CHAPTER 8

INTEGRATED STUDY OF GBS AND HUMAN UREAPLASMAS – THEIR EVOLUTIONARY "WISDOM"

Fanrong Kong, Gwendolyn L. Gilbert

Centre for Infectious Diseases and Microbiology, Insititute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, NSW 2145, Australia

Statement of Joint Authorship

Kong, F. (candidate)
Did all the bioinformatic work, interpreted the data and wrote the manuscript.
Gilbert G. L. (supervisor)
Supervised the overall project. Assisted in the analysis and interpretation of data, and made a significant contribution to the manuscript.

8.1. Integrated study GBS and human ureaplasmas together.

8.1.1. Microbiology – in an era of paradigm shifts.

In microbiology, we are in an era of paradigm shifts (Bull *et al*., 2000), one of which is the popular use of bioinformatics. In bioinformatics, the research strategy is based upon data collection and mining (retrieval and integration) to generate knowledge (the understanding of what is important about a situation) from information (the sum of everything we know about that situation) (Bull *et al*., 2000). In the process, integrated analysis of different types of data may provide solutions, which analysis of either separately could not provide (van Belkum *et al*., 2001). For example, the combination of comparative genomic and evolutionary data – an integrated phylogenomic approach – is one successful case (Fraser *et al*., 2000; Eisen & Fraser, 2003).

This study aimed to determine whether several major, seemingly unrelated, aspects of two clinically and phylogenetically related, urogenital tract and perinatal conditional pathogens – group B streptococcus (GBS, *S. agalactiae*) and human ureaplasmas (*U. parvum* and *U. urealyticum*) – could be accounted for in a simple and unified manner (Almogy *et al*., 2002). In particular, we aimed to show whether an integrated study of the evolution of GBS and human ureaplasmas could reveal aspects of both that may not otherwise be apparent.

8.1.2. GBS and human ureaplasmas – many clinical similarities and differences.

GBS and ureaplasmas are two of the most common microorganisms associated with obstetric and perinatal infections (Razin *et al*., 1998; Schuchat, 1998). Ureaplasma colonization rates in pregnant women are normally higher (up to 80%) than those of GBS (up to 40%), and higher than in non-pregnant women (~20%) because of the possible effects of oestrogens (Reid *et al*., 1993) and higher urea levels in the vagina

during pregnancy (Kenny & Cartwright, 1977). GBS colonization rates are similar in pregnant and nonpregnant women (Hoshina *et al*., 1991; Manning, 2003).

The pathogenic potential of GBS and ureaplasmas does not guarantee that disease will occur with their presence (Larsen & Monif, 2001). In fact, they rarely cause disease – most colonized people are healthy. Although GBS is more virulent than ureaplasmas and is a leading cause of perinatal and maternal septicemia, the incidence of disease is grossly disproportional to that of colonization (Larsen & Monif, 2001). Studies of ureaplasma pathogenesis are more controversial (Razin *et al*., 1998). The high colonization rates and low morbidity and mortality rates of GBS and especially of ureaplasma infections suggest that they are not "true" pathogens but conditional pathogens (Goncalves *et al*., 2002; Razin *et al*., 1998; Schuchat, 1998).

8.2. GBS and human ureaplasmas are phylogenetically closely related.

Many studies, especially recent genome-based studies, suggest that *Streptococcus* spp. (in particular GBS) and human ureaplasmas are phylogentically closely related.

8.2.1. Low G+C Gram positive bacteria.

Analysis of full genome sequences confirmed the findings of previous conventional studies, that *U. parvum* (Glass *et al*., 2000) and GBS (Glaser *et al*., 2002; Tettelin *et al*., 2002) belong to the low G+C group of Gram positive bacteria – although ureaplasmas are not typically Gram positive, because they lack a cell wall (Razin *et al*., 1998). The G+C content of *U. parvum* (25.5%) is lower than that of closely related mollicutes, *M. genitalium* (32%) (Fraser *et al*., 1995) and *M. pneumoniae* (40%) (Himmelreich *et al*., 1996). Similarly, the G+C content of GBS (36%) is lower than those of *S. pyogenes* (38.5%) (Nakagawa *et al*., 2003) and *S. pneumoniae* (40%) (Hoskins *et al*., 2001; Glaser *et al*., 2002). The relatively low G+C content of both, within their corresponding families, supports the close relationship between

GBS and ureaplasmas. It is consistent with the hypothesis that ureaplasmas arose by reductive evolution (see below) from GBS and *M. pneumoniae* from *S. pneumoniae* or *S. pyogenes.* It is supported by the observation that human pathogenic mollicutes and *Streptococcus* spp. that affect the respiratory tract have higher G+C contents (*S. pneumoniae,* ~40% and *M. pneumoniae,* 40%) than those that affect the urogenital tract (*U. parvum,* 25.5%, *M. genitalium* 32% and GBS, ~36%). This suggests at least partially site-specific evolution and ecology (Chen & Zhang, 2003; Hurst & Merchant, 2001; Knight *et al*., 2001; Sandberg *et al*., 2003).

8.2.2. Genome size reorganization – evolution by reduction.

Phylogenomic studies show that, when bacterial lineages make the transition from free-living to obligate host-associated bacteria, they undergo a major loss of genes and DNA (Moran, 2002). Ureaplasmas (especially *U. parvum*) evolved from their low G+C content ancestor by a drastic reduction in genome size – from about 2.2 Mbp to about 0.8 Mbp – mainly through loss of many cell components (Rocha & Blanchard, 2002). Pre-genomic studies (mainly based on PFGE) showed that the GBS genome size vary from 2,030 to 2,290 kb (Dmitriev *et al*., 2002). The genome sizes of the four serovars of *U. parvum* are 760 kbp and of the ten serovars of *U. urealyticum,* 840 to 1,140 kb (Robertson *et al*., 1990). The full genome sequences of *U. parvum* (Glass *et al*., 2000) and GBS (Glaser *et al*., 2002; Tettelin *et al*., 2002) support the accuracy of previous PFGE genome size predictions (Robertson *et al*., 1990). In the process of "evolution by reduction" (Razin *et al*., 1998), *U. urealyticum* is less advanced, and 80-380 kbp longer, than *U. parvum*. We hypothesize that *U. urealyticum* has not discarded all of its "hostile" genes, which may explain clinical and pathological differences between *U. parvum* and *U. urealyticum* – the latter is a less common commensal, but more likely to be associated with some disease syndromes, such as urethritis (Povlsen *et al*., 2002; Deguchi *et al*., 2004).

Table 8.1. General features of GBS and ureaplasma genomes.

	GBS NEM316	**GBS 2603 V/R**	***U. parvum***
Length (bp)	2, 211, 485	2, 160, 267	751, 719
Genes (no.)	2, 118	2, 175	613
Gene density (gene no./kbp)	0.958	1.006	0.816
G+C (%)	35.6	35.7	25.5
tRNA (no.)	80	80	30
rRNA (no.)	7	7	2
rRNA gene order	16-23-5	16-23-5	16-23-5
rRNA located regions	455kbp	406kbp	203kbp
***cps* genes**	III-3	V	*epsG*
Major protein antigen	Alp2	Rib	MBA
Mobile genetic elements	IS*Sag2*	IS*1381*-IS*861*-IS*1548*-GBSil-IS*Sag1*-IS*Sag2*	None
Hemolysin genes	*cylE*	*cylE*	*hlyC, hlyA*
Urease genes	-	-	urease gene complex

Note.

epsG **–** exopolysaccharide biosynthesis, glycosyltransferase gene.

Other characteristics that support the concept of "evolution by reduction" are summarized in Table 8.2. Briefly, they are:

- all 7 GBS rRNA operons are within a fairly short region on the right replichore, whereas those of other *Streptococcus* spp. are on the both right and left replichores (including *S. pneumoniae,* in which rRNA operons are also confined to a short region 400 kbp) (Nakagawa *et al*., 2003; Tettelin *et al*., 2001). Both *U. parvum* rRNA operons are also on the right replichore (*M. pneumoniae* and *M. genitalium* have single rRNA operon on right replichore) (Fraser *et al*., 1995; Himmelreich *et al*., 1996) .
- ratios of 16S rRNA operon copy number to full genome size are higher for GBS and *U. parvum* than for other *Streptococcus* spp. and *Mycoplasma* spp., respectively.
- GBS has a significantly greater number of tRNA genes than the other *Streptococcus* spp. (Hoskins *et al*., 2001; Tettelin *et al*., 2001; Glaser *et al*., 2002; Tettelin *et al*., 2002; Nakagawa *et al*., 2003), most of the tRNA genes of GBS and *S. pyogenes* (but not *S. pneumoniae*) are on the positive strand. *U. parvum* has slightly fewer tRNA operons than *M. pneumoniae* and *M. genitalium*. Like GBS, the first tRNA operons are located on the positive strand, although they are not the majority. All of those located on the positive strand (the first 9 and one other) have corresponding tRNA operons among those of GBS that are on the positive strand. This suggests they may come from the same ancestor/source.

8.2.3. Indels.

Bacterial evolution is associated with continuous generation of novel genetic variants, resulting from point mutations, genetic rearrangements and horizontal gene transfer (Ziebuhr *et al*., 1999). Gene insertions and deletions or indels (Britten *et al*., 2003) are the major events underlying the emergence and evolution of bacterial pathogens and symbionts (Britten *et al*., 2003; Ochman & Moran, 2001).

Table 8.2. Some similar characteristics of GBS and *U. parvum* and their closely related species that support the concept of reductive evolution from *Streptococcus* to *Mollicutes*.

	Genome size (kbp)	16S rRNA		tRNA	
		Right or left replichores	N/genome size	N=	N= positive strand
GBS	2,030-2,290	7R (455 kbp)	0.0032	80	first 69
S. pyogenes	1,852-1,901	4R+1L or 4R+2L (750 kbp)	0.0026-0.0032	57-60	first 42-49
S. pneumoniae	2,039-2,161	3R+1L (400 kbp)	0.0018-0.0019	58	first 4-5
U. parvum	752	2R (203 kbp)	0.0026	30	first 9[a]
U. urealyticum	840-1,140	-	0.0018-0.0023	-	-
M. pneumoniae	816	1R	0.0012	35	first 4-5, negative strand
M. genitalium	580	1R	0.0017	32	-

Note.

a. Although a minority of *U. parvum* tRNA genes was in the positive strand, all 10 (including the first 9) were among the first 69 *S. agalactiae* tRNA genes.

The analysis of indels shows excellent correlation with phylogeny based on 16S rRNA for nearly all species (Gupta & Griffiths, 2002). Indel analysis clearly places *Streptococcus* spp. and *U. parvum* in the low G+C Gram-positive group (Gupta & Griffiths, 2002) (See Figure 8.1., which was kindly provided by Dr Gupta).

Different functional categories of genes evolve at significantly different rates. It is suggested that nonessential genes drive evolutionary diversification (Jordan *et al*., 2002). If so, the “backbone” of the genome (including essential genes) should be conserved. This means that, after *M. genitalium*, *U. parvum* would be more useful than GBS in defining the minimum genome (Fraser *et al*., 1995), because more of its genes are “essential” (Jordan *et al*., 2002).

Long-term processes leading to the development of new species or subspecies are termed macroevolution, and short-term developments, which occur during days or weeks, are considered to be microevolution (Morschhauser *et al*., 2000). Indels apparently contribute to both macroevolution and microevolution (Gupta & Griffiths, 2002). In particular, it appears that significant deletions or "evolution by reduction" (macroevolution) led to *U. parvum* and *U. urealyticum* evolving from a GBS/ureaplasma common ancestor. In addition both deletions and insertions (especially pathogenicity islands) (Glaser *et al*., 2002) would have contributed to microevolutionary diversity and possibly the development of virulent clones of GBS (Jones *et al*., 2003).

8.2.4. Supertree.

Although indels and lateral gene transfer (LGT) (Ziebuhr *et al*., 1999) contribute to significant intra-species differences (Morschhauser *et al*., 2000), analysis of the highly conserved core genes or backbone of a species is likely to be more useful to infer bacterial phylogeny, as shown in our GBS comparative genomic study (Daubin *et al*., 2002). The “supertree” based on core genes confirms the positions of

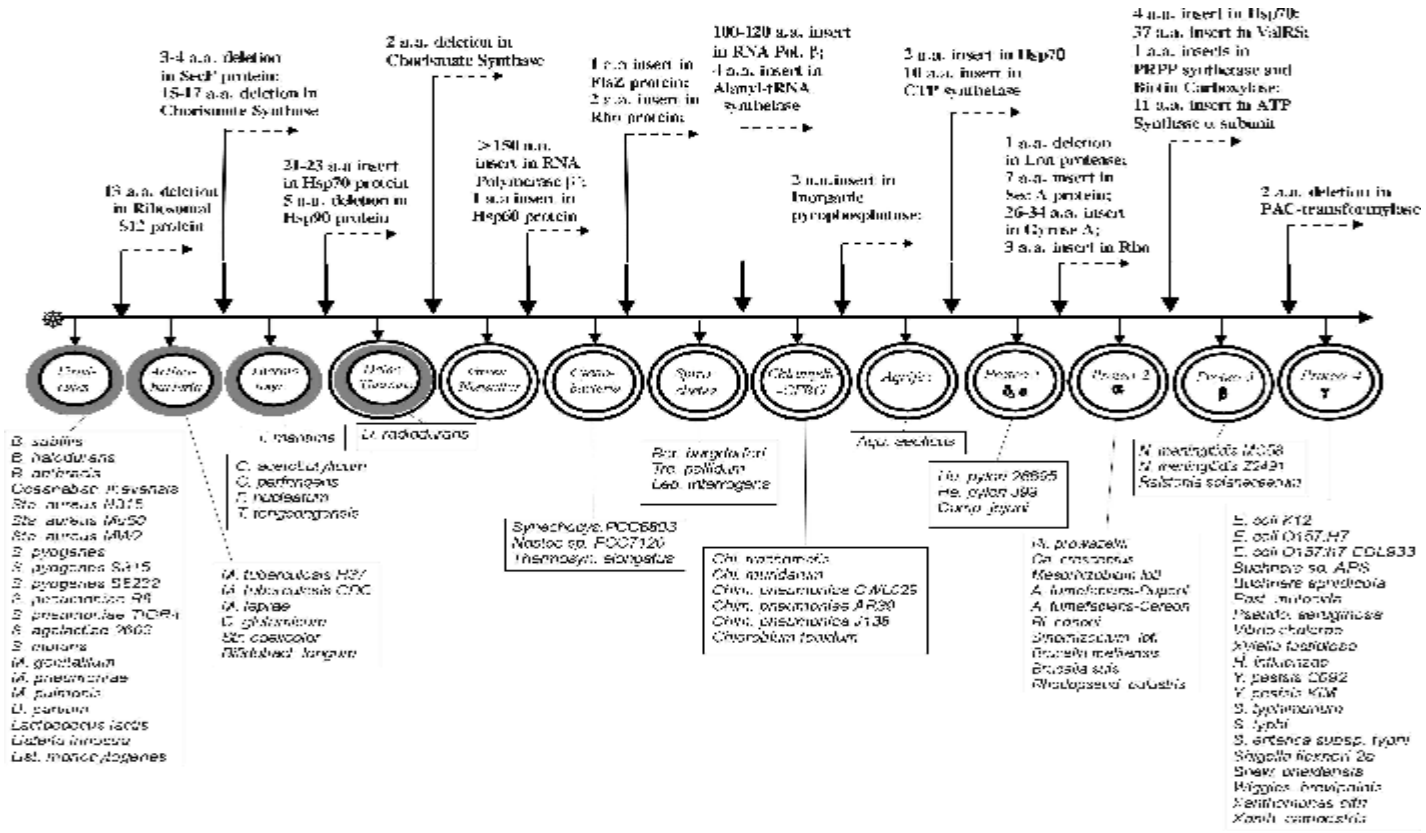

Figure 8.1. (according to Daubin *et al*., 2002) The branching order of the main groups within *Bacteria* based on conserved indels present in various proteins. The thick arrows above the line show the evolutionary stages where the identified indels are postulated to have been introduced. The phylogenetic placements of species from completed bacterial genomes based on these signatures are as shown. The model predicts that once a signature has been introduced in the main lineage, all groups to the right of it should contain the signature whereas all groups preceding it should lack it. The model is strongly supported by the sequence data for completed bacterial genomes with only 11 exceptions from the predicted pattern observed in 1,842 observations. Abbreviations in the protein names are: PRPP synthetase – phosphoribosyl pyrophosphate synthetase; PAC-transformylase –5-aminoimidazole-4-carboxamide formyltransferase.

Streptococcus spp. (in particular GBS – personal communication with Dr Daubin suggested that GBS and *S. pyogenes* is in the neighbour branch) and *Mollicute* spp. (including *U. parvum*) in two neighbouring clusters of the low G+C Gram-positive bacteria (Daubin *et al*., 2002).

8.3. Evolution – not only useful for genome mining?

In the postgenomic era, much has been gained from combined and integrated genomic and evolutionary studies of human pathogenic bacteria (Fraser *et al*., 2000; Eisen & Fraser, 2003). Could we go another step further? Analysis of the evolution of GBS and human ureaplasmas may help us to understand how and perhaps why similarities and differences – presence or absence of genes, DNA substitution in noncoding regions and global patterns of synteny (conserved gene order) – have arisen across species (Eisen & Fraser, 2003).

8.3.1. Can we attribute some "wisdom" to GBS and human ureaplasmas?

Improving our understanding of the biomedical significance of GBS and human ureaplasmas may require changing the way we look at them (Woolhouse *et al*., 2002). This may show how a wide range of major, seemingly unrelated issues in the study of GBS and human ureaplasmas may be accounted for in a simple and unified manner, from an evolutionary perspective (Almogy *et al*., 2002). Researchers also need to show "wisdom", because the evolutionary perspective is itself a theory or hypothesis rather than "fact" or "truth". If we postulate that organisms – specifically GBS and human ureaplasmas – have some "wisdom" (at least as imposed by us), the hypothesis may be much easier to understand. An analysis of evolutionary "wisdom" can be used to understand bacterial metabolism, pathogenicity, physiology and behavior (Eisen & Fraser, 2003). For example, although co-evolution between host and pathogen has been difficult to prove rigorously, in practice it can have a major influence on the interpretation of genetic variation in biomedically important traits (Woolhouse *et al*., 2002).

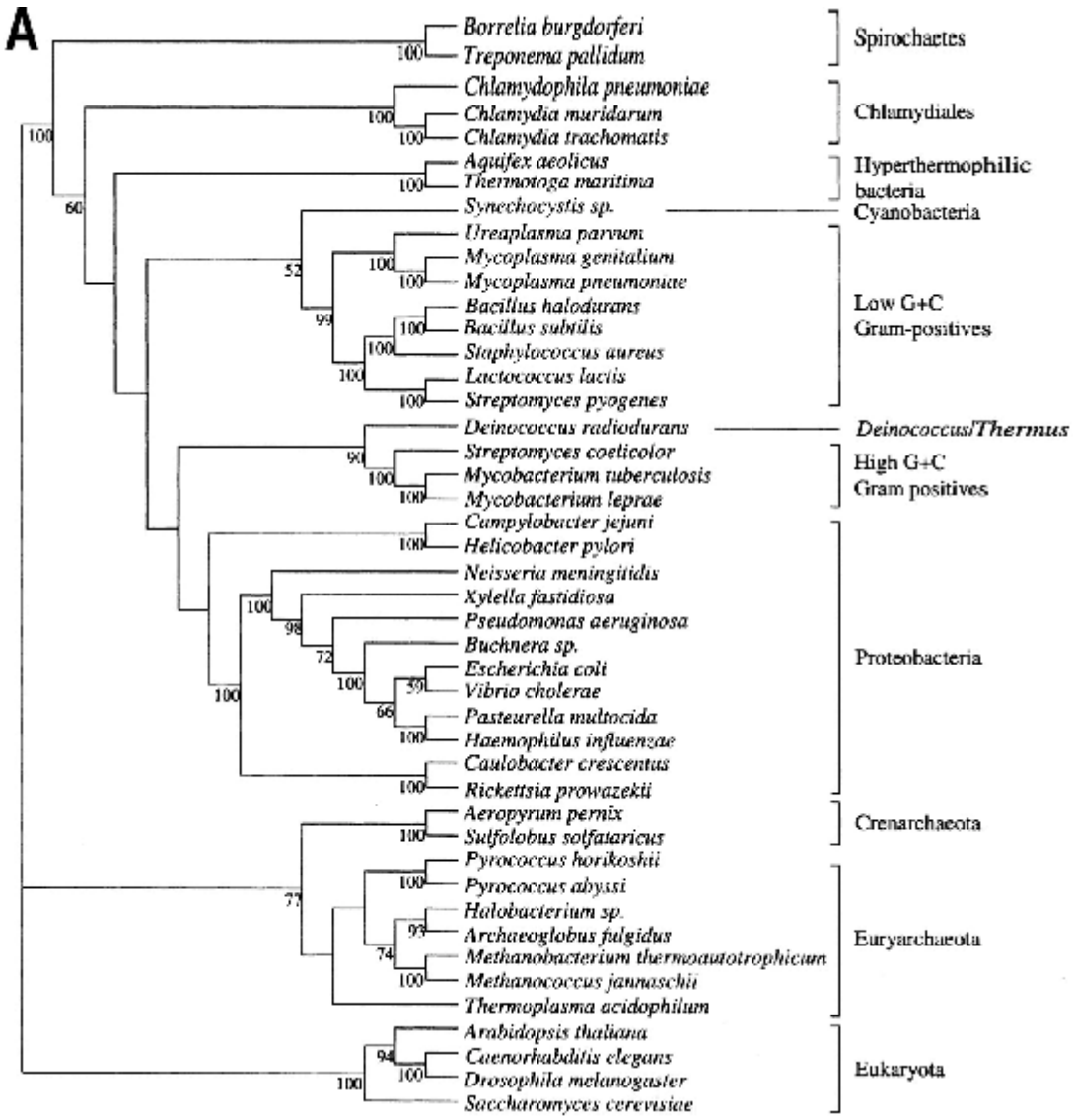

Figure 8.2. Supertrees of 45 species constructed with 730 trees. Supertree based on trees made by BIONJ and a gamma distribution estimation of evolutionary rate heterogeneity. Personal communication with Dr Daubin suggested that ***Streptococcus agalactiae*** and ***Streptococcus pyogenes*** is in the neighbour branch.

8.3.2. Understanding the "wisdom" of GBS and human ureaplasmas through their population genetic structures.

We hypothesize that the evolutionary potential of GBS and human ureaplasmas is reflected in their population genetic structure (McDonald & Linde, 2002). We could even consider using population genetic structure to evaluate the "wisdom" of their evolution. In relation to pregnant women, human ureaplasmas apparently are more successful than GBS, because they colonize a much greater proportion of the population – up to 80% compared with 40% or less for GBS. *U. parvum* is a more common colonizer than *U. urealyticum* – their relative ratio is about 4 to 1 (Kim *et al*., 2003; Schuchat, 1998) – indicating that *U. parvum* is more "successful" than *U. urealyticum* in coevolution (Woolhouse *et al*., 2002).

8.3.3. Population structure of pathogens: the role of immune selection.

It is interesting to consider how the interplay between a pathogen (Almogy *et al*., 2002) and host immune responses, particularly to conserved and variable antigens, shapes the pathogen's population structure (Gupta & Anderson, 1999). At one extreme, immune selection against polymorphic determinants can cause pathogen populations to self-organize spontaneously into discrete antigenic types that may either be maintained over long periods or undergo cyclical or chaotic fluctuations. At the other extreme, diversity may be drastically reduced by competition, induced by a strong immune response against a conserved determinant, because the many different "clones" can be eliminated in the mean and very short time (Gupta & Anderson, 1999).

GBS is more virulent, stimulates a stronger host immune response and has a relatively lower population size in pregnant women, than ureaplasmas. The parasitic life style and molecular mimicry adopted by ureaplasmas, to avoid stmulating the host immune attack, means that they colonise a higher proportion of the population (Baseman & Tully, 1997). The N-terminal ends of the major surface antigens of

GBS (C alpha and Rib) and human ureaplasmas (MBA) are more important than their variable regions. GBS hide their variable regions by deletion of tandem repeat elements (Madoff *et al*., 1996) and ureaplasmas hide theirs by phase variation (Monecke *et al*., 2003), ureaplasmas (at least U. *parvum* serovar 3) also mimic human host antigens as well as by varying the number of repeats (Baseman & Tully, 1997).

The repetitive units of the major surface antigens of both GBS and human ureaplasmas have multiple different formats (Zheng *et al*., 1995; Zheng *et al*., 1996; Lachenauer *et al*., 2000). Because they are located at surface sites and recognised by the host (Zheng *et al*., 1994; Lachenauer *et al*., 2000), changing their format repeatedly allows them to to evade immune attack (Lachenauer *et al*., 2000; Monecke *et al*., 2003).

8.3.4. Common themes but different strategies – two interesting case studies?

Analysis of genomes of microbial pathogens has provided common themes relating to virulence, host adaptation and evolution – including lateral gene transfer (LGT), genome decay and antigenic variation among pathogens (Wren, 2000).

"Genome decay" is the process by which it is hypothesized that GBS and human ureaplasmas developed from a common ancestor. Loss of about half of the genome size (about 2.2 Mbp) of a common ancestor resulted in *U. urealyticum*, and loss of another small portion (80-380 kbp) in *U. parvum* (Razin *et al*., 1998). This resulted in loss of ureaplasma virulence genes, which are not essential for a parasite-like organism (Ochman & Moran, 2001; Razin *et al*., 1998).

"Lateral gene transfer (LGT)" in GBS is shown by the presence of mge and pathogenic islands (PIs) and is the process by which it can increase its virulence (Ochman & Moran, 2001; Glaser *et al*., 2002; Tettelin *et al*., 2002). The availability of complete genome sequences for multiple strains of GBS hopefully will allow

insight into what determines that certain bacterial strains are more pathogenic than others. Already, for GBS, the availability of two compete genomes, and others near completion, allow us to begin intraspecies comparative genomics (Kruger & Baier, 1997; Fraser *et al*., 2000; Whittam & Bumbaugh, 2002).

8.4. Repetitive sequences.

In contrast to the smaller genomes of obligate host-associated bacteria, like human ureaplasmas (Glass *et al*., 2000; Rocha & Blanchard, 2002), those of free-living bacteria, like GBS, often carry phages and other repetitive sequences that mediate genomic rearrangements (Glaser *et al*., 2002; Tettelin *et al*., 2002).

8.4.1. Repetitive sequences in protein antigens.

Genomic studies indicate that tandem-repeat gene polymorphisms are more common than is generally believed (Andersson *et al*., 2002). A catalogue of putative polymorphic repeats within transcribed sequences comprises a large set of potentially phenotypic or disease-causing loci (Andersson *et al*., 2002; Rocha & Blanchard, 2002). The major protein antigens of both GBS (family defined by Bca and Rib) and ureaplasmas (MBA) contain tandem-repeat regions (see below). Repetitive sequences are also found in other GBS but not *U. parvum* surface located proteins (Glass *et al*., 2000; Glaser *et al*., 2002; Tettelin *et al*., 2002).

8.4.2. Other genomic repeats.

Comparative genomic studies have shown that, among *Streptococcus* spp., *S. pyogenes* (3 genomes) contains 4-6 phages, which occupy 7.0-12.4% of genome length (Banks *et al*., 2002; Nakagawa *et al*., 2003) and *S. pneumoniae* contains many BOX (about 25 copies) (Martin *et al*., 1992) and RUP repetitive sequences (108 copies in serotype 4 genome) (Oggioni & Claverys, 1999; Hoskins *et al*., 2001; Tettelin *et al*., 2001). By contrast, although *S. agalactiae* contains several genomic islands (GIs) including possible pathogenic islands (PIs) (Glaser *et al*., 2002;

Tettelin *et al*., 2002), but no typical phages or repetitive sequences. This suggests that *S. agalactiae* is relatively more stable (or less flexible) than *S. pyogenes* and *S. pneumoniae* (Glaser *et al*., 2002; Tettelin *et al*., 2002; Nakagawa *et al*., 2003).

Rocha and Blanchard established a bioinformatic strategy to detect the major recombination "hot-spots" in the four published mollicute genomes (Rocha & Blanchard, 2002). *U. parvum* has fewer genomic repeats than *M. pneumoniae* (Himmelreich *et al*., 1996) and *M. genitalium* (Fraser *et al*., 1995; Rocha & Blanchard, 2002) – which suggests that it is less flexible than the two human *Mycoplasma* spp. (Mira *et al*., 2002).

The relative lack of repetitive sequences and genome stability, compared with their respective close relatives, is another common feature of GBS and human ureaplasmas.

8.5. Genotypes and diseases.

As in many other bacterial infections, there are a variety of disease manifestations and clinical outcomes for GBS and ureaplasma infections, the basis of which is not well understood. One factor that can contribute to occurrence, and differences in severity, of disease is the variation in virulence that exists among strains of a bacterial population (Whittam & Bumbaugh, 2002). The intra-species variability in gene content, genomic organization, and gene expression may account for variation in the severity of disease and for the diverse clinical outcomes of infection (Finlay & Falkow, 1997; Whittam & Bumbaugh, 2002) .

8.5.1. Why genotyping study?

GBS serotypes are based on capsular polysaccharide and serosubtypes on major surface proteins; human ureaplasmas serotypes are based on surface proteins (see below). Both GBS and human ureaplasmas have multiple serotypes and, for both,

the traditional serotyping methods**,** based on panels of antisera, are not very practical for routine use. Therefore, we developed molecular serotype identification systems.

Previously, the bases of GBS protein antigen "serosubtyping**"** and ureaplasma serotyping targets were not very clear (Zheng *et al*., 1995; Zheng *et al*., 1996; Lachenauer *et al*., 1999; Lachenauer *et al*., 2000). To explain them more clearly, we also studied the GBS (Kong *et al*., 2002b) and ureaplasma protein antigen gene targets (Kong *et al*., 2000b).

The relationships between species (for ureaplasmas) and/or serotype (for both GBS and ureaplasma) and pathogenesis/virulence still have not been resolved, partly because of the difficulties associated with conventional serotyping (Abele-Horn *et al*., 1997; Heggie *et al*., 2001; Manning, 2003).

For GBS, a molecular serotyping method would be valuable and practical alternative for use in disease surveillance and future vaccine design (Jones *et al*., 2003).

8.5.2. Our genotyping study.

Our genotyping studies (Kong *et al*., 2002a; Kong *et al*., 2002b; Kong *et al*., 2003) showed that there were differences in the distribution of GBS serotypes based on geographic area and patient age. GBS serosubtype III-2, in our system, is the high virulence clone previously recognised by others (Musser *et al*., 1989; Takahashi *et al*., 1998; Kong *et al*., 2003). Differences in distribution of ureaplasma serotypes are mainly determined by species; for example, *U. parvum*, especially serovars 3 and 6, are the predominant species and serovar (Kong *et al*., 2000; Knox *et al*., 1997 & 2003).

8.5.3. Considerations for future genotyping.

Now that more specific and discriminatory genotyping methods are available, further clinical studies will be required to elucidate the relationships between

genotype and pathogenicity, for both human ureaplasmas and GBS (Heggie *et al*., 1994; Heggie *et al*., 2001; Schuchat, 1998). Although serotype III (III-2 in particular) has been shown to be the most virulent serotype, all serotypes can cause infection. More discriminatory tools are needed to identify more specific virulence markers, based on data mined from multiple GBS genomes to develop genechip microarrays (or equilvalent technology) (Glaser *et al*., 2002; Tettelin *et al*., 2002). Because human ureaplasmas are less virulent than GBS, it is more difficult to draw conclusions about the relationship between serotypes and virulence; less definite associations, compared with GBS, are to be expected. How to deal with this problem is still a challenge. Host factors are important in pathogenesis of infection due to both organisms, but probably more so in ureaplasma infections.

8.6. Comparative genomics – GBS and human ureaplasmas?

While analysis of a single genome provides tremendous biological insights into any given organism, comparative analyses of multiple genomes will provide substantially more information (Fraser *et al*., 2000). As comparative genomics shifts from inter-species to intra-species differences, could we shift to comparing clusters of strains causing different types of infection (perinatal, urogenital tract, respiratory tract, etc.)? The transition from free-living or facultatively parasitic life cycles to permanent associations with hosts involves a major loss of genes and DNA (Rocha & Blanchard, 2002). By mining the respective genomes of ureaplasmas and the bacteria from which they are believed to have evolved, could we relate the loss of genome content to their reduced virulence? In future, the availability of the *U. urealyticum* genome should identify the reason for the 80-380 kbp size difference between it and *U. parvum*, and may allow to explain other differences between them.

8.6.1. Major surface protein antigens.

The major surface protein antigens of human ureaplasmas (MBA) and GBS (Bca/Rib/Alp2-5) are structurally similar (Zheng *et al*., 1995; Lachenauer *et al*., 2000; Schuchat, 1998), but there are differences in their repetitive sequences. A sequence homology search showed that the repetitive units of *U. parvum* serovar 3 MBA repetitive units have far more similarities to mammalian collagen/collagen-like proteins than other human ureaplasma and GBS surface proteins. Up to 100 matched sequences were found in *U. parvum* serovar 3, compared with 0-12 in *U. urealyticum* serovar 10 MBA and GBS Bca/Rib/Alp2-3 repetitive units. A previous study in our laboratory showed that monoclonal antibody against *U. parvum* serovar 3 MBA repetitive units, but not against those of other ureasplasma serovars, cross-reacted with normal rabbit lung tissue (Kirsty Hannaford-Turner, 2002 PhD thesis, University of Sydney). This provided indirect evidence supporting the sequence search results (Hance *et al*., 1976).

As previously discussed, many more pregnant women carry ureaplasmas than GBS. The ratio of *U. parvum* and *U. urealyticum* carriage is 4 to 1 and *U. parvum* serotype 3 is the predominant serotype; *U. urealyticum* serotypes 7, 11 and 9 are rare. We hypothesise that the similarities between surface protein repetitive units and human connective tissue antigens have developed during coevolution and have influenced ureaplasma genetic population structure. This hypothesis is supported by the greater similarity of the predominant *U. parvum* serovar 3, and the fact that *U. urealyticum* serotypes 7, 11 and 9, which lack repetitive sequences, have the least similarity with human proteins.

8.6.2. Capsule and capsular polysaccharide.

GBS capsular polysaccharide is its major virulence factor. Human ureaplasmas also have a capsule-like structure or exopolysaccharide, which contains glucosyl-like resides (Whitescarver *et al*., 1975; Robertson & Smook, 1976). In ureaplasmas, *epsG* encodes glycosyltransferase, which is involved in exopolysaccharide biosynthesis. There is no *cps* gene cluster, similar to that in GBS, in *U. parvum* genome (Glass *et al*., 2000), but 40-68% homology between *eps* and *cps* genes has

been reported (Stingele *et al*., 1996). Again, this supports the concept of "degenerate or reverse evolution" of human ureaplasmas from gram-positive/*Streptococcus* spp. (Razin *et al*., 1998). Unlike GBS, the capsule-like polysaccharide of ureaplasmas is not the serotype-specific antigen.

Table 8.3. Sequence comparison of human ureaplasma and GBS major surface protein antigens.

Protein name	GenBank No.	Whole length	Length of N-terminal	Length of repetitive unit
MBA-Up-3	L202329	409 aa	151 aa	6 aa x 42.5
MBA-Uu-10	U50459	487 aa	147 aa	8 aa x 42.3
Bca	M97256	1020 aa	172 aa	82 aa x 9
Rib	U58333	1231 aa	172 aa	79 aa x 12
Alp2	AF208158	786 aa	172 aa	76 aa x 3
Alp3	AF245663	865 aa	172 aa	79 aa x 5

8.6.3. Mobile genetic elements (mge).

Comparative genomics has shown that *S. agalactiae* contains a large number of mge – including insertion sequences (IS), group II introns (Kong *et al*., 2003) and prophage-like structures (Glaser *et al*., 2002; Tettelin *et al*., 2002). No mge were found in the *U. parvum* genome (Glass *et al*., 2000) but a Tn*1545*-like transposon (Tn*916*) was found in tetracycline resistant ureaplasmas that carry *tetM* gene (de Barbeyrac *et al*., 1996; Taraskina *et al*., 2002). It would be of interest to establish whether the extra 80-380 kbp of *U. urealyticum* contains mge (de Barbeyrac *et al*., 1996).

8.7. GBS and human ureaplasma – taxonomy case studies supporting reconsideration of species definition.

8.7.1. The case for separation of *U. parvum* and *U. urealyticum*.

DNA-DNA hybridization shows only 60% homology between *U. parvum* and *U. urealyticum* (Chrisstiansen *et al.*, 1981; Harasawa *et al.*, 1991) and they have also different genome sizes (Robertson *et al*., 1990). Divergent nucleotide sequences of several highly conserved genes, which are identifiable by biovars/species-specific PCR targeting 16S rRNA, the 16S-23S rRNA intergenic regions, the genus-defining urease or the serovar-defining multiple-banded antigen (MBA) genes, also attest to their phylogenetic distinctiveness. Further evidence for establishment of two separate species – *U. parvum* and *U. urealyticum* – is provided by:

- distinctive RFLP and randomly amplified polymorphic DNA [RAPD] typing patterns;
- phenotypic differences (including clustering of antigenic types, polypeptide patterns of whole-cell preparations, differential inhibition by manganese, and polymorphism among their ureases, pyrophosphatases, diaphorases) (Robertson *et al*., 2002).

8.7.2. The case for combination of *S. agalactiae* and *S. difficile.*

Genotypic studies, including our study of the partial *cps* gene cluster, serotype Ib-specific PCR positive result (Kong *et al.*, 2002a), and phenotypic studies, including whole-cell protein electrophoretic analysis and serological studies, showed that *S. difficile* is actually an atypical *S. agalactiae* (GBS) type Ib (Vandamme *et al.*, 1997).

These two examples highlight some controversies in definitions of bacteria species (Robertson *et al.*, 2002; Vandamme *et al.*, 1997) and suggest that some issues are still unresolved. The polyphasic theory is an important contribution and deserves wider acceptance (Vandamme *et al.*, 1996).

8.8. Prevention – eradicate or conserve to prevent disease?

Overwhelming evidence has demonstrated that GBS can be harmful to humans and there is some evidence that ureaplasmas are also harmful in some circumstances, although less hostile than GBS (Razin *et al.*, 1998; Schuchat, 1998). Because of their conditional pathogenicity, the wisdom of attempting to eradicate colonization has been questioned because:

1) Only a very small proportion of GBS and human ureaplasma colonized persons will develop significant disease.
2) GBS and human ureaplasma strains are highly diverse at a genetic level and vary in virulence.
3) The antiquity of GBS and human ureaplasma infection in humans and their co-evolution suggests that both, but especially ureaplasmas, are commensal to humans. Their eradication may replace benign or even beneficial strains with more harmful ones and may provoke other problems.

It remains for carefully designed prospective studies, rather than hypotheses, to determine whether the potential risks of eradication of commensals, which are also

potential pathogens, outweigh the benefits (Hunt *et al*., 2001). Careful review of the literature confirms that GBS infection is a serious pathogen, albeit in a minority of those colonized. Intrapartum antibiotic prophylaxis (IAP) will be replaced by vaccines to prevent neonatal GBS infection (Paoletti & Madoff, 2002). Ideally, vaccines (or antibiotic prophylaxis) would be targeted at the small proportion of GBS carriers whose infants are most at risk.

For human ureaplasmas the issue is more difficult. There is no suitable bactericidal antibiotic and vaccination is unlikely to be appropriate because of the cross reactivity or mimicry between MBA and human antigens (although a DNA vaccine could theoretically bypass this problem) (Mabanta *et al*., 2003).

8.9. Conclusions.

Human ureaplasmas (*U. parvum* and *U. urealyticum*) (Razin *et al*., 1998) and group B streptococcus (GBS, *S. agalactiae*) (Schuchat, 1998) are both common potential perinatal pathogens. Studying them together provides more insight into their evolution and the pathogenesis of infection than studying them separately (van Belkum *et al*., 2001; Almogy *et al*., 2002).

Evolutionary and population genetics suggest that human ureaplasmas are "cleverer" than GBS (Razin *et al*., 1998) considering their ratio of "investment" (e.g. genome size) to "outcome" (e.g. different prevalence and bacterial population levels in pregnant women). During their coevolution with human beings, ureaplasmas apparently thrived in their "parasite-like" lifestyle and have finally reached a successful co-existence, with high population levels for their limited genomic resources. This required reduction in virulence – by loss of most of their capsule and mimicry by MBA repetitive units of host antigens – and becoming more parsiminonious and "friendly" – by using a host waste product (urea) as their energy source. This parasitic or symbiotic relationship contributes to controversy associated with study of human ureaplasma infection (Razin *et al*., 1998).

GBS have adopted another strategy. They have increased their virulence by retaining and acquiring new virulence molecules: capsule; surface protein family and IgA binding protein (C beta); mge and a variety of other specific virulence factors (Glaser ***et al.***, 2002). But considering their genome size and their population, they are not as wise (efficient or successful) as human ureaplasmas.