Integrated study of group B streptococcus and human ureaplasmas – the paradigm shifts

Fanrong Kong

A thesis submitted to the University of Sydney in fulfilment of the requirement for the degree of Doctor of Philosophy.

Department of Medicine University of Sydney

2004

Declaration

This thesis is less than 100, 000 words in length and contains no material that has been accepted for the award of any other degree or diploma in any other universities.

I declare that the author, with the following exception, conducted all the practical work incorporated into this thesis:

- All the New Zealand GBS strains were provided by Dr Diana Martin (ESR, New Zealand) and GBS conventional serotyping was mainly performed by Sonia Gowan (ESR, New Zealand). Fourteen selected isolates were serotyped by Abbie Weisner (PHLS, United Kingdom).
- Ureaplasma and mycoplasma reference strain culture was partially performed by Susannah Gordon (CIDM, Westmead Hospital).
- Gregory James (CIDM, Westmead Hospital) provided all the needed molecular equipment, reagents, softwares and some protocols.
- Zhenfang Ma assisted in some ureaplasma PCR and data analysis.
- Mark Wheeler (WMI, Westmead Hospital) performed sequencing.
- Professor Lyn Gilbert (supervisor) supervised the overall project. Assisted in research design, in the analysis and interpretation of data, and made a significant contribution to the project.

Preface

Someone told me “a PhD thesis is just like a **story**”. From my viewpoint, the **story** formed the major part of my past six-year life, and they will significantly affect my future research career and even my whole life. I tried to make the story something that would be interesting and useful to be read. On the other hand, I tried to cut its length to save the possible readers precious time.

As a PhD student, besides to **master** or to learn the techniques and methodologies as a **Master** student or a **technician**, to think about the **philosophy** or to explain the findings (including that of the others besides myself) would be also or even more important for the future related research fields. You would find the thesis in some extent (in particular, in section 3) reflected the above considerations.

From the teacher side, the selection of **good** teachers would be very important for a **good** student. I was so lucky to be a student of Professors Lyn Gilbert and Tania Sorrell. I also want to thank so many “informal teachers”, I may only know many of them from their excellent publications (I tried my best to include their contributions in my reference sections though I might still have lost many for cutting the volume reason), and these teachers are also highly appreciated. Molecular lab – I located myself there in my past six years, all the staffs there were so kind and the conditions there were so good – the experience there will be kept in my mind as a very memorable life in my personal history.

The two **tiny** microorganisms that I selected, as my study objectives – GBS and ureaplasmas – were also my very **great** teachers! They showed me that they also had some kind of “wisdom”, which I tried to expose partial in the thesis (especially in section 3) but far from enough. No doubt, their “wisdom” deserves to be further studied – especially based on their invaluable genome resources and benefit from the new bioinformatic concepts, theory and research tools.

The thesis contains three sections. The first section of general introduction provided the background and some useful concepts of the next two sections. The second section mainly contained those so called **"traditional"** molecular microbiology study of the two microorganisms. It included mainly the genotyping studies, which included most of my previous publications (also see appendices). My past six year publications (in PDF format) had been used as appendices and were put at the end of the thesis. Their inner relationship was also given at the beginning of the appendices. So only selected parts (after modification and reorganization) that were believed to be important to explain the outline of the **story** were put into the body of the thesis (mainly in section 2). The third section of the thesis contained some so called **"novel"** genomic and bioinformatics based study of the two organisms. Because we are facing the dramatic paradigm shift in their post genomic era, the ideas contained in the study may not be absolutely correct; even for the correct ones, some of them may still look a little bit **ugly**. But at least I would like to use them to show the potential values for doing this kind of **"novel"** study.

It was not easy to be a PhD student, especially considering the era of knowledge explosion, and the so many paradigm shifts. In this case, I really feel that what I had done was far from enough. However, to get the PhD or to pass the **milestone** is my long-term dream (even from my early childhood). So could I dare to use this thesis to have a **try**?

If the thesis can be seen as a **very tiny** drop of but **useful** water comparing with the endless sea of the **truth**, I will be more than happy**!**

Acknowledgments

The work described in this thesis was completed at:

Molecular Laboratory, Reference Laboratory Section, Centre for Infectious Diseases and Microbiology (CIDM), Institute of Clinical Pathology and Medical Research (ICPMR), Westmead Hospital, Westmead, NSW 2145, Australia. All the staffs in Molecular Laboratory and Reference Laboratory Section are highly appreciated.

The project was partially supported by Australian Postgraduate Awards 2000 (APA 2000) and Westmead Millennium Institute (WMI) Initiating Grant.

This thesis couldn't become true without:

Professor Lyn Gilbert, a great and very knowledgeable person, my highly respected supervisor – I was so lucky to be her student. Her excellent supervision, patient editing, and many other supports made my past six years productive and enjoyable. Professor Tania Sorrell, a very kind person, my associate supervisor – she always prepared to give her kind help to me when I needed.

I thank Mark Wheeler for his precious help in sequencing and Glenys Connor for some GBS data collection. Dr Hui Wang, in the third year of my PhD, while I was training him, gave me a great deal of help, was much appreciated.

I would like to thank all my family members; their continuous encouragements and supports made me feel full of energy. Zhenfang Ma, my deeply beloved wife, my another half, the most valuable **gift** I could get from my God – she gave me the greatest and invaluable supports, any words can't express my thanks to her – I **love** you! At the final stage of my PhD thesis writing, my lovely son (Alan, Xiangyu Kong) joined with us – welcome you my little honey and hope you can make my future **story** more exciting!

Abstract

Group B streptococcus (GBS, *S. agalactiae*) and human ureaplasmas (*U. parvum* and *U. urealyticum*) are two clinically and phylogenetically related, potential perinatal pathogens. Their relationships between genotypes and pathogenesis of GBS and ureaplasma infection were still not well understood, one of the reason is that both of them are still short of a very practical genotyping system. In the study, to solve the above problem we developed genotyping systems for the organisms (the second section). For human ureaplasmas, based on four genes/gene clusters (rRNA gene clusters, the elongation factor Tu genes, urease gene complexes and multiple banded antigen genes), we designed many primer pairs suitable for developing species identification assays for the two newly established human ureaplasma species (*U. parvum* and *U. urealyticum*). Further, based on the heterogeneity of ureaplasma multiple banded antigen gene (which contains species- and serovar-specific regions), we developed genotyping methods for each ureaplasma species. For GBS, based on three sets of molecular markers (capsular polysaccharide synthesis gene clusters, surface protein antigen genes and mobile genetic elements), we developed a genotyping system. The primary evaluation of the genotyping systems showed that the genotyping systems were practical alternative assays for the conventional serotyping and they will be useful to further explore the relationships between genotypes and pathogenesis of GBS and ureaplasma infection. In the study, we introduced novel data and tools into GBS and ureaplasma studies especially from genomic- and bioinformatics-based molecular microbiology

(the third section). For two newly established human ureaplasma species, based on the *U. parvum* serovar-3 genome, and using the above four important genes/gene clusters, we exposed some interesting problems in the understanding of new ureaplasma taxonomy especially in the post genomic era. For GBS, we studied the two published full genomes and exposed some new problems or possible future new research fields. In particular we found the two finished and one ongoing GBS genomes were all non-typical and suggest that future genomic project had better have genetic population structure viewpoint. Finally, we suggested that integrated studies of the two potential or conditional perinatal pathogens, from the viewpoint of evolution, would provide a new understanding angle of the pathogenesis of the two organisms. Studies suggested that during coevolution, human ureaplasmas (especially *U. parvum*) became friendlier than their ancestors to their human host (by losing most of its virulence genes); however, GBS tried to increase its invasive abilities (by getting more virulence genes) to fight against the human host attack.

Table of Contents

List of Figures

List of Tables

Abbreviations

ABI	Applied Biosystems
Alp2	C alpha-like 2 antigen
alp2	C alpha-like 2 antigen gene
Alp3	C alpha-like 3 antigen
alp3	C alpha-like 3 antigen gene
Alp4	C alpha-like 4 antigen
alp4	C alpha-like 4 antigen gene
Alp5	C alpha-like 4 antigen
alp5	C alpha-like 4 antigen gene
AmiC	Amidase family protein
ANGIS	Australian National Genomic Information Service
ATCC	American Type Culture Collection (Manassas, VA, USA)
Bac	C beta antigen or IgA binding protein
bac	C beta antigen or IgA binding protein gene
Bca	C alpha antigen
bca	C alpha antigen gene
BPS	group B protective surface protein
bp	base pair
BspA	A cell surface associated leucine-rich repeat protein involved in adhesion to fibronectin and fibrinogen
CAMP	(discovered by) Christie, Atkins, and Munch-Petersen
CDC	Centers for Disease Control and Prevention
CDS	Coding sequences
CGH	Comparative genome hybridization
CI	Confidence interval
CLD	Chronic lung disease (of prematurity)
Clp	Clp ATPase family of molecular chaperones
CNS	Central nervus system
CbpD	Choline binding protein D
CpdB	Cyclo-nucleotide phosphodiesterase
CPS	Capsular polysaccharide
cps	Capsular polysaccharide synthesis (gene cluster)
CS	Conventional serotyping/serotype
CSF	Cerebrospinal fluid
DNA	Deoxyribonucleic acid

dNTP	Deoxynucleotide triphosphate
EaeH	EaeH of *Escherichia coli* O157:H7
EF-TU	Elongation factor Tu
EOD	Early onset disease (infection occurring within the first week of life)
FbsA	A fibrinogen receptor from group B streptococcus
G+C	Guanine plus Cytosine
GBS	Group B streptococcus or group B streptococci or *S. agalactiae*
g	Gram
GIs	genomic islands
Human ureaplasmas	**–** *U. parvum* and *U. urealyticum*
h	Hour
Hsa	(antigen that recognition of) sialic acid-containing host receptors
IAP	Intrapartum antibiotic prophylaxis
Indels	Insertations and deletions
IS	Insertion sequence(s)
Lmb	Laminin-binding protein
lmb	Laminin-binding protein gene
LOD	Late onset disease (infection occurring between 1 week and 2 to 3 months of age)
LSU rRNA	Large submit rRNA
l	Liter/Litre
M	Molar
MBA	Multiple banded antigen
mba	Multiple banded antigen gene
mge	Mobile genetic element(s)
min	Minute
ml	Microlitre/Microliter
MLEE	Mutiple locus enzyme electrophoresis
MLST	Multiple locus sequence typing
MS	Molecular serotype(s)
NanA	Sialic acid lyase (catalyzes the hydrolysis of sialic acid into pyruvate and N-acetylmannosamine)
NCTC	National Culture Type Collection (Colindale, UK)
No.	Number(s)
NGU	Non-gonococcal urethritis
nra	Encode a response regulator (no response to atmospheric conditions)
OR	Odds ratio
ORF	Open reading frame(s)

Pas The surface protein antigen I/II of *Streptococcus intermedius*

PCR Polymerase chain reaction

PFBP *Streptococcus pyogenes* fibronectin-binding protein

PFGE Pulsed-field gel electrophoresis

pgp Protein gene profile(s)

PIs Pathogenicity islands

PrtS Serine proteinase, subtilase family

PspC Pneumococcal surface protein C

PulA Alkaline amylopullulanase

R28 *Streptococcus pyogenes* surface protein R28

R5 (or BPS protein) – group B protective surface protein

RAPD Random amplified polymorphic DNA

RFLP Restriction length fragment polymorphisms

rgg Encode a response regulator

Rib Rib antigen

rib Rib antigen gene

RNA Ribonucleic acid

rof Encode a response regulator

s Second

ScpB Serine protease and C5a peptidase

Sec10 Surface exclusion protein

SpaA Streptococcal protein antigen A of *Streptococcus sobrinus*

Sip Surface immunogenic protein

SrpA Periplasmic linker protein

Srt Sortase

SSH Suppressive subtractive hybridization

Ssp5 Agglutinin receptor

sst Serosubtype(s)

SSU rRNA Small subunit rRNA

***T*a** Annealing temperature

***T*m** Melting temperature

tuf Elongation factor Tu gene

U Unit

UAB University of Alabama at Birmingham

UP *Ureaplasma parvum*

Ureaplasmas – *U. parvum* and *U. urealyticum*

UU *Ureaplasma urealyticum*

Ureaplasma broth – 10B broth (Shepard, 1970)

WebANGIS – Website of Australian National Genomic Information Service (**http://www1.angis.org.au/pbin/WebANGIS/wrapper.pl**)

YbgE Putative branched-chain aminotransferase

YfkN 2`,3`-cyclic-nucleotide 2`-phosphodiesterase