diseases and the measurement of genetic diversity in relation to relevant biological properties such as pathogenicity, drug resistance etc. (van Belkum *et al.*, 2001).

Genotyping techniques are considered superior to phenotyping methods, which are often less reproducible and discriminatory (Manning, 2003).

C1.2. Molecular epidemiology.

The term "molecular epidemiology" involves both "molecular" – the use of the techniques of molecular biology – and the “epidemiology” – the study of the distribution and determinants of disease occurrence in human populations (Foxman & Riley, 2001).

Molecular epidemiology can be used to examine disease patterns, investigate outbreaks, describe transmission and population dynamics, identify risk factors, understand evolution and disease pathogenesis, monitor the efficacy of control programmes (Foxman & Riley, 2001; Manning, 2003).

C2. Genotyping of GBS and ureaplasmas.

C2.1. Serotypes of GBS and ureaplasmas.

Human ureaplasmas include two species: *U. parvum* and *U. urealyticum*; *U. parvum* contains 4 serovars (1, 3, 6 and 14) and *U. urealyticum* contains 10 serovars (2, 4, 5, 7-13) (Robertson & Stemke, 1982).

S. agalactiae is the only species of group B streptococcus (GBS). Nine distinct capsular serotypes, Ia, Ib, and II to VIII, have been identified (Chaffin *et al*., 2000). GBS also can be serotyped or serosubtyped by surface protein antigens: C alpha, C beta, R and X proteins (Ferrieri, 1988; Lachenauer *et al*., 2000).

C2.2. Serotyping targets of GBS and ureaplasmas.

C2.2.1. Capsular polysaccharide.

GBS capsular polysaccharide antigens have been studied intensively – they are GBS major serotyping target antigens. The capsular polysaccharide synthesis (*cps*) gene

clusters of six serotypes (Ia, Ib, II, III, IV, V, and VI) have been sequenced and analysed (Chaffin *et al*., 2000).

Ureaplasmas also have capsule-like structures (extramembranous carbohydrates) and contain glucosyl-like residues (Whitescarver & Furness, 1975; Robertson & Smook, 1976), but their encoding genes have not been studied as systemically as those of GBS (Chaffin *et al*., 2000; Glass *et al*., 2000).

C2.2.2. Protein antigens.

The major surface antigens of GBS are a family of variable proteins (Bca/Rib/Alp2/Alp3) (Lachenauer *et al*., 2000). All the family members contain N- and C-terminal conserved regions and the middle repetitive unit region (Lachenauer *et al*., 2000). The repetitive units occupy 60-80% of the gene length (Lachenauer *et al*., 2000). GBS C beta protein is an IgA binding protein, which also contains repetitive units (Berner *et al*., 2002). These GBS surface protein antigens are the basis for serosubtyping separate from, but related to CPS serotyping (Lachenauer *et al*., 2000).

Ureaplasma major surface proteins are the multiple banded antigens (MBA) , which contain both species and serotype definition sites (Zheng *et al*., 1995). Their gene repetitive units also occupy over 60% of the gene length (Zheng *et al*., 1996).

C2.3. Genotyping is required and important for GBS and ureaplasmas.

From the point of methodology, the traditional serotyping methods for both GBS and ureaplasmas are not very practical (Stemke & Robertson, 1985; Manning, 2003). The reason is that both of them are multiple serotype species and the antisera are expensive and not widely available (Stemke & Robertson, 1985; Manning, 2003).

The accurate genetic bases of GBS (capsular polysaccharide and protein antigen) (Chaffin *et al.*, 2000; Lachenauer *et al.*, 2000) and ureaplasma (MBA) serotypes (Zheng *et al.*, 1995) are still not very clear.

The relationship between GBS and ureaplasma serotype and pathogenesis has not been defined partly because of the impractical nature of conventional serotyping (Manning, 2003; Stemke & Robertson, 1985) and partly because of the relatively poor discriminatory ability of conventional serotyping (Manning, 2003). A prerequisite for solving the problem is to develop practical species identification and typing (for serotype, serosubtype and genotype) assays (Abele-Horn *et al.*, 1997; Manning, 2003).

C3. Molecular epidemiology of GBS and ureaplasmas.

Traditionally, serotyping has been used to study GBS and ureaplasmas (Robertson *et al.*, 1986; Harrison *et al.*, 1998). Among the nine distinct capsular polysaccharide serotypes (Ia, Ib, and II to VIII) (Chaffin *et al.*, 2000), serotypes Ia, III, and V are currently the most common isolates from the United States and many other countries. Serotype III is the most prevalent serotype associated with neonatal disease and serotype V is an important emerging GBS serotype worldwide (Manning, 2003). There are four *U. parvum* and ten *U. urealyticum* serovars (Robertson *et al.*, 1986). The distribution of species (previous biovars) and serovars differs in different populations, but generally the four serovars of *U. parvum* are more common than ten *U. urealyticum* serovars (Knox *et al.*, 2003; Ren & Zhu, 2003).

The phenotypic serotyping methods have contributed to our understanding of GBS and ureaplasma infections, but they are limited since they do not reveal information regarding genetic identity (Bidet *et al.*, 2003; Echahidi *et al.*, 2002). Many investigators now use genotypic techniques alone or in conjunction with certain phenotypic techniques (Manning, 2003; Martinez *et al.*, 2001). GBS and ureaplasma

studies demonstrate that more molecular epidemiological information is needed to understand the pathogenesis of disease and to design alternative prevention strategies (Bidet *et al.*, 2003; Zheng *et al.*, 1992). For instance, comparing strains that cause disease with common colonization strains is important to identify bacterial factors that may be important in disease pathogenesis (Lopardo *et al.*, 2003; Povlsen *et al.*, 2002).

C4. Integrated study of infection, population genetics and evolution.

C4.1. Population genetics.

The analysis of bacterial collections, which accurately represent the natural population, by genotyping provides data that can be used both to investigate the population structure of bacterial pathogens and for the molecular characterization of bacterial isolates. A spectrum of possible population structures exists – with few bacterial species occupying the extremes of highly clonal and completely non-clonal, most containing both clonal and non-clonal elements (Spratt & Maiden, 1999).

At present, two main kinds of population structure can be distinguished in natural microbial populations:
(a) Species that are not subdivided into discrete phylogenetic lineages (panmictic species or basically sexual species with occasional bouts of short-term clonality fall into this category);
(b) Species that are strongly subdivided by either cryptic speciation or clonal evolution.

Improvements in available statistical methods are required to refine these distinctions and to better quantify the actual impact of gene exchange in natural microbial populations. Moreover, a codified selection of markers with appropriate molecular clocks (in other words: adapted levels of resolution) is sorely needed to

answer distinct questions that address different scales of time and space: experimental, epidemic, and evolutionary. The problems raised by natural genetic diversity are very similar for all microbial species, in terms of both basic and applied science (Tibayrenc, 1996).

C4.2. Evolutionary genetics.

Evolutionary studies address how genetic changes are induced, persist, and become fixed. At least four fundamental mechanisms can give rise to variation used to define evolutionary genetics: mutation, hypermutation, genetic recombination, and selection (van Belkum *et al*., 2001). Evolutionary genetics, apart from improving our basic knowledge of the taxonomy and evolution of microbes, can also greatly contribute to applied research in microbiology. Evolutionary genetics provides convenient guidelines for better interpreting genetic and molecular data dealing with microorganisms.

The three main potential applications of evolutionary genetics in microbiology are:
(a) Epidemiological follow-up (with the necessity of evaluating the stability of microbial genotypes over space and time);
(b) Taxonomy in the broad sense (better definition and sharper delimitation of presently described taxa, research into hidden genetic subdivisions);
(c) Evaluation of the impact of the genetic diversity of microbes on their relevant properties (pathogenicity, resistance to drugs, etc).

C4.3. Virulence clones.

A common theme demonstrated by these analyses is that distinct bacterial clones are responsible for disease outbreaks and increases in infection frequency. Unique combinations of virulence genes or alleles of virulence genes characterize many of these clones. Because substantial interclonal variance results in relative differences in virulence, molecular population genetic studies have led to the concept that the unit of bacterial pathogenicity is the clone or cell line. Continued new insights into

host parasite interactions at the molecular level will be achieved by combining clonal analysis of bacterial pathogens with large-scale comparative sequencing of virulence genes (Musser, 1996).

Understanding the population structure and evolution of pathogens is important (Spratt & Maiden, 1999; McDonald & Linde, 2002). In many bacterial infections there is a variety of disease manifestations and clinical outcomes. One factor that can contribute to differences in severity is the variation in virulence that exists among strains of a bacterial population. Genomic studies reveal that bacterial populations are comprised of isolates that show a surprisingly wide spectrum of genetic diversity at the DNA level (Whittam & Bumbaugh, 2002). It is clear that during bacterial evolution, gene acquisition and loss contributes substantially to the variation in virulence and to the overall genetic variation harboured in pathogen populations (Whittam & Bumbaugh, 2002). It is possible that the integrated study of GBS and ureaplasmas from the viewpoint of infection, population genetics and evolution could provide us some new clues for their further study (van Belkum *et al*., 2001).

Infectious microbial epidemiology, population genetics of microorganisms, and evolutionary dynamics all rely on genotyping data for discrimination between microbial genotypes, which made the integrated study of infection, population genetics and evolution of microorganisms possible (Tibayrenc, 1999; van Belkum, 2003). It is believed that these integrated studies will be widely used in future microbiological studies because of their efficiency in solving problems (Gurtler & Mayall, 2001; Wiedmann, 2003).

D. WHY GENOMICS AND COMPARATIVE GENOMICS?

D1. Some useful concepts or glossary.

D1.1. Genomic.

The capacity to sequence and assemble entire genomes of bacteria, pathogenic protozoans, and fungi in a rapid and cost-effective way has energized every aspect of microbial science (Fraser *et al*., 2000). Genomic technologies could be defined as those used to manipulate and analyse genomic information (Galas & McCormack, 2003). So far, more than 200 bacterial genomes have been fully sequenced, including most important human pathogenic species (http://www.tigr.org). The genomic era brings many advantages and new hopes for many research areas; however, the acquisition and analysis of genome sequence data is not an end in itself, but a starting point for generating testable hypotheses for future research (Wren, 2000).

D1.2. Comparative genomics.

Recently, the trend in the sequencing of closely related genomes, including the sequencing of more than one strain from a single pathogenic species means that the comparative genomics has truly come of age (Whittam & Bumbaugh, 2002; Wren, 2000). While analysis of a single genome provides tremendous biological insights into any given organism, comparative analysis of multiple genomes provides substantially more information (Fraser *et al*., 2000).

Comparative genomic studies of selected important human pathogens have contributed to our understanding of bacterial diversity, pathogenesis/virulence, evolution and the development of improved disease therapeutics (Fraser *et al*., 2000).

D1.3. Bioinformatics.

Bioinformatics is composed of many different interrelated scientific fields such as genomics, proteomics, and transcriptional profiling (Bull *et al*., 2000; Luscombe *et*

al., 2001). For bioinformatics, the search strategy is based upon data collection and storage, and mining of databases in order to generate knowledge (Bull *et al*., 2000). Bioinformatic databases include DNA (genes and genomes), RNA, and protein sequences, proteomes, macromolecular structures, chemical diversity, biotransformations, metabolic pathways (metabolomes), biodiversity, and systematics, etc. (Bull *et al*., 2000). Thus, innovative "experiments" can be made *in silico* rather than *in vivo* or *in vitro* and bioinformatic technology will accelerate the paradigm shift from bench-top to desk-top (Bull *et al*., 2000; Katoh, 2002). Bioinformatics is more successful than its competitors in solving problems in search and discovery (Bull *et al*., 2000).

D1.4. Genomic islands (GIs).

Many genomic islands have been discovered in a variety of pathogenic as well as non-pathogenic bacteria (Hentschel & Hacker, 2001). Depending on the functions which are encoded by genomic islands, these may also be called pathogenicity islands (PIs), symbiosis, metabolic or resistance islands (Osborn & Boltner, 2002). GIs are acquired by horizontal gene transfer and are now recognised as important contributors to bacterial adaptation and evolution (Hentschel & Hacker, 2001).

D1.4.1. GIs share a set of unifying but highly conserved features (Osborn & Boltner, 2002):

(i) GIs are present in the genomes of many bacteria but absent from the genomes of closely related strains.

(ii) They are often large (10-500 kb); however, smaller inserts (1-10 kb) have also been found and may be termed genomic islets.

(iii) They usually differ in G+C content and codon usage from the rest of the chromosome.

(iv) They are flanked by specific sequences (direct repeats), which may be generated following integration into the host genome via recombination.

(v) They are usually associated with tRNA loci, which presumably act as targets for the integration of foreign DNA.

(vi) They often possess genes or cryptic pseudogenes coding for genetic mobility such as phage genes, insertion sequences (IS) (Mahillon *et al.*, 1999), integrases, transposases and origins of replication.

(vii) They are frequently but not necessarily unstable.

D1.4.2. Pathogenicity islands (PIs).

PIs have been discovered in many pathogens of humans, animals and plants (Osborn & Boltner, 2002). PIs encode clusters of genes whose products contribute to virulence (Hentschel & Hacker, 2001; Strauss & Falkow, 1997). They range from adherence factors, to mechanisms required for entry into the host cell or acquisition of limiting metabolites, to overt toxins. Frequently, PIs encode type III and IV secretion systems, which are responsible for modulation of host contact, interference with the host signal transduction pathways, promotion of apoptosis and entry into non-phagocytic cell.

D1.5. Taxonomy.

Taxonomy (*taxis*: arrangements or order; *nomos*: law), also known as (bio)systematics, is the study of diversity of organisms and their relationships, comprising classification, nomenclature and identification (Hofling *et al.*, 1997). Theoretical studies of organism classification, which involves the clustering of units into groups, identification and development of coherent nomenclature, are essential (van Belkum *et al.*, 2001). Classification means the arrangement of organisms in groups; nomenclature is the attribution of correct international scientific names to organisms; identification is the inclusion of unknown strains in groups derived from classification (Moreno, 1997). Vandamme *et al.* advocate so-called polyphasic taxonomy – taxonomy based on a combination of data obtained by various laboratory techniques (Vandamme *et al.*, 1996; Moreno, 1997), which should be

widely accepted in future (Moreno, 1997).

D1.5.1. Species of bacteria.

The definition of species is a primary underlying concept; however, it is controversial and is undergoing continuous refinement (Moreno, 1997). Among microbial taxonomists, there is general agreement that the species concept currently in use – a bacterial species is defined as an entity in which members have a DNA-DNA homology value of at least 70% – is useful, pragmatic and universally applicable within the prokaryotic world (Murray & Schleifer, 1994; Moreno, 1997). A species can be seen as a "group with a common origin that is composed of the smallest diagnosable cluster of individual organisms within which there is a parental pattern of ancestry and descent". A species can also be considered as "condensed nodes" in an "otherwise cloudy, confluent taxonomic space" (Moreno, 1997).

Phylogeny – Development of systems based on evolutionary relationships rather than general resemblance (***phylon***: race or tribe; ***genesis***: generation or origin) (van Belkum ***et al.***, 2001).

D1.6. Microarray.

Recently developed GeneChip technology provides efficient access to genetic information using miniaturized, high-density arrays of DNA or oligonucleotide probes. Such microarrays are powerful tools to study the molecular basis of many aspects on a scale that would be impossible using conventional analysis (Gabig & Wegrzyn, 2001; Magee ***et al.***, 2001).

D1.6.1. Microarrays have been used:

- to disclose differences in gene content between taxonomically related strains (Schoolnik, 2002) and infer their phylogeny (Whittam & Bumbaugh, 2002);

- to quantify the extent of genetic variability within natural populations at the gene level of resolution (Fitzgerald & Musser, 2001);
- to identify crucial differences between pathogen and commensal (Schoolnik, 2002);
- to study host-pathogen interactions, mainly by identifying genes from pathogens that may be involved in pathogenicity and by surveying the scope of the host response to infection (Israel *et al*., 2001);
- to analyse genetic polymorphisms of specific loci associated with resistance to antimicrobial agents; this has the potential to identify targets for drug design (Hamels *et al*., 2001; Israel *et al*., 2001).

Although many challenges lie ahead, the usability of microarrays means that they should accelerate the advance of genetic epidemiology on multiple fronts (Dalma-Weiszhausz *et al*., 2002).

D2. Ureaplasma and GBS genomic studies.

There is one overarching lesson: completion of the genomic sequence of any species answers many questions, while at the same time it invites totally new ones (Subramanian *et al*., 2001). The acquisition and analysis of ureaplasma (Glass *et al*., 2000) and GBS genome sequence data (Glaser *et al*., 2002; Tettelin *et al*., 2002) are not ends in themselves, but starting points for generating testable hypotheses relating to their pathogenesis (Subramanian *et al*., 2001).

D2.1. *U. parvum* genomic study.

D2.1.1. *U. parvum* genome.

The complete genome sequence of *U. parvum* was released in 2000; it was the third mycoplasma to be sequenced (Glass *et al*., 2000). The genome size is 751,719 bp, which is smaller than any other sequenced microbial genome except *M. genitalium*

(Fraser *et al*., 1995). The genome contains 613 predicted protein-coding genes and 39 genes that code for RNAs (Glass *et al*., 2000). *U. parvum* G+C nucleotide content is 25.5%; it is much more A+T-rich than any of the other prokaryotic genomes sequenced to date (Glass *et al*., 2000). The feature that almost all (95%) ATP synthesis is the result of urea hydrolysis make ureaplasmas unique among mycoplasmas and all bacteria (Glass *et al*., 2000) and its urease is 30- to 180-fold more efficient than reported for other bacterial ureases (Glass *et al*., 2000).

D2.1.2. Combinatorial and comparative genomics.

Five ureaplasma proteins, urease, immunoglobulin-A1 (IgA1) protease, phospholipases A and C, and MBA have been proposed as virulence factors (Robertson *et al*., 2002). The pathogenic effect of urease is caused by its generation of ammonia. IgA1 protease might give ureaplasmas the capacity to invade the upper urogenital tract by degrading human IgA1 (Robertson *et al*., 1994). Ureaplasma phospholipases may be responsible for premature labour by altering prostaglandin biosynthesis. However, neither an IgA1 protease nor phospholipase A or C genes could be identified in the *U. parvum* genome (Glass *et al*., 2000). Pollack identified the similar problem when comparing genomic and enzymatic activity data (Pollack, 2001). He suggested a combinatorial analysis involving available evidence of genomic sequence, transcription, translational phenomena, structure and enzymatic activity, which will give the best picture of the organism (Pollack *et al*., 2002). Other *U. parvum* genome based studies are limited and often as a part of a larger study or as a comparison with the other genomes (Razin *et al*., 1998). For example, the comparison of the genomes of *M. genitalium*, *M. pneumoniae* and *U. parvum*, has helped define the essential functions of a self-replicating minimal cell (Razin *et al*., 1998). Rocha and Blanchard established a bioinformatic strategy to detect the major recombination hot-spots (in particular genomic repeats) in the genomes of *U. parvum* and other 3 mycoplasma species (Rocha & Blanchard, 2002).

D2.1.3. *U. urealyticum* genome.

To clarify some problems in *U. parvum* genomic studies and to make comparative genomic studies possible, the *U. urealyticum* genome will be valuable. This would help in better understanding of the new taxonomy of human ureaplasmas (Robertson *et al.*, 2002). A study by Robertson *et al.* showed that genome sizes were 760 kbp for the four serovars of *U. parvum* and 840-1,140 kbp for ten serovars (eleven strains) of *U. urealyticum* (Robertson *et al.*, 1990). If *U. urealyticum* genome were available, it would help explain the extra 80 or more kbp genomic fragments. However, because *U. urealyticum* genome will not be sequenced in the foreseeable future, we must adopt alternative strategies to help understand their new taxonomy, for example, a study of core genes has proved to be promising (Daubin *et al.*, 2002).

D2.2. GBS genomic studies.

D2.2.1. GBS genomes.

GBS genome projects have recently moved from the study of distantly related organisms to within-species comparisons of multiple strains (Glaser *et al.*, 2002; Tettelin *et al.*, 2002; Whittam & Bumbaugh, 2002) (http://www.tigr.org/). Complete sequences of two GBS genomes have been completed and a third is about to be released (Glaser *et al.*, 2002; Tettelin *et al.*, 2002) (http://www.tigr.org/).

D2.2.1.1. Serotype III strain NEM316 genome.

A serotype III strain NEM316 genome that was responsible for a fatal case of septicaemia was sequenced at the Pasteur Institute by Dr Glaser's group (http://www.pasteur.fr/english.html) (Glaser *et al.*, 2002). The most interesting finding is 14 genomic islands, which contain known and putative virulence genes, mostly encoding surface proteins, as well as a number of genes related to mobile elements. Some of these islands could therefore be considered to be pathogenicity

islands (PIs) and they may have an important role in virulence acquisition and genetic diversity (Glaser *et al*., 2002).

D2.2.1.2. Serotype V strain 2603 V/R genome.

Tettelin *et al.* from The Institute for Genomic Research (http://www.tigr.org/) sequenced the complete genome sequence of a *S. agalactiae* serotype V strain 2603 V/R. Their comparative genome hybridization (CGH) experiments between the 2603 V/R and 19 *S. agalactiae* strains from several serotypes using whole-genome microarrays, revealed the genetic heterogeneity among *S. agalactiae* strains, even of the same serotype (Tettelin *et al*., 2002). They also found 15 genomic regions as possible GIs.

D2.2.2. GBS comparative genomic.

Besides these two completed GBS genomes, a serotype Ia reference strain A909 genome at The Institute for Genomic Research (http://www.tigr.org/) is about to be released (http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NC_004128). While a single genome analysis provides tremendous biological insights into GBS, intraspecies comparative genomics of multiple serotypes or strains will provide substantially more information (Fraser *et al*., 2000). Strains often differ in their ability to cause disease. DNA microarrays, and bioinformatics, as powerful comparative genomic research tools, are uncovering novel virulence determinants, hidden aspects of pathogenesis, and new targets for vaccine development (Fraser *et al*., 2000; Whittam & Bumbaugh, 2002).

D2.3. Link GBS genomic data with other study findings.

Dmitriev *et al.* constructed physical chromosomal maps of *S. agalactiae* serotype II/(alpha + beta) and III/alpha strains. Their estimated genome sizes varied from 2,030 to 2,290 kb (Dmitriev *et al*., 2002), which proved to be very accurate when

compared with the sizes of the three full GBS genomes (http://www.ncbi.nlm.nih.gov/entrez/viewer.fcgi?val=NC_004128) (Glaser *et al.*, 2002; Tettelin *et al.*, 2002). But their study found only six copies of ribosomal operons, which was incorrect. The possible reason was that two copies are directly connected together, which made them difficult or impossible to identify as two copies by the PFGE methods they used (Dmitriev *et al.*, 1998).

Bohnsack *et al.* constructed the complete physical map of RDP III-3, which was a virulent subtype of serotype III. Their findings suggest that the genetic variation that distinguishes the RDP type III-3 from other serotype III strains occurs largely within localized areas of the genome containing known or putative virulence genes (Bohnsack *et al.*, 2002). Their finding was supported by analysis of the two known GBS genomes (Glaser *et al.*, 2002; Tettelin *et al.*, 2002) and will be further confirmed in future by BM110 genome (also RDP III-3).

Schubert *et al.* recently described the ***fbsA*** gene, which encodes FbsA, a fibrinogen receptor in GBS. Their study suggested that FbsA is the major fibrinogen receptor in this strain (Schubert *et al.*, 2002). Their sequencing of ***fbsA*** from five different GBS strains revealed significant variation in the number of repeat-encoding units (Schubert *et al.*, 2002). Although it was not specifically annotated in NEM316 genome, closer examination of the genome should reveal its presence (Glaser *et al.*, 2002; Schubert *et al.*, 2002).

These are only very limited examples to show the “link” studies. No doubt, this kind of study in future will be more fruitful, efficient, and also mutually beneficial for both genome mining and conventional bench-top study than either alone (Pollack, 2001).

E. THESIS ARRANGEMENT

The first section above is the general introduction of the thesis.

The second section relates to molecular microbiology and includes genotyping studies of ureaplasma and GBS, which will cover most of my previous ureaplasma and GBS publications with re-organization and/or modification (see appendices). The thesis-related publications (in PDF format) have been added as appendices. The inner relationship of the publications is explained at the beginning of the appendices. Only selected parts, which are believed to be important to explain the "story"**,** have been included in the body of the thesis.

The third section relates to genomic- and bioinformatics-based microbiology and contains the genomic-based ureaplasma taxonomy and GBS intra-species comparative genomic studies. I also would examine evolutionary aspects of ureaplasma and GBS at the end of the thesis.

F. HYPOTHESIS TO BE TESTED BY THE STUDY

The relationship between GBS and ureaplasma types (serotype, subtype, genotype) and disease is still not well understood. One hypothesis is that there are variations in virulence among strains of a bacterial population, some of which more easily cause disease (Whittam & Bumbaugh, 2002). In order to test the hypothesis, well-developed genotyping systems for both of them are needed in addition to conventional serotyping methods. Hopefully, the genotyping systems are closely related to conventional serotyping, but more practical and with higher discriminatory abilities (van Belkum *et al.*, 2001).

Now that both GBS and urealasmas are in the post genomic era, my hypothesis is that the genome resources and bioinformatic research tools and theories together can

help resolve previous problems more efficiently. For ureaplasmas, we would like to test whether *U. parvum* genome based core gene analysis could help better understanding of the new ureaplasma taxonomy. For GBS, we hope to further expose the pathogenicity-related GBS heterogeneity, based on intraspecies comparative genomic study.

Thus far, the most exciting work on the genomics of pathogens has clearly been in the area of evolution (Sassetti & Rubin, 2002). Both GBS and ureaplasmas are common potential perinatal pathogens and phylogenetically related. GBS is more virulent than ureaplasmas. We would like to test whether the integrated study of GBS and ureaplasmas together can help us understand them as perinatal pathogens better than study either alone. To test the hypothesis, we will examine them from an evolutionary viewpoint.