ABSTRACT

Title of Document:	MOLECULAR AND BIOINFORMATICS APPROACHES TO REDEFINE OUR UNDERSTANDING OF UREAPLASMAS: MOVING BEYOND SEROVARS
	Vanya Paralanov, Doctor of Philosophy, 2014
Directed By:	Prof. Jonathan Dinman, Cell Biology and Molecular Genetics, University of Maryland College Park Prof. John I. Glass, Synthetic Biology, J. Craig Venter Institute

Ureaplasma parvum and *Ureaplasma urealyticum* are sexually transmitted, opportunistic pathogens of the human urogenital tract. There are 14 known serovars of the two species. For decades, it has been postulated that virulence is related to serotype specificity. Understanding of the role of ureaplasmas in human diseases has been thwarted due to two major barriers: (1) lack of suitable diagnostic tests and (2) lack of genetic manipulation tools for the creation of mutants to study the role of potential pathogenicity factors.

To address the first barrier we developed real-time quantitative PCRs (RT-qPCR) for the reliable differentiation of the two species and 14 serovars. We typed 1,061 ureaplasma clinical isolates and observed about 40% of isolates to be genetic mosaics, arising from the recombination of multiple serovars. Furthermore, comparative genome analysis of the 14 serovars and 5 clinical isolates showed that the *mba* gene, used for serotyping ureaplasmas was part of a large, phase variable gene system, and some serovars shown to express different MBA proteins also encode *mba* genes associated with other serovars. Together these data suggests that differential pathogenicity and clinical outcome of an ureaplasmal infection is most likely due to the presence or absence of potential pathogenicity factors in individual ureaplasma clinical isolates and/or patient to patient differences in terms of autoimmunity and microbiome.

To address the second barrier we are adapting the traditional molecular biology and novel synthetic biology tools to Ureaplasma, such as creation of oriC plasmids, use of transposons, and most prominently the engineering bacterial genomes cloned as yeast centromeric plasmids followed by genome transplantation to make ureaplasma mutants programmed by the genomes manipulated in yeast. This will allow for the creation of targeted single or multiple mutants that will greatly increase the understanding of ureaplasma pathogenicity. Efforts to transplant the genomes of bacteria, outside the *mycoides* group have been thwarted due to recombination between the donor and recipient cell genomes. We are exploring the use of the DNA cross-linking drug, Mitomycin C, to inactivate the recipient cell genomes and thus prevent false positive transplantation results and potentially increase the genome transplantation efficiency.

.

MOLECULAR AND BIOINFORMATICS APPROACHES TO REDEFINE OUR UNDERSTANDING OF UREAPLASMAS: MOVING BEYOND SEROVARS

By

Vanya Metodieva Paralanov

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2014

Advisory Committee:
Professor Jonathan Dinman, Chair
Professor John I. Glass, Co-Chair
Professor Daniel Stein
Professor Philip Bryan
Professor Rose Viscardi
Dr. Elisabet Caler

Dedication

To all the little ones born too soon and fighting for their lives, all the little angels gone too soon and their families!

"Sometimes the smallest things take up the most room in your heart!"

-Winnie the Pooh

Just For *By Unknown Author*

Just for this morning, I am going to step over the laundry,

And pick you up and take you to the park to play.

Just for this morning, I will leave the dishes in the sink,

And let you teach me how to put that puzzle.

Just for this afternoon, I will unplug the telephone and keep the computer off,

And sit with you in the backyard and blow bubbles.

Just for this afternoon, I will not yell once, not even a tiny grumble when

You scream and whine for the ice cream truck, and I will buy you one if he comes by.

Just for this afternoon, I won't worry about what you are going to be

When you grow up, or second guess every decision I have made

Where you are concerned.

Just for this afternoon, I will let you help me bake cookies,

And I won't stand over you trying to fix them.

Just for this afternoon, I will take us to McDonald's

And buy us both a Happy Meal so you can have both toys.

Just for this evening, I will hold you in my arms and tell you a story

About how you were born and how much I love you.

Just for this evening, I will let you splash in the tub and not get angry.

Just for this evening, I will let you stay up late

While we sit on the porch and count all the stars.

Just for this evening, I will snuggle beside you for hours,

And miss my favorite TV shows.

Just for this evening when I run my finger through your hair as you pray,

I will simply be grateful that God has given me the greatest gift ever given.

I will think about the mothers and fathers,

Who are searching for their missing children,

The mothers and fathers who are visiting

Their children's graves instead of their bedrooms,

And mothers and fathers who are in hospital rooms

Watching their children suffer senselessly,

And screaming inside that they can't handle it anymore.

And when I kiss you goodnight I will hold you a little tighter, a little longer.

It is then, that I will thank God for you and ask him for nothing,

Except one more day.....

Acknowledgements

I have come so far with the encouragement, support, and mentorship of so many important people that it only makes sense to acknowledge everyone chronologically! My dear grandmother, who I call Mami, was always sure and used to tell me: “One day you will be a Doctor of the sciences!” This is a word for word translation from what she used to often tell me in Bulgarian when I was still in elementary school and all throughout my school and university career. I am so grateful that she never doubted my abilities and her voice has always been in the back of my head when I needed a dose of encouragement!

My parents, Rossitza and Metodi, who are both physicist, instilled in me curiosity and love of science and research. They nurtured my science interests and followed my lead! The most fun part for me was visiting often their laboratory and exploring with them. They have been very supportive and selflessly provided academic, moral, and financial support, including countless days of childcare for my sweet son, Christopher! I will be eternally grateful for the sacrifices they have made and continue to make! Another person, who has also always believed in me and has been proud of me, is my dear sister, Radostina! We have been through a lot together, helping each other. I am eternally grateful for all the sacrifices she has made to help me and our parents, and take care of our grandparents!

I am proud to say that I have known my best friend, who is also my wonderful husband, for 15 years or half of my life! I honestly don’t remember and don’t want to know what life is like without him! He has been extremely supportive throughout my student life, even when it seemed it will never end and he was getting a bit inpatient!

He has always been there for me, believing in me, loving, reminding me that I can do it, and telling me how proud of me he is! He was there to help me through countless sleepless nights, making sure that I don't forget to eat, don't drink too much caffeine, and get some rest and sleep as soon as possible. Thank you for providing for us all the necessities, love and much more, so I can concentrate on my school and research! But most of all thank you for Christopher and being a wonderful, loving and engaged father! We will always love you unconditionally!

I would also like to acknowledge the support, advice, and help from my husband's family, Mariana, Ljuben, Blagovest, and Grandma Sava or my second mom, dad, brother and grandmother, who have taken interest in my research and were proud of my every academic milestone and achievement!

My passion for science, and in particular biology, was reignited when I came to the USA and met my college professors. I was amazed and inspired by their dedication, passion, and enthusiasm for research and teaching. They were approachable by their students and eager to answer student's questions! These were all qualities that my University professors in Europe, unfortunately, failed to demonstrate. I would like to say that I am so happy and grateful I met Dr. Thomas Burkett, my Biotechnology degree advisor and professor. I owe a lot of my success today and how far I have gone to his enthusiasm and passion for science along with his guidance, advising and encouragement. Dr. Burkett introduced me to the Molecular Biology, Biochemistry and Bioinformatics (MB3) Program at Towson University and informed me about and strongly recommended that I apply to the summer internship program at the J. Craig Venter Institute.

After I graduated the Biotechnology Program, I transferred to the MB3 Program at Towson University, where I continued to meet wonderful professors, inspiring me to continue my University career. My program advisor and course professor, Dr. James Saunders was extremely supportive. His office door was always open for the students in the MB3 program and other students. He made sure that we all felt like one big family! Dr. Saunders fostered my leadership skills and further encouraged me to pursue a doctoral degree. I greatly appreciate our conversations and discussions, his optimism, care, and encouragement!

Dr. Elisabet Caler and Dr. John Glass, who I met at the J. Craig Venter Institute, became an integral part of my academic career and I like to think of them as my scientific parents! Lis provided me with a bioinformatics internship position and was my mentor along with Hernan Lorenzi. I learned a lot from them and they provided a lot of help and advice during my graduate career. I appreciate tremendously Lis not only as a scientific advisor but also as a friend!

Most importantly, I would like to extend endless thanks Dr. John Glass, my dissertation advisor, who adopted me in his synthetic biology team and provided me with a project that turned into my doctoral dissertation. I am eternally grateful that he turned my qualifying exam idea into a real grant and never stopped believing in me (even through the long periods of negative results)! He is the best advisor and supervisor I could have ever asked for my graduate career! For the total of 7 years I have worked with him as research assistant and later as a graduate student, he has helped me grow as a scientist and as a person. No words exist that are capable of describing how grateful I am for his understanding, patience, encouragement, support

and sometimes even lending a shoulder to cry on! I appreciate that he was always there for me, always caring, through the emotional rollercoaster my personal life has been: from very sick, dear, family members to welcoming our son into the world!

I would like to also thank my dissertation committee members Dr. Jonathan Dinman, Dr. Rose Viscardi, Dr. Philip Bryan, and Dr. Daniel Stein for their time, support, critical advice, guidance and enthusiasm! I would also like to thank them for being patient and understanding during the difficult times I have had!

I think it is safe to say that it takes a village of scientists to raise a graduate student and my home village will always be the JCVI synthetic biology group! Everyone in the group has contributed to shaping my scientific thinking! I would like to thank for the extensive scientific and personal support of some members of the group who I think of not only as colleagues, but also as friends: Michael Montague, Nacyra Assad-Garcia, Suchismita Chandran, Sanjay Vashee, Radha Krishnakumar and Michael Algire. Last but not least, I would like to thank Olivia Chao, who has worked with me as an intern for over a year and a half. She is incredibly smart and motivated and I respect her as a colleague and a friend! I have no doubt that she will succeed in anything she begins! It was my pleasure to work with her and I appreciate all her hard work and perseverance through countless experiments! I am looking forward to seeing all the great things she will do!

Table of Contents

Dedication ii

Table of Contents viii

List of Tables xii

List of Figures xiii

List of Abbreviations xiv

Chapter 1. Introduction 1
- 1.1 Microbiology of Ureaplasma species 1
- 1.2 Epidemiology 3
- 1.3 Clinical Importance 3
  - 1.3.1 Ureaplasma and Adverse Pregnancy Outcomes 4
  - 1.3.2 Infertility and other genitourinary tract diseases 9
  - 1.3.3 Ureaplasmal extragenital diseases 10
- 1.4 Diagnostic Methods 10
  - 1.4.1 Conventional techniques 10
  - 1.4.2 Molecular techniques 12
  - 1.4.3 Choosing culture versus molecular-based tests for UU detection 21
- 1.5 Pathogenicity: Reasons for differential clinical outcome 23
  - 1.5.1 Serovars 23
  - 1.5.2 Host immune response 24
  - 1.5.3 Potential Virulence Factors of Ureaplasma Species 25
- 1.6 Antimicrobial Susceptibility of Ureaplasma 28
- 1.7 Major Barriers to Understanding Pathogenicity 30

Chapter 2. Comparative Genome Analysis of *19 Ureaplasma urealyticum* and *Ureaplasma parvum* strains 33
- 2.1 Abstract 33
- 2.2 Introduction 34
- 2.3 Materials and Methods 35
  - 2.3.1 Sequencing methods for ATCC and 4 clinical isolates 35
  - 2.3.2 Annotation 35
  - 2.3.3 Comparative genomics 38
  - 2.3.4 MBA locus 38
  - 2.3.5 PLC, PLA, and IgA protease genes 39
  - 2.3.6 PLC assay 40
  - 2.3.7 Phylogenetic trees 41
- 2.4 Results and Discussion 42

2.4.1 Genome sequencing of 19 *U. urealyticum* and *U. parvum* strains 42
2.4.2 Genome Analysis of UU clinical isolates that could not be genotyped to the serovar level 44
2.4.3 Gene content analysis of all 19 Ureaplasma genomes 45
2.4.4 Ureaplasma phylogenetic tree 54
2.4.5 Recombination and Integration of DNA 62
2.4.6 Potential pathogenicity genes 65
2.5 Conclusions 83

Chapter 3. Detection and Characterization of Human Ureaplasma Species and Serovars by Real-time PCR 88
3.1 Abstract 88
3.2 Introduction 89
3.3 Material and Methods 92
3.3.1 Bacterial strains and clinical specimens. 92
3.3.2 DNA preparation for PCR. 94
3.3.3 PCR Primers and Probes. 95
3.3.4 PCR conditions. 99
3.4 Results 102
3.4.1 Analytical specificity and sensitivity of Ureaplasma species-specific multiplex PCR. 102
3.4.2 Comparison of species-specific multiplex PCR with traditional PCR and culture. 104
3.4.3 Analytical specificity and sensitivity of serovar-specific real-time PCR assays. 105
3.5 Discussion 107

Chapter 4. Extensive horizontal gene transfer in UU questions the utility of serotyping for diagnostic purposes 114
4.1 Abstract 114
4.2 Introduction 114
4.3 Materials and Methods 119
4.3.1 Bacterial isolates 119
4.3.2 DNA preparation 121
4.3.3 Genotyping of clinical isolates by PCR 121
4.3.4 Separation of isolates containing multiple serovars 121
4.3.5 Quantification of each serovar in clinical isolates to distinguish mixtures from hybrids 122
4.3.6 DNA Sequencing 123
4.3.7 Statistical Analysis 124
4.4 Results and Discussion 124
4.4.1 Species and serovar distributions of isolates from different patient populations 124
4.4.2 Isolates that were negative for all serovar-specific assays 127
4.4.3 Isolates containing multiple serovars and conflicting serovars 127

4.4.4 Separation of isolates with multiple serovars 129
4.4.5 Verifying hybrid isolates by sequencing multiple gene-coding, polymorphic loci 129
4.4.6 Prevalence of hybrid ureaplasmas 134
4.5 Conclusions 137
Supporting materials 142

Chapter 5. Expanding the Genetic Toolbox for Ureaplasmas 143
5.1 Abstract 143
5.2 Introduction 143
5.3 Materials and Methods 147
5.3.1 Ureaplasma cultures 147
5.3.2 Extracting intact genomic DNA 149
5.3.3 Restriction Digest and CHEF Gel Analysis of gDNA Plugs 150
5.3.4 Minimum Inhibitory Concentration Assay: 150
5.3.5 Testing for Nucleases Present on the cell surface of UU cells 151
5.3.6 Preparation of Crude Cellular Extracts 152
5.3.7 Crude Cellular Extract Restriction Enzyme Assay 152
5.3.8 Transformation/PEG Concentration Optimization: 152
5.3.9 Plasmids Used and Their Construction 154
5.4 Results and Discussion 158
5.4.1 Development of traditional tools for genetic manipulation of UU genomes 158
5.4.2 Testing UU cells for the presence of surface and restriction nucleases 160
5.4.3 Minimum Inhibitory Concentration (MIC) Assay: 162
5.4.4 Development of a Transformation Protocol for Ureaplasma species 162
5.4.5 Isolation and analysis of intact genomic DNA 169
5.5 Conclusions 170
5.6 Future Directions 174
5.6.1 Adapt synthetic biology tools to create desired Ureaplasma mutants 174
5.6.2 Transplantation of UUR9 into UPA or *U. diversum* recipient 176
5.6.3 Clone the tetracycline resistant UUR9 genome in yeast 180
5.6.4 Delete a gene of interest and transplant the deletion mutant to a recipient cell 183
5.6.5 Suggested modifications to the ureaplasma genome 185

Chapter 6. Selecting against Recombination Events during Genome Transplantation by Crosslinking the Genome of the Recipient Cell with Mitomycin C .. 188
6.1 Introduction 188
6.1.1 Genome transplantation 188
6.1.2 Mitomycin C 190
6.2 Materials and Methods 193

6.2.1 MMC Solutions and Storage 193
6.2.2 Minimum Inhibitory Concentration 193
6.2.3 Determining the MMC concentration and conditions to inactivate a grown culture 194
6.2.4 Preparation of intact circular genomes in agarose plugs 195
6.2.5 Digests of genomes treated with MMC 196
6.2.6 Genome Transplantation 196
6.3 Results and Discussion 198
6.3.1 Susceptibility of different mycoplasmas to MMC 198
6.3.2 Inactivation of the recipient cells genomic DNA with MMC 200
6.3.3 Transplanting genomes without antibiotic selection markers 208
6.4 Discussion 212

Chapter 7. Conclusions and Future Directions 215
7.1 Ureaplasma clinical importance 215
7.2 Development of diagnostic tools 215
7.3 What is a serovar? 216
7.4 Changing how we understand serovars 217
7.5 Shift in literature to host-pathogen studies with animal models and cell lines 221
7.6 Need to develop genome manipulation tools to support the understanding of pathogenicity and increase the power of using animal models. 223

Bibliography 229

List of Tables

Table 1-1. Antimicrobial Susceptibilities of Ureaplasmas 29

Table 2-1. Overview of UUR and UPA genomes..... 43
Table 2-2. Anticodon Table of tRNAs Showing count of tRNAs used by Human Ureaplasmas. 46
Table 2-3. Role Category Breakdown of Genes 47
Table 2-4. Pan genome of different *Ureaplasma* species sets 48
Table 2-5. Serovar to Serovar Difference Based on Whole Genome Comparison on the Nucleotide Level Expressed in Percent 50
Table 2-6. Number of Clusters of Orthologous Genes (COGs) per functional category present only in UUR or UPA serovars. 53
Table 2-7. Proteases Conserved in All 14 ATCC Serovars. 69
Table 2-8. Clusters of Orthologous Genes Potentially Involved in the MBA Phase Variable System of Ureaplasmas. 74
Table 2-9. Tandem Repeating Units (TRUs) Identifies in the *mba* locus. 75

Table 3-1. Bacterial species used to validate real-time PCR assays..... 93
Table 3-2. Primers and probes for UU species-specific multiplex PCR assays 95
Table 3-3. Primers and probes for UU serovar-specific monoplex PCR assays 97
Table 3-4 Master mix components of serovar-specific monoplex PCR assay 100
Table 3-5. Amplification program of monoplex serovar-specific PCR assays 101

Table 4-1. Description of *Ureaplasma* clinical isolates..... 120
Table 4-2. Species and serovar distributions of clinical isolates 126
Table 4-3. DNA sequencing results of polymorphic loci, *mba* gene, and serovar marker regions 133
Table 4-4. Types of *Ureaplasma* hybrids 136

Table 5-1. Distribution of UU Clinical Isolates..... 145

Table 6-1. Susceptibility of Different Mycoplasmas and Ureaplasmas to MMC..... 198
Table 6-2. Efficiency of Genome Transplantation with MMC Treated Cells 202
Table 6-3. Efficiency of Genome Transplantation with MMC Treated Cells under different conditions 203

Table S 1. PCR primers for sequencing of polymorphic loci between two serovars. 142

List of Figures

Figure 2-1. Percent GC Distribution Among Genes of The Ureaplasma Pan Genome (19 Strains). 49
Figure 2-2. Dotplot of the genomes of UUR4 and UUR12. 51
Figure 2-3. Phylogenetic tree based on the gene content of the pan genome of ureaplasmas. 55
Figure 2-4. MBA Based Phylogenetic Tree of 19 Ureaplasmas. 56
Figure 2-5. Phylogenetic tree of 19 ureaplasma strains based on 82 housekeeping genes. 58
Figure 2-6. Phylogenetic trees based on the nucleotide sequence of multiple concatenated genes: 59
Figure 2-7. Phylogenetic trees based on the combination of two of the four groups of concatenated genes. 60
Figure 2-8. Phylogenetic trees based on the combination of tree of the four groups of concatenated genes. 61
Figure 2-9. *Ureaplasma parvum* Multiple Banded Antigen Locus. 77
Figure 2-10.*Ureaplasma urealyticum* Multiple Banded Antigen Locus. 78
Figure 2-11.The MBA Locus in UUR4, UUR12, and UUR13. 80

Figure 3-1. Ureaplasma species-specific multiplex real-time PCR assays 103
Figure 3-2. Monoplex serovar-specific PCR assays 106

Figure 4-1. Clinical isolates containing multiple serovars and untypable serovars.. 128
Figure 4-2. DNA sequencing determining *Ureaplasma* hybrid or mixed isolates. .. 131
Figure 4-3. Prevalence of hybrid serovars in clinical isolates 135

Figure 5-1. Distribution of 271 Mixed Ureaplasma Clinical Isolates. 145
Figure 5-2. *Ureaplasma* colonies on 10B Agar. 148
Figure 5-3. pMOD-2 custom EZ-tn5 plasmid. 155
Figure 5-4. PCR for the presence of the Tn5_UUR9*tuf*_TetM after PEG transformation reaction. 167
Figure 5-5. Inter- and Intra-species Distribution of 75 Chimeric Ureaplasma Isolates 177
Figure 5-6. Moving a bacterial genome into yeast, engineering it, and transplanting it back into a bacterium. 181
Figure 5-7. TREC method for seamless deletions of target genes in genomes cloned and maintained in yeast 184

Figure 6-1. Adducts of MMC and DNA formed by reductive activation. 191
Figure 6-2. Restriction Digest of gDNA from MMC Treated Cells. 206
Figure 6-3. Transplanting Genomes Extracted from a Tetracycline Resistant *M. mycoides* strain without tetracycline selection. 209
Figure 6-4. Transplanting a Tetracycline Resistant *M. mycoides* Strain Genome Extracted from Yeast without tetracycline selection. 210

List of Abbreviations

5-FOA -5-fluoroorotic acid

A - Adenine

ABI – Applied Biosystems

AMV-RT - Avian Myeloblastosis Virus Reverse Transcriptase

ATCC – American Type Culture Collection

BLAST - Basic Local Alignment Search Tool

BLASTn - Basic Local Alignment Search Tool nucleotides

BLASTp - Basic Local Alignment Search Tool protein (amino acids)

bp –base pair

BPD – Bronchopulmonary Dysplasia

BV – Bacterial Vaginosis

C – Cytosine

CCU – Color Changing Unit

CD – Conserved Domain

CD43, CD44, CD34, CD45 - Cluster of Differentiation (CD), used as cell markers in immunophenotyping

cDNA - complementary DNA (DNA synthesized from mRNA template)

CFU – Colony Forming Unit

CHEF - Clamped Homogeneous Electric Fields

CHP – Conserved Hypothetical Protein

COG – Cluster of Orthologous Genes

COX-2 – Cyclooxygenase-2

Cp – Cross point

CSF – Cerebrospinal Fluid

Ct –Cycle threshold

DMSO - Dimethyl Sulfoxide

DNA – Deoxyribonucleic Acid

dNTPs – Deoxynucleotide

DTT – Dithiothreitol

EC - Enzyme Commission

EDTA - Ethylenediaminetetraacetic acid

ELC –Electrochemiluminescence

ELISA - Enzyme-Linked Immunosorbent Assay

FRET – Fluorescence Resonance Energy Transfer

G – Guanine

GA – Gibson Assembly

gDNA – Genomic Deoxyribonucleic Acid

GFP – Green Fluorescent Protein

GO – Gene Ontology

GT –Genome Transplantation

HGT – Horizontal Gene Transfer

HMM – Hidden Markov Model

HRM PCR –High Resolution Melt Polymerase Chain Reaction

IgA – Immunoglobulin A

IL- Interleukin

IRDB – Inverted Repeat Database

ISH – *in-situ* hybridization

IVH - Intraventricular Hemorrhage

JCVI – J. Craig Venter Institute

Kbp – Kilo-base pair (1’000 base pairs)

LAMP – Loop Mediated Isothermal Amplification

LB –Luria Broth

LCR – Ligase Chain Reaction

LPS -Lipopolysaccharide

MAbs – Monoclonal Antibodies

MBA – Multiple Banded Antigen

MBB –Mycoplasma Broth Base

Mbp – Mega-base pair (1’000’000 base pairs)

MGAT –Multi Genome Annotation Tool

MIC – Minimal Inhibitory Concentration

Mim- Macrophage Infection Mutant

MMC –Mitomycin C

Mmc –*Mycoplasma mycoides* subsp. *capri*

mPCR/RLB – multiplex Polymerase Chain Reaction combined with Reverse Line hybridization Blotting

mRNA – messenger Ribonucleic Acid

MSA – Multiple Sequence Alignment

NABD – Nucleic Acid Based Detection

NADPH - Nicotinamide Adenine Dinucleotide Phosphate reduced

NANA – N-Acetylneuraminic Acid

NASBA – Nucleic Acid Sequence-Based Amplification

NEB – New England Biolabs

NEC - Necrotizing Enterocolitis

NF-κB - nuclear factor kappa-light-chain-enhancer of activated B cells

NGU – Nongonococcal Urethritis

NICU – Neonatal Intensive Care Unit

Nt - Nucleotide

ORF – Open Reading Frame

oriC - origin of replication chromosomal

PBS - Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

PEG - Polyethylene Glycol

PFGE – Pulse Field Gel Electrophoresis

PID – Pelvic Inflammatory Disease

PLA – Phospholipase A

PLC – Phospholipase C

PLD – Phospholipase D

PMSF - Phenylmethanesulfonyl Fluoride

p-NPPC – p-Nitrophenylphosphorylcholine

PPLO –Pleuro Pneumonia-Like Organisms

qPCR – quantitative Polymerase Chain Reaction

RM – Restriction Modification

RMS – Restriction Modification System

RNA – Ribonucleic Acid

rRNA – ribosomal Ribonucleic Acid

SDA – Strand Displacement Amplification

SP-A – Surfactant Protein A

SPC – Simple Probe Chemistry

SSCP – Single Stranded Conformation Polymorphism

STD – Sexually Transmitted Disease

T – Thymine

TAR – Transformation Associated Recombination

TBE - Tris-Borate-EDTA

TE – Tris -EDTA

Tet - Tetracycline

TGFβ1 – Transforming Growth Factor beta 1

TIGR – The Institute of Genomic Research

TLR – Toll Like Receptor

TMHMM – Transmembrane domain Hidden Markov Model

tmRNA - transfer-messenger Ribonucleic Acid

TNFα - Tumor Necrosis Factor Alpha

TR- Tandem Repeat

TRDB – Tandem Repeat Database

TREC – Tandem Repeat coupled with Endonuclease Cleavage

TRIS - tris(hydroxymethyl)aminomethane

tRNA – transfer Ribonucleic Acid

TRU – Tandem Repeating Unit

TSB – Tryptic Soy Broth

tuf – transcription elongation factor

UAB – University of Alabama at Birmingham

UNG – Uracil-DNA-Glycosylase

UPA – *Ureaplasma parvum*

UPA3 - *Ureaplasma parvum* serovar 3

ure – the urease genes

UU – Ureaplasmas (including *Ureaplasma urealyticum* and *Ureaplasma parvum*)

UUR – *Ureaplasma urealyticum*

UUR2 - *Ureaplasma urealyticum* serovar 2

YAC – Yeast Artificial Chromosome

YCp – Yeast Centromere Plasmid

Chapter 1. Introduction

1.1 Microbiology of Ureaplasma species

Ureaplasmas are among the smallest self-replicating organisms capable of axenic growth. Shepard provided the first description of ureaplasmas in 1954, initially known as "T-strain" mycoplasmas, when he cultivated them in vitro from the urethras of men with nongonococcal urethritis (NGU) [1]. The genus Ureaplasma, was established in 1974 [2], comprising those members of the family *Mycoplasmataceae* in the class *Mollicutes* that hydrolyse urea and use it as a metabolic substrate for generation of ATP. There are five species in the class *Mollicutes* that are human pathogens. The best known is *Mycoplasma pneumoniae*, which is a respiratory pathogen that is an agent of "walking pneumonia." The other four, *Mycoplasma genitalium*, *Ureaplasma parvum* (UPA), *Ureaplasma urealyticum* (UUR), and *Mycoplasma hominis* are all urogenital pathogens. This genus *Ureaplasma* currently has seven recognized species that have been isolated from humans and various animals: *U. canigenitalium* (dogs), *U. cati* (cats), *U. diversum* (cattle), *U. felinum* (cats), *U. gallorale* (chickens), *U. parvum* (UPA) (humans), and *U. urealyticum* (UUR) (humans). Numerous other ureaplasmas of animal origin have been described, but they have not been given species designations [3]. Human ureaplasmas were considered to belong to a single species, *U. urealyticum*. Patient isolates and corresponding patient sera were collected and each patient isolate was reacted with each individual serum sample. These identified 14 groups of reactions, therefore the patient isolates were classified into 14 serovar groups [4-6]. Therefore, UUR

comprised of 14 serovars distributed between two genetically distinct biovars until 2002 when the two biovars were reclassified as two distinct species, UPA and UUR, based on the sequences of 16S rRNA gene, the 16S-23s-rRNA intergenic region, the urease gene and DNA-DNA hybridization experiments [7]. UUR comprises 10 serovars – UUR2, UUR4, UUR5, UUR7-13, and UPA includes 4 serovars – UPA1, UPA3, UPA6, UPA14 [7].

Genomes of UPA and UUR are very small (760-1140 kilobase (kb) pairs), have an extremely low GC content of 25.5% and utilize a non-standard genetic code [8, 9]. Their limited biosynthetic capacity is responsible for many of the biological characteristics and fastidious growth requirements of human *Ureaplasma* spp. It is believed that *Ureaplasma* spp. have evolved from Gram-positive bacteria by degenerative evolution, which included losing the peptidoglycan cell wall [4]. Lack of a rigid cell wall prevents them from staining by Gram stain, confers pleomorphic cell structure, and makes them very susceptible to dehydration, thereby limiting them to a parasitic existence in association with eukaryotic cells of their host. Ureaplasmas are coccoid cells approximately 0.2 - 0.3 µm in diameter with a generation time of about one hour. The organisms can be cultivated overnight in liquid media such as Shepard's 10 B broth containing a mycoplasma broth base supplemented with 20% horse serum, 10% urea, DNA, yeast extract, and L-cysteine [10]. They will produce granular brown urease-positive colonies approximately 15 -60 µm in diameter on A8 agar after 24-48 hours of incubation at 37°C and 5% CO_2 due to calcium chloride incorporated into the media [10].

1.2 Epidemiology

As many as 40% to 80% of healthy adult women may harbor ureaplasmas in their cervix or vagina and they are readily transmitted venereally as well as vertically; with a transmission rate to infants born to colonized mothers as high as 90% [11]. As more and more studies of ureaplasma prevalence in different countries are being performed, it is evident that prevalence of ureaplasma colonization varies based on sexual activity status, geographical location and ethnicity, socioeconomic status, and hormonal changes (menopause and pregnancy) [11-20]. Iwasaka et al report *Ureaplasma* spp. can be detected in the vaginal flora of 25% post-menopausal women, 67% of sexually active women, and 40% of sexually inactive women [14, 21, 22]. Ureaplasma colonization during pregnancy varies from 35% to as high as 90% [14, 23, 24]. Their occurrence is somewhat less in the lower urogenital tract of healthy men (approximately 20-29%) [25, 26]. *U. parvum* is considerably more common than *U. urealyticum* as a colonizer of the male and female urogenital tracts and in the neonatal respiratory tract [11].

1.3 Clinical Importance

Ureaplasmas are common opportunistic pathogens, as they don't cause disease in some healthy individuals, however they are implicated in a variety of clinical outcomes including but not limited to: non-gonococcal urethritis, pelvic inflammatory disease, infertility, adverse pregnancy outcomes, chorioamnionitis, and bronchopulmonary dysplasia in neonates [11]. Ureaplasmas reside primarily on the mucosal surfaces of the urogenital tracts of adults or the respiratory tracts in infants.

There are no reports of intracellular ureaplasmas. They are capable of attaching to a variety of cell types such as urethral epithelial cells, spermatozoa, and erythrocytes [12]. Non-pregnancy related ureaplasma infections are usually easily treated with administration of proper antibiotics (discussed later). The greatest burden of ureaplasma infections is from adverse pregnancy outcomes and neonatal sequelae. Pregnant women colonized with ureaplasma are more likely to deliver preterm infants [27-29], requiring prolonged stay in hospital neonatal intensive care units (NICU). If a preterm infant is diagnosed with respiratory ureaplasma infection, the chance of it developing bronchopulmonary dysplasia is significantly elevated [30-34], requiring further medical care and could have long-term health implications [35].

1.3.1 Ureaplasma and Adverse Pregnancy Outcomes

Ureaplasma spp. are the micro-organisms most frequently isolated from amniotic fluid, and placentae in women who deliver prematurely, either with preterm premature rupture of membranes or in preterm labor with intact membranes and from the cord blood, respiratory tract and cerebrospinal fluid of infants born prematurely who develop bronchopulmonary dysplasia and developmental disabilities [11, 22, 36]. These organisms have been isolated in the amniotic fluid as early as the 16th week, and can result in a clinically silent chronic and progressive infection where delivery does not occur for several weeks [37]. Detection of *Ureaplasma* sp. by the PCR assay in second-trimester amniotic fluid from asymptomatic women was highly correlated with subsequent preterm labor and preterm delivery [27]. Isolation of *Ureaplasma* spp. from amniotic fluid or placenta has been consistently associated

with histologic chorioamnionitis and this relationship is inversely related to birth weight [11]. There are more than 50 studies on the relationship of ureaplasma colonization and adverse pregnancy outcomes, including preterm labor, premature preterm rupture of membranes, and chorioamnionitis. The essence of these studies has been summarised in a review by Capoccia [14]. A direct causal relationship of these organisms with adverse pregnancy outcomes has been elusive due to the polymicrobial nature of cervicovaginal colonization and most intra-amniotic infections [38] and study design. However, there is growing experimental evidence of the capacity of ureaplasmas to cause adverse pregnancy outcomes using different animal models and *in vitro* ureaplasma infections. [27, 36, 39]. The strongest evidence is the induction of chorioamnionitis, fetal inflammation and preterm labor in the Rhesus macaques after intra-amniotic injection of UPA [25]..

1.3.1.1 Contribution of *Ureaplasma* species to development of BPD

BPD is defined as oxygen dependence at 28 days or at 36 weeks postmenstrual age [35]. Sung *et al* observed that ureaplasma respiratory colonization was inversely related to gestational age. Sixty-five per cent of infants delivered before 26 weeks gestation were culture or PCR-positive compared to 31% of infants delivered at or after 26 weeks gestational age. Their findings also show that UPA serovars are the most common [40]. The studies examining the relationship of ureaplasma colonization and BPD are controversial. Some studies report that there is association [30-34], while other did not find an association [41-44]. This controversy may be attributed to different patient populations, divergent ventilator management practices, non-uniform ureaplasma detection techniques, and publication bias [45].

Three meta-analyses were performed with stringent entry criteria in order to analyse the relationship of ureaplasma and BPD. The studies were published in 1995 [46], and 2014 [47] and observed a significant association between *Ureaplasma* spp. respiratory tract colonization and BPD.

The major insights into the pathogenesis of ureaplasma-mediated lung injury have been contributed by epidemiological studies of human preterm infants and experimental intrauterine infection models in mice, sheep and non-human primates[35]. Observations from the animal studies have been confirmed by lung histological examinations of infants dying with ureaplasma pneumonia sharing characteristics of moderate-to-severe fibrosis; increased myofibroblasts; disordered elastin accumulation; and increased number of tumour necrosis factor-α (TNF-α) and transforming growth factor β1 (TGFβ1) immunoreactive cells [22, 35]. Similar findings were observed in the antenatally infected animal models, the baboon, rhesus macaques, and sheep with UPA. When the immature, UPA infected baboon was exposed postnatally to ventilation and oxygen it developed chronic inflammatory and profibrotic immune response that contributes to lung injury, altered developmental signaling, and fibrosis [48]. Duration of intrauterine UPA exposure greater than 10 days leads to lung fibrosis and thickened alveolar walls in the rhesus macaques, thus the severity of fetal lung injury, influx of inflammatory cells and epithelial necrosis was dependent on the length of intrauterine exposure to the infection [27]. Acute intrauterine exposure to either UPA3 or lipopolysaccharide (LPS) alone or combined leads to differential expression of TGFβ signalling components and elastin deposition in the ovine fetal lung [49]. These data confirm the association of antenatal

ureaplasma exposure and BPD due to its contribution to lung inflammation, altered lung development and lung fibrosis, which are characteristic of the BPD phenotype [35].

1.3.1.2 Contribution of Ureaplasma species to development of NEC

The intestinal barrier development may be altered due to direct exposure of the fetal intestinal tract to microbes and amniotic fluid containing inflammatory mediators. This may stimulate an inflammatory response and lead to increased intestinal permeability and potential bacterial translocation [35]. Okogbule-Wonodi *et al* demonstrated that the risk for NEC was increased 2-fold for infants born before 33 weeks gestation with respiratory *Ureaplasma* spp. infection. The risk for NEC increased 3.3-fold for infants born before 28 weeks gestation. The same study also showed that among infants suffering NEC, those that were ureaplasma positive had significantly higher levels of IL-6 and IL-1β in their cord blood compared to those that were ureaplasma negative [50]. These observations were confirmed in the sheep intrauterine infection model, where fetal intestinal exposure to amniotic UPA3 isolate for up to 14 days prior to preterm delivery caused intestinal inflammation, reduced enterocyte proliferation and villous atrophy. Improvement of the condition was seen by including recombinant interleukin-1 receptor antagonist in the treatment, suggesting that IL-1 signalling is likely involved in ureaplasma-mediated intestinal injury [51].

1.3.1.3 Ureaplasmal Invasive Disease in the neonate

Ureaplasma spp. occurrence in blood, cerebrospinal fluid (CSF) and brain tissue is at a lower rate than in the lungs with only 12.6%–23.6% [52-54] . Nevertheless, this rate suggests the potential of ureaplasmas for invasive disease. Viscardi *et al* found that the risk of severe IVH was 2.5-fold higher in infants whose serum was positive for ureaplasma by PCR, than PCR-negative infants [52]. Interestingly, ureaplasma detection in CSF was usually asymptomatic and normal CSF parameters [22, 52], except in one study where six out of eight infants with ureaplasma positive CSF developed severe IVH, followed by posthaemorrhagic hydrocephalus or death [55]. Experimental intrauterine infection models provide further evidence for the connection of invasive ureaplasma infections and fetal brain injury. In the murine model intrauterine UPA3 infection lead to microglia activation, delayed myelination and disturbed neuronal development in fetal and neonatal brains[56]. In the rhesus macaque model an antenatal UPA infection resulted in culture positive CSF and fetal brains for 20% of the animals, confirming the invasive capability of ureaplasmas [27]. A potential link between invasive ureaplasma, cytokinemia, and neonatal brain injury has been suggested based on the observation that when serum IL-1β is elevated in combination with ureaplasma-positive serum the risk of IVH increases 5-fold. Thus, cytokine activation of the central nervous system immune response is likely responsible for the brain injury associated with an ureaplasma infection [22, 35].

1.3.1.4 Long-term Outcomes of Perinatal Ureaplasma Exposure

Isolation of ureaplasmas from the upper respiratory tract in infants and children less than 3 years of age has been associated with wheezing [22]. Maternal *Ureaplasma* spp. vaginal colonization during pregnancy was associated with a 2-fold increased risk for infants [22]. At 2 years of age, infants born earlier than 33 weeks gestation who were exposed to intrauterine ureaplasma infection had higher rate of cerebral palsy and lower psychomotor development index scores on the Bayley Scales of Infant Development at 2 years adjusted age compared with non-exposed infants [57]. As of today there are not enough studies on the long-term outcomes of perinatal ureaplasma exposure. Viscardi brings up the need to include long-term pulmonary and neurodevelopmental assessments in studies of potential therapeutic interventions for perinatal ureaplasma exposure [35].

1.3.2 Infertility and other genitourinary tract diseases

UPA and UUR are most often found in the genitourinary tract of sexually active adults as common opportunistic pathogens that can cause many different infections. There is no doubt that they can cause non-gonococcal urethritis (NGU) [58] in both men and women. An association of ureaplasma colonization and bacterial vaginosis (BV)[3, 59], cervicitis [3], and endometritis[3] has been reported; however it is not clear yet if ureaplasma itself is causing the pathology due to the multibacterial nature of such diseases. Infection can spread to the upper female genital tract and it can lead to pelvic inflammatory disease (PID) [60], or if it happens during pregnancy – to chorioamnionitis and further pregnancy complications as

outlined above. Furthermore, PID is a common cause of infertility, chronic pain and ectopic pregnancy [61]. The possible influence of ureaplasmas on human fertility is still not clear. However, more studies are emerging reporting association between ureaplasma colonization and subpar fertility parameters or infertility as the development of diagnostic procedures, especially more common use of PCR methods, expands. Major contributors to infertility in patients colonized with ureaplasma are low sperm quality and recurrent miscarriage [62-70].

1.3.3 Ureaplasmal extragenital diseases

Ureaplasma has been reported to cause arthritis in persons with hypogammaglobulinemia [11]. There are case reports of Ureaplasma spp. as a cause of meningitis, aortic graft infection, and intrarenal abscess in adults [71-73]. There have been few reports of Ureaplasma spp. as a cause of sternal wound infection, a life-threatening complication after cardiac surgery associated with high morbidity and mortality [74-77].

1.4 Diagnostic Methods

1.4.1 Conventional techniques

Culture is considered the reference standard for detection of *Ureaplasma* spp. in clinical specimens and remains the most widely used method for laboratory diagnosis. Some laboratories prefer to prepare their own liquid and solid media to reduce costs, but commercial sources are available in many countries [78]. If specimens have to be shipped to an off-site reference laboratory for culture,

particular care must be taken to ensure the organisms remain viable, which means the specimen must be inoculated into an appropriate transport medium and shipped frozen on dry ice. Detailed methods and quality control procedures for culture-based detection of ureaplasmas are available in reference texts [10, 79]. Due to their relatively rapid growth rates, confirmed culture results for human ureaplasmas are usually available within 2 to 5 days, exclusive of specimen transport time and shipment if an offsite reference laboratory is used. Human ureaplasmas can be positively identified to genus level by their colonial morphology and urease production on A8 agar, but genotypic or serologic procedures are necessary for species designation. For routine diagnostic and patient management purposes it is not necessary to determine which species is present in a clinical specimen. A single culture procedure utilizing 10 B broth and A8 agar can detect *Ureaplasma* spp. as well as *Mycoplasma hominis,* which is also a pathogen of the human urogenital tract but it uses arginine as an energy source istead of urea. Culture has an additional advantage in that it provides an isolate that can be studied further, and on which antimicrobial testing can be performed if needed. Up to 50% of *Ureaplasma* isolates in the United States may contain *tetM* and exhibit tetracycline resistance [11].

While serology has been an important tool in diagnosis of *Mycoplasma pneumoniae* infections for many years, it has not proven useful for ureaplasmal infections, largely because of ubiquity of these organisms in healthy persons makes interpretation of antibody titers difficult and the mere existence of antibodies alone cannot reliably distinguish among colonization, current or prior infection [10]. Even though serological assays for detection of antibody responses to human ureaplasmas

have been applied in research settings, such assays have never been standardized and are not generally practical for diagnostic purposes.

1.4.2 Molecular techniques

The advances in molecular biology over the past 25 years have greatly improved diagnostic methods for infectious diseases. As a result fast, sensitive, and specific techniques for their detection and in some instances, further characterization have been developed. Turnaround time has significantly reduced. However, the expense has increased due to the requirement for having complex molecular biology laboratory equipment and personnel with sufficient education and training to employ these techniques correctly. The most important molecular-based methods for ureaplasma detection and characterization are described in subsequent sections.

1.4.2.1 Nucleic acid amplification methods

1.4.2.1.1 PCR

PCR was first described in its modern form in 1988 [80]. By the end of the decade the widespread use of PCR was bringing about a revolution in clinical diagnostics. No area of diagnostic medicine was more enabled by PCR than clinical microbiology. Shortly after the publication of the *U. urealyticum* urease gene sequences in 1990 [81] and 1991, Willoughby and colleagues reported the first use of PCR to detect ureaplasmas [82].

As PCR evolved as a molecular biology technique, numerous variations of the original assay have developed. PCR assays have become available in several

formats, such as reverse-transcription PCR, real-time PCR, and multiplex PCR. Real-time PCR is capable of amplifying and simultaneously detecting and quantifying the target DNA molecule as it accumulates during the reaction in *real time* after each amplification cycle. Real-time PCR is frequently combined with reverse transcription to quantify messenger RNA (mRNA). Multiplex-PCR uses several pairs of primers annealing to different target sequences and allows the simultaneous analysis of multiple targets in a single sample. According to the detection strategy, all PCR methods can be classified into conventional PCR and real-time PCR. Conventional PCR measures the end stage PCR products using gels or other methods while real-time PCR detects and quantifies the products simultaneously with amplification.

Various PCR methods as well as their respective gene targets have been reported to detect *Ureaplasma* spp. The earlier gel-based conventional PCR assays targeted sequences of 16S rRNA gene and 16S rRNA -23S rRNA intergenic spacer regions [83-87], the *ure* gene subunits [81, 88], and the *mba* genes [86, 89-92]. Published real-time PCR assays have targeted primarily the *ure* gene [93-95] and the *mba* gene [96, 97]. A summary of published PCR assays and gene targets used for *Ureaplasma* PCR assays for detection, speciation, and serotyping of the organisms is provided in [3].

Published PCR assays to characterize the *Ureaplasma* spp. at the serovar level have mainly focused on primers based on the *mba* gene and its 5′-end upstream regions. In combination with direct sequencing or restriction enzyme analysis, these assays were capable of distinguishing the 4 serovars of *U. parvum* and dividing the

10 serovars of *U. urealyticum* into different subgroups [89-91, 96, 98]. Because of limited sequence variation in the *mba* genes, earlier PCR-based methods lacked the capacity for complete separation of all 14 serovars. Moreover, whole genome sequencing of all 14 serovars presented in chapter 2 has shown *mba* to be part of a large gene family present in many variations in different serovars and the gene is phase variable [99]. Thus, while the *mba* gene is a valid target for serovar-specific PCRs, attention must be paid to make certain that targets do not span sites at which chromosome rearrangements occur during phase variation.

In Chapter 3 we describe the development of a real-time PCR assay for *Ureaplasma* speciation and serotyping. We evaluated the complete genome sequences for all 14 serovars and designed primers and probes targeting new specific regions for *Ureaplasma* speciation and serotyping by real-time PCR. One multiplex species-specific PCR assay and 14 serovar-specific monoplex PCRs were developed and shown to accurately distinguish between the two species and among all 14 serovars without cross-reactivities [97]. The *UU063* gene (NP_077893), which encodes a conserved hypothetical protein that is identical in all 4 *U. parvum* serovars, was used as target for *U. parvum*; a 15,072 bp open reading frame (ORF), UUR10_0554, that is almost perfectly (>99.97%) conserved in all 10 *U. urealyticum* serovars (serovar 10, GenBank ID ACI59931.1), was selected as target for *U. urealyticum.* Serovar-specific PCR assays to study large numbers of clinical specimens and/or *Ureaplasma* isolates were developed to answer the question of differential pathogenicity.

At least one molecular-based assay that includes detection of *Ureaplasma* spp. in clinical specimens is commercially available in some countries, but not in the United States. STD6 ACE Detection by Seegene, Inc. (Rockville, MD, USA, http://www.seegene.com) simultaneously detects *Trichomonas vaginalis, Mycoplasma hominis, Mycoplasma genitalium, Chlamydia trachomatis, Neisseria gonorrhoeae,* and *U. urealyticum* in endocervical/urethral swabs. The novel feature of the Seegene technology is a dual priming oligonucleotide system that contains two separate priming regions linked by a polydeoxyinosine spacer. The polydeoxyinosine has no binding specificity or priming, but instead forms a bubble or tiny D-loop that results in more stable priming. The Seegene STD6 ACE kit should work with any thermocycler, and the post PCR assay is designed for either manual or automated gel electrophoresis [100]. The ureaplasma assay is not species specific. It amplifies a 130 bp region of the *ureD* gene cassette [101]. Seegene now offers a new version of their kit for detecting urogenital pathogens that differentiates *U. urealyticum* from *U. parvum* based on the *U. urealyticum ureD* and *U. parvum ureC* genes called AnyplexTM II STI-7 Detection Kit. A recent study compared the 3 assays available for detection of *Ureaplasma* spp: Anyplex II, Seeplex, and Mycoplasma IST. Anyplex II and Seeplex are PCR based assays, while Mycoplasma IST is a culture based assay. The PCR assays were significantly more accurate than the culture based assay, and Seeplex was more accurate than Anyplex II, however it cannot differentiate between UPA and UUR [102].

A high-resolution melt (HRM) PCR assay for the differentiation of the four serovars of UPA in a single step directlyfrom clinical samples was developed

recently [103]. In HRM PCR, UPA strains are separated into four serovars using the conserved coding region of the multiple-banded antigen (MBA) gene by PCR amplification and HRM analysis to generate reproducible and distinct melt profiles that distinguish clonal representatives of the four UPA serovars. Furthermore, this HRM PCR assay could classify DNA extracted directly from 16 clinical swabs. An important limitation of this assay is that it cannot identify the composition of mixed serovar samples. In the presence of a clinical sample containing multiple serovars, this assay can only tell if the sample contains more than one UPA serovar. This is an important limitation, because our genotyping study of 1061 clinical isolates of ureaplasmas showed that 40% of samples contained more than one serovar and 30% contained more than two serovars. Furthermore, many mixed clinical samples contained mixtures of UPA+UUR serovars [104], in which case this assay will show only the UPA constituent of the sample, thus requiring that this assay is run along an additional assay capable of detecting UUR. As the connection between ureaplasma infection and development of disease gets better established, we can expect to see more diagnostic assays developed to include detection of ureaplasmas. A multiplex PCR assay capable of detecting four venereal pathogens simultaneously, one of them UUR/UPA, was recently developed by Yao *et al* [105].

1.4.2.1.2 Novel PCR –based nucleic acid-based detection systems.

Molecular-based testing used in clinical laboratories relies primarily on PCR; however PCR can be used in combination with other nucleic acid based detection (NABD) techniques. NABD techniques are advantageous because of their speed, sensitivity, and specificity, especially when the target organism grows slowly and/or

is difficult to culture. NABD techniques used in combination with PCR for *Ureaplasma* detection and species identification include reverse line hybridization blotting combined with multiplex PCR (mPCR/RLB) [106] and microarrays [107]. These methods require a visualization step based on specific interaction between biotin and conjugated-streptavidin. The mPCR/RLB technique integrates multiplex PCR (mPCR) using biotin-labeled primer pairs or biotin-dNTPs to generate labeled PCR products that hybridize with highly specific, membrane-bound, amine-labeled oligonucleotide probes by reverse line blot hybridization [108]. To visualize the hybridized PCR products, the membrane is incubated with peroxidase-labeled streptavidin and chemiluminescent substrate, e.g. electrochemiluminescence (ECL) detection liquid. Chemiluminescence results can be detected by a light sensitive film or a lumino-imager [108].

The microarray used for the detection of ureaplasma and some other microorganisms is based on immobilizing probes on a glass surface that hybridize to their complimentary biotin labeled-DNA targets produced by previous PCR procedure. The microarray is then incubated with gold-conjugated streptavidin. The interaction of biotin and streptavidin leads to silver precipitation onto the steptavidin-bound nanogold particles. This is visualized as black spots on the microarray [107].

1.4.2.1.3 Non-PCR-based nucleic acid-based detection systems

There are other nucleic acid amplification based techniques that do not use PCR and have been used for detection of pathogenic organisms like loop-mediated isothermal amplification (LAMP)[109], ligase chain reaction (LCR)[110], nucleic

acid sequence-based amplification (NASBA)[111-113], strand displacement amplification (SDA)[114] and *in-situ* hybridization (ISH)[115]. In some cases their chief attribute is that their technology is not subject to patents surrounding PCR.

Loop-mediated isothermal amplification (LAMP) allows amplification of target DNA to the magnitude of 10^9 copies in less than an hour. The technique uses DNA polymerase and four specific primers that recognize six distinct sequences on the target. An internal primer is designed to include sense and anti-sense strand sequence of the target DNA. The annealing of this primer to target DNA initiates the LAMP reaction. In the following step an outer primer is used to begin strand displacement DNA synthesis, as a result of which a single-stranded DNA molecule is released. This molecule serves as a template, primed by a second inner primer annealing to the other end of the target DNA. The newly synthesized DNA strand is displaced by a second outer primer, and forms a stem-loop DNA structure. The subsequent hybridization of one inner primer to the loop initiates cycling displacement DNA synthesis resulting in the production of new stem-loop DNA structures. The final products of LAMP are stem-loop DNAs with multiple loops between alternately inverted repeats of the target DNA[109, 116].

Ligase chain reaction (LCR) is based on the ligation of two adjacent synthetic oligonucleotides by thermostable ligase. The reaction uses total of four oligonucleotides. Two are designed to be complementary to the sense strand of the DNA target and the other two to the anti-sense strand. The reaction is heated to 95°C to denature the target DNA and then cooled to allow oligonucleotide hybridization

and ligation. The ligated oligonucleotides serve as template DNA in subsequent cycles to amplify the target DNA[110].

Nucleic acid sequence-based amplification (NASBA) is an isothermal (~40°C) amplification technique primarily used for amplification and detection of mRNA. The technology has been used also on DNA templates. The process involves two primers and the enzymes avian myeloblastosis virus reverse transcriptase (AMV-RT), *Escherishia coli* RNase H and T7 RNA polymerase. The method can amplify the target sequence more than 10^{12}-fold in 90 minutes. The first primer binds to the RNA template and initiates the synthesis of a DNA:RNA hybrid molecule by AMV-RT. The RNA strand of the hybrid is hydrolysed by RNase H and the second primer can anneal to the single DNA strand and is elongated by AMV-RT to produce double stranded DNA molecule (cDNA). The first primer contains target specific sequence and a T7 RNA polymerase promoter site. Thus a new now anti-sense RNA copy can be produced to start a cycle in which now the second primer initiates the production of the DNA:RNA hybrid molecule. NASBA can be used in a real-time and multiplex format. The real-time format of NASBA involves the use of a molecular beacon, which is a hairpin-shaped oligonucleotide. One end of the beacon has a fluorescent group and the other a quencher. When the beacon anneals to the target sequence the quencher is situated farther away from the fluorophore, therefore light emission can be detected.[111-113, 117, 118]

Strand displacement amplification (SDA) is another isothermal (~40°C) technique, which can amplify a target sequence 10^{8}-fold in a 2 hour incubation period. The template DNA is heat denatured. Four primers are needed for this

technique: Two forward and two reverse primers. The inner forward and reverse primers include 5′ restriction enzyme site. The outer forward and reverse primers do not contain restriction enzyme sites. After heat-denaturation of the template the reaction is cooled down to ~40°C and primers anneal to the target sequences; follows addition of restriction enzyme and exonuclease deficient DNA polymerase. All four primed regions are extended, which causes the displacement of the strands synthesized by the extension of the internal pair of primers. The displaced primers serve as templates. Eventually, fragments with restriction enzyme sites are generated. The restriction enzyme nicks the DNA at those sites and the 3′-end at the nick is extended by DNA Polymerase, displacing the stand. Primers hybridize to the displaced strand to repeat the process, which leads to exponential amplification of the target sequence. SDA has been used in multiplex and real-time formats as weel.[119, 120].

In situ hybridization allows visualization of the localized gene expression within the context of tissue morphology. ISH was shown to be useful in detection of *Ureaplasma* in lung tissue samples collected from intranasally inoculated newborn mice. After fixation of the tissue samples, the slides were incubated with biotinylated DNA probes specific for an internal nucleotide sequence within the urease gene. The hybridization signal was amplified by binding of peroxidase-conjugated streptavidin to biotin and byotinyl tyramide incubations. A chromogenic dye producing a brown precipitate at the hybridization sites was used to visualize the hybridization signal [115].

1.4.3 Choosing culture versus molecular-based tests for UU detection

While there is no question that molecular-based tests have been extremely important for research studies, investigating *Ureaplasma* pathogenesis, epidemiology, and disease over the past two decades, culture still plays a significant role in routine diagnostic testing of clinical specimens obtained from patients in whom a clinically significant *Ureaplasma* infection is suspected. The discussion above makes a strong and clear-cut case for the many advantages of real-time PCR over conventional PCR for detection and speciating ureaplasmas. Real-time PCR is more expensive than conventional PCR in terms of cost of reagents and materials necessary to perform the assay; however the one-day turnaround time compared to 2-3 days for conventional PCR offsets these costs when considering the personnel time involved. Thus, the decision is quite straightforward in selecting real-time PCR over conventional PCR if financial resources permit. It is somewhat more complex deciding whether PCR or culture-based detection should be employed on a routine basis. It is impossible to state that one method or the other is always preferred in every circumstance. Even though real-time PCR is theoretically more sensitive than culture and can provide same day results, it may not be a practical methodology for daily runs in a laboratory that receives only an occasional specimen submitted for *Ureaplasma* detection or which does not have suitable physical facilities, molecular diagnostic equipment or personnel with the necessary education and training to perform NABD techniques. Each PCR assay requires reagents for positive and negative and internal controls as well as the specimen to be tested, so the cost per test decreases if more tests are performed simultaneously. The LightCycler 2.0by Roche

has a throughput of 32 samples per run, including the necessary controls. The cost of culture is essentially the same whether one specimen or multiple specimens are processed at the same time. The reagent cost of running a single real-time PCR assay for *Ureaplasma* detection and speciation is approximately 4-fold greater than the costs for liquid and solid media for culture-based diagnosis, assuming the media is prepared on site. If purchased commercially, the cost of media is greater but is still considerably less than the cost of running a single PCR assay.

The first culture methods developed for human ureaplasmas in 1950s were further refined over the next several years so that basic and clinical studies could be performed to investigate their pathogenic role in disease. Commercial marketing of agar and broth media based on the original formulations described by researchers facilitated adoption of ureaplasma cultures for diagnostic purposes by larger hospital laboratories and reference laboratories, so now it is a reasonably commonly performed procedure in the United States. Widespread availability of diagnostic kits in several European and Asian countries has also contributed to more broad use of culture-based diagnosis for ureaplasmal infections. The development of PCR-based detection systems have been obligatory for species such as *Mycoplasma genitalium* and are now the tests of choice for *M. pneumoniae*. However, need for PCR-based detection of *Ureaplasma spp*. has not been as critical because of their relative ease of in vitro cultivation and rather rapid growth. PCR-based studies have been necessary to provide data to support the separation of *U. parvum* and *U. urealyticum* as individual species and to enable complete characterization of all 14 serovars in a rapid and reproducible manner. Since the question of differential pathogenicity of the

two human *Ureaplasma* species and their respective serotypes has not been answered conclusively, there are no current recommendations for identification of ureaplasmas to the species or serovar level for diagnostic purposes. We anticipate that molecular-based technology is likely to become much more widely available and simplified in the future, making it appeal to a broad range of clinical laboratories. If more diagnostic companies choose to include gene targets for *Ureaplasma* spp. in commercial PCR kits, especially if such products can be used to detect other pathogens in the same assay, their use is bound to increase. Availability of the complete genomic sequence data for all 14 serovars will surely facilitate development of such additional assays.

1.5 Pathogenicity: Reasons for differential clinical outcome

1.5.1 Serovars

It has been proposed that differential pathogenicity may be dependent on the specific serovar colonizing the individual. Studies focusing on correlating serovars to specific diseases and disease severity have failed to make specific connections of serovars to disease. We genotyped 1061 clinical isolates to analyze the potential relationship of specific serovars to specific clinical outcomes; however to our surprise we found extensive horizontal gene transfer and high occurrence of samples containing more than one serovars, making it impossible to establish such relationships (work described in Chapter 4 of this dissertation).

1.5.2 Host immune response

There is increasingly more evidence that the severity and type of disease is greatly influenced by the immune response of the patient. An intact humoral immune response appears to be important in limiting invasion and dissemination of ureaplasma beyond mucosal surfaces. This is demonstrated by their tendency to cause chronic respiratory infections in persons with hypogammaglobulinemia, and to cause invasive disease in preterm neonates [7]. A recent study [121], demonstrated that mice deficient the innate host defense molecule, surfactant protein A (SP-A) had delayed clearance of ureaplasma from their lungs, increase in inflammatory cells and pro-inflammatory cytokine expression, compared to wild type mice. SP-A is deficient in preterm infants, and may at least partially explain the vulnerability of this population to more severe outcomes of an ureaplasma infection.

Ureaplasma infection induces the stimulation of inflammatory cytokines, which is believed to be a causative link between intrauterine infection and organ injury [35]. The release of TNF-α, IL-1β, IL-8, monocytechemoattractant-1, TGFβ1 and other mediators has been demonstrated by various cell types *in vitro*. Increased concentration of these cytokines in tracheal aspirates of infants during their first week of life is associated with ureaplasma colonization and development of BPD [122]. The activation of these networks leads to inflammatory cell recruitment and activation and subsequent local tissue injury, leading to alterations in normal developmental pathways during critical periods of development [122].

Toll-like receptors (TLRs) can mediate microbial recognition by the host innate immune system. UPA lipoproteins, including the MBA, have been shown to activate NF-κB in reporter cell lines through TLR1, 2 and 6 signaling [123]. A study by Peltier *et al*. showed this signaling to be mediated through TLR2 or TLR4 receptors [124]. SNPs in these TLR6 have been reported to make the host less susceptible to ureaplasma colonization and development of diseases [125]. This confirms that differences in the host immune system response and genetic makeup play roles in the difference observed in clinical outcome among different patients. It is interesting to note that chronic intra-amniotic colonization by ureaplasmas profoundly diminished responses to intra-amniotic lipopolysaccharide in the preterm fetal sheep [126]. Such endotoxin tolerance can be beneficial to the host due to decreased inflammation and thus decreased organ injury [45].

Although we are gaining much better knowledge of the molecular mechanism in the host leading to specific diseases in the presence of ureaplasma infections, still little is known of the molecular mechanism of this opportunistic pathogen leading to these host responses.

1.5.3 Potential Virulence Factors of Ureaplasma Species

Ureaplasmas are capable of attaching to a variety of cell types such as urethral epithelial cells, spermatozoa, and erythrocytes [12]. The adhesins of ureaplasmas have not been characterized completely, but current evidence suggests the receptors are sialyl residues and/or sulphated compounds [7]. As described above, the severity and type of pathology varies greatly. Differential pathogenicity

may be dependent on the presence of specific virulence factors in the ureaplasma strain colonizing the individual. One of the major goals of our comparative genome analysis of 19 ureaplasma strains, described in Chapter 2, was to find genes that may potentially be involved in virulence and pathogenicity. Although there is no evidence that ureaplasmas produce toxins, they do possess several potential virulence factors.

A major family of phase-variable, surface proteins, the multiple banded antigens (MBA), is immunogenic during ureaplasmal infections. The MBA is a surface lipoprotein consisting of an N-terminal conserved domain including a signal peptide, lipoprotein attachment site, and transmembrane domain and a C-terminal variable domain which usually consists of tandem repeating units. MBAs have been used as a basis for the development of reagents for diagnostic purposes and for serotyping [62-65]. Furthermore, the variability of the number of tandem repeats of the variable domain requires interaction with the host immune system as demonstrated in the sheep intrauterine infection model. Only chronic infections (69 days) elicited variability of the number of tandem repeats of the MBA TRU [127, 128]. The MBA family contains at least 22 different TRUs. The variable number of tandem repeats of the TRU of the *mba* and the potential capacity of the organism to switch the TRUs has been suggested as a mechanism to evade the host immune system response. It is possible that the different types of TRUs may lead to or contribute to differential pathogenicity. Furthermore, the *mba* is one of the genes we saw to be exchanged via HGT among two ureaplasma strains, co-infecting a patient.

Immunoglobulin A (IgA) protease activity has been demonstrated in all tested Ureaplasma strains representing 13 of the 14 serovars (UUR13 was not tested)

[66, 67]. IgA protease has been considered as one of the major factors contributing to the pathogenic potential of ureaplasmas [66, 67]. It is expressed in bacterial pathogens especially if they colonize a mucosal surface [68]. This can provide them with an advantage in evasion of the host-defenses. It is interesting to note that the Neisseriae commensal species do not express this enzyme [69]. Our computational analysis of all ureaplasma sequenced genomes could not identify an IgA protease gene, suggesting that this may be a novel type of IgA protease (see Chapter 2).

Ureaplasmas have also been reported to have phospholipase A1, A2 and C activities [70-73]. Such activity when the infection reaches the amnion or placenta can lead to production of free arachidonic acid, which may activate the synthesis of prostaglandins, thus inducing labor prematurely. Our computational analysis of the 19 sequenced ureaplasma genomes could not identify any of these phospholipases. Our attempts to replicate the results of the studies reporting these activities were unsuccessful. During our attempts we carefully analyzed the design of the original radioactivity assay and the commercial Amplex Red PLC detection assay (Invitrogen). We determined that both assays would measure the activity of both PLC and PLD, thus are not specific for PLC. The Amplex Red PLC assay can be modified to measure PLD activity only by omitting the first step in the assay. The PLD activity can then be subtracted as background from the PLC+PLD activity measured by the assay. Using this strategy we were unable to detect PLC activity in ureaplasmas infecting humans. Recently, PLC activity was reported in *Ureaplasma diversum*, however, the Amplex Red assay was used to detect it. The authors did not modify the assay to measure and subtract the potential PLD activity of the cells. We

did identify a phospolipade D domain containing gene in the sequenced UUR and UPA genomes, which may be present in *U. diversum*. Therefore, further experiments are needed to confirm if the phospholipase activity observed in *U. diversum* is PLC, PLD or maybe both. A detailed discussion of additional potential pathogenicity factors, as well as the ones listed above is provided in Chapter 2.

1.6 Antimicrobial Susceptibility of Ureaplasma

Ureaplasmas are naturally resistant to β-lactams as they lack peptidoglycan. The absence of folic acid synthesis makes sulfonamides inefficient against them to [14]. However, ureaplasmas are generally sensitive to tetracyclines, fluoroquinolons, macrolides and clindamycin, although resistance to some of these has been reported. Table 1-1 summarizes the susceptibility of ureaplasma isolates collected during pregnancy and compiled from different studies[14]. Recent studies in animal models have evaluated novel treatment plans during pregnancy to prevent premature birth and neonatal complications. Grigsby *et al* showed that azithromycin maternal therapy can eradicate ureaplasmas from amniotic fluid, placenta and fetus, prolonging the pregnancy and reducing the severity of neonatal complications [129, 130]. Keelan *at al* explored the use of solitromycin for maternal administration in the pregnant sheep model [131]. They found that its rate of transplacental passage is relatively high and its efficiency of transfer was greater than existing macrolides. This may be one of the reasons they observed that a single maternal dose is sufficient to establish therapeutic antibiotic levels in the fetal and amniotic compartments.

Table 1-1. Antimicrobial Susceptibilities of Ureaplasmas
Reused with permission from [14]

Antimicrobials	*U. urealyticum* MIC (μg/ml)
Erythromycin	0.2–2–4
Tetracycline	0.05–2
Ciprofloxacin	0.1–16
Ofloxacin	0.2–25
Clarithromycin	≤0.004–2
Josamycin	0.5–4
Pristinamycin	0.1–1
Azithromycin	0.5–4
Doxycycline	0.02–1

1.7 Major Barriers to Understanding Pathogenicity

Understanding of the role of ureaplasmas in human diseases has been thwarted thus far due to two major barriers: 1) Lack of suitable diagnostic tests that can identify and characterize the two species and their respective serotypes that could account for differential pathogenicity; 2) Lack of genetic manipulation tools for the creation of deletion mutants and their complementation to study the role of potential pathogenicity factors.

A major effort in the field of ureaplasma research over the past 20 years has been focused on overcoming the first barrier. Since 1990, nearly 300 peer-reviewed publications have described basic and clinical investigations involving ureaplasmas using PCR and various other nucleic acid-based techniques. Based on the genome sequences of the 14 ATCC type strains of each ureaplasma serovar, our group has developed species and serovar specific real-time PCR assays to detect ureaplasmas in patient isolates [132], which were applied to a collection of 1,061 clinical isolates [133]. The results of this large scale serotyping study are discussed in detail in Chapter 3 of this dissertation. Briefly our study showed 39% of clinical isolates contained more than one serovar, and 28% of the evaluated mixed isolates were chimeric organisms containing markers from up to 4 serovars (Figure 4-1 and Table 4-4). This suggested that ureaplasmas are subject to extensive horizontal gene transfer. Therefore it is possible that ureaplasmas do not exist as stable serovars in their hosts, but rather as a dynamic population exchanging genes, most likely not only among themselves, but also with other bacteria in their environment. Based on these

findings, the long sought after serotyping assays are of limited value, and there is no utility in trying to associate individual serovars with specific disease. The differential clinical outcome of ureaplasmal infection may be due primarily to presence or absence of potential pathogenicity factors in the colonizing ureaplasma strain [134] and/or patient to patient differences in terms of autoimmunity and microbiome.

The second major barrier to the understanding of ureaplasma pathogenicity is the lack of molecular tools for creation of deletion to study the role of potential pathogenicity factors. This problem has not been explored nearly as exhaustively as the development of serotyping assays. The capacity to create deletion mutants has been a prerequisite in understanding host-pathogen interaction and factors involved in pathogenicity in many well studied pathogens.

Approximately twenty years ago efforts to develop genetic manipulation systems for ureaplasma were unsuccessful (personal communication with Kevin Dybvig). This is one reason little use was made of the first ureaplasma genome sequence, which was published early in the genomic era. New technologies such as TN4001, OriC plasmids, expression of *E. coli* RecA via a suicide plasmid, and the technology developed for synthetic biology offer new possibilities for development of genetic tools for ureaplasma. The development of tools for genetic manipulation of *Ureaplasma* spp. will tremendously facilitate the study and understanding of factors leading to differential pathogenicity of Ureaplasma strains. Mutants can be extremely valuable for study in existing animal models for ureaplasmal infections in humans. Furthermore, the increased understanding of pathogenicity factors can aid in development of better screening methods, drug targets and vaccines to decrease the

morbidity associated with ureaplasmal infections. The results of this effort are discussed in detail in Chapter 5 of this dissertation.

Chapter 2. Comparative Genome Analysis of *19 Ureaplasma urealyticum* and *Ureaplasma parvum* strains

2.1 Abstract

Ureaplasma urealyticum (UUR) and *Ureaplasma parvum* (UPA) are sexually transmitted bacteria among humans implicated in a variety of disease states including but not limited to: nongonococcal urethritis, infertility, adverse pregnancy outcomes, chorioamnionitis, and bronchopulmonary dysplasia in neonates. There are 10 distinct serotypes of UUR and 4 of UPA [5, 7, 83, 86, 87, 135, 136]. Efforts to determine whether difference in pathogenic potential exists at the ureaplasma serovar level have been hampered by limitations of antibody-based typing methods, multiple cross-reactions and poor discriminating capacity in clinical samples containing two or more serovars.

We determined the genome sequences of the American Type Culture Collection (ATCC) type strains of all UUR and UPA serovars as well as four clinical isolates of UUR for which we were not able to determine serovar designation. UPA serovars had 0.75-0.78 Mbp genomes and UUR serovars were 0.84-0.95 Mbp. The original classification of ureaplasma isolates into distinct serovars was largely based on differences in the major ureaplasma surface antigen called the multiple banded antigen (MBA) and reactions of human and animal sera to the organisms. Whole genome analysis of the 14 serovars and the 4 clinical isolates showed the *mba* gene was part of a large superfamily, which is a phase variable gene system, and that some serovars have identical sets of *mba* genes. Most of the differences among serovars are

hypothetical genes, and in general the two species and 14 serovars are extremely similar at the genome level.

Comparative genome analysis suggests UUR is more capable of acquiring genes horizontally, which may contribute to its greater virulence for some conditions. The overwhelming evidence of extensive horizontal gene transfer among these organisms from our previous studies combined with our comparative analysis indicates that ureaplasmas exist as quasi-species rather than as stable serovars in their native environment. Therefore, differential pathogenicity and clinical outcome of a ureaplasmal infection is most likely not on the serovar level, but rather may be due to the presence or absence of potential pathogenicity factors in an individual ureaplasma clinical isolate and/or patient to patient differences in terms of autoimmunity and microbiome.

2.2 Introduction

We sequenced the 14 ATCC UPA and UUR serovars as an effort to aid the development of serotyping methods and to enhance the study of the suggested differential pathogenicity [10] and ureaplasma biology. Based on these sequences real-time PCR genotyping assays were developed that detect the 14 ATCC serovars without cross-reactions [12]. Surprisingly, the application of these assays to 1,061 clinical isolates failed to correlate specific serovars with different clinical outcomes. Our inability to correlate patient disease outcomes with specific serovars was at least in part because a large fraction of these patient samples were classified as genetic hybrids. This result was based our serotyping PCR assays. DNA sequencing of parts

of dome of the hybrid genomes showed that serotype specific markers were transferred horizontally among ureaplasmas [24]. Combining these findings with the comparative genome analysis of the 14 ureaplasma ATCC serovars has allowed us to better understand the potential mechanisms and reasons for these observations among clinical isolates. We report on genes that may contribute to the virulence of ureaplasmas, including the MBA and its putative mechanism of phase variation.

2.3 Materials and Methods

2.3.1 Sequencing methods for ATCC and 4 clinical isolates

Ureaplasmas were grown in 10B medium and phenol chloroform extracted as described previously [8]. We randomly fragmented through shearing the purified genomic DNA from the 14 ATCC type strains and generated 1–2 kbp and 4–6 kbp fragment libraries. Using Sanger chemistry and ABI 3730 DNA sequencers, each serovar was sequenced to 8-12X redundancy. In order to obtain data to complete the genome sequence of Serovar 2, the Sanger data were supplemented with 454 pyrrosequencing (Roche) data. We sequenced the 4 clinical isolates only using 454 chemistry. Genome sequences produced with Sanger chemistry were assembled using the Celera Assembler. The 454 data were assembled using the Newbler Software Package for de novo genome assembly.

2.3.2 Annotation

All 14 ureaplasma strains were annotated using the JCVI Prokaryotic Annotation Pipeline followed by manual quality checks and manual curration to

enhance the quality of annotation before being submitted to NCBI. Annotation was done on various levels, the individual protein level, the pathways and the multiple genome comparisons. The annotation pipeline has two distinct modules: one for structural annotation and the other for functional annotation. The structural annotation module predicts an extensive range of genomic features in the genome. Glimmer3 [137] was used to predict the protein coding sequences whereas, tRNAs, rRNAs, cDNAs, tRNA and ribozymes are predicted based on matches to Ram libraries, a database of non-coding RNA families [138]. The programs tRNA scan [139] and ARAGORN [140], which is a program that detects tRNA and tmRNA genes.

For functional annotation, JCVI uses a combination of evidence types which provides consistent and complete annotation with high confidence to all genomes. The automated annotation pipeline has a functional annotation module (AutoAnnotate), which assigns the function to a protein based on multiple evidences. It uses precedence-based rules that favor highly trusted annotation sources based on their rank. These sources (in rank order) are TIGRFAM HMMs [141] and Pfam HMMs, best protein BLAST match from the JCVI internal PANDA database and computationally derived assertions (TMHMM and lipoprotein motifs). Based on the evidences, the automatic pipeline assigns a functional name, a gene symbol, an EC number and Gene Ontology domains [142], which cover cellular component, molecular function and biological process(es). The assigned domains are related to evidence codes for each protein coding sequence with as much specificity as the underlying evidence supports. The pipeline also predicts the metabolic pathway using Genome properties [143], which are based on assertions/calculations made across

genomes for the presence or absence of biochemical pathways. Genome properties incorporate both calculated and human-curated assertions of biological processes and properties of sequenced genomes. A collection of properties represents metabolic pathways and other biological systems and these are accurately detected computationally, generally by the presence/absence of TIGRFAMs and Pfam HMMs. This is the basis for the automatic assertions made for the presence of the whole pathway/system in any genome. Finally a curator checked for consistency and quality of annotation, deleting spurious assertions and inserting any missed ones. This resulted in the manual merging of some genes, primarily the MBA genes, which were problematic for the automated pipeline due to the nature of their repeats. JCVI's internal Manual Annotation tool (MANATEE) was used extensively to annotate these genomes. Manatee is a freely available, open-source, web-based annotation and analysis tool for display and editing of genomic data. The genome comparisons and annotation transfer were done using the Multi Genome Annotation Tool (MGAT) which is an internally developed tool integrated within MANATEE to transfer annotations from one gene to other closely related genes. The clusters are generated based on reciprocal best BLASTP hits determined by Jaccard-clustering algorithm with a BLASTP identity >= 80%, a P value <= 1e-5 and a Jaccard coefficient threshold of 0.6. The clusters are composed of genes both within the genome and across different ureaplasma genomes. The same clusters are used in the genome comparisons generated by SYBIL (http://sybil.sourceforge.net/), which is also an open source web based software package for comparative genomics [144].

2.3.3Comparative genomics

The 19 genomes were compared using a variety of bioinformatics tools. Sybil [144] was used to generate clusters of orthologous genes (COGs), Jaccard clusters (paralogous gene clusters) and identify genes specific for each strain (singletons). The information generated with Sybil was used to deduce the pan genome for all 19 sequenced ureaplasma strains and different subsets of strains. PanSeq version 2.0 [145] was used to identify unique areas in the clinical UUR isolates that could not be serotyped. The functional annotation of genes in those areas was examined using Manatee [146] . The percent difference table between pairs of genomes was generated by mapping pairs of ureaplasma genomes to each other using BLASTN; that is, contigs in genome 1 were searched against the sequences in genome 2. The BLASTN results were processed to compute the mean identity and fraction (of contig) covered for each contig in genome 1. These values were totaled to give the final value of mean identity and fraction covered when mapping genome 1 to genome 2. All 182 comparisons were carried out. In the mapping process, no attempt was made to compute a one-to-one mapping between genome 1 and genome 2, and thus, multiple regions in genome 1 can map to a region in genome 2. The mean percent difference was calculated from the generated data and reported in Table 3.

2.3.4MBA locus

The nucleotide sequence of all genomes was uploaded to the Tandem Repeats Database (TRDB) and the Inverted Repeats Database (IRDB) [147] and was analyzed using the tools in the database to find all tandem and inverted repeats.

Genomes were analyzed one at a time and the main tandem repeating unit of the MBA of the serovar was located and the genomic area around it was inspected for other tandem repeats. This approach identified the presence of tandem repeats in the close vicinity to the MBA, that when compared through the Basic Local Alignment Search Tool (BLAST) [148] against the rest of the serovars' genomes matched the MBA's tandem repeating units of other serovars. The putative recombinase recognition sequence was identified by analyzing inverted repeats detected with the IRDB tools and close examination of the MBA loci of serovars 4, 12, and 13, which have the same set of tandem repeating units in different rearrangements. Dotplots were generated for these serovars using Dotter [149] and BLASTn [148] to help identify the conserved sequence that may serve as a recombinase recognition site. To identify other genes of the MBA phase variable system all COGs generated by the Sybil [144] computes that had participating genes annotated as MBA were examined and organized into Table 2-7.

2.3.5 PLC, PLA, and IgA protease genes

Tools used to search the genomes were BLAST [148, 150] and Hidden Markov Models (HMMs) [151] deposited in PFAM [152]. We set up databases of all human ureaplasma open reading frames, proteins and full genome sequences. BLASTn and BLASTp [148, 150] were used initially to search the open reading frames and protein databases with known PLC, PLA1, and PLA2 genes and protein sequences. Using this approach we were not able to identify any significant hits. To

make sure that the gene was not missed by the gene predicting software, we used tBLASTn [150] to search the ureaplasma full genomes translated nucleotide database.

2.3.6 PLC assay

Amplex® Red Phosphatidylcholine-Specific Phospholipase C Assay Kit (Invitrogen Cat.No.A12218) was used to detect activity of the enzyme in whole cell lysates, membrane, cytosolic, and media fractions of exponential and stationary phase cultures. The Amplex® Red Assay provides lecithin as substrate for PLC that when cleaved forms phosphocholine. Phosphocholine is modified to choline by alkaline phosphatase, which in the presence of choline oxidase produces betaine and H_2O_2. The Amplex red reagent in turn reacts in the presence of H_2O_2 and horseradish peroxidase to produce the red fluorescent compound resorufin. However, if the test sample contains PLD, PLD will cleave lecithin to produce choline, which bypasses the alkaline phosphatase step of the assay's cascade; therefore, this assay would give a combined readout of PLC and PLD. Due to the potential presence of a PLD gene in ureaplasmas, to make the assay PLC specific we modified the assay by repeating it for each test sample, but omitting alkaline phosphatase from the reaction, in order to be able to subtract any activity by the putative PLD enzyme in the ureaplasma genomes. Everything else followed the manufacturer's assay protocol. ATCC UPA3 and UUR8 cultures were grown in 10B or Trypticase Soy Broth to exponential phase. Cells were harvested through centrifugation and subjected to osmotic lysis. Cell membranes were collected through ultracentrifugation. The cleared cell lysates and

the cell membranes were tested for PLC activity with the Amplex Red assay and with the previously published assay by DeSilva and Quinn [153-155].

2.3.7 Phylogenetic trees

Multiple sequence alignments (MSA) and phylogenetic tree constructions were performed using ClustalX 2.1 [85]. [156]. Phylogenetic trees were visualized with Dendroscope [157]. Multi-gene phylogenetic trees were generated by aligning the nucleotide sequences of 82 genes: the 7 genes encoding the urease subunits (ureA-G), 47 genes encoding ribosomal proteins, 12 genes encoding RNA and DNA polymerase subunits, and 16 genes encoding tRNA ligases. The MSAs of all genes were concatenated and edited with Jalview 2.6.1 [158] to remove the non-informative positions (100% conserved in all 19 genomes) from the alignment. This was needed because the extreme similarity among the strains generated multiple sequence alignments containing approximately 5% informative positions. Although these informative positions were enough to separate the two species, they were not enough to resolve the relationship among serovars/strains within each species. The removal of the non-informative positions increased the bootstrap values but did not affect the structure of the clades. The phylogenetic tree was generated with ClustalX 2.1 neighbor-joining bootstrap option. The gene content tree was generated using the information from the formed clusters of orthologous genes (COG) to generate a table with a serovar on each row and a COG in each column. The presence of a gene in a serovar for each COG was marked with the number 0–6 (0 = none, 1–6 = number of copies of the gene in the serovar). Singletons were added to the table to increase the

informative data. The core genome COGs (genes conserved in all 19 genomes) were removed from the dataset, since they are non-informative. To be able to use ClustalX 2.1 to generate the tree the numbers were turned to letters: (0 =C, 1 = S, 2=T, 3=P, 4=A, G=5, N=6). The table was turned into a multifasta formatted file and loaded into ClustalX 2.1. The sequences did not need to be aligned with ClustalX 2.1, since they were already aligned. The tree was constructed using the bootstrap, neighbor joining method. The root for all trees is a poly-A sequence of similar size, since only the relationship within ureaplasmas was of interest.

2.4 Results and Discussion

2.4.1 Genome sequencing of 19 *U. urealyticum* and *U. parvum* strains

Subsequent to the publication and annotation of the complete genome of a clinical isolate of UPA3 by Glass and colleagues [8], sequencing of all 14 serovar type strains deposited in the ATCC was begun to study differences among them and examine them for virulence factors. The intent was to completely sequence the ATCC UPA3, which is the reference strain for UPA, and UUR8, which is the reference strain for UUR. The genomes of those serovars were completed along with UUR2 and UUR10. The sequencing coverage for each genome varied between 7X to 14.5X (Table 2-1).

Table 2-1. Overview of UUR and UPA genomes

Serovar	ATCC	GenBank Accession no	PFGE size in Kbp*	Genome size (bp)	Contigs	ORFs	Hypothetical Proteins	G+C%	Sequence coverage
1	27813	NZ_ABES00000000	760	753,674	8	604	212	25%	14.6X
3	27815	NC_010503	760	751,679	1	609	219	25%	10.2X
3	700970	NC_002162	Patient Isolate	751,719	1	614	154	25%	-
6	27818	NZ_AAZQ00000000	760	772,971	5	619	221	25%	11.4X
14	33697	NZ_ABER00000000	760	749,965	7	594	199	25%	14.5X
2	27814	NZ_ABFL00000000	880	861,061	1	664	248	26%	10.7X
4	27816	NZ_AAYO00000000	910	835,413	4	654	206	26%	7.0X
5	27817	NZ_AAZR00000000	1140	884,046	18	677	252	26%	8.5X
7	27819	NZ_AAYP00000000	880	875,530	4	660	246	26%	8.3X
8	27618	NZ_AAYN00000000	890	874,381	1	673	232	26%	9.9X
9	33175	NZ_AAYQ00000000	950	947,165	10	711	244	26%	8.6X
10	33699	NC_011374	890	874,478	1	657	232	26%	12.1X
11	33695	NZ_AAZS00000000	840	876,474	6	644	236	27%	10.0X
12	33696	NZ_AAZT00000000	870	873,466	2	650	234	25%	9.0X
13	33698	NZ_ABEV00000000	900	846,596	5	655	234	25%	11.1X
2033	Unkn. serovar	AJFX00000000	Patient Isolate	804,560	16	646	190	26%	39.0X
2608	Unkn. serovar	AJFY00000000	Patient Isolate	856,546	14	667	258	26%	60.0X
4155	Unkn. serovar	AJFZ00000000	Patient Isolate	858,890	18	684	225	26%	73.0X
4318	Unkn. serovar	AJGA00000000	Patient Isolate	844,666	16	661	181	26%	52.0X

Genome sizes of UPA serovars were between 0.75-0.78 Mbp and of UUR serovars between 0.84-0.95 Mbp. We sequenced the genomes of four UUR clinical isolates that were negative for all of our serovar genotyping real-time PCR assays [104]. All of the isolates' genomes had some minor genome rearrangements, regions that were deleted, and some regions that were inserted and are new for the urealyticum group when compared to the ATCC reference strains. Whether we can assign new serovar numbers to any of the unidentifiable isolates is a matter of clarifying the requirements for an ureaplasma to be considered a specific serovar.

2.4.2 Genome Analysis of UU clinical isolates that could not be genotyped to the serovar level

There were 67 (6%) isolates that were negative for all serovar-specific assays from the collection of 1061 clinical isolates of human ureaplasmas from diverse patient populations that we typed using our species and serovar specific real-time PCR assay. These serovars could not be assigned to any of the 14 known serovars by PCR [104]. Initial analysis of dotplots comparing each isolate's genome to each ATCC serovar showed that isolates 2033 and 2608 were most closely related to serovars 12 and 4, which are the closest related serovars among the urealyticums (data not shown). Isolate 4155 was most similar to serovar 11, whereas isolate 4318 was most similar to serovar 2. All of the isolates' genomes had some minor genome rearrangements, areas that were deleted, and some areas that were inserted and are new for the urealyticum group when compared to the ATCC reference strains. The reason these isolates were not positive for any of our serovar specific PCR assays was

that the targets were either missing completely or some of the target was missing or modified so that one of the primers would not bind. However, it is clear that these isolates have changes in other areas of the genome as well. Whether we can assign new serovar numbers to any of the unidentifiable isolates is a matter of clarifying the requirements for a ureaplasma to be considered a specific serovar. The genomes of the four isolates were examined for regions that were not present in any of the ATCC serovars using PanSeq [145]. Such areas were identified in isolates 2608 (3217nt long; genomic region coordinates: 4318-7534), 2033 (989nt long; genomic region coordinates: 585519-586507), and 4318 (3217nt long; genomic region coordinates: 21576-24792), but not in 4155. Isolates 2608 and 4318 had a 99% identical region that was novel for ureaplasmas. All of these areas encoded hypothetical proteins of unknown function. The unique region of isolate 2033 contained one 825 nt long gene (ORF00459). The gene encoded a signal peptide which suggests that it may be secreted. The unique area of isolate 2608 unique area contains 3 genes: ORF00004 (840nt), ORF00005 (561nt), and ORF00006 (1125nt). The unique area of isolate 4318 also contains 3 genes: ORF01501, ORF01502, and ORF01503.

2.4.3 Gene content analysis of all 19 Ureaplasma genomes

All strains had the expected two rRNA operons and tRNA coding genes (Table 2-2). UPA serovars have an average of 608 genes, of which 201 encode hypothetical proteins on average, and UUR serovars have an average of 664 genes, of which 230 encode hypothetical proteins on average (Table 2-3). The ureaplasma pan genome

Table 2-2. Anticodon Table of tRNAs Showing count of tRNAs used by Human Ureaplasmas.

The number next to each anticodon sequence represents the number of tRNA gene copies in ureaplasmas. Green marks the start anticodon and red marks the 2 stops. The UGA codon (a stop codon in the standard genetic code) codes for tryptophan and is read by tTrpUCA in yellow on the table.

First Position (5' end)	Second Position: A			Second Position: G			Second Position: U			Second Position: C			Third Position (3' end)
A	AAA	0	tPhe	AGA	0	tSer	AUA	0	tTyr	ACA	0	tCys	A
G	**GAA**	**1**		GGA	0		**GUA**	**1**		**GCA**	**1**		
U	**UAA**	**1**	tLeu	**UGA**	**1**		UUA	-	Stop	**UCA**	**1**	tTrp	
C	**CAA**	**1**		CGA	0		CUA	-		**CCA**	**1**		
A	AAG	0	tLeu	AGG	0	tPro	AUG	0	tHis	**ACG**	**1**	tArg	G
G	GAG	0		GGG	0		**GUG**	**1**		GCG	0		
U	**UAG**	**1**		**UGG**	**1**		**UUG**	**1**	tGln	**UCG**	**1**		
C	CAG	0		CGG	0		CUG	0		CCG	0		
A	AAU	0	tIle	AGU	0	tThr	AUU	0	tAsn	ACU	0	tSer	U
G	**GAU**	**1**		GGU	0		**GUU**	**1**		**GCU**	**1**		
U	UAU	0		**UGU**	**1**		**UUU**	**1**	tLys	**UCU**	**1**	tArg	
C	**CAU**	**3**	tMet	CGU	0		**CUU**	**1**		CCU	0		
A	AAC	0	tVal	AGC	0	tAla	AUC	0	tAsp	ACC	0	tGly	C
G	GAC	0		GGC	0		**GUC**	**1**		**GCC**	**1**		
U	**UAC**	**1**		**UGC**	**1**		**UUC**	**1**	tGlu	**UCC**	**1**		
C	CAC	0		CGC	0		CUC	0		CCC	0		

Table 2-3. Role Category Breakdown of Genes

Role Category Breakdown	Total ORFs:	Assigned function	Unknown function	Hypothetical proteins	Unclassified, no role category	Amino acid biosynthesis	Purines, pyrimidines, nucleosides, and nucleotides	Fatty acid and phospholipid metabolism	Biosynthesis of cofactors, prosthetic groups, and carriers	Central intermediary metabolism	Energy metabolism	Transport and binding proteins	DNA metabolism	Transcription	Protein synthesis	Protein fate	Regulatory functions	Signal transduction	Cell envelope	Cellular processes	Mobile and extrachromosomal element functions	Disrupted reading frame
UPA1	604	394	46	141	0	2	20	9	8	15	29	82	57	18	104	24	7	1	54	11	0	1
UPA3	609	390	41	164	0	2	20	8	9	13	32	78	54	18	105	26	6	1	50	12	2	0
UPA6	619	398	38	160	0	2	20	7	9	13	31	81	57	17	106	26	7	2	52	11	2	1
UPA14	594	396	42	144	0	2	20	8	9	14	30	79	56	18	105	25	5	1	53	11	0	2
UUR2	664	431	40	161	2	1	21	8	8	23	31	86	63	18	105	25	4	1	70	14	3	0
UUR4	654	829	67	174	2	1	20	8	8	14	30	76	52	18	111	29	5	1	64	12	7	3
UUR5	677	425	36	177	1	1	20	9	8	16	32	86	56	21	108	29	6	1	65	15	5	4
UUR7	660	414	39	181	1	1	20	8	8	16	30	84	55	18	108	27	5	1	65	13	4	4
UUR8	673	423	40	186	0	1	21	8	8	23	30	88	59	18	105	25	4	1	67	14	3	2
UUR9	711	469	35	176	2	1	21	8	7	18	30	79	66	20	112	30	6	1	75	18	26	5
UUR10	657	425	34	172	1	1	22	8	9	20	30	83	57	17	113	29	5	1	60	10	6	7
UUR11	644	413	41	169	1	1	21	8	8	14	30	80	58	16	105	25	4	0	65	13	4	5
UUR12	650	416	36	170	1	1	20	8	7	14	30	73	53	18	112	27	4	1	68	12	6	3
UUR13	655	421	39	163	0	1	20	8	8	13	30	78	59	19	105	25	6	2	67	16	7	3
2033	646	398	41	190	18	1	19	9	11	11	32	47	46	15	113	25	5	2	58	10	8	0
2608	667	351	41	258	18	1	19	9	11	11	32	47	48	16	115	25	6	2	6	11	6	0
4155	684	401	40	225	20	1	19	9	11	11	32	46	48	16	114	25	6	2	56	11	9	0
4318	661	454	47	181	18	2	22	8	11	13	33	43	59	16	117	25	6	2	61	12	5	0
Average	**652**	**436**	**41**	**177**	**5**	**1**	**20**	**8**	**9**	**15**	**31**	**73**	**56**	**18**	**109**	**26**	**5**	**1**	**59**	**13**	**6**	**2**

based on all 19 sequenced ureaplasma genomes contains 1020 protein coding genes of which 758 genes have orthologs in at least one other ureaplasma strain, and 515 genes are universally conserved among all 19 strains (ureaplasma core genome). The number of genes identified only in the genome of single serovars (singletons) is 262. The average number of singletons per genome is 14, however the range is wide (0 singletons in ATCC UPA3 and 68 in ATCC UUR9). Table 2-4 compares the pan genomes of different sets of ureaplasma species.

Table 2-4. Pan genome of different *Ureaplasma* species sets

	All 19 strains	14 ATCC Serovars	*U. urealyticum* (14 strains)	*U. parvum* (5 strains)
Pan genome	1020	971	938	688
Core genome	515	523	553	538
Singletons	262	246	216	77
Clusters of Orthologous Genes(COGs)	758	725	722	688

It has been suggested that genes that are not affected by the selective pressure on mycoplasmas gradually mutate at a faster rate than genes whose sequences are highly conserved to a higher AT content and eventually are lost [8]. Therefore, the %GC content may point out which genes are important for ureaplasmas or have recently been acquired horizontally. We evaluated the percent GC content of all genes across the 19 sequenced strains. Genes encoding hypothetical surface proteins conserved across all ureaplasma strains with high GC content may play an important role for ureaplasmas in processes like adherence to mammalian cells and colonization. A histogram of the distribution of %GC values of the ureaplasma pan genome shows that core genome genes with assigned function generally have a higher

GC content than hypothetical genes (Figure 2-1). The median for the core genome was 27%GC, therefore genes with %GC higher than 27 are likely to be essential and/or acquired. The median for the hypothetical proteins was 24%GC. Considering that the ureaplasma genomes have an overall 25%GC content, it is likely that genes with GC content below 25% may be non-essential and on their way to be lost. The lowest GC content is of a hypothetical protein with only 13%GC content.

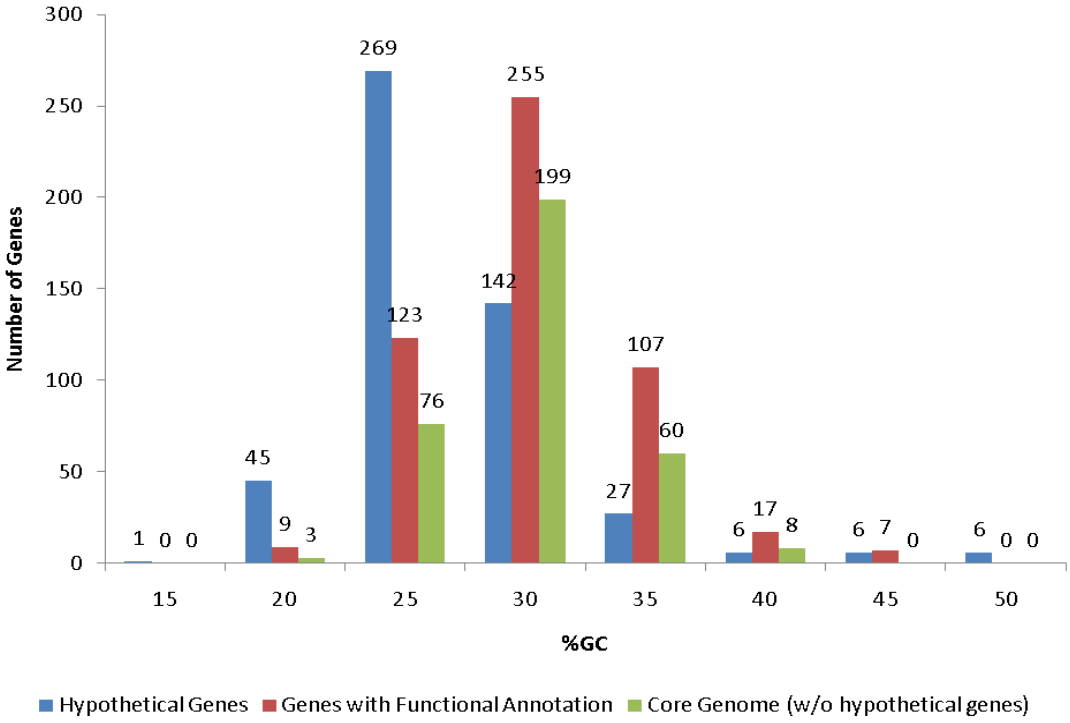


Figure 2-1. Percent GC Distribution Among Genes of The Ureaplasma Pan Genome (19 Strains).

For genes that are present in two or more genomes the average %GC of the orthologs was used. Values were grouped in bins (example: bin 20 contains genes with %GC from 15 to 20%). %GC of singleton genes was also included in the histogram.

The genomes of the 14 sequenced ATCC ureaplasma serovar strains showed extreme similarity between the two species and 14 serovars. The comparison of the finished genomes shows synteny on the gene level and not many rearrangements. We obtained percent difference values by whole genome comparison on the nucleotide level. The average intra-species percent difference was 0.62% with the least difference between UUR4 and UUR12 of only 0.06%, and the greatest difference between UUR9 and UUR13 of 1.27%. On the inter-species level the average percent difference was 9.5%, with the greatest difference between UPA1 and UUR9 of 10.2% (Table 2-5). Figure 2-2 shows the minor differences between the genomes of serovars 4 and 12.

Table 2-5. Serovar to Serovar Difference Expressed in Percent

The percent difference was obtained by whole genome comparison on the nucleotide level.

	1	3	6	14	2	4	5	7	8	9	10	11	12	13
1		0.66	0.52	0.75	9.90	9.99	9.68	9.78	9.66	10.23	9.84	9.70	9.93	9.79
3	0.70		0.49	0.35	9.93	9.67	9.33	9.43	9.33	10.01	9.43	9.36	9.66	9.84
6	0.62	0.52		0.50	9.82	9.82	9.40	9.49	9.38	9.95	9.53	9.42	9.76	9.75
14	0.83	0.33	0.45		9.92	10.01	9.59	9.69	9.57	9.99	9.70	9.60	9.95	9.83
2	9.82	9.87	9.58	9.81		0.86	0.74	0.78	0.76	1.25	0.74	0.77	0.86	0.84
4	9.90	9.60	9.57	9.83	0.94		0.69	0.64	0.69	0.82	0.88	0.66	0.07	0.80
5	9.72	9.31	9.25	9.52	0.72	0.60		0.15	0.13	0.66	0.56	0.16	0.58	0.66
7	9.72	9.32	9.25	9.52	0.82	0.60	0.16		0.15	0.66	0.53	0.11	0.60	0.67
8	9.76	9.35	9.27	9.54	0.71	0.59	0.08	0.10		0.61	0.51	0.11	0.59	0.65
9	10.90	9.83	9.60	9.71	1.21	0.72	0.63	0.62	0.60		0.85	0.63	0.75	1.08
10	9.79	9.35	9.29	9.56	0.70	0.81	0.51	0.48	0.51	0.87		0.46	0.80	0.43
11	9.73	9.33	9.25	9.52	0.80	0.61	0.16	0.11	0.16	0.67	0.51		0.60	0.64
12	9.85	9.58	9.52	9.79	0.93	0.06	0.67	0.64	0.69	0.85	0.87	0.65		0.80
13	9.70	9.74	9.47	9.66	0.97	0.86	0.79	0.76	0.75	1.27	0.56	0.74	0.86	

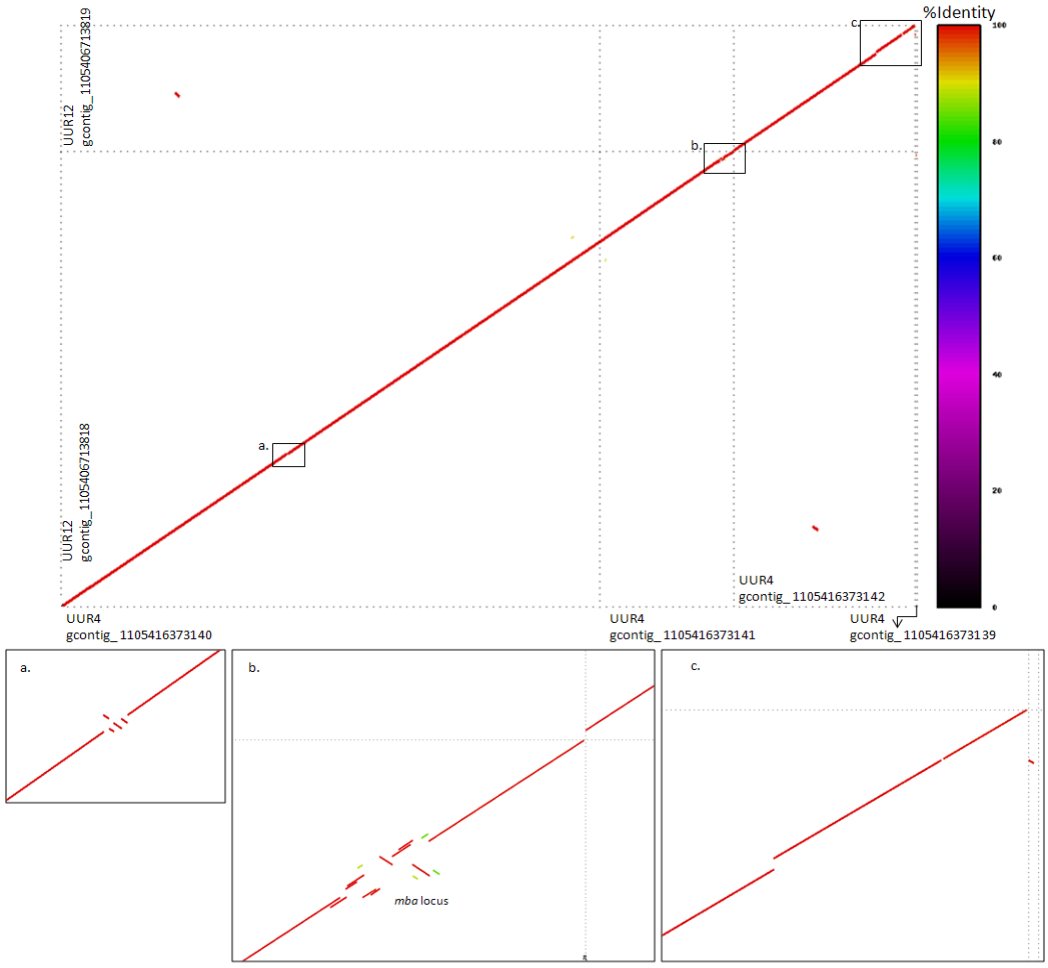


Figure 2-2. Dotplot of the genomes of UUR4 and UUR12.

Boxes a, b, and c are enlarged views of the 3 areas with minor rearrangements or differences between the two genomes. Box b shows the rearrangement of the *mba* locus.

As mentioned earlier, UUR serovars have about 118 Kbp (13.5%) larger genomes than UPA serovars. As a result UUR serovars have on average 58 genes more than UPA serovars. Fifty percent of these extra genes encode hypothetical proteins, the rest are spread among different functional categories (Table 2-3).

Table 2-6 shows the predicted genes present only in UUR serovars or only in UPA serovars. As it is seen in Table 2-3, UUR had more genes encoding cell surface proteins, DNA restriction modification enzyme genes and remnants of transposons (truncated genes or genes with unverified frameshifts). Furthermore, there are subtle differences in the predicted activities of proteins encoded by various reductase genes among serovars, which may facilitate unequal resistance of different ureaplasmas to oxidative stress during colonization and infection.

Table 2-6. Number of Clusters of Orthologous Genes (COGs) per functional category present only in UUR or UPA serovars.

Ureaplasma urealyticum

Only in UUs (none in UPs)	COGs
conserved hypothetical protein	67
putative lipoprotein	8
multiple banded antigen	6
putative membrane protein, degenerate	6
site-specific recombinase, TndX family protein	3
transposase, degenerate	3
DNA primase	2
AAA family ATPase	1
ABC transport ATPase component	1
chromosome segregation SMC protein homolg	1
DNA topoisomerase IV, B subunit	1
ISSod4, transposase, interruption	1
putative conjugative mobilization protein	1
putative merozoite surface protein 3	1
putative single-strand binding protein	1
restriction endonuclease family protein	1
superfamily II DNA and RNA helicase	1
TolAhomolog	1
TraG/TraD family	1

Ureaplasma parvum

Only in UPs (none in UUs)	COGs
hypothetical protein	14
type I restrictiomodification system	3
integrase - recombinase protein	2
protein of unknown function	2
nucleoside -2 deoxyribosyltransferase superfam	1
type I site specific deoxyribonuclease, HsdR	1

2.4.4 Ureaplasma phylogenetic tree

Constructing an accurate phylogenetic tree that resolves the relationship of ureaplasma serovars has been difficult due to the extreme similarity of these organisms on the genome level. Several methodologies exist for the construction of phylogenetic trees: single gene trees, trees based on concatenated gene sequences, gene content trees, and gene order trees. Phylogenetic trees based on single genes are unlikely to provide an accurate lineage of the serovars because of horizontal gene transfer among ureaplasmas. We find extensive horizontal gene transfer among clinical isolates relative to the 14 ATCC type strains [104]. Another challenge of building intra-species phylogenetic trees based on a single gene is that the primary nucleotide sequences of the genes conserved among all ureaplasma serovars/strains have such a high percentage of identity that there are not enough informative positions in the multiple sequence alignment to provide a resolution capability with high confidence. A gene content tree is based on a multiple sequence alignment in which each sequence (line) represents the genome of a strain and each position (column) in the multiple sequence alignment signifies the presence or absence of a gene in the strain. Therefore, such a tree has a binary nature (presence=1, absence=0). The pan genome of ureaplasmas generates a relatively short multiple sequence alignment: 1020 positions for 1020 genes in the pan genome. Therefore, a gene content tree of ureaplasma strains does not have the fine resolution capability of a phylogenetic tree based on nucleotide sequences. This can be noted in the low bootstrap values of the deep nodes of the gene content tree based on the pan genome that we constructed (Figure 2-3). We did not attempt to construct a gene order tree,

because the majority of the genomes are in multiple pieces, thus making it hard to judge the gene order in these genomes.

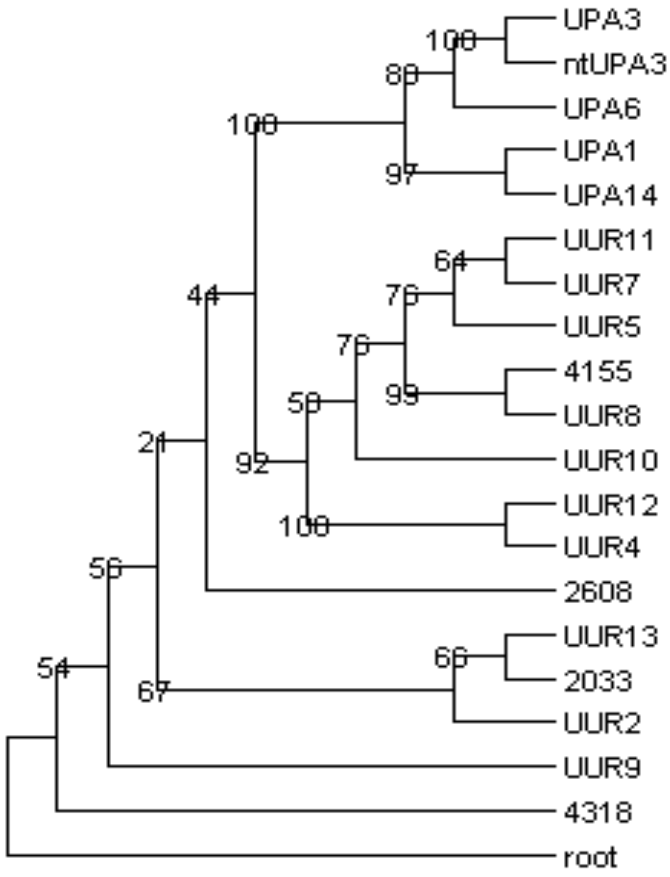


Figure 2-3. Phylogenetic tree based on the gene content of the pan genome of ureaplasmas.

Phylogenetic trees of ureaplasmas have been published previously, showing clear separation of the *parvum* and *urealyticum* species [159, 160]. The conserved domain of the *mba* genes has been used to generate a phylogenetic tree to resolve the relationship of serovars [98, 135]. We reconstructed the *mba* conserved domain tree using the first 430 nucleotides of the *mba* gene of all 19 strains (Figure 2-4).

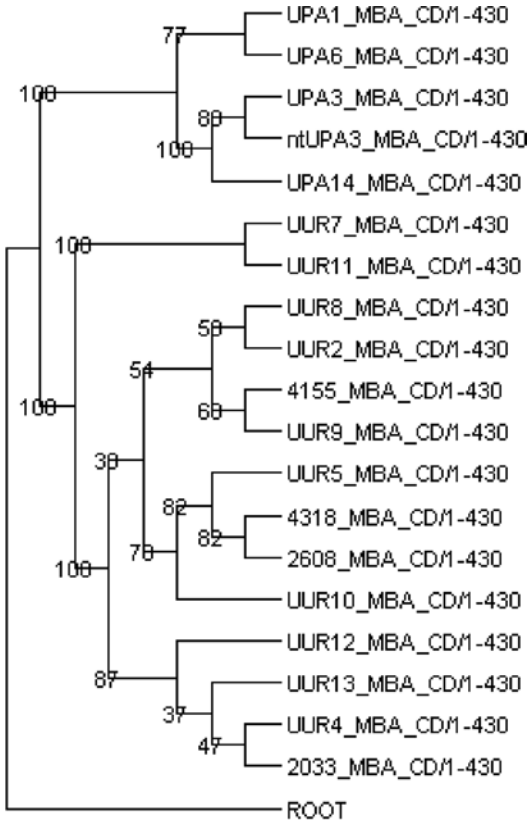


Figure 2-4. MBA Based Phylogenetic Tree of 19 Ureaplasmas.

The tree is based on the nucleotide sequence of the conserved domain of the *mba* (1-430 nt).

We also present a phylogenetic tree (Figure 2-5) based on the information of the nucleotide sequence of 82 housekeeping genes forming four groups: 1) 16 tRNA ligase genes 2) 12 RNA and DNA polymerase genes, 3) 47 ribosomal protein genes, and 4) 7 ureases. The clades of the multigene tree are very similar to the clades of the previously published *mba* based tree; however, the deep nodes of the two trees show some differences. These differences may be due to differences in the gene acquisition events that are averaged in a phylogenetic tree based on multiple genes versus a single gene tree. Similar differences in the deep tree nodes can be seen in the phylogenetic trees resulting from the concatenated alignments of the genes of each of the four groups and the trees resulting from different combinations of the groups (Figure 2-6, Figure 2-7, and Figure 2-8). However, as more genes are used to construct the trees, the clade and node structure of the trees becomes more consistent.

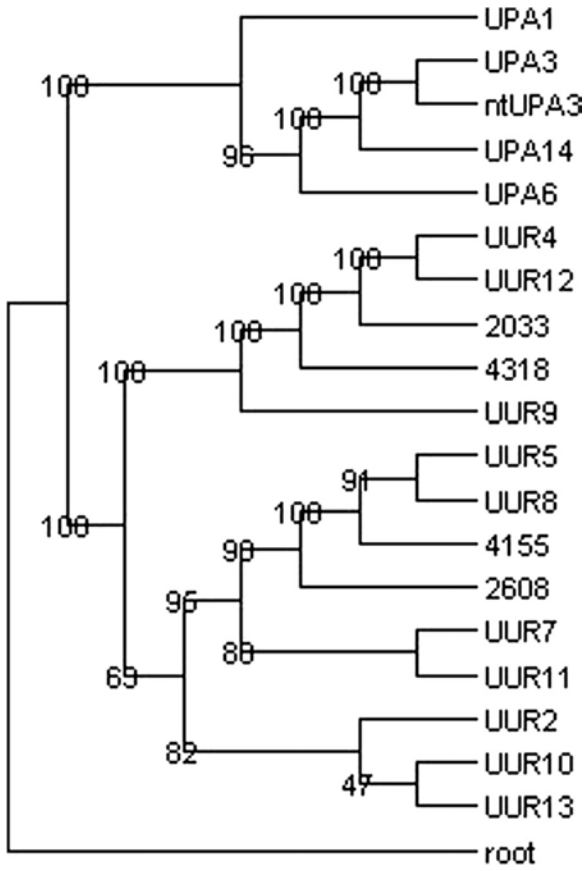


Figure 2-5. Phylogenetic tree of 19 ureaplasma strains based on 82 housekeeping genes.

ATCC type strains are labeled with tree letters (species) followed by a number (serovar). UUR=Ureaplasma urealyticum; UPA=Ureaplasma parvum; ntUPA3= clinical isolate sequenced in 2000; 2033, 2608, 4155, and 4318 are clinical isolates of Ureaplasma urealyticum that cannot be serotyped. The tree is based on the concatenated alignment of 82 housekeeping genes (16 tRNA ligase genes, 12 DNA and RNA polymerase genes, 47 ribosomal protein genes, and 7 urease genes). The non-informative positions were removed from the alignments. The removal of the non-informative positions increased the bootstrap values.

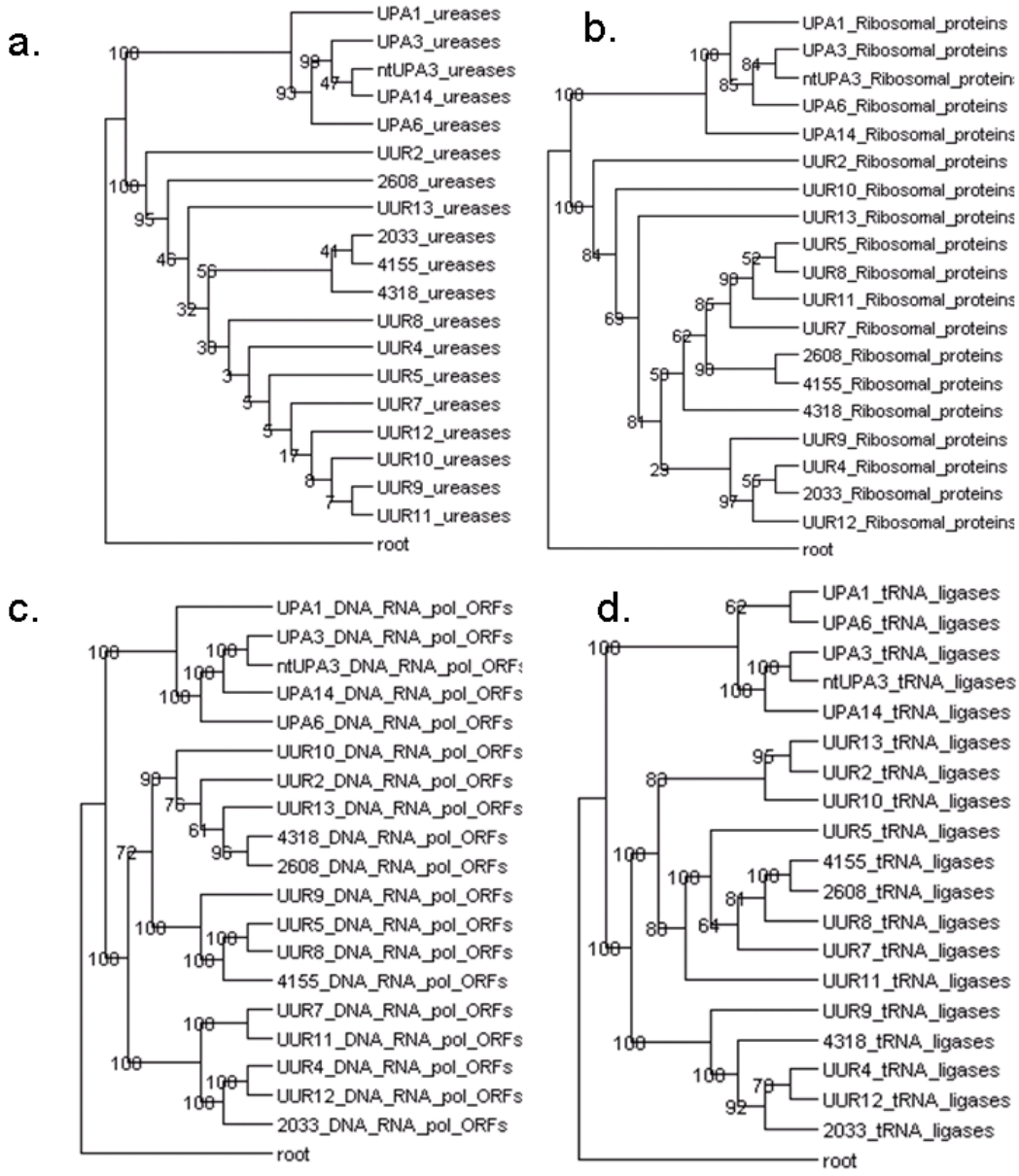


Figure 2-6. Phylogenetic trees based on the nucleotide sequence of multiple concatenated genes:

a. 7 urease genes; b. 47 ribosomal protein genes; c. 12 RNA and DNA polymerase genes; d. 16 tRNA ligases

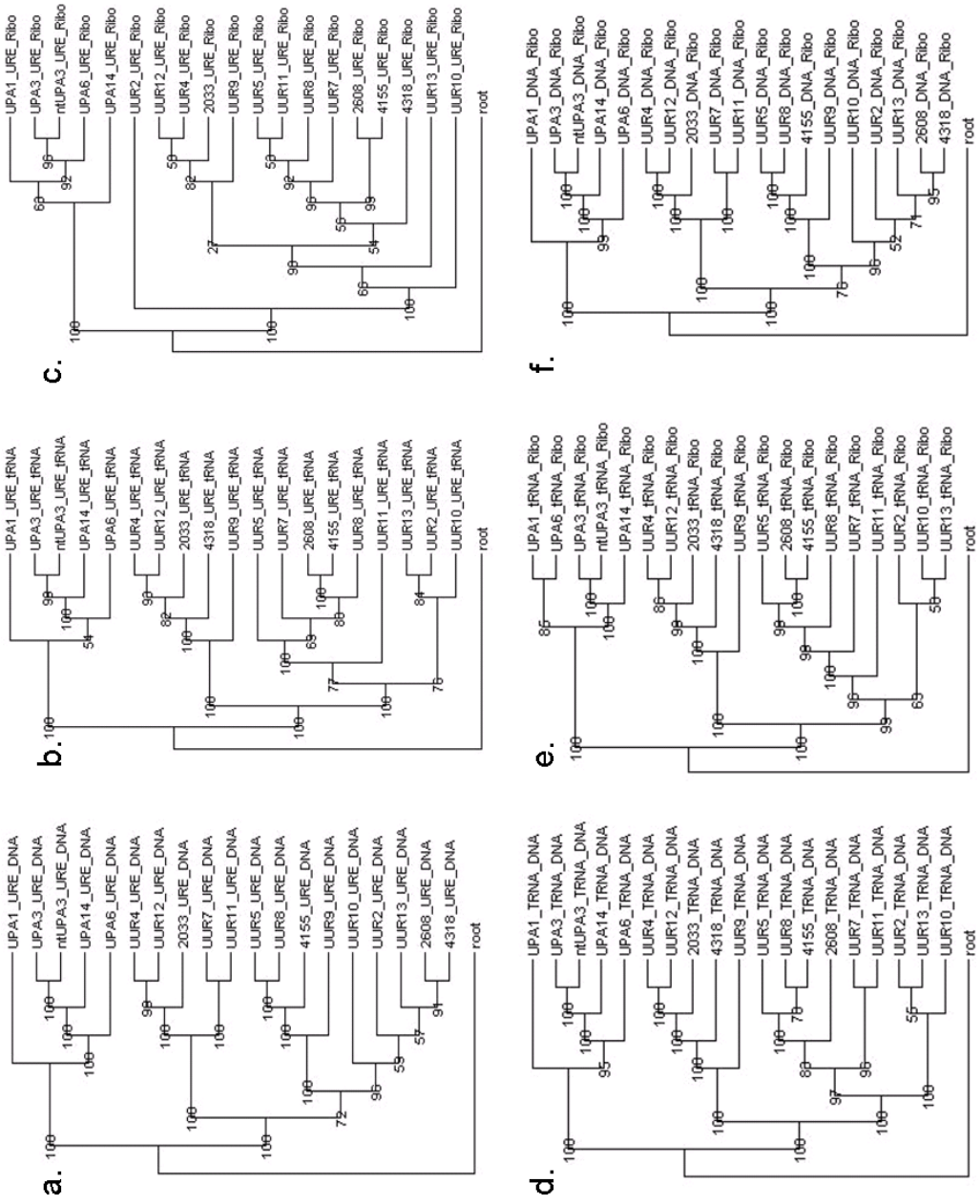


Figure 2-7. Phylogenetic trees based on the combination of two of the four groups of concatenated genes.

The nucleotide sequence of the genes was used; a. tree based on 7 usease genes and 12 RNA and DNA polymerase genes; b. tree based on 7 usease genes and 16 tRNA ligase genes; c. tree based on 7 urease genes and 47 ribosomal protein genes; d. tree based on 16tRNA ligase genes and 12 RNA and DNA polymerase genes; e. tree based on 16 tRNA ligase genes and 47 ribosomal protein genes; f. tree based on 12 RNA and DNA polymerase genes and 47 ribosomal protein genes.

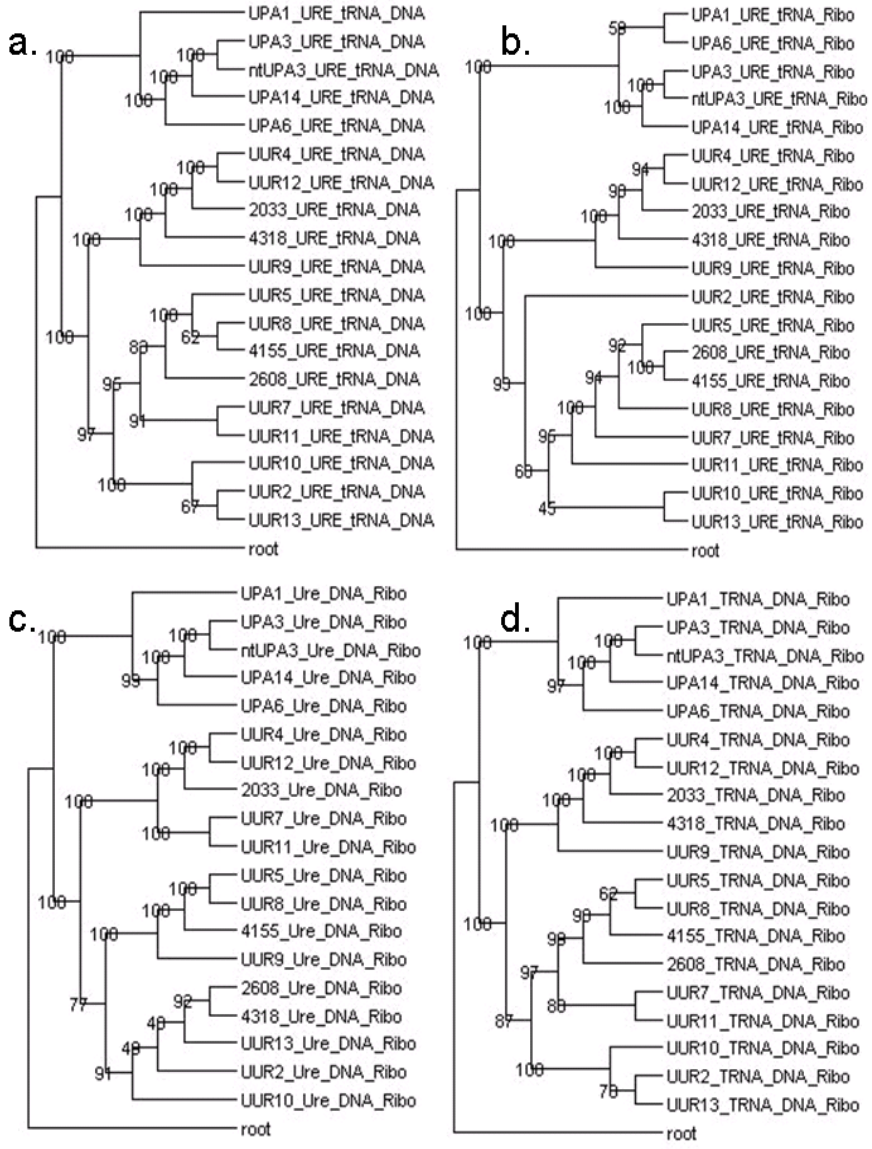


Figure 2-8. Phylogenetic trees based on the combination of tree of the four groups of concatenated genes.

The nucleotide sequence of the genes was used; a. tree based on 7 usease genes, 12 RNA and DNA polymerase genes and 16 tRNA ligase genes; b. tree based on 7 usease genes and 16 tRNA ligase genes, and 47 Ribosomal protein genes; c. tree based on 7 urease genes, 12 RNA and DNA polymerase genes, and 47 ribosomal protein genes; d. tree based on 16tRNA ligase genes, 12 RNA and DNA polymerase genes, and 47 ribosomal protein genes;

2.4.5 Recombination and Integration of DNA

All ureaplasma serovars contained one or more integrase-recombinase genes and some serovars contained transposases, or remnants of transposases, and some phage related proteins. Most of the recombinases were site-specific tyrosine recombinases, which are present also in other mycoplasmas and firmicutes. The highest number and variety of such genes was observed in serovar 2, and in general, UUR serovars had higher number of these genes than UPA serovars. However, insertion events represented only a small portion of the average 118 Kbp difference between the two species. A site-specific integrase-recombinase was adjacent to the phase variable locus of the MBA in 12 of the 14 serovars. This recombinase was likely involved in the rearrangements of the *mba* locus resulting in the variation of the C-terminal of this surface antigen. The presence of transposases suggested that foreign mobile DNA elements have been inserted in the genomes of ureaplasma serovars. Some of the transposases have truncations or unverified frameshifts indicating that the mobile element that they were part of was most likely no longer mobile. It was no surprise to find transposon related genes in serovar 9, which had acquired tetracycline resistance. The *tetM* gene was identified as part of a Tn916 transposon, based on the genes around it. Although tetracycline-resistant ureaplasma were probably less frequent when serovar 9 was isolated, now they comprise 25-35% of all patient isolates. A report covering the years 2000-2004 from several states in the USA showed that 45% of unique clinical isolates of *Ureaplasma* spp. contain *tetM* and are tetracycline-resistant [11]. Further evidence of genome integrated transposons were some of the site-specific recombinases found in the genomes:

TnpX, required for the excision of Tn4451 [11] and TndX, which was the first member of the large-resolvase subgroup of the resolvase/invertase family of site-specific recombinase shown to be able to mediate the insertion and excision of a conjugative transposon, more specifically Tn5397 [161].

A TraG/D family protein was recognized in serovars 9 and 13 (UUR9_0186 [GenBank: ZP_03079565] and UUR13_0031 [GenBank: ZP_02932006]). The TraG/D (transport) family genes aid the transfer of DNA from the plasmid into the host bacterial chromosome [162, 163], mediate the interactions between the DNA processing (Dtr) and mating pair formation (Mpf) systems during conjugation. Another suggestion for the capacity of horizontal gene transfer in at least some serovars is the presence of relaxases/mobilization proteins (UUR9_0148 [GenBank: ZP_03079581] and UUR13_0045 [GenBank: ZP_02696018]). Such proteins are required for the horizontal transfer of genetic information contained on plasmids that occurs during bacterial conjugation [164]. Aligning the genomes of the 14 ATCC ureaplasma genomes made evident two major insertion events. The first one was consistent with a transposon insertion, due to the repeat of some host sequence on both sides of the inserted region. At the time of insertion a short part of the 3' end of the *ruvB* was duplicated, so that the insertion was located between the full length *ruvB* gene and its short duplication. The insertion has been inherited by UPA1, 3, and 14 from a common ancestor. Some of the genes present in this insertion had orthologs in UUR serovars. The inserted DNA fragment was 11,822 bp long in UPA3 and 14, and 12293 bp in UPA1. It contained 8 genes, which encoded 6 hypothetical proteins, one hypothetical protein containing a subtilase domain, and one Type I specificity

subunit restriction protein. The second insertion was present in 9 of the 14 serovars (UPA3, and 6, UUR4, 5, 7, 8, 10, 11, and 12) and had a size of about 20 Kb.

Based on the fact that there were three phage genes in the insert, we believe that this event is due to a phage insertion into the genomes. The first gene of the insertion encodes an integrase-recombinase protein that contains a phage integrase domain (UPA3_0153 [GenBank: YP_001752228]). A phage recombination protein Bet (UPA3_0162 [GenBank: YP_001752237] is encoded further downstream of the integrase and the final gene in the insert encodes a phage terminase, large subunit, of the pbsx family (UPA3_0176 [GenBank: YP_001752251]. The rest of the genes are hypothetical proteins, however some of them encode one or more transmembrane domains and/or signal peptides, suggesting that they may play a role on the surface of the ureaplasma cell. It is important to note that the same exact insertion regions have been identified through a comparative genomic microarray analysis of 10 UPA clinical strains [134]. In this comparative genome microarray study these two insertions were present in some isolates of the same serovar and absent in other isolates of the same serovar. The authors suggest the phage insertion might be a putative pathogenicity island. Although the C+G content of the insertion is less than 1% higher than the rest of the genome, Momynaliev and colleagues [134] found that GCGC and CGCG tetranucleotides, that are present in ureaplasma DNA fragments, were missing in the inserted DNA fragment, thus providing another clue of the foreign character of the inserted DNA fragment.

Examining the putative restriction-modification (RM) genes in the 14 serovars suggests that, although each serovar has from six to twelve RM genes, most RM

systems are incomplete. Serovars 3, 5, 7, 8, 10, and 11 may have a complete type III RM system, serovar 9 may have a complete type I and type II RM system, whereas serovars 1, 14, 2, 12, and 13 appear to have only remnants of RM systems. RM systems are used in general by organisms to protect themselves from foreign DNA like viruses. Although phages that infect ureaplasmas have not been reported, the existence of these RM systems, as well as the presence of either intact or remnants of RM systems in the other urogenital mycoplasmas *M. genitalium* and *M. hominis* suggests that there are phages that infect these obligate parasites. In organisms like *Chlamydia* spp., which are obligate intracellular parasites and have no identifiable infecting viruses, there are no functional RM systems [165].

2.4.6 Potential pathogenicity genes

2.4.6.1 Phospholipase C, A1, A2

Phospholipase C, A_1, and A_2 (PLC, PLA1, PLA2) activity was reported in Ureaplasma serovars 3, 4, and 8 by DeSilva and Quinn [153-155]. It is important to note that the assay used by DeSilva measures combined activity of PLC and phospholipase D (PLD) because both cleavage products are in the soluble fraction and the radioactively labeled hydrogen would be found in both cleavage products [166]. PLC activity has been reported in *Ureaplasma diversum* cells as well, and has been suggested to play a role in ureaplasma invasion in mammalian cells [167]. However, the detection method used the artificial substrate p-nitrophenylphosphorylcholine (p-NPPC), which can be hydrolyzed by several other enzymes that can hydrolyze phosphate esters, including PLD [168]. All 14 ATCC

ureaplasma serovar genomes and the genome of the previously sequenced clinical isolate of UPA3 were extensively evaluated for the presence of PLC, PLA1, and PLA2 genes. No genes showed significant similarity to known sequences of PLC, PLA1, or PLA2 in any of the genomes. HMMs developed for known PLC, PLA1, and PLA2 did not detect any ureaplasma genes with significant similarity. This suggested that ureaplasma may encode phospholipases that are either very degenerate or have evolved separately from known phospholipases as previously suggested by Glass et al [8], or that no phospholipase genes are present in *Ureaplasma* spp. It is interesting to note that a PLD domain containing protein was easily identified. In all serovars this protein is annotated as cardiolipin synthase (UPA3_0627 [GenBank YP_001752673]).

We used two PLC assays to test ureaplasmas for PLC activity: Invitrogen's Amplex® Red Phosphatidylcholine-Specific Phospholipase C Assay Kit, which detects also PLD activity, and the original PLC assay published by DeSilva and Quinn. We were not able to detect PLC or PLD activity in ureaplasma cultures of serovars 3 and 8. Our attempts to repeat De Silva and Quinn's PLC assay using L-a-dipalmitoylphosphatidylcholine – (choline-methyl-3H) with UPA3 and UUR8 cultures grown to exponential phase and processed to collect the cell membranes and cleared cell lysates as described in their original publications [153-155] failed to replicate the specific activity levels they reported in ureaplasma cultures. Because we were not able to find PLC, either computationally or experimentally, we believe that this gene is not present in ureaplasmas. However, a study done by Park et al. suggests implication of PLD in the signaling cascade that activates COX-2, leading to

production of prostaglandins and initiation of labor [169]. Since all ureaplasma serovars and the four sequenced clinical isolates contain a gene with PLD domains, a future functional characterization of this gene would be of interest. We have not been able to find computationally the genes encoding PLA1 and PLA2 in ureaplasmas.

2.4.6.2 IgA Protease

In the mammalian immune system, a primary defense mechanism at mucosal surfaces is the secretion of immunoglobulin A (IgA) antibodies. Destruction of IgA antibodies by IgA specific protease allows evasion of the host defense mechanism. In *Neisseria gonorrhoeae* the IgA protease doubles as a LAMP-1 protease to allow it to prevent fusion of the phagosome with the lysosome [170]. IgA protease activity was demonstrated in ureaplasma serovars [171, 172]. All sequenced human ureaplasma genomes were evaluated for IgA protease genes with the same methods as the phospholipases gene search. We could not computationally identify an IgA protease gene. It is possible that the IgA protease activity detected in all ATCC serovars could be due to the expression of a protease capable of cleaving the IgA antibody, as a its secondary function. Therefore, this protease does not have the typical IgA protease amino acid sequence and conserved domain architecture and has another primary protease function. We obtained the complete lest of Pfam families from the Pfam database, which holds a large collection of protein families, each represented by multiple sequence alignments and HMMs [173]. We used this list to create a new list containing all Pfam families that described any proteases or peptidases. This resulted in 166 pfam families at the time of our search in 2011. We used this list to search our ureaplasma genome annotation database to find all ORFs in all14 ureaplasma

genomes which sequences matched these pfam families. We found a total of 236 ORFs which sequences matched at least one of these pfam families. We used our comparative genome analysis and BLAST to organize the 236 genes in COGs. Since all UU serovars showed IgA protease activity we filtered out all ORFs that were not present in all 14 serovars. This filter narrowed our result to 10 potential ORFs that may be genes encoding a protease with potentially secondary function as an IgA protease. These genes could be cloned and expressed in *E. coli* or another bacteria that does not express an enzyme capable of cleaving IgA antibodies and the potential capability of these proteases to also cleave IgA antibodies could be explored. Ureaplasmas, utilize a non-standard genetic code in which the UGA codon codes for the amino acid tryptophan. When this codon is present in the candidate gene it will need to be changed to UGG through site directed mutagenesis in order to be translated properly in bacteria that utilize the standard genetic code. Table 2-7 contains the 10 protease genes that are conserved in all 14 sequenced.

Table 2-7. Proteases Conserved in All 14 ATCC Serovars.

These genes could be further investigated for IgA protease activity. The UGA codon in ureaplasmas codes for tryptophan instead of a stop codon and would need to be corrected for proper expression in *E. coli* or other bacteria with standard genetic code.

Ureaplasma urealyticum serovar 8 (NCBI locus tag)	Description	No. of UGA codons	Comments
UUR8_0075	Oligoendopeptidase	10	broad specificity
UUR8_0116	ATP-dependent Metalloprotease FtsH	3	transmembrane domain
UUR8_0251	Methionine Aminopeptidase	3	
UUR8_0282	Glutamyl Aminopeptidase	1	annotated as cellulase in some UU serovars
UUR8_0321	Glycoprotease family protein	0	
UUR8_0361	ATP-dependent Lon protease	2	Serine protease region
UUR8_0462	O-sialoglycoprotein endopeptidase	1	*P. haemolytica* pathogenicity factor
UUR8_0596	ABC transporter	2	transmembrane domain and peptidase C39 family
UUR8_0614	Oligoendopeptidase F	9	
UUR8_0627	M24 Metallopeptidase family	2	

2.4.6.3 Nucleases

Nucleases have been reported as potential pathogenicity factors in other organisms as well [174]. Ureaplasmas belong to a group of organisms that import nucleotides for DNA and RNA synthesis. Therefore it is likely that they have secreted or surface bound nucleases that may also play a role in pathogenicity. We identified 15 potential nucleases, of which two had a predicted signal peptide, and thus are likely to be secreted or surface bound. These nucleases may be an interesting target for further studies of their potential involvement in pathogenicity.

2.4.6.4 Putative O-sialoglycoprotein peptidase

All 14 ureaplasma serovars contained a gene annotated as an O-sialoglycoprotein endopeptidase (UPA3_0428 [GenBank: ACA33260]). This enzyme has been shown to cleave human erythrocyte glycophorin A in other bacteria [175]. The same study showed that the specificity of this peptidase is limited to O-glycosylated membrane glycoproteins, and it cannot cleave N-glycosylated proteins. Abdullah et al. [175] suggest that the potential targets of this enzyme in the host are sialoglycoproteins of the mucosal epithelial cells or on the cell surfaces of macrophages. In fact the O-sialoglycoprotein peptidase of *Mannheimia haemolytica* cleaves from the surface of the human cell line KGla the CD43-leukosialin and other human O-sialoprotein antigens like the progenitor cell-restricted antigen CD34, the hyaluronate receptor CD44, and the leukocyte common antigen tyrosine phosphatase CD45 class of molecules [175]. If the ureaplasma putative O-sialoglycoprotein peptidase is capable of cleaving such targets, this could be a mechanism for evasion

of the host immune system, colonization of the host, and eventually establishment of an infection. In *M. haemolytica* isolates the presence of this gene is associated with the capacity of the bacteria to cause pneumonia in calves [175].

2.4.6.5 Macrophage infection mutant protein, MimD

UUR2 contained a gene annotated *mimD* (UUR2_0526 [GenBank: ZP_03771352]) standing for macrophage interaction mutant D. *Mycobacterium marinum* is a fish, amphibian, and human pathogen that may be able to survive and replicate in macrophages. A study of macrophage infection *D. marinum* mutants identified a mutation in a hypothetical protein that resulted in this phenotype [176]. The exact function of this gene in interactions with macrophages is not yet defined; however the ureaplasma annotated *mimD* gene (183 aa) had 40% identity and 68% similarity over 179 aa long alignment with the *M. marinum mimD* gene (731 aa). Further characterization of MimD in other systems and possibly ureaplasma would be interesting.

2.4.6.6 Resisting hostile environment

Bacteria are known to produce substances that give them competitive advantages over other bacteria in their environment. Some of these substances are bacteriocins (like mutacin produced by *Streptococcus mutans*) and H_2O_2 to inhibit the growth of other bacteria [177]. UUR13 has two of the three suggested genes involved in immunity to mutacin, *mutE* and *mutG* [178]. A glutathione peroxidase gene [GenBank: ACA33207.1] was found in all UPA serovars. Peroxiredoxin [GenBank: ZP_03772062] was present in all the UUR serovars. These genes could

play a role in resisting oxidative stresses and bacteriocins produced by the rest of the bacteria on the mucosal surfaces they occupy. We detected a thioredoxin reductase system in all 19 genomes [GenBank: ACA33034 and NP_078428]. The thioredoxin reductase system has been described previously in mycoplasmas and functions as a detoxifying system to protect the organism from reactive oxygen compounds [179]. The presence or absence of such genes in an individual ureaplasma strain may contribute to the difference of pathogenic potential of the strain.

2.4.6.7 Multiple Banded Antigen (MBA) Superfamily

The original classification of ureaplasma isolates into distinct serovars was largely based on differences in the major ureaplasma surface antigen called the multiple banded antigen (MBA) (8-10, 12). MBA consists of an N-terminal conserved domain and a C-terminal variable domain. The conserved domain contains a signal peptide, lipoprotein attachment site, and one transmembrane domain. While the conserved *mba* domains for all 14 serovars had been sequenced previously, for most serovars sequencing of the variable domain, which was thought to be serovar specific, was only partial [180-182]. Our whole genome data confirmed that variable regions usually consist of tandem repeating sequence/units (TRU). Only in UUR13 is the conserved domain attached to a variable domain that does not contain any tandem repeats. The same variable domain is found also in UUR12 and UUR4; however it is not attached to the conserved domain of the *mba* in these serovars. The MBA is recognized by the Toll-like receptors 1, 2, and 6, and is capable of inducing the cytokine, NF-κB and antibody production [123]. It is conceivable that ureaplasmas would have evolved strategies to vary the MBA in order to evade this response.

Ureaplasma isolates can vary the number of the tandem repeats of their *mba* gene in response to challenge with antibodies presumably by slipped strand mutagenesis [183]. Furthermore, *mba* can phase vary with neighboring genes, and UPA3 was recently shown to produce a chimeric gene though phase variation by fusing the N-terminal part of an *mba* paralogous gene (UU172 [GenBank: CBI70486] to its neighboring gene (UU171 [GenBank: NP_078003]) [99, 184]. These findings suggest that *mba* and some *mba* paralogous genes might be involved in strategies for evading the host immune system employed by ureaplasmas.

One of the surprises of our whole genome analysis and comparison of the 14 ATCC serovars showed the *mba* genes to be part of a large complex gene superfamily comprising 183 UPA and UUR genes and 22 subfamilies (Table 2-8). There were a limited number of unique variable domains as shown in Table 2-9.

Table 2-8. Clusters of Orthologous Genes Potentially Involved in the MBA Phase Variable System of Ureaplasmas.

This table contains the NCBI locus tags for genes potentially involved in the MBA phase variable system . Genes with tandem repeats are highlighted in green. A red box is drawn around the MBA genes expressed in the ATCC type strains.

	HP	HP	HP	HP	MBA-N	MBA	HP	HP	HP	MBA-N	HP	HP	MBA-N	MBA-N	MBA	MBA	MBA	MBA	MBA		
Serovar	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	Total
UPA1 (gus03)	UPA1_G0402 UPA1_C0001 UPA1_D0001 UPA1_I0001 UPA1_I0002	UPA1_G0112	UPA1_G0118	UPA1_G0420	UPA1_G0174	UPA1_F0003 UPA1_B0001 UPA1_G0002		UPA1_F0002 UPA1_E0002	UPA1_G0001 UPA1_F0001 UPA1_J0001 UPA1_A0001														18
UPA3 (gus01)	UPA3_0179	UPA3_0501	UPA3_0505	UPA3_0196	UPA3_0563	UPA3_0390	UPA3_0391																7
UPA6 (gus04)	UPA6_A0161	UPA6_A0511	UPA6_A0515	UPA6_A0178	UPA6_A0571	UPA6_A0411		UPA6_A0412	UPA6_A0413														8
UPA14 (gus05)	UPA14_G0051	UPA14_E0115	UPA14_E0119	UPA14_G0068		UPA14_G0263 UPA14_B0001		UPA14_E0001 UPA14_D0001 UPA14_G0264															9
UUR2 (gus06)	UUR2_0166	UUR2_0529	UUR2_0533	UUR2_0183	UUR2_0604	UUR2_0396	UUR2_0393			UUR2_198	UUR2_0094	UUR2_0096	UUR2_0093	UUR2_0657	UUR2_0399	UUR2_0394	UUR2_0544						15
UUR4 (gus07)	UUR4_0061	UUR4_0393	UUR4_0388	UUR4_0044	UUR4_0314	UUR4_0673	UUR4_0675			UUR4_0030	UUR4_0134	UUR4_0132	UUR4_0135	UUR4_0263	UUR4_0671	UUR4_0672	UUR4_0374						15
UUR5 (gus08)	UUR5_G0187	UUR5_J0019	UUR5_K0003	UUR5_G0204	UUR5_K0071	UUR5_E0067	UUR5_E0064			UUR5_G0219	UUR5_G0119	UUR5_G0121	UUR5_G0118					UUR5_E0070					12
UUR7 (gus09)	UUR7_0167	UUR7_0539	UUR7_0543	UUR7_0184	UUR7_0613	UUR7_0421				UUR7_0198	UUR7_0091	UUR7_0093	UUR7_0090	UUR7_0666					UUR7_0422	UUR7_0424			13
UUR8 (gus02)	UUR8_0167	UUR8_0542	UUR8_0546	UUR8_0184	UUR8_0619	UUR8_0422 UUR8_0006	UUR8_0418 UUR8_0002			UUR8_0198	UUR8_0097	UUR8_0099	UUR8_0096	UUR8_0672			UUR8_0558		UUR8_0426				16
UUR9 (gus10)	UUR9_0558	UUR9_0177	UUR9_0172	UUR9_0541	UUR9_0686	UUR9_0631	UUR9_0629			UUR9_0527	UUR9_0248	UUR9_0246		UUR9_0378			UUR9_0747 UUR9_0213						13
UUR10 (gus11)	UUR10_0163	UUR10_0540	UUR10_0544	UUR10_0180	UUR10_0615	UUR10_0415	UUR10_0414			UUR10_0194	UUR10_0094	UUR10_0096	UUR10_0093	UUR10_0669	UUR10_0419	UUR10_0416	UUR10_0556						15
UUR11 (gus12)	UUR11_0162	UUR11_0527	UUR11_0531	UUR11_0179	UUR11_0599	UUR11_0411				UUR11_0193	UUR11_0090	UUR11_0092	UUR11_0089	UUR11_0650					UUR11_0410	UUR11_0413			13
UUR12 (gus13)	UUR12_A0163	UUR12_A0527	UUR12_A0531	UUR12_A0180	UUR12_A0597	UUR12_A0418	UUR12_A0419			UUR12_A0194	UUR12_A0089	UUR12_A0091	UUR12_A0088	UUR12_A0648	UUR12_A0422	UUR12_A0420					UUR12_A0421		15
UUR13 (gus14)	UUR13_0347	UUR13_0021	UUR13_0017	UUR13_0330	UUR13_0595	UUR13_0161				UUR13_0317	UUR13_0415	UUR13_0413	UUR13_0416	UUR13_0542	UUR13_0159	UUR13_0162					UUR13_0136		14
Total	18	14	14	14	13	18	9	6	5	10	10	10	9	9	5	5	4	2	2	2	2	2	183

Table 2-9. Tandem Repeating Units (TRUs) Identifies in the *mba* locus.

The name of each TRU consists of the mba gene name followed by the period size (bp) of the repeating unit. Different sequences of the same period size are marked by "." and a version number (ex. mba18.1 and mba18.2). Observed minimum and maximum copy number of the TRU is shown in the third column. Column 6 shows the serovar in which the conserved domain was associated with each TRU. Note that the conserved region of the UPA1 mba was found linked to two different TRUs (highlighted).

Name	Period Size (bp)	Copy # in sequenced ATCC	Serovars	Thought to be unique for serovar	Conserved domain attached in serovar (clinical isolate)	Clinical Isolates of UU; unknown serovar
***mba*12bp**	12	60.8	6	6	6	-
***mba*18bp.1**	18	36.7-53.7	1	1	1	-
***mba*18bp.2**	18	40.6	3	3	3	-
***mba*21bp**	21	29.5-32.0	14	14	14	-
***mba*24bp.1**	24	20.2-33.5	2,5,8	5	5 (2608, 4318)	2608, 4318, 4155
***mba*24bp.2**	24	34.6	10	10	10	-
***mba*30bp**	30	17.2-26.2	4,12,13	4	4 (2033)	2033
***mba*42bp**	42	7.6-11.6	7,10,11	11	11	-
***mba*45bp**	45	2.0-10.0	2,5,8,9	9	9	4155
***mba*213bp.1**	213	3.0-4.0	4,10,12,13	-	-	2033
***mba*213bp.2**	213	2.8-3.9	2,5,8	2	2	4155
***mba*213bp.3**	213	1.9	2	-	-	-
***mba*231**	231	2.8-3.9	7	7	7	-
***mba*252bp.1**	252	1.9-5.9	8,9,11	8	8	4155
***mba*252bp.2**	252	2.1-4.1	4,10,12,13	12	12	-
***mba*252bp.3**	252	2.0-3.0	2,5	-	-	-
***mba*276bp**	276	2.0-3.8	2,8,9	-	(4155)	2608, 4318
***mba*327bp**	327	2.3-4.0	1	-	1	-
***mba*330bp**	330	4	10	-	-	2608
***mba*333bp**	333	3.0-4.0	4,12,13	-	-	2033, 4318
***mba*336bp**	336	2.9	6	-	-	-
***mba*579bp**	579	1.9	5	-	-	-

We found that all UUR serovars and UPA1 and 6 had more than one tandem repeating unit type in their *mba* locus. Although some of these TRUs have not yet been observed to be attached to the conserved domain of the *mba*, they are surrounded by inverted repeats that contain a putative recombinase recognition site. This suggested that these TRUs were involved with the *mba* and contributed to surface antigen variation. The UPA serovars had a simpler MBA phase variation systems than the UUR serovars: the UPA conserved domain was surrounded by inverted single base pair repeats, containing the 25 base pair putative recombinase recognition site and a site-specific recombinase gene (UPA3_0388 [GenBank: YP_001752455]) was located near the *mba* locus (Figure 2-9 and Figure 2-10). The inverted repeats and the site-specific recombinase were potentially involved in inverting the orientation of the transcriptional promoter and conserved domain in order for expression to occur with one or the other TRU. It is interesting to note that one TRU was short and had a high copy number (18 nt – UPA1, 12 nt – UPA6, repeated >30X) and the other one was long and had a low copy number (327 nt – UPA1, 336 nt – UPA6, repeated <5X). Rearrangements of the *mba* locus were evident in the smaller contigs of unfinished serovar genomes (Figure 2-9 and Figure 2-10).

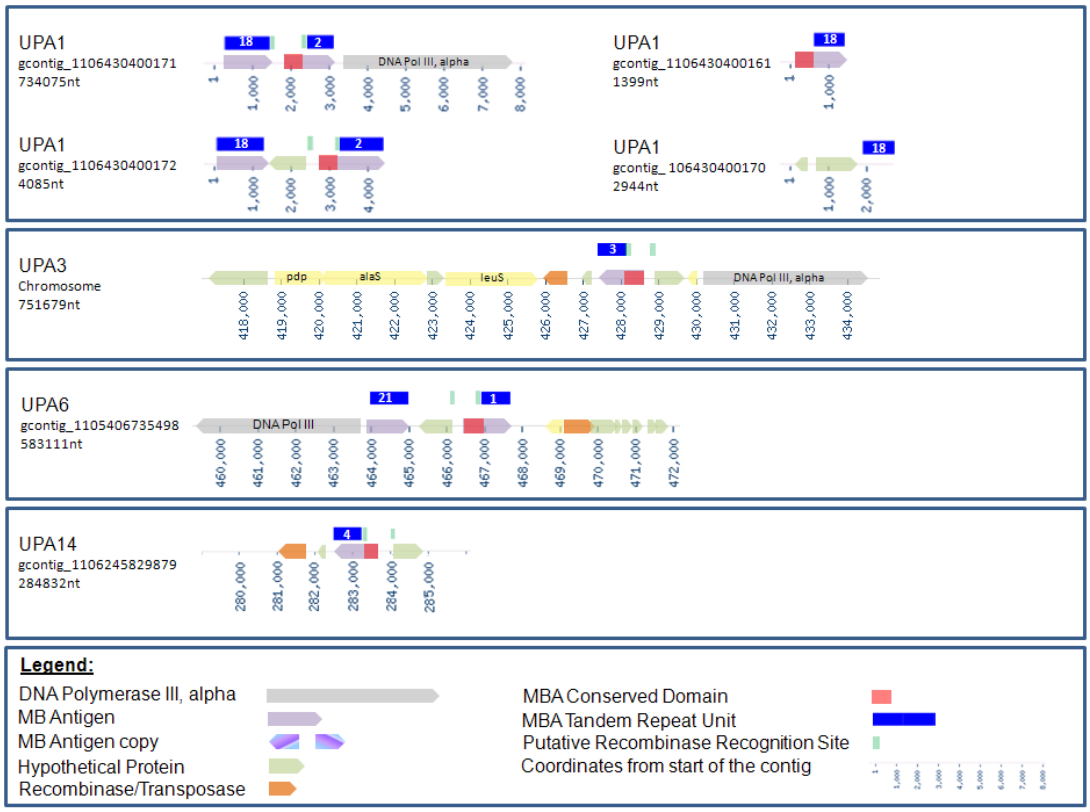


Figure 2-9. *Ureaplasma parvum* Multiple Banded Antigen Locus.

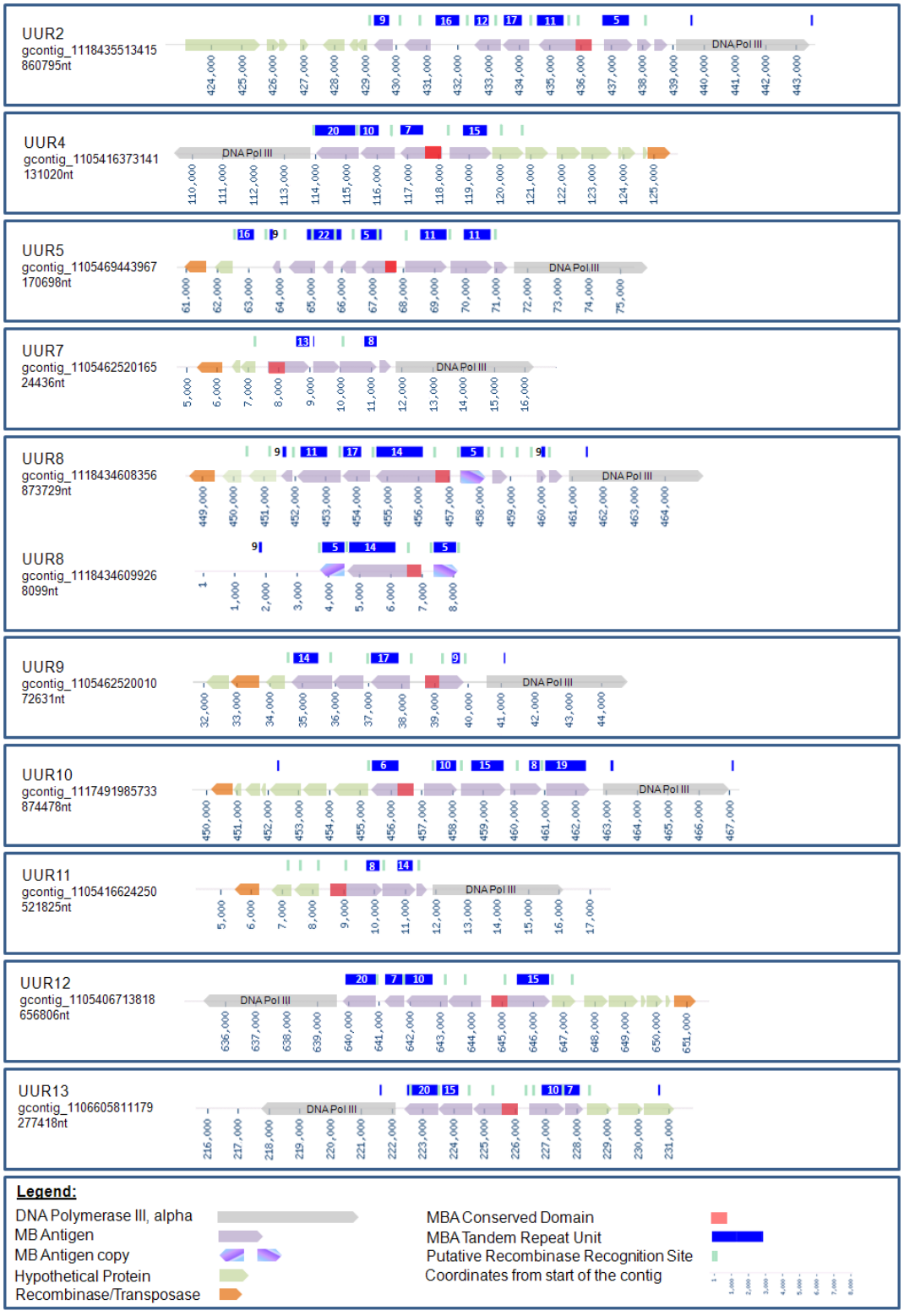


Figure 2-10. *Ureaplasma urealyticum* Multiple Banded Antigen Locus.

UPA1 genome sequencing data clearly shows a sub-population in which the conserved domain of the *mba* is attached to the alternative TRU ([GenBank: NZ_ABES01000008] – assembly 17, [GenBank: NZ_ABES01000003] – assembly 20; Figure 2-9 & Table 2-8) and another subpopulation in which another gene is present between the two TRUs ([GenBank: NZ_ABES01000002] – assembly 21). The high repeat number of the *mba* TRUs, and the existence of a subpopulation in the culture being sequenced that has a rearrangement of the *mba* locus, represent an ambiguity for the assembly software, resulting in the generation of smaller alternative contigs that cannot be assembled into the chromosome. The alternative 327 nt *mba* TRU of UPA1 is on a 1399 nt long contig [GenBank: NZ_ABES01000008] that contains only this gene, and it ends truncating the 327 nt TRU at only 2.3 repeats compared to 4 repeats on the main contig. Furthermore, comparing the two variations of the *mba* locus makes evident the break-points where the flip of the conserved domain occurred. This coincides with the sites of the inverted repeats suspected to be part of the mechanism for MBA phase-variation. This represents sequencing evidence that this serovar could express both variations of the MBA at different times.

All UUR serovars have more than two TRUs in close proximity to each other. Serovars UUR7 and UUR11 have only 2 TRUs each, whereas UUR2 and UUR5 have 6 TRUs each, which is the maximum number of TRUs observed. The largest *mba* loci are around 10 KB and have 6 TRUs and some non-TRU *mba* genes. Each *mba* locus contains only one conserved domain. The loci are always located adjacent to the DNA pol III alpha subunit (except UPA14) and on the other side of the loci there is a putative Xer-C site-specific recombinase. Next to each TRU there is a putative 25 nt

recombinase recognition sequence

[ACTTT(T/C)TCT(G/C)TTTGATAATT(C/A)AAAT]. The same recognition site is located next to some non-TRU genes in the loci, therefore making them likely to be involved in this phase variable superfamily. Furthermore, serovar 13 has a non-TRU variable domain fused to the conserved domain of the *mba*, confirming that the variable unit does not necessarily require tandem repeats. An interesting observation is that UUR4, 12 and 13 have the same *mba* locus composition in 3 different rearrangements (Figure 2-11).

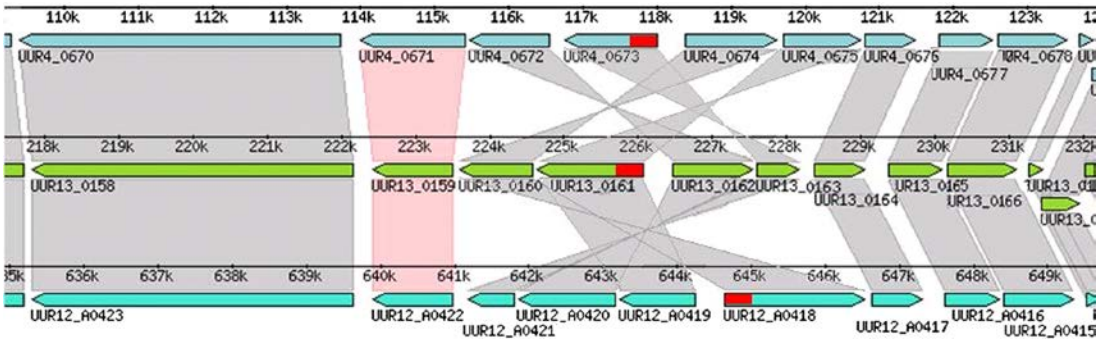


Figure 2-11. The MBA Locus in UUR4, UUR12, and UUR13.

Genes in each genome are represented as directional blue or green boxes. Orthologous gene clusters (COGs) are represented by gray or pink bands spanning across the tree genomes. The COG with a pink band represents the first *mba* gene in the MBA locus. The locus includes the next 4 genes following the gene in the pink labeled COG (all tree genome have 5 *mba* genes each). The conserved domain of the *mba* is marked by a red box. Rearrangements of the genes are visible by following the twisting of the connecting bands.

Most TRUs were found to be present in more than one serovar. By carefully analyzing small contigs in unfinished ureaplasma genomes, we identified variations of the *mba* loci. For example, on a small contig of UUR8 [GenBank: NZ_AAYN02000001] we saw a partial *mba* locus arranged alternatively by duplicating one of the TRUs in the locus. Examining the sequencing and assembly data of such contigs confirms that these contigs are not misassembled, but rather represent a subpopulation of the sequenced culture. The proposed mechanism for variation of the ureaplasma *mba* locus resembles the previously reported variable loci of *Mycoplasma bovis*: *vis*, *Mycoplasma pulmonis: vrs* and *Mycoplasma agalactiae: vis*-like [185] in that a site-specific Xer-1 recombinase and inverted repeats are likely to be involved and a putative recombinase recognition site has been determined. The *mba* locus resembles the *M. pulmonis vrs* locus in that it has only one promoter and one conserved domain per *mba* locus, which needs to be moved in front of a variable domain to make a functional surface MBA.

Examination of the *mba* loci of the four sequenced UUR clinical isolates that cannot be assigned to a serovar shows that the *mba* conserved domain is UUR specific. Due to the repetitive nature of the *mba* TRUs the loci are broken into multiple contigs, making it impossible to determine the exact order of the genes in the *mba* loci without further sequencing. Isolate 2033 had 4 identifiable TRUs (*mba*333bp, *mba*213bp.1, *mba*30bp and the non-tandemly repeating unit of UUR13; see Table 2-9). Of these, *mba*30bp was found attached to the conserved domain of the MBA and is the equivalent of the active TRU in UUR4. The same TRU was also present in the mba loci of UUR12 and UUR13. Isolate 2608 contained 3 identifiable

TRUs (*mba*24bp.1, *mba*267bp, and *mba*330bp). The conserved domain was found attached to *mba*24bp.1, as in UUR5; this TRU was also present in UUR2 and UUR8. Clinical isolate 4318 had 3 identifiable TRUs (*mba*24bp.1, *mba*276bp, and *mba*333bp). The conserved domain was attached to *mba*24bp.1. Isolate 4155 had 5 identifiable TRUs (*mba*24bp.1, *mba*45bp, *mba*213bp.2, *mba*252bp.1, and *mba*276bp). The conserved domain was attached to *mba*276bp; this TRU had not been previously seen attached to a conserved domain in any of the 14 ATCC type strains, including the clinical UPA3 described by Glass et al. [8]. This is a further confirmation that the TRUs found in the *mba* locus are part of this phase variable system, which trough recombination should be capable to present on the surface of the ureaplasma cell different TRUs at different times. It would be interesting to investigate whether some TRUs are more immunogenic than others and therefore may contribute to differential pathogenicity. As mentioned earlier the *mba* variable domain has been used as one of the determinants of serovar classification. It is interesting to note that serovars 4 and 12, which have an identical set of MBA genes, have a percent difference at the nucleotide level in a whole genome comparison (Table 2-5) of only 0.06 or 0.07% (value depends on which genome is used as reference sequence). In addition, we observed two chimeric *U. parvum* strains in a clinical isolate that had exchanged through horizontal gene transfer their *mba* genes [104]. Taken together, these observation suggest that the *mba* locus is dynamic and can comprise of a different set of variable domains at different times, therefore making this gene an unsuitable target for serovar differentiation unless serovar designation is solely based on the currently expressed MBA variant.

2.5 Conclusions

Ureaplasmas have been associated with many different clinical outcomes; however, they have been detected also in healthy individuals. Due to their differential pathogenicity, effort has gone into assignment of patient isolates into serovars and attempting to correlate specific serovars with specific clinical outcomes. Analysis of ureaplasma samples obtained from patients in the 1970s identified 14 different serovars based on patient and animal antiserum reactions. The expanded serotyping scheme developed by Robertson and Stemke in 1979 is based on antiserum generated by injecting rabbits with emulsified preparations of cell suspensions of each strain separately [186]. Studies were not done at this time to determine the antigen that the sera antibodies were recognizing. In a later study, Watson et al. (1990) reported the finding of an antigen recognized by infected humans that contains serovar-specific and cross-reactive epitopes. This antigen presented a multiple banded pattern on immunoblots, wherefore, it was named multiple banded antigen (MBA). The same study tested only 4 patient sera in blocking experiments with monoclonal antibodies; therefore, it is not possible to deduce the exact antigens for all serovars involved in the serotyping of the 14 serovars. Because of the suggested serovar-specific epitopes of the MBA, this protein has been used in attempts to develop better serotyping techniques. However, the cross-reactivity between serovars still could not be eliminated. Comparing the 14 genomes of the ATCC type serovars enabled us to better understand why there is cross-reactivity when attempting to use anti-MBA antibodies for serotyping. This is due to the fact that all ATCC serovars have more than two possible MBAs, each expressed at different times, through a phase variable

gene system. There was a limited number of unique variable domains, however, it was showed that one such unique variable domain unit was exchanged/acquired by horizontal gene transfer [104], suggesting that the *mba* locus is dynamic and can acquire or lose variable domains. Therefore the MBA genes are not suitable for a serotyping tool.

Ureaplasmas have been shown to adhere to different eukaryotic cells although their adhesins have not been identified. Experiments done to gain a better understanding of the adhesion properties of ureaplasma showed that cytadherence involves N-acetylneuraminic acid (NANA) as a ligand receptor molecule. The same study showed that ureaplasma adherence was significantly lower, but not inhibited by neuraminidase treatment, therefore, there are additional receptors that do not involve NANA [187]. Our comparative genome analysis of the 14 ATCC serovars showed that ureaplasmas have a great variety of genes coding for surface proteins and lipoproteins, however the majority of these genes could not be assigned a function, since they were orthologous of genes coding for proteins of unknown function or the predicted gene did not have an ortholog outside of the *Ureaplasma* genus. It is possible that such genes will have a higher GC content, due to their importance for the organism, in which case the %GC table could be used together with signal peptide and transmembrane domain prediction to identify candidate genes to study for adherence properties. The MBAs are part of the surface proteome of the ureaplasmas and have been shown to be recognized by the toll like receptors (TLR) and induce NF-κB production [123]. Recognition by the TLR can elicit the release of inflammatory chemokines and cytokines that in turn trigger prostaglandin production

in the amnion, chorion, deciduas and myometrium, leading to uterine contractions and eventually may lead to pre-term labor. The variety of MBA variable domains and the capacity of the organism to vary their sizes and switch between variable domains could mean that different MBAs, when recognized by the TLRs, may have a different capacity to activate the innate immune system [36]. The fact that the MBA variable domain is recognized by patient antibodies and antibody pressure leads to phase variable switch in their size or the variable domain [183] suggests that the different variable domains could be used for host immune system evasion.

Although we expected to find evidence of differential pathogenicity on the serovar level, the majority of the differences among the two species and the serovars are in genes encoding proteins for which we could not assign functions. There are a limited number of potential pathogenicity factors that could be recognized computationally. The previously shown activity of IgA protease in all 13 tested serovars [171, 172, 188] can be an important tool for host immune system evasion in the mucosal surfaces, however we could not identify the gene responsible for this enzyme activity computationally. The ureaplasmal IgA protease may be a novel IgA protease. We believe that one of the predicted genes, which contain a protease functional domain in their sequence may be responsible for the observed protease activity.

PLC, PLA1 and PLA2 activity was also demonstrated previously [153-155] and has been thought to be a potential pathogenicity factor and contributor in adverse pregnancy outcomes. None of the genes encoding these enzymes was found in the 14

ureaplasma genomes computationally. Our attempts to detect PLC activity with a PLC commercial assay and by repeating the original experiments were unsuccessful.

Studies involving clinical isolates of ureaplasma have revealed hyper-variable DNA regions that may potentially harbor genes aiding the pathogenicity of ureaplasmas [134] and chimeric ureaplasma isolates revealing overwhelming evidence of extensive horizontal gene transfer in these organisms [104], which can explain the cross-reactivity of sera. Taken together these findings suggest that there might be innumerable serovars or strains based on different combinations of horizontally transferred genes. Our comparative genome study has identified genes that could support horizontal gene transfer. These genes combined with the observed chimeric clinical isolates of ureaplasma suggest that these organisms possess active recombination mechanisms. Therefore, it is possible that ureaplasmas do not exist as stable serovars in their host, but rather as a dynamic population.

We do know that UUR causes non-gonococcal urethritis in males and pelvic inflammatory disease (PID) and/or endometritis in pregnant women more frequently than UPA; however no other clinical outcome is significantly more associated with either species or a particular serovar [104, 189-194]. We cannot identify any clear gene or constellation of genes that might account for greater UUR virulence in some situations; although we do note a difference in the genes whose products are associated with resistance to H_2O_2, a known microbial pathogenicity factor. The widely different clinical outcomes of ureaplasmal infection could be the result of the presence or absence of potential pathogenicity factors in the colonizing ureaplasma strain. Alternatively, it may be more likely that the different clinical outcomes are

either all or in part the result of patient to patient differences in terms of autoimmunity and microbiome.

Future studies of ureaplasma biology should concentrate on the development of molecular tools for the generation of ureaplasma gene knock-out mutants for example, in order to study genes potentially involved in pathogenicity. The sequenced genomes can aid in the development of such tools, by identifying transposons, integrated phage genomes, and genes involved in horizontal gene transfer. To aid the identification of potential pathogenicity factors, the large collection of clinical isolates should be explored for presence/absence of candidate genes. Considering the low cost of sequencing nowadays, the genomes of isolates from patients with different conditions should be sequenced and their comparison should further aid the identification of genes involved in differential pathogenicity.

Acknowledgements

This work was supported by NIAID award 5R01AI072577. I gratefully acknowledge the direct contribution to this work by Jin Lu, Lynn B Duffy, Donna M Crabb, Susmita Shrivastava, Barbara A Methé, Jason Inman, Shibu Yooseph, Li Xiao, Gail H Cassell, Ken B Waites, and John I Glass as well as the assistance and contributions to this project by our J. Craig Venter Institute colleagues, Michael Montague, Elisabeth Caler, Sanjay Vashee, Mikkel Algire, Nacyra Assad-Garcia, Diana Radune, Jessica Hostetler, Scott Durkin, Jonathan Crabtree, and Jonathan Badger.

Chapter 3. Detection and Characterization of Human Ureaplasma Species and Serovars by Real-time PCR

3.1 Abstract

We designed primers and probes for detection and discrimination of Upa and UUR and their 14 serovars by real-time PCR. Analytical sensitivity and specificity for the multiplex species-specific PCR were determined by testing corresponding American Type Culture Collection (ATCC) type strains, 47 other microbial species, and human genomic DNA. The limits of the multiplex PCR were $2.8x10^{-2}$ CFU/μl PCR mixture for detecting *U. parvum* and $4.1x10^{-2}$ CFU/μl PCR mixture for detecting *U. urealyticum*. Clinical specificity and sensitivity were proven by comparison with culture and traditional PCR. For detection of any *Ureaplasma* species, the clinical sensitivity and specificity of real-time PCR were 96.9% and 79.0%, respectively, using culture as reference. Multiplex real-time PCR was more sensitive than traditional PCR in discriminating the two *Ureaplasma* species in culture-positive subcultures. Each of the 14 monoplex serovar-specific PCR assays was specific for the corresponding ATCC type strain serovar. The serovar-specific PCRs were compared with previous immunoblotting results using monoclonal antibodies and results were in complete or partial agreement with serological methods for 35 (70%) of 50 isolates tested. This new speciation PCR is specific and sensitive in detection of ureaplasma species in clinical specimens and the serovar-specific PCR assays are the first set of complete genotypic assays to differentiate all 14 known *Ureaplasma*

serovars. These assays provide quick and reliable means for investigating the epidemiology and pathogenicity of ureaplasmas at the serovar level.

3.2 Introduction

Shepard provided the first descriptions of T-strain mycoplasmas, now known as ureaplasmas, in 1954 when he cultivated them *in vitro* from the urethras of men with nongonococcal urethritis [1], but the genus *Ureaplasma* was not formally designated until 1974 [195]. *Ureaplasma urealyticum*, eventually shown to be comprised of two biovars, was considered to be the only species of this genus known to infect humans until 2002 when its two biovars were reclassified as two distinct species, *U. parvum* and *U. urealyticum*, based on genome size, sequences of the 16S rRNA gene, the 16S-23S rRNA intergenic region, enzyme polymorphisms, DNA-DNA hybridization, differential growth response to manganese, and differences in the multiple-banded antigen (*mba*) genes [7]. At least 14 serovars have been identified in human ureaplasmas; *U. parvum* contains serovars 1, 3, 6, and 14, while *U. urealyticum* contains the remaining 10 serovars 2, 4, 5, 7, 8, 9, 10, 11, 12, and 13 [7].

The two *Ureaplasma* species can be isolated singularly or simultaneously from clinical specimens [11]. They are common opportunistic pathogens colonizing the urogenital tract in healthy persons, yet they are also implicated in invasive diseases, such as urethritis, post-partum endometritis, chorioamnionitis, spontaneous abortion, as well as premature birth, and low birthweight, pneumonia, bacteremia, meningitis, and chronic lung disease in prematurely born infants [11]. The reasons why *Ureaplasma* spp. produce clinically significant infections in some instances and

do not in others are still unknown although an intact functional humoral immune system appears to be very important as evidenced by the frequent occurrence of systemic ureaplasmal infections such as septic arthritis in adults and children with antibody deficiencies [11]. Speculations about the association of particular *Ureaplasma* species or serovars with certain diseases have provided incentive for investigations to develop accurate methods for determining the serovars of clinical isolates [11].

Typing methods for *Ureaplasma* serovar designation using polyclonal or monoclonal antibodies (MAbs) directed against whole cells or purfied antigens have included growth inhibition tests [196, 197], metabolic inhibition tests [186], immunofluorescence of colonies on agar [198-200], immunoperoxidase [201], enzyme-linked immunosorbent assay (ELISA) [6], and immunoblots [180, 202]. These antibody-based techniques are time-consuming, cumbersome assays that often yield results that are not reproducible, difficult to interpret and inconclusive because of multiple cross-reactions and poor discriminating capacity in clinical samples containing two or more serovars. Lack of commercial availability and standardization generally limited serotyping to those laboratories that developed the serological reagents.

New rapid molecular methods based on genotypic instead of phenotypic markers are logical choices to replace the conventional antibody-based serotyping methods. PCR-based methods are also becoming an important alternative to conventional culture for initial detection of ureaplasmas in clinical specimens and have the additional advantage of discrimination between the two *Ureaplasma* species

[83]. The earlier gel-based traditional PCR assays for detection and speciation of human ureaplasmas targeted sequences of the 16S rRNA gene and 16S-23S rRNA intergenic spacer regions [83, 84, 86, 87], the urease gene subunit [81, 88], or the multiple-banded antigen (*mba*) genes [86, 89-92]. Major issues remaining to be resolved for traditional PCR are its specificity and analytic sensitivity, as well as potential for contamination. Real-time PCR assays targeting the urease gene [94-96], *mba* gene [93] and 16S rRNA [203] have recently been developed to distinguish the two *Ureaplasma* species. Compared to traditional PCR and culture, real-time PCR speciation assays are quantitative, as well as more rapid, specific, sensitive and less subject to contamination [96]. However, some of previously published real-time PCR assays available for ureaplasma speciation require two separate tests [95, 96] or do not have satisfactory sensitivity in one combined test [94].

To characterize the *Ureaplasma* species at the serovar level by PCR, genotyping primers based on the *mba* gene and its 5′end upstream regions have been designed previously and partial serovar identification was achieved. In combination with direct sequencing or restriction enzyme analysis, these assays were capable of distinguishing the 4 serovars of *U. parvum* and divided the 10 serovars of *U. urealyticum* into different subgroups [89, 90, 96, 98, 204, 205]. Among these PCR serotyping methods, only one was real-time based [96]. Because of limited sequence variation in the *mba* genes, all PCR-based methods reported so far have lacked the capacity for complete separation of all 14 serovars. Moreover, our personal experience with whole genome sequencing of all 14 serovars has shown *mba* to be part of a large gene family present in many variations in different serovars and the

gene is phase variable [99]. Thus, while the *mba* gene is still in some cases a valid and in fact necessary target for serovar specific PCRs, attention must be given to the locations of phase variable forms of these genes so that PCR amplicons do not span recombination sites within these genomes. We evaluated the complete genome sequences for all 14 *Ureaplasma* serovars and designed primers and probes targeting new specific regions for speciation and serotyping by real-time PCR assays. One species-specific multiplex PCR and 14 serovar-specific monoplex PCRs were developed and assessed for analytical and clinical specificity and sensitivity.

3.3 Material and Methods

3.3.1 Bacterial strains and clinical specimens.

Fourteen type strains for the *Ureaplasma* serovars designated in Table 3-1 were obtained from the American Type Culture Collection (ATCC) and used to evaluate specificity and sensitivity of the multiplex speciation and monoplex serotyping assays. Twenty-one *Mycoplasma* and *Acholeplasma* species, 21 other bacterial species, 1 *Candida* species, 4 viruses, as well as human gDNA were tested by real-time PCR to determine specificity of the multiplex UU speciation assay. These organisms were selected based on probability to be in human samples and availability in our laboratory. Most of these microorganisms were from ATCC (live strains or DNA only), but some were clinical isolates. *Mycoplasma amphoriforme* was obtained from M.F. Balish (Miami University, Oxford, OH, USA). Viral DNA was obtained from Dr. Fred Lakeman (Department of Pediatrics, UAB). Human genomic DNA was obtained from Roche Diagnostics, (Indianapolis, IN).

Table 3-1. Bacterial species used to validate real-time PCR assays

Name	Source	Name	Source
Ureaplasma species		*Mycoplasma muris*	ATCC 33757
U. parvum serovar 1	ATCC 27813	*Mycoplasma neurolyticum*	ATCC 15049
U. parvum serovar 3	ATCC 27815	*Mycoplasma primatum*	ATCC 15497
U. parvum serovar 6	ATCC 27818	*Mycoplasma pulmonis* A	ATCC 19612
U. paruvm serovar 14	ATCC 33697		
U. urealyticum serovar 2	ATCC 27814	Other Bacterial species	
U. urealyticum serovar 4	ATCC 27816	*Bordetella pertussis*	ATCC BAA-589D
U. urealyticum serovar 5	ATCC 27817	*Burkholderia cepacia*	ATCC 25416
U.urealyticum serovar 7	ATCC 27819	*Chlamydophila pneumoniae*	clinical isolate
U. urealyticum serovar 8	ATCC 27618	*Corynebacterium diphtheriae*	ATCC 10701
U. urealyticum serovar 9	ATCC 33175	*Enterobacter cloacae*	clinical isolate
U. urealyticum serovar 10	ATCC 33699	*Enterococcus faecalis*	ATCC 29212
U. urealyticum serovar 11	ATCC 33695	*Escherichia coli*	ATCC 25922
U. urealyticum serovar 12	ATCC 33696	*Haemophilus influenzae*	ATCC 23533
U. urealyticum serovar 13	ATCC 33698	*Haemophilus parainfluenzae*	ATCC 7901
		Helicobacter pylori	clinical isolate
Human Mycoplasma species		*Klebsiella pneumoniae*	ATCC 33495
Acholeplasma laidlawii PG 8	ATCC 23206	*Legionella pneumophila*	ATCC 33152
Acholeplasma occuli	ATCC 27350	*Moraxella catarrhalis*	ATCC 25240D
Mycoplasma amphoriforme	clinical isolate	*Neisseria gonorrhoeae*	ATCC 49226
Mycoplasma buccale	ATCC 23636	*Proteus mirabilis*	ATCC 2798
Mycoplasma faucium	ATCC 25293	*Pseudomonas aeruginosa*	ATCC 27853
Mycoplasma fermentans PG 18	ATCC 19989	*Staphylococcus aureus*	ATCC 29213
*Mycoplasma genitalium*PG37	ATCC 33530	*Streptococcus agalactiae*	ATCC 13813
Mycoplasma hominis PG21	ATCC 23114	*Streptococcus pneumoniae*	ATCC 49619
Mycoplasma lipophilum	ATCC 27790	*Streptococcus pyogenes*	ATCC 19615
Mycoplasma orale	ATCC 23714	*Streptococcus salivarius*	ATCC 13419
Mycoplasma penetrans	ATCC 55252	*Candida* species	clinical isolate
Mycoplasma pirum BER p9	clinical isolate		
Mycoplasma salivarium PG 20	ATCC 23064	Viruses	
Mycoplasma spermatophilum	ATCC 49695	Adenovirus	clinical isolate
		Buffalopox virus BP4	clinical isolate
Non-human Mycoplasma species		Herpes simplex virus type 1	clinical isolate
Mycoplasma arthritidis PG 27	ATCC 23192	Herpes simplex virus type 2	clinical isolate
Mycoplasma collis	ATCC 35278		
Mycoplasma hyorhinis	clinical isolate	Human genomic DNA	Roche Diagnostics

Three groups of clinical specimens or microorganisms maintained in the culture collection of the UAB Diagnostic Mycoplasma Laboratory were used in this investigation. They consisted of: (1) 132 neonatal nasal or endotracheal aspirate specimens obtained between 2004 and 2007; (2) 194 low passage (≤ 3) *Ureaplasma*-positive broth subcultures of neonatal nasal or endotracheal aspirates obtained between 1994 and 2003; and (3) 50 low passage (≤ 3) *Ureaplasma*-positive broth subcultures of placental specimens obtained between 1987 and 1991 and serotyped by immunoblotting using serotype-specific MAbs described previously [202]. Bacterial subcultures, clinical specimens, and ureaplasmal serological reagents were kept frozen at -80^{o}C until processed for speciation and serotyping PCR assays. Cultures were performed as described by Waites, et al [206].

3.3.2 DNA preparation for PCR.

Genomic DNA was extracted by the proteinase K method as described previously [207]. Briefly, 200 μl of broth subculture or 1 ml of clinical specimens was centrifuged at 14,000 g for 20 minutes at 4^{o}C; pellet was digested with 200 μl proteinase K (1 mg/ml) lysis buffer for 1 hour at 60^{o} C. Proteinase K was then inactivated by incubation at 95° C for 5 minutes. In case of inhibition, proteinase K digested samples were purified using QIAamp® DNA Blood Mini Kit (Qiagen, Valencia, CA). Genomic DNAs of 14 *Ureaplasma* serovar ATCC type strains were purified by both methods for sensitivity tests. Prepared DNA samples were stored at -80°C.

3.3.3 PCR Primers and Probes.

Primer/probe sets were designed using Roche LightCycler Probe Design Software 2.0. A multiplex real-time PCR was designed to differentiate the two *Ureaplasma* species simultaneously. The primers anneal to genes present only in *U. parvum* or *U. urealyticum* (Table 3-2). The *U. parvum* primer/probe set anneals to the 477 bp UP063 gene (NP_077893), which encodes a conserved hypothetical protein that is identical in all 4 *U. parvum* serovars. The *U. urealyticum* primer/probe set anneals to a 15,072 bp open reading frame (ORF) that is almost perfectly (>99.97%) conserved in all 10 *U. urealyticum* serovars (serovar 10 Genbank ID ACI59931.1). The amplicon size of the two species was the same (152 bp).

Table 3-2. Primers and probes for UU species-specific multiplex PCR assays

Species	Target	Name	Sequence (5' to 3')
U. parvum	UU063	UP063#1F	Tgcggtgtttgtgaact
		UP063#1R	Tgatcaaactgatatcgcaattataga
		UP063#1 Probe 1	tggtttaacgtgtttttgaagtgctacaaaat-FL
		UP063#1 Probe 2	LC Red 640-cccatttcagccatggtgccatca-phosphate
U. urealyticum	UUR10_0554	UU1271F	Cagtagcaaatcgtgcttaca
		UU127#1R	Tcattaaaatcatttgcactagtcaaaata
		UU127#1 Probe 1	gataataacacttggacaatttttaaccaaagcga-FL
		UU127#1Probe 2	LC Red 705-aaggattagagttttgttgccatggtagtcaaa-phosphate

To design the serovar-specific primers/probes, all 14 genomes were computationally searched to identify 25-35 base pair candidate regions that were unique to each genome. Each of the candidate region/loci sequences in a genome was then searched against all the other genomes and only those that had <80% identity matches were kept. The subsequent number of the primer loci to test varied in the different serovars, from 12 (serovar 12) to 1887 (serovar 9). Unique regions of the genomes comprised of mobile genetic elements, such as the transposon containing a tetracycline resistance marker in serovar 9, were not used as targets. We used the Roche LightCycler Probe Design Software 2.0 to analyze the regions around some of the unique 30 base pair regions and primers and probes were generated. Each possible primer was checked against all 14 UUgenomes using BLAST and less than 1 in 20 of the unique small regions yielded acceptable serovar-specific primers based on computational analysis. The resulting primers and probes shown in Table 3-3 were tested with the LightCycler using the 14 ATCC type stains. Most serovar-specific primers or primer/probe sets annealed to conserved hypothetical genes. The specificity of 8 serovars (serovars 2, 3, 6, 7, 8, 9, 13, and 14) came from the 2 primers while the other 6 were from probes, either Simple Probe Chemistry (SPC) probes (serovars 1, 5, 10, and 11) or fluorescence resonance energy transfer (FRET) probes (serovars 4 and 12). Serovar 5 primers may also amplify the other 9 *U. urealyticum serovars,* however, the amplicon is much larger (668 bp versus 121 bp in serovar 5) and the probe has a large unmatched loop (547 bp) so this should not affect specificity. A short extension time that only allows complete amplification of serovar 5 fragment but not the others enables the distinction. A melting curve analysis should

also be able to tell the differences of the products if unexpected amplification occurred. Urease-based primers and ^{32}P-labled probes for *Ureaplasma* speciation by traditional PCR were described previously [208]. All primers were synthesized by Invitrogen (Carlsbad, CA) and probes were manufactured by Roche Diagnostics.

Table 3-3. Primers and probes for UU serovar-specific monoplex PCR assays

Primer names contain serovar number (ex. UP1- *U. parvum* serovar 1); MBA: Multiple Banded Antigen; CHP: Conserved Hypothetical Protein;

Primer Name	**Sequence (5' to 3')**	**Description**
UP1F1-2	GATAATTTGAATTATCAAACAGAAAAAGTG	MBA UPA1_G0002 gcontig:1106501367851 Amplicon: 116 bp
UP1R1-2	TGTTCTTTACCTGGTTGTTGT	
UP1 probe 1	Fluorescein-SPC-GAACCAAAAGAAAATGGTGGAGAACAACC-Phosphate	
UU2_F_1	CTTGTTGATAAGCAAAAGGATTAC	putative lipoprotein UUR2_0261 (gcontig_1118448734166) Amplicon: 76 bp
UU2_R_1	GTTTTGGTTCTGGATTGTTTGAC	
UP3F1-2	AATAATGCAAATTATGATGTTAAATTAACC	CHP UPA3_0391/UU376 (NC_010503) Amplicon: 63 bp
UP3R1-2	TGTTTCTATGTTAACATTTAACAATTTAGC	
Uu04_1F	CTAAAGCGTGCTTAATGTGT	Intergenic gcontig_1105428130534 Amplicon: 162 bp
Uu04_1R	GTTAGTTTGGGAACCACCT	
Uu04_1 probe 1	TGTTATTATTATAGAAATTGTTGTAAGAATCAAAG-FL	
Uu04_1 probe 2	LC Red 610-ATAATTTCGTATAATGTCATGTTGTTCACCTCCTT-Phosphate	
Uu05-3F	ATTATGAAAAATTAAACTCTCATTACTCG	ATP-dependent RNA helicase UUR5_G0006 (gcontig_1105469453572) Amplicon: 121 bp
Uu05-3R	TTCTCATATTGAAAAGAAAATGAATGCTC	
Uu05-3 probe	Fluorescein-SPC-AGGAAGAATAAAACATTTTAATTTATATCCACGAA-Phosphate	
UP6F1	GAACTTTGAAACAGCTCCG	MBA UPA6_A0411

UP6R1	CCTGGTTCTTTACCTGGTTCTTTA	(gcontig_1105428157138) Amplicon: 60 bp
UU7_F_1	CAAACAAAGCATTAACAGCTTCAAAA	MBA UUR7_0421 (gcontig_1104407586372) Amplicon: 58 bp
UU7_R_1	TTATTAACTTCTTCCTTTGTTGTAAGTGTAGC	
UU8_F_1	AGTTTTAATTATTTCGTTGTTTAAGTAGC	Intergenic (gcontig_1118436613429) Amplicon: 60 bp
UU8_R_1	AGTTTTTACGTGGTAAGTGGTT	
UU9_F_1	CGAAGCGGAAGTCGCAGGT	FtsK/spoIIIE family protein UUR9_0160 (gcontig_1105462525206) Amplicon: 51 bp
UU9_R_1	TATTGCCACACCAGCCAGCA	
UU10_F_4	TGCATCAACATCGCTAAATTCA	MBA UUR10_0418 (NC_011374) Amplicon: 95 bp
UU10_R_4	ATACGCCTACTCCGACT	
UU10_P_4	Fluorescein-SPC-TTGGTGTAGGTGTTGGTTGTGG-Phosphate	
UU11_F_1	AACCTTATCAATTCTATTAATCATAGTTCA	Intergenic gcontig_1105428123667 Amplicon: 126 bp
UU11_R_1	TGAAAACAAAAAACACGCTCC	
UU11_P_1	Fluorescein-SPC-AAACCAAACTAACACATTAAGCACGC-Phosphate	
Uu12_1F	AATTGGTCAAACAACATATCGTG	CHP UUR12_A0390 (gcontig_1105428175482) Amplicon: 167 bp
Uu12_1R	ATTTAACACGATTAACATCTTCACGTTTA	
Uu12_1 probe 1	TTTGCGAAGCATTACCGCCAAATAC-FL	
Uu12 _1 probe 2	LC Red 705-CGCTTAGTTATGAACCAGCTGTTATTAATATAGCG-Phosphate	
13_F_1	CCAAGACCAGCACCTACTGCC	F_1: CHP R_2: putative SSB protein gcontig_1106518107722 Amplicon: 90 bp
13_R_1	AGAAAACGAGGAGGAATAAATAATGGATT	
UP14F1	TCCACACTACGGTAAGTAGTTT	Intergenic gcontig_1106438328052 Amplicon: 60 bp
UP14R1	CGCGCAGACCCTTGAATA	

3.3.4 PCR conditions.

Real-time PCR assays were performed on a Roche LightCycler 2.0 instrument. An asymmetric PCR condition was used for the multiplex speciation PCR: PCR master mix contained less forward primers (0.3 μM of UP063#1F and 0.2 μM of UU127#1F) than reverse primers (0.5 μM). Other components were: 0.15 μM of each UP probe, 0.2 μM of each UU probe, 0.5U of uracil-DNA glycosylase (UNG), 3mM of $MgCl_2$, and 4 μl of 5X Multiplex DNA Master HybProbe buffer (Roche Diagnostics). A volume of 2 μl of the specimen was added to the master mix to reach a total reaction volume of 20 μl. PCR program was: 10 min pre-incubation at 40°C and 95°C, then followed by 45 cycles of amplification: 95°C, 15s; 55°C, 10s, with single measurement; 72°C, 9s. Melting curves were generated by 95°C, 0s; 65°C, 30s; 95°C, with slope rate 0.1°C/s and continuous acquisition mode. Finally the machine was cooled down to 40°C for 30s. PCR conditions for 14 monoplex serovar-specific PCRs are summarized in Table 3-4 and Table 3-5. All 20 μl reaction volumes contained 0.5U of uracil-DNA-glycosylase (UNG) and 2 μl of specimen. Serovars 1, 10, and 11 PCRs were asymmetric (Table 3-4). Every PCR program included a pre-incubation procedure and final cooling procedure described above. Touchdown programs were designed for serovars 3, 5, and 8 to ensure specificity (Table 3-5). All PCR runs included a positive control that was the designated species or serovar being evaluated and a negative control (water). For speciation PCR assays, external amplification controls (DNA mixture from serovars 3 and 10) were also included in

each run. Urease gene-based traditional PCR conditions have been described previously [207].

Table 3-4 Master mix components of serovar-specific monoplex PCR assay

	Forward Primer (µM)	Reverse Primer (µM)	Probe 1 (µM)	Probe 2 (µM)	PCR Buffer[a]	$MgCl_2$ (mM) [b]	UNG (U) [c]
Serovar 1	0.3	0.5	0.2		2		0.5
Serovar 2	0.5	0.5			1	3.0	0.5
Serovar 3	0.25	0.25			1	3.0	0.5
Serovar 4	0.5	0.5	0.2	0.2	2		0.5
Serovar 5	0.4	0.5	0.2		2		0.5
Serovar 6	0.4	0.4			1	3.0	0.5
Serovar 7	0.25	0.25			1	3.0	0.5
Serovar 8	0.25	0.25			1	3.0	0.5
Serovar 9	0.25	0.25			1	3.0	0.5
Serovar 10	0.3	0.5	0.15		2		0.5
Serovar 11	0.2	0.5	0.2		2		0.5
Serovar 12	0.5	0.5	0.2	0.2	2		0.5
Serovar 13	0.4	0.4			1	3.0	0.5
Serovar 14	0.5	0.5			1	3.0	0.5

a1 = LightCycler® FastStart DNA Master SYBR Green I, 2 µl
a2 = LightCycler® FastStart DNA MasterPLUS HybProbe, 4 µl
b Addition MgCl2 added
c UNG = uracil-DNA glycosylase

Table 3-5. Amplification program of monoplex serovar-specific PCR assays

Serovar	Amplification Program	Melting Range
1	95°C, 1"; 55°C, 10", "Single" acquisition; 72°C, 4". 45 cycles.	50°C - 70°C
2	95°C, 0"; 57°C, 4"; 72°C, 2", "Single" acquisition. 45 cycles.	65°C - 95°C
3	Amplification I: 95°C, 0"; 57°C, 5". 15 cycles. Amplification II: 95°C, 0"; 57°C, 3"; 2nd target 53°C, 1°C/step, delay 1 cycle; "Single" acquisition. 45 cycles.	65°C - 85°C
4	95°C, 10"; 55°C, 6", "Single" acquisition; 72°C, 5". 45 cycles.	50°C - 80°C
5	Amplification I: 95°C, 2"; 57°C, 6"; 72°C, 2". 10 cycles. Amplification II: 95°C, 5"; 57°C, 8"; Ramp rate=5°C/s; 2nd target 54°C, 1°C/step, delay 1 cycle; "Single" acquisition; 72°C, 3 ". 45 cycles.	50°C - 70°C
6	95°C, 0"; 58°C, 2"; 72°C, 1", "Single" acquisition. 40 cycles.	63°C - 90°C
7	95°C, 0"; 55°C, 1", "Single" acquisition. 40 cycles.	65°C - 95°C
8	Amplification I: 95°C, 0"; 66°C, 5". 15 cycles. Amplification II: 95°C, 0"; 66°C, 2"; 2nd target 58°C, 1°C/step, delay 1 cycle; 72°C, 1", "Single" acquisition. 40 cycles.	66°C - 95°C
9	95°C, 0"; 62°C, 1", "Single" acquisition. 33 cycles.	70°C - 95°C
10	95°C, 2"; 56°C, 5", "Single" acquisition; 72°C, 2". 45 cycles.	55°C - 75°C
11	95°C, 2"; 53°C, 10", Ramp rate=5°C/s, "Single" acquisition; 72°C, 4". 50 cycles.	54°C - 75°C
12	95°C, 10"; 59°C, 5", "Single" acquisition; 72°C, 6". 45 cycles.	50°C - 80°C
13	95°C, 0"; 57°C, 2"; 72°C, 2", "Single" acquisition. 40 cycles.	63°C - 90°C
14	95°C, 0"; 55°C, 4"; 72°C, 2", "Single" acquisition. 45 cycles.	70°C - 90°C

3.4 Results

3.4.1 Analytical specificity and sensitivity of Ureaplasma species-specific multiplex PCR.

Using genomic DNA of each of 14 ATCC *Ureaplasma* serovar type strains as template, our multiplex real-time PCR run on a LightCycler 2.0 instrument differentiated *U. parvum* and *U. urealyticum* in two different channels without any cross reactions. All serovars within the two respective species were recognized (Figure 2-1). Species-specific amplification was successful in the presence of high copy numbers of the other *Ureaplasma* species (1:100 ratio, data not shown). No cross reactivity was detected in any other microorganisms listed in Table 3-1.

To test the detection limits of the assay, genomic DNA was prepared from 1:10 serial dilutions of broth cultures for *U. parvum* and *U. urealyticum*, using *U. parvum* serovar 3 and *U. urealyticum* serovar 10 as representatives, respectively. The detection limits for the *U. parvum* assay were 2.8×10^{-2} CFU/μl PCR mixture; for *U. urealyticum*, the numbers were 4.1×10^{-2} CFU/μl PCR mixture.

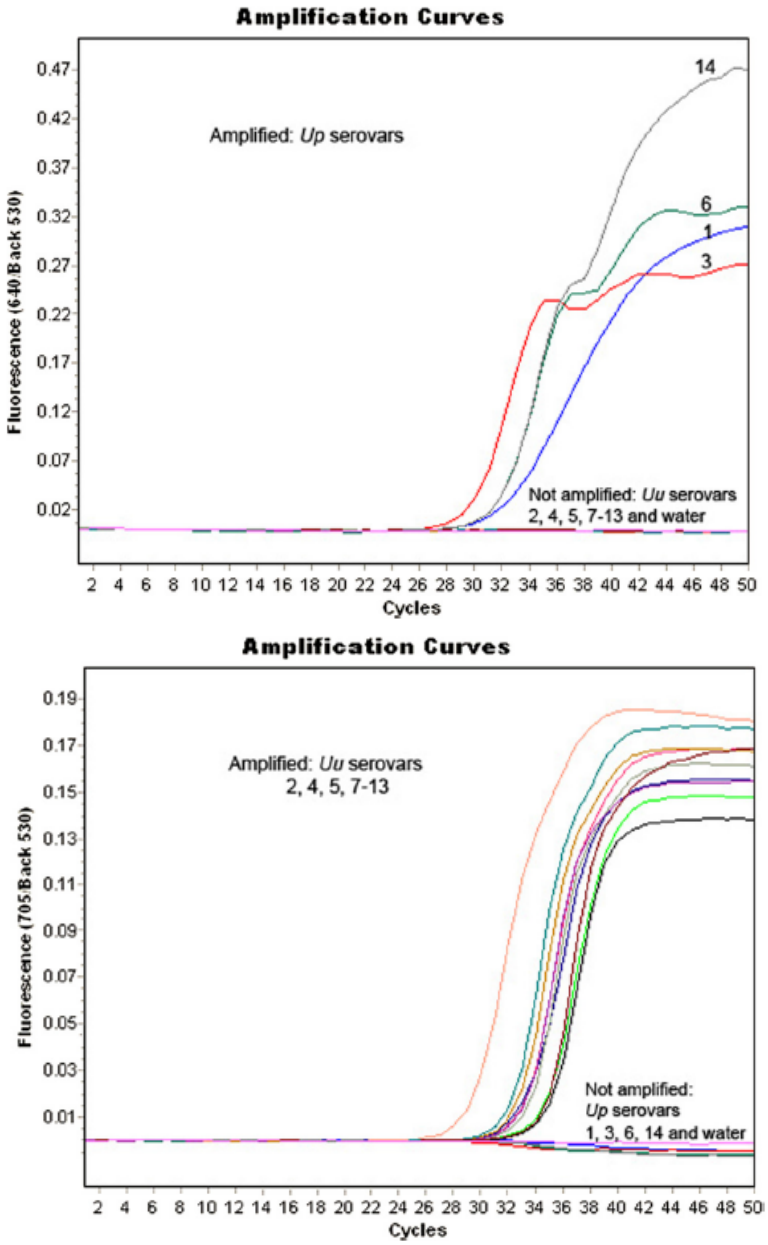


Figure 3-1. Ureaplasma species-specific multiplex real-time PCR assays

Amplification curves for Ureaplasma species-specific multiplex real-time PCR assays tested against 14 ATCC serovar type strains demonstrating simultaneous distinction between *U. parvum* (UPA) and *U. urealyticum* (UUR) and detection of all serovars of each species.

3.4.2 Comparison of species-specific multiplex PCR with traditional PCR and culture.

The species-specific multiplex PCR assay was evaluated in two groups of archived clinical specimens/subcultures in order to include comparison with both culture and traditional PCR. Among 132 clinical specimens in group 1, there were 32 (24.2%) culture-positive for *Ureaplasma* spp. and 100 (75.8%) culture-negative. The multiplex speciation PCR detected 52 (39.4%) UU-positive samples and designated 80 (60.6%) as negative. Thus, PCR detected 20 (15.2%) more positive samples than culture. There was only one culture-positive specimen that was negative by real-time PCR and the numbers of organisms in that specimen were < 10 CFUs/ml. Using culture as the reference standard, real-time PCR had a sensitivity of 96.9% and a specificity of 79.0%. Among 194 UU-positive subcultures from clinical specimens, traditional PCR detected 176 positives and 18 negatives, whereas the multiplex real-time PCR detected 188 positives and 6 negatives. Thus, the multiplex real-time PCR reduced the false-negative rate compared to traditional PCR from 9.3% to 3.1%. The specimens' DNA used for this study was originally isolated in 2004, so it is possible that degradation occurred over time to account for the fact that some culture-positive specimens were PCR-negative when retested for the present evaluation. Among the 175 subcultures positive by both PCR methods, 116 (66.3%) contained only UPA 54 (30.9%) contained only UUR, and 5 (2.9%) contained both species by traditional PCR; for multiplex speciation PCR, 115 (65.7%) contained only UPA, 48 (27.4%) contained only UUR, and 12 (6.9%) contained both species. Overall, real-time PCR recognized 6 more UPA and 1 more UUR than traditional PCR.

3.4.3 Analytical specificity and sensitivity of serovar-specific real-time PCR assays.

Real-time PCR assays specific for 14 UU serovars were developed for the ATCC type strains and the specificity of each PCR was tested against the 13 other serovar type strains. Each individual PCR assay amplified the designated serovar but none of the other 13 serovars (Figure 3-2). For PCRs with primers only (serovars 2, 3, 6, 7, 8, 9, 13, and 14), the primer concentrations were between 0.25-0.5 μM; for PCRs that required a combination of primers and probes, asymmetric PCR conditions were used (except for serovars 4 and 12). The annealing time of all programs was kept as short as possible, especially for the 8 primers only programs, which could be as short as 1 second (serovars 7 and 9). The annealing temperatures were kept as high as possible to ensure the specificity of each PCR. Touch-down programs containing a short amplification step at stringent conditions prior to the major amplification program were developed for serovars 3, 5 and 8 to improve the specificity. After amplification, a melting curve analysis was performed for each PCR providing additional proof of differentiation power. The detection limits of the serovar-specific PCR assays ranged from 15 (serovar 6) to 3,461 (serovar 9) DNA molecules/μl of PCR mixture. No evaluation was performed directly on broth subcultures to determine the limit of CFU detection as was done for the species specific multiplex PCR assay, and these assays were not tested directly for specificity against other microbial species since their utility is primarily in the testing of isolates or clinical specimens known to contain UPA or UUR.

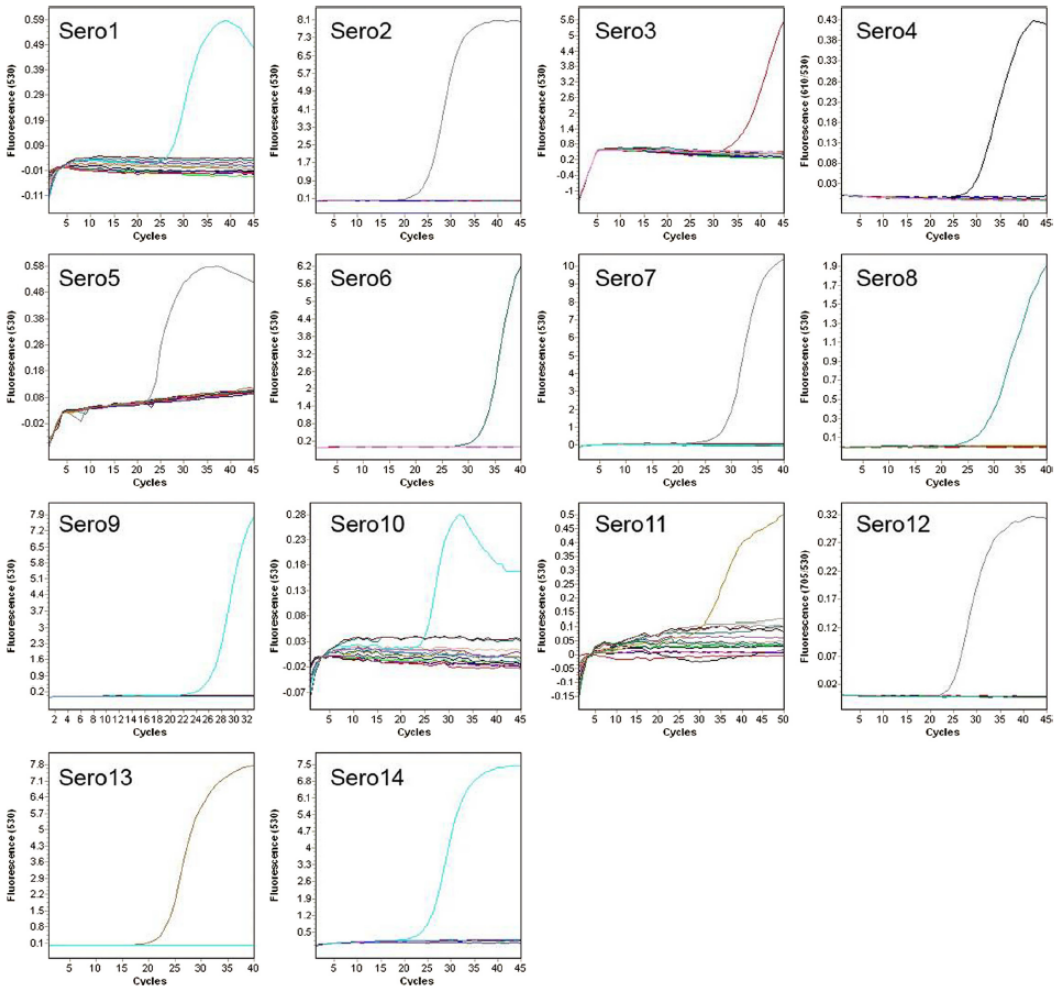


Figure 3-2. Monoplex serovar-specific PCR assays

Amplification curves for monoplex serovar-specific PCR assays tested individually with all 14 Ureaplasma sp. serovar (Sero) type strains demonstrating specific amplification of each

3.5 Discussion

Taking advantage of the genomic sequence data now available for all 14 *Ureaplasma* serovars, our multiplex real-time PCR assay, for simultaneous detection and discrimination of the two UU species, targets two conserved hypothetical protein gene sequences (UU063 in UPA strains and UUR10_00554 in UUR strains), each present only in UPA or UUR. This differs from previously reported assays for *Ureaplasma* speciation that target small differences within genes conserved in both species. The specificity of this assay was ensured computationally and experimentally verified by BLAST analysis and running the assay against a broad range of other microorganisms.

Compared to culture and traditional PCR, our findings using a single step multiplex assay format agreed with those of Cao et al. [93] who found the real-time PCR to be the most sensitive in detection of UU species. The detection threshold of our assays is well below the level necessary to detect ureaplasmas in clinical specimens and our multiplex PCR detected UU spp. in 15.2 % more clinical specimens than culture in a direct comparison using optimum cultivation methods developed and validated in our laboratory over several years [206]. Even though culture is often considered the reference method for detection of ureaplasmas in clinical specimens, improved PCR assays may be inherently more sensitive and capable of detecting the organisms in very low numbers. Thus, the PCR-positive and culture-negative specimens we encountered may in fact be true positives rather than false-negatives. Whether a PCR assay is more sensitive than culture for detection of

ureaplasmas in clinical specimens depends on the PCR target, assay conditions, and the method of culture used. Numerous studies have shown clearly that PCR assays can be superior to culture [11, 207, 209-211]. However, even though PCR has the added potential advantages of providing UU species identification and same-day turnaround without the necessity of maintaining organism viability, culture is relatively simple and can often provide results in 24 to 48 hours. Compared to traditional PCR, the new multiplex real-time PCR greatly reduced the false negative rate in detection of UU species as defined by culture positivity, making this assay even more attractive than traditional PCR in this setting.

Yi et al. [94] developed a real-time PCR assay to simultaneously detect and discriminate the two UU species in a single test with one pair of common primers for both species and two species-specific TaqMan probes. This assay, however, lacked analytical sensitivity and the clinical sensitivity was lower than traditional PCR. In contrast, our multiplex real-time PCR found a total of 6 more UPA alone or in combination than with our traditional PCR, and 1 additional UUR. No other studies have provided a quantitative comparison of species differentiation using real-time PCR versus traditional PCR.

Molecular genotyping methods based on the *mba* gene and its 5′-end have been explored as a replacement of the antibody-based phenotyping methods [86, 89, 90, 96, 98, 212, 213] that have been in use for more almost three decades since the original description of the 14-member serotyping scheme by Robertson and Stemke in 1982 [5]. However, using the sequence data available for previous genotyping methods, only partial serovar identification was possible. The 4 serovars of *U.*

parvum were readily differentiated from each other, by traditional PCR with different sets of primers [89] or by two multiplex real-time PCRs [96]. However, to distinguish the 10 serovars of *U. urealyticum* is still challenging. Using single-stranded conformation polymorphism (SSCP) analysis, the 10 serovars were divided into 2 groups: A (serovars 2, 5, 8, and 9) and B (serovars 4, 7, 10, 11, 12, and 13) [213]. By PCR and sequencing, Kong et al. [86] classified the 10 serovars into 3 subgroups: 1 (serovars 2, 5, 8, and 9), 2 (serovars 4, 10, 12, and 13), and 3 (serovars 7 and 11). The same investigators then provided a better discrimination by dividing the 10 serovars into 5 MBA genotypes with the added individual separation of serovars 9 and 10: A (serovars 2, 5, and 8), B (serovar 10), C (serovars 4, 12, and 13), D (serovar 9) and E (serovars 7 and 11) [89]. This approach using just the sequence of an *mba* gene specific to each serovar should be reconsidered in light of the whole genome sequences that show all UU serovars encode multiple members of the phase variable *mba* gene family.

In designing our serovar-specific real time PCR assays, we sought to avoid the *mba* gene family because of its ubiquity within the ureaplasmas and because of its phase variability. This intent was thwarted by the high inter-serovar identity among the 14 serovars. The average difference among the 4 UPA genomes is 0.56%, and among the 10 UUR genomes the average is 0.63%. While it is straightforward to design PCRs for discrimination of any pair of serovars, discrimination of one serovar from the other 13 is often very challenging. We sought to have all assays employ primers and probes; however in many cases this was not possible. We were even forced in several cases to obtain serovar identification to target *mba* genes or *mba*

paralogous gene family genes as shown in Table 3. In those instances, we made sure our PCR targets did not span sites at which chromosome rearrangements took place during phase variation. Those sites could be identified by analyzing whole genome shotgun sequencing assemblies. By computationally searching thousands of unique loci in each serovar and testing with 14 ATCC type stains, we created PCRs specific to all 14 serovars without any cross reactions. Many PCRs predicted to be serovar-specific failed when tested and were abandoned. Ultimately, each serovar primer/probe set shown in Table 3 contains at least one serovar specific primer or probe. In the cases where the set could cross-react, leading to significantly larger amplicon, the PCR conditions were adjusted to prevent it. The PCRs are sensitive in detecting each specific serovar with detection limit of 15-3,500 molecules/PCR mixture. The strategy to multiplex the 14 PCR assays which should be more efficient was not used because each PCR condition was unique and combination with another assay compromised specificity and sensitivity of one or both assays (data not shown).

The real-time PCR-based typing results were in partial or full agreement with antibody-based results from immunoblotting for 35 of 50 (70%) clinical isolates analyzed by both methods. Discrepant results could be classified into four major groups: Group 1 included 6 isolates showing PCR-positive serovars but immunoblots were negative or non-typeable. The PCR serovars all belonged to their corresponding species determined by the multiplex speciation PCR suggesting that PCR typing is more sensitive than serological typing. Group 2 discrepant results involved those serovars determined by the two methods to belong to different species. One isolate was pure serovar 3 (UPA) by immunoblot but contained serovars 8, 10, 11, and 12 by

PCR. PCR speciation showed it was *U. urealyticum*. Another isolate was pure serovar 6 (UPA) by immunoblot but contained PCR serovars 4 and 8. PCR speciation also indicated it was UUR. The reasons for this complete non-agreement could not be explained other than the possibility that several years of storage could have influenced PCR assays performed at a later time. In these cases, our new real-time multiplex PCR speciation and real-time monoplex PCR typing still agreed with each other, indicating the reliability of our real-time PCR methods. Group 3 discrepant results involved detection of different, non-overlapping serovars within the same species. Six isolates (3 UPA and 3 UUR), showed such differences. Group 4 discrepant results were those found in 9 isolates in which both methods partially agreed with each other, i.e., isolates containing multiple serovars had at least one common serovar detected by two methods, but the rest were different. These findings might be explained by the changes of relative bacterial populations in mixed cultures stored for long periods or differences in the specificity and sensitivity of the PCR versus antibody-based methodologies [213].

Considering the discrepancy between the typing methods, it is important to point out that the reference strains used to develop MAbs and PCR primers/probes were different. The MAbs were produced against reference strain collections that were different from the type strains used for PCR development that were purchased directly from ATCC in 2006. The original whole genome sequences of serovar 3 from a representative UAB strain [8] and ATCC serovar 3 type strain 27815 were slightly different. Genetic variations of clinical isolates from reference strains might be sufficient to lead to different specificities and this was apparent when we retested the

previously developed MAbs to serotype the 50 placental isolates against the 14 ATCC serovar type strains. For example, the UAB strain serovar 1-specific MAb, cross reacted with ATCC serovars 3 and 10 (ATCC 37815 and ATCC 33699), (data not shown). The complexity of variations in nucleotide sequences of a certain gene or region, which were the bases of molecular genotyping methods, is not the same as that of antigen-antibody reactions, which were the bases of serological typing methods. Thus, a complete correlation between the two methods could not be expected. This was also the same in PCR-single strand conformation polymorphism analysis typing [213].

A single clinical isolate of UUR that was determined by multiplex speciation PCR was not typeable by either method, possibly representing a new serotype. The old 14-member-serotyping scheme may need to be revised if additional unknown subtypes are found by the new 14-PCR-based genotyping method when larger numbers of clinical isolates are analyzed. Sequence analysis may provide more information about such untypeable isolates.

In conclusion, our new multiplex real-time PCR for detection and discrimination of UU species is robust, ready to replace the traditional PCR for clinical diagnostic purposes, and is a suitable alternative to culture. We have shown for the first time that all 14 serovars of UU can be clearly differentiated from each other using real-time PCR technology. This should greatly facilitate future investigations of differential pathogenicity of ureaplasmas at the species and serovar level in clinical conditions when studying clinical isolates of UU species. Our data regarding the monoplex serovar-specific PCR assays were limited to evaluations of

sensitivity and specificity using type strains and clinical isolates. We did not evaluate clinical specimens directly. Ideally, a serotyping method should be sufficiently sensitive to use in this manner. The analytical sensitivities we obtained suggest our assay should meet the requirements for direct testing of clinical specimens that may contain serovars alone or in combination. Proof of this will require additional clinical studies.

Acknowledgements

This work was supported by federal funds under grants RO1A1072577 and RO1209 and contract NO1-AI-30071 from the National Institute of Allergy and Infectious Diseases and grant AI 28279 from the National Institute of Child Health and Human Development.I gratefully acknowledge Li Xiao for the majority of the work described in this chapter. John Glass, Shibu Yooseph and I provided primer and probe design, bioinformatics analysis of primers and probes, analysis of assay data and troubleshooting of PCR reactions. I also greatfully acknowledge the direct contributions of Gail H. Cassell, Lynn B. Duffy, and Ken B. Waites as well as the technical support provided by Donna Crabb and Amy Ratliff

Chapter 4. Extensive horizontal gene transfer in UU questions the utility of serotyping for diagnostic purposes

4.1 Abstract

UPA and UUR are sexually transmitted, opportunistic pathogens of the human urogenital tract. There are 14 known serovars distributed between the two species. For decades it has been postulated based upon limited data that virulence is related to serotype specificity. Results were often inconclusive due to small sample size and extensive cross reactivity between certain serovars. We developed real-time quantitative PCRs that allow reliable differentiation of the type strains of each of the 14 serovars. To investigate species and serovar distribution, we typed 1,061 clinical isolates of human ureaplasmas from diverse patient populations. There was only a tenuous association between individual *Ureaplasma* serovars and certain patient populations. This may in part be explained by the fact that almost 40% of the isolates were genetic mosaics, apparently arising from recombination of multiple serovars. This explains the extensive cross-reactivity based upon serotyping and the lack of consistent association of given serotypes with disease.

4.2 Introduction

It has been speculated for many years that individual *Ureaplasma* species or serovars might be associated with certain diseases more than others, as is the case for bacteria such as *Streptococcus pneumoniae* and *Haemophilus influenzae* [214, 215].

Although several studies reported that *U. urealyticum* is more pathogenic than *U. parvum* [84, 216-220], conflicting results have been found by others [208, 221], so it is possible that differential pathogenicity might exist at the serovar level rather than at the species level. Inconsistent results implicating specific serovars (or serovar groups) with various clinical conditions [4, 196, 205, 222-225] coupled with frequent detection of *Ureaplasma* isolates comprised of more than one serovar have also been frequently reported [222-224, 226, 227]. Differing results among investigations could also be related to inadequate or imprecise typing methods. The possibility that individual *Ureaplasma* isolates may express multiple serovar specificities has been suggested, but never investigated in a systematic manner due to the technical difficulty of separating all of the 14 serovars [222, 227].

Two major methodologies to classify *Ureaplasma* isolates to the species and serovar levels have been described. Antibody-based phenotyping methods included growth/metabolism inhibition tests [186, 196, 197] antibody-linked epi-immunofluorescence or color reaction assays[199, 201, 228, 229]. These methods yielded inconclusive results because of multiple cross-reactions and poor discriminating capacity. Cross-reactions have been observed even in certain serovar reference strains [5]. Molecular genotyping methods were more rapid and accurate, readily separating the two *Ureaplasma* species [83, 88, 93-95]. However, due to limited sequence variation in the PCR targets, only partial separation of serovars was achieved [86, 89, 90, 96, 98]. Recently we developed a set of *Ureaplasma* species- and serovar-specific real-time PCR assays that separate completely all 14 ATCC serovars type strains without cross-reactions [230].

The improved specificity of real-time PCR compared to conventional PCR is mainly due to the use of a third oligonucleotide called a probe that binds to the target sequence. Usually, the probe is designed to correspond to a sequence located between the forward and reverse primers. Generally, the probe is labeled with a fluorophore and a quencher. When the quencher is in close proximity to the fluorophore, the fluorescence is inhibited or emitted at a different wavelength. Utilizing a fluorescently labeled probe allows a signal to be detected by the instrument only if the probe has annealed to its complementary target sequence and the DNA polymerase has amplified the target region therefore increasing the distance between the quencher and the fluorophore. The use of a labeled probe minimizes the probability of cross-reaction and detection of undesired amplicons. Such probes are also specifically useful when differentiating very closely related organisms is desired, as is the case of determining which of the 14 Ureaplasma serovars are present in a clinical specimen. Another feature of real-time PCR that contributes to the specificity of the assay is that the amplicon melting temperature is determined at the end of the assay, and can be used to verify whether the desired PCR product is being detected.

Analytical sensitivity, can also be improved when using real-time PCR assays. PCR sensitivity depends on many factors including DNA extraction method, selection of primers, selection of visualization method, and reaction conditions. In addition to those factors, real-time PCR diagnostic assays are often designed to produce shorter amplicons with an optimal length of under 300bp compared to conventional PCR [85]. The smaller the amplicon the more efficient is the PCR reaction and the lower probability for the target templates to be sheared during handling. Based on careful

primer selection and optimization of reaction conditions, an increased sensitivity of real-time PCR may observed compared to assays designed for conventional PCR [81].

Real-time PCR is faster than conventional PCR mainly because of the elimination of post-cycling amplicon handling and more efficient instrumentation. The fact that real-time PCR is a closed reaction from beginning to end minimizes the risk of errors due to amplicon contamination causing false positive assays. Data generated in the real-time PCR assay is analyzed automatically by computer software specially designed for the instrument being used such that no gel electrophoresis is necessary. These features have contributed to the introduction of real-time PCR into clinical laboratories much faster than conventional PCR [84]. Furthermore, the current real-time PCR platforms usually have faster temperature ramps (as fast as 20ºC/sec in the Roche LightCycler) than conventional thermocyclers, which additionally shortens the time required for the procedure. Conventional PCR methods for detection of Ureaplasma spp. can take between 2-3 days, about the same turnaround time as culture, whereas real-time PCR can potentially provide results the same day a specimen is received. Since organism viability does not have to be maintained for PCR-based detection, specimen collection, handling, and transport for Ureaplasma detection by PCR is much simpler than for culture.

An added bonus of real-time PCR over conventional PCR is that real-time PCR can be used to quantitate the starting amount of DNA molecules and therefore the pathogen load of the patient can be estimated. This is especially useful when detecting organisms such as ureaplasmas that can be part of the human flora but can

be harmful when their load is higher than normal. In such cases a qualitative positive result from conventional PCR may not be sufficient to decide whether a person needs medical treatment. Theoretically, each commensal organism can have its own established cycle threshold (Ct) or cross point (Cp) values range in which it would be considered harmless based on accumulated clinical data. The lower the Ct or Cp value the higher the load of the organism in the patient.

Horizontal gene transfer (HGT) plays an important role in microbial adaptation, speciation, and evolution [231]. Although mycoplasmas and ureaplasmas are characterized by their minimal genomes and are thought to have undergone regressive evolution, which usually is not favorable for active DNA acquisition [232], evidence has shown that HGT occurs among phylogenetically distinct mycoplasmal species sharing the same ecological niche [233] or within the same species [234]. Comparative genomic analyses have indicated possible HGT between *U. parvum* and *Mycoplasma hominis* [235]. *Ureaplasma* spp. have also been shown to form biofilm [236], structures thought to promote DNA exchange among other bacteria. Furthermore, recombinases, transposases and putative conjugative transposon mobilization proteins have been identified in genomes of the 14 *Ureaplasma* serovars[237]. These findings raise the question of whether HGT occurs among *Ureaplasma* serovars and its potential implications for the hypothesis of differential pathogenicity at the serovars level.

Using *Ureaplasma* species- and serovar- specific real-time PCR assays that we recently developed and validated [230], we analyzed a large number of clinical isolates from different geographic regions and from different patient populations to

classify them to species and serovar in order to investigate differential pathogenicity at these two levels. Initial evaluation of the results suggested the likelihood of HGT between species and among serovars, and prompted additional investigations.

4.3 Materials and Methods

4.3.1 Bacterial isolates

Reference strains for the 14 *Ureaplasma* serovars were obtained from the American Type Culture Collection (ATCC) and used as the controls in serovar-specific PCR assays. The 1061 unique clinical isolates evaluated for species and serovar distributions were obtained from cultures collected and stored frozen at -80°C or in lyophilized form dating from the late 1970s to 2009. Isolates originated from Alabama and various other states within the continental United States and from Alberta, Canada in the patient groups described in Table 4-1. Ureaplasmas recovered from clinical specimens were initially identified to genus level by standard methods including colony morphology and urease production on A8 agar.

Table 4-1. Description of *Ureaplasma* clinical isolates

Specimen Type	**Group Description**	**No. Isolates**
Control groups		
vaginal swabs	healthy pregnant females	169
placental tissue collected at cesarean section in women with intact fetal membranes	females without histologic chorioamnionitis	42
catheterized urine	male with neurogenic bladder without urethritis	25
endotracheal aspirate	preterm infants without bronchopulmonary dysplasia (BPD)	108
Diseased groups		
endometrial biopsy tissue	females with pelvic inflammatory disease (PID) and/or post-partum endometritis	85
placental tissue collected at cesarean section in women with intact fetal membranes	females with histologic chorioamnionitis	18
endotracheal aspirate	preterm infants with bronchopulmonary dysplasia (BPD)	88
urethral swab or urine	males with non-gonococcal urethritis (NGU) from Canada	421
urethral swab or urine	males with non-gonococcal urethritis (NGU) from USA	81
blood, cerebrospinal fluid, synovial fluid, pleural fluid, lung tissue	invasive isolates from various patient groups including adults and children	24
Total		**1061**

4.3.2 DNA preparation

Genomic DNA was extracted from all 1061 clinical isolates by the proteinase K method as described previously [207]. Inhibited samples were further purified by QIAamp® DNA Blood Mini Kit (Qiagen, Valencia, CA). Prepared DNA samples were stored at -80°C unless submitted immediately for PCR assay.

4.3.3 Genotyping of clinical isolates by PCR

The clinical isolates were first classified to species level using a multiplex species-specific real-time PCR assay for which primers, probes, reagents and PCR conditions have been previously described [230]. UPA and UUR isolates were then typed for their corresponding serovars by a series of serovar-specific real-time PCR assays [230]. In the event that isolates were negative for any serovars within their corresponding species, additional PCR assays for serovars of the other species were performed. A designated ATCC type strain control and a negative control (distilled water) were included in every PCR run. Untypable isolates were subjected to a secondary PCR assay targeting the *urease* gene [88]. The species- and serovar-specific PCR assays were performed using a LightCycler 2.0 (Roche, Indianapolis, IN)

4.3.4 Separation of isolates containing multiple serovars

Four isolates shown to contain multiple serovars by real-time PCR were thawed and incubated in 10B broth overnight. The broth cultures were then filtered through a 0.2 μm filter, inoculated onto A8 agar and incubated 24-48 hours until

colonies were readily visible under a stereomicroscope at 126X magnification. At least 10 single colonies from each isolate were removed with a sterile needle or pipette tip for individual overnight cultures in 10B broth. DNA was then isolated from each broth culture and prepared for PCR assay as described above.

4.3.5 Quantification of each serovar in clinical isolates to distinguish mixtures from hybrids

To quantify each serovar marker in isolates containing multiple serovars, a universal control plasmid, pUC19-UU, which carries one copy of each serovar marker (except for markers of serovars 4 and 5, which were not able to be stably incorporated into the plasmid), was constructed and used to generate an external quantification standard curve. Briefly, the RT-PCR assay target for each serovar was PCR amplified from extracted gDNA so that the amplicons overlapped each other in order to be assembled and cloned into the multi cloning site of pUC19 using the GA method. For this each to forward primer an overhang sequence was added complementary to the end of the upstream target, so that the total primer length is 60nt. Serovar-specific PCRs were performed on 271 randomly selected isolates containing multiple serovars (except serovars 4 and 5) and quantities of each serovar were calculated. When comparing two serovars, based on the largest calculated quantity difference of the pUC19-UU plasmid control between PCR runs, an arbitrary differentiation standard was made: an equal or less than 5X difference was considered a hybrid; a difference of 5-10X was considered to represent hybrid or mixture

(hyb/mix); any difference greater than 10X was considered to represent a mixed culture; if both serovars were in low quantity, it was designated as undetermined.

4.3.6 DNA Sequencing

A total of 7 clinical isolates, each containing two serovar markers were selected for polymorphic loci analysis by Sanger DNA sequencing. Isolates 10902 and 97078 contained serovars 1 and 6; isolates 10901 and 8510 contained serovars 3 and 6; isolates 24318 and 25353 contained serovars 9 and 10. For each pair of serovars, 7-10 loci throughout the genome that contained multiple base pair polymorphisms between each serovar pair were selected. PCR primers were designed to amplify from both serovars across the region containing the polymorphism of about 700-800bp. Loci were selected to be gene coding, have at least 90% identity and 20 or more mismatched nucleotides by BLASTn analysis (Table S 1).The *mba* gene of UPA3 and UPA6 was sequenced by primers flanking the whole *mba* gene region. Serovar-specific PCR primer regions of serovars 9-12 were also amplified and sequenced. PCR was performed using a Veriti 96 well thermal cycler (Applied Biosystems, CA). DNA sequencing was performed in the University of Alabama at Birmingham Heflin Center Genomics Core Facility. DNA sequences were analyzed using CLC DNA workbench 5.

We sequenced the genomes of four *U. urealyticum* patient isolates that we could not identify on the serovar level using 454 pyrosequencing (454 Life Sciences, Branford, CT). The genomes were assembled using the Newbler Assembler (454 Life Sciences). We then compared the four isolates to each of the ATCC reference

urealyticum genomes by generating dot plots. Sequencing was done at the J. Craig Venter Institute.

4.3.7 Statistical Analysis

Fisher exact and χ^2 test were performed to compare the distributions of UU species and serovars in different patient populations. A p value of < 0.05 was considered statistically significant. Analyses were conducted using SAS software (SAS Institute Inc., Cary, NC) and SPSS 16.0 (SPSS Inc., Chicago, IL).

4.4 Results and Discussion

4.4.1 Species and serovar distributions of isolates from different patient populations

Of 1061 unique clinical isolates that were typed, UPA was detected in 508 (48%) isolates, UUR was detected in 406 (38%) isolates and both species were detected in 140 (13%) isolates. The remaining 7 isolates (1%) could not be typed to species level by real-time PCR. However, one of them was positive for UUR using a PCR targeting the *urease* gene [88]. The prevalence of UUR was significantly increased in isolates from endometrial biopsy tissues from women with pelvic inflammatory disease (PID) and/or endometritis (33% *vs* 14%, $p \leq 0.01$) in comparison to vaginal swabs from healthy pregnant women. It was also more common in American men with NGU when compared to men without urethritis (68% *vs* 48%, $p < 0.05$). No association of either UU species was found in placentas of women with *vs* without chorioamnionitis. There was also no difference in infants with

or without chronic lung disease of prematurity or among the 24 invasive isolates. Canadian men with NGU had a significantly higher percentage of mixed species than US men (26% *vs* 6%, $p = 0.001$); however, the difference might merely reflect methodology/isolation procedures.

The 14 serovars were distributed unevenly in all groups: serovar 3 was most common (332) and serovar 5 was least (19) when counting the presence of any serovars in all 1061 isolates (Table 4-2). When comparing the prevalence of serovars in diseased groups with their corresponding controls, no consistent patterns implicating individual serovars were observed. Serovars with significantly increased prevalence in diseased groups were: serovars 5, 8 and 11 in the PID/endometritis group, serovar 6 in the chorioamnionitis group, serovar 9 in US men with NGU, and serovar 10 in neonates in the BPD group. Among the 24 invasive isolates, serovar 3 was most common, occurring in 7 (29%) isolates. The serovar distribution between NGU isolates from US and Canadian men was significantly different, indicating geographic factor may play an important role.

Table 4-2. Species and serovar distributions of clinical isolates

Notes: Numbers in parenthesis are percentage; [a] $p \leq 0.01$; [b] $p < 0.05$, compared to the corresponding controls; [c] one isolates was *U. urealyticum* determined by a secondary PCR test; [*] $p = 0.001$; Up, *U. parvum;* Uu, *U. urealyticum;* spp., species; sv, serovar; PID, pelvic inflammatory disease; NGU, nongonococcal urethritis; BPD, bronchopulmonary dysplasia; US, United States; Ca, Canada.

Species/ Serovar	PID and Endometritis		Chorioamnionitis		NGU			BPD		Invasive Isolates	Total
	Neg	Pos	Neg	Pos	Neg	Pos (US)	Pos (Ca)	Neg	Pos		
Total isolates	169	85	42	18	25	81	421	108	88	24	1061
Up	137 (81%)	56 (66%)	30 (71%)	14 (78%)	13 (52%)	21 (26%)	99 (24%)	68 (63%)	57 (65%)	13 (54%)	508
Uu	24 (14%)	28 (33%)[a]	11 (26%)	4 (22%)	12 (48%)	55 (68%)[a]	210 (50%)	28 (26%)	23 (26%)	11 (46%)	406
Up+Uu	7 (4%)	1 (1%)	0 (0%)	0 (0%)	0 (0%)	5 (6%)	111 (26%)[a]	10 (9%)	6 (9%)	0 (0%)	140
Nag spp.	1 (1%)	0 (0%)	1 (2%)	0 (0%)	0 (0%)	0 (0%)	1 (<1%)	2 (2%)	2 (2%)[c]	0 (0%)	7[c]
sv1	33 (20%)	12 (14%)	4 (10%)	2 (11%)	3 (12%)	5 (6%)	52 (12%)	16 (15%)	13 (15%)	3 (13%)	143
sv3	86 (51%)	31 (37%)	18 (43%)	4 (22%)	6 (24%)	15 (19%)	97 (23%)	34 (32%)	34 (39%)	7 (29%)	332
sv6	41 (24%)	19 (22%)	8 (19%)	9 (50%)[a]	4 (16%)	10 (12%)	126 (30%)	41 (38%)	27 (31%)	3 (13%)	288
sv14	6 (4%)	1 (1%)	0 (0%)	1 (6%)	1 (<1%)	3 (4%)	8 (2%)	8 (7%)	7 (8%)	0 (0%)	35
sv2	6 (4%)	2 (2%)	1 (2%)	2 (11%)	2 (8%)	14 (17%)	89 (21%)	11 (10%)	6 (7%)	3 (13%)	136
sv4	6 (4%)	3 (4%)	2 (5%)	0 (0%)	1 (4%)	14 (17%)	83 (20%)	3 (3%)	2 (2%)	1 (4%)	115
sv5	1 (1%)	4 (5%)[a]	1 (2%)	2 (11%)	0 (0%)	1 (1%)	10 (2%)	0 (0%)	0 (0%)	0 (0%)	19
sv7	9 (5%)	4 (5%)	1 (2%)	2 (11%)	2 (8%)	8 (10%)	103 (25%)	13 (12%)	4 (5%)	2 (8%)	148
sv8	0 (0%)	3 (4%)[a]	3 (7%)	0 (0%)	2 (8%)	17 (21%)	55 (13%)	4 (4%)	4 (4%)	0 (0%)	88
sv9	19 (11%)	5 (6%)	4 (10%)	0 (0%)	0 (0%)	14 (17%)[a]	30 (7%)	7 (7%)	4 (5%)	5 (21%)	88
sv10	7 (4%)	6 (7%)	1 (2%)	1 (6%)	3 (12%)	10 (12%)	116 (28%)	5 (5%)	11 (13%)[a]	5 (21%)	165
sv11	14 (8%)	17 (20%)[a]	3 (7%)	3 (17%)	4 (16%)	21 (26%)	24 (6%)	19 (18%)	12 (14%)	2 (8%)	119
sv12	5 (3%)	4 (5%)	3 (7%)	1 (6%)	0 (0%)	9 (11%)	24 (6%)	6 (6%)	4 (5%)	0 (0%)	56
sv13	10 (6%)	5 (6%)	1 (2%)	0 (0%)	0 (0%)	6 (7%)	15 (4%)	10 (9%)	2 (2%)	5 (21%)	54
Nag sv.	3 (2%)	4 (5%)	2 (5%)	0 (0%)	3 (12%)	8 (10%)	40 (10%)	2 (2%)	4 (5%)	1 (8%)	67

4.4.2 Isolates that were negative for all serovar-specific assays

There were 67 (6%) isolates that could not be assigned to any of the 14 known serovars by PCR. To ascertain why these isolates were negative in all of our serovar-specific assays, we did whole genome shotgun sequencing of 4 of these isolates. Genome analysis showed isolates 2033 and 2608 are most closely related to serovar 12 and 4. The ATTC serovars 12 and 4 were the closest related serovars among the urealyticum group. Isolate 4155 was most similar to serovar 11, whereas isolate 4318 was most similar to serovar 2. Relative to the ATCC reference strains, all of the isolates' genomes had some minor genome rearrangements, areas that were deleted, and some areas that were inserted and are new for the urealyticum group. Analysis of the target areas for the serovars-specific PCR assays showed that the target was either missing completely, or some of the target was missing or modified so that one of the primers would not bind. However, it is clear that these isolates have changes in other areas of the genome as well. Whether we can assign new serovar numbers to any of the unidentifiable isolates is a matter of clarifying the requirements for a ureaplasma to be considered a specific serovar.

4.4.3 Isolates containing multiple serovars and conflicting serovars

Multiple serovar markers were detected in 413 (39%) isolates, predominantly *U. urealyticum* (201) and mixed species (124) (Figure 4-1). Among them, 223 (21%) isolates contained 2 serovar markers, 91 (9%) isolates contained 3 serovar markers, and 99 isolates (9%) contained 4 or more serovar markers. The maximum number of serovar markers detected in a single isolate was 10. The distribution of serovar

marker numbers per isolate did not show significant differences between control and diseased groups, except US men with NGU.

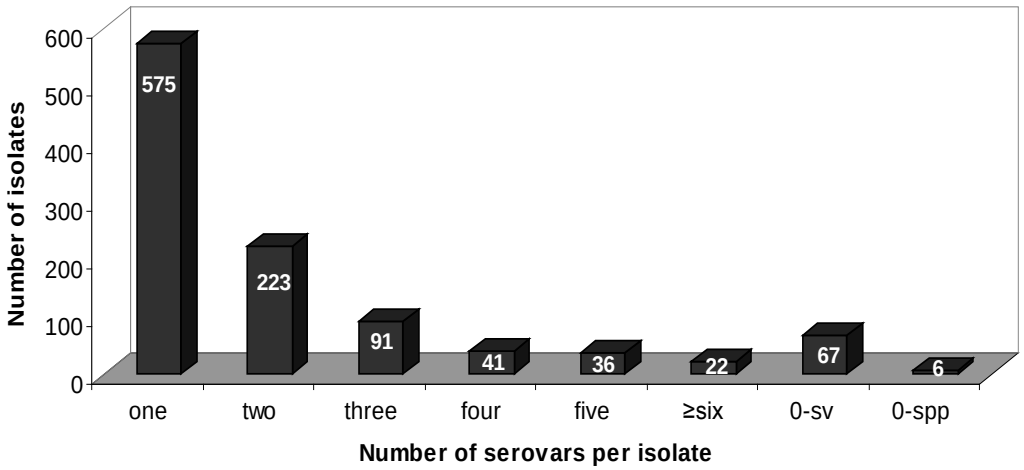


Figure 4-1. Clinical isolates containing multiple serovars and untypable serovars.

Clinical isolates containing multiple serovars and untypeable serovars. A total of 1061 clinical isolates were typed to species and then serovar level by real-time PCR. In the event isolates were determined to be negative for any serovars within their designated species, additional PCR assays for serovars of the other species were performed. Untypeable isolates (0-sv, negative for any serovar; 0-spp, negative for any species) were subjected to a secondary PCR assay targeting the urease gene to confirm the species.

Discordant typing results were observed in 42 isolates. Three isolates typed as *U. parvum*; however their serovar markers belonged to *U. urealyticum*; on the other hand, 15 isolates typed as *U. urealyticum* but contained *U. parvum* serovar markers. One *U. urealyticum* isolate showed serovar markers of both species. One isolate negative for the species-specific real-time PCR contained 2 different *U. urealyticum* serovar markers; while 22 isolates positive for both species showed serovar markers from only one.

4.4.4 Separation of isolates with multiple serovars

In an effort to isolate pure serovars from clinical isolates containing multiple serovars, we found that the serovars in 4 isolates could not be separated after filtering and selecting single colonies for subculture. Two isolates contained serovars 3 and 14; one isolate contained serovars 1 and 14; and one contained serovars 3 and 6. As a control, a mixture of equal amounts of ATCC type strains of serovars 1 and 6 was made and the two serovars were completely separated after the same separating procedures. We suspected that these isolates and many of the others containing multiple serovars might not be true mixtures, but hybrid organisms carrying multiple serovar markers.

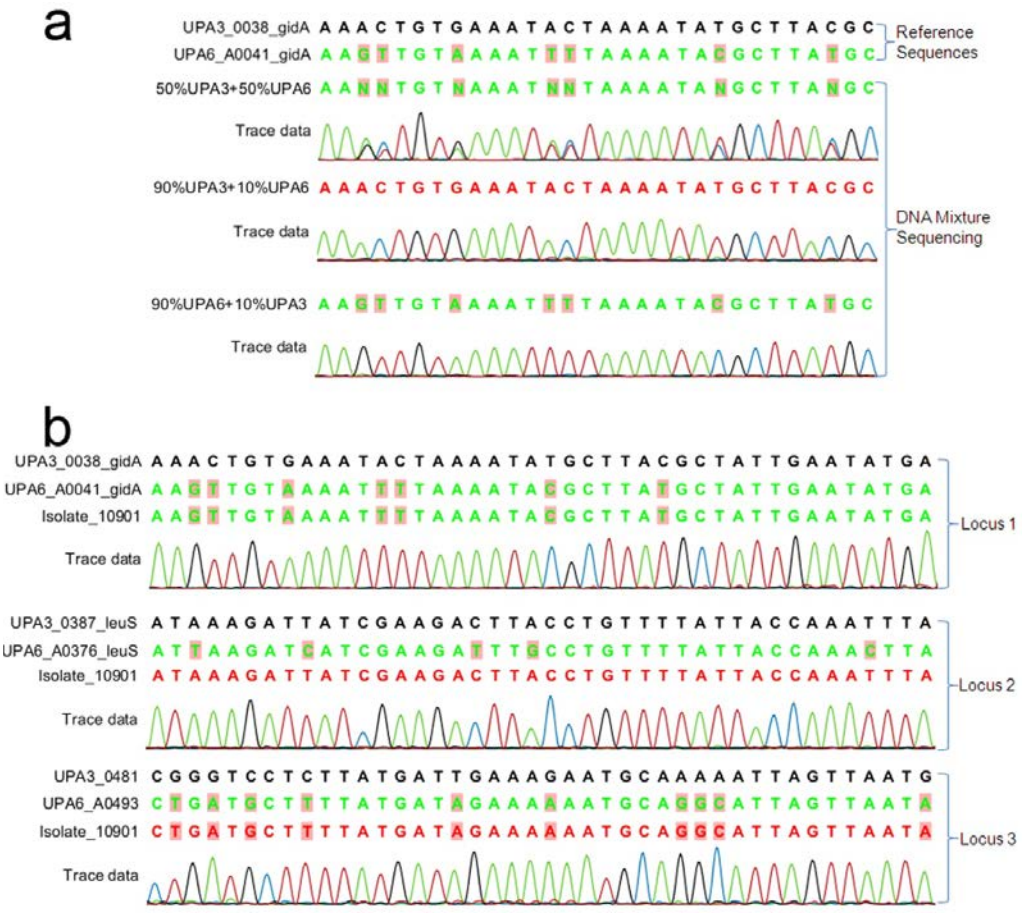
4.4.5 Verifying hybrid isolates by sequencing multiple gene-coding, polymorphic loci

To simplify the test to determine whether the isolates were mixtures or hybrids, we focused on isolates that were positive for two serovar markers and analyzed them by

DNA sequencing (primers are listed in Table S 1 at the end of this chapter). We chose to sequence 7-10 gene-coding loci that are widely distributed around the two genomes and contain multiple base-pair polymorphisms. Control mixtures of equimolar genomic DNA concentration of ATCC type serovars 3 and 6 had visible double nucleotide peaks in the sequencing chromatograms as shown in Table 4-2a. Differentiating mixed versus pure isolates was limited when the DNA concentration ratio reached 9: 1 and clean single peaks representing the predominant DNA type appeared (Figure 4-2a). A hybrid would be expected to show single peaks at polymorphic sites representing the sequence of one serovar in some loci and the other serovar in the rest of the loci (Figure 4-2b). In all tested loci, ATCC type strains of corresponding serovars were amplified and sequenced as controls.

Figure 4-2. DNA sequencing determining *Ureaplasma* hybrid or mixed isolates.

DNA sequencing of multiple polymorphic loci throughout the genome. DNA sequencing trace files were aligned by using CLC genomic workbench. (a) Sequencing of DNA obtained from artificial mixtures of U. parvum serovar 3 and U. parvum serovar 6. Reference sequences of gidA orthologous genes of U. parvum serovars 3 and 6 are shown on the first two lines. The type of DNA mixture for each reaction is provided on the left of the sequences. Polymorphisms are marked with shaded boxes. Trace data indicated double peaks in the 1:1 mixture and single peaks in 9:1 mixtures, representing the predominant DNA. (b) Sequencing of multiple polymorphic loci in isolate 10901. A representative window of three of seven sequenced loci is shown. Polymorphic sites are marked with shaded boxes. Isolate 10901 showed single peaks at the polymorphic loci. Loci 1 and 3 showed characteristics of serovar 6, while locus 2 was the same as serovar 3.



Seven clinical isolates suspected to contain a hybrid organism of two serovars were analyzed. Six isolates (10902, 97078, 10901, 8510, 24318 and 25353) clearly showed some polymorphic loci characteristic of one serovar and some characteristic of another (Table 4-3). Seven polymorphic loci were examined in two isolates containing serovars 1 and 6. Isolate 10902 showed an obvious hybrid pattern: 4 loci were from UPA6 and 3 loci were from UPA1. In isolate 97078, 6 out of 7 loci were UPA6 specific, and 1 locus was closest to UPA14 (blast showed 99.55% identity to UPA14). In order to examine the reason for having serovar-specific real-time PCR signal for both UPA1 and UPA6, we proceeded to sequence the two serovars' real-time PCR targets. Both PCR assays targeted the *mba* gene of each serovar. The *mba* sequencing chromatogram depicted only MBA1. Since real-time PCR is more sensitive than DNA sequencing, and DNA sequencing failed to recognize mixtures of 9:1 or lower ratios, we hypothesized that isolate 97078 contained two hybrids in a ratio of at least: 90% or more of a UPA6 with a UPA1 specific MBA hybrid, and 10% or less of a UPA1 with a UPA6 specific MBA hybrid. Another seven loci were tested for serovars 3 and 6 and hybrid patterns were observed in isolates 10901 and 8510. Isolates 24318 and 25353 were serial isolates from the same patient. They were positive by real-time PCR for serovars 9 and 10 and had ten polymorphic loci from UUR10 and two specific loci from UUR9. Sequencing of the real-time PCR targets indicated that the two isolates contained serovar markers of both UUR9 and UUR10. Taken together, all data suggested that these two isolates are hybrids.

Table 4-3. DNA sequencing results of polymorphic loci, *mba* gene, and serovar marker regions

Compared serovars	Isolate	Polymorphic Locus	Polymorphic character	MBA/Serovar marker	Conclusions
UPA1_6	10902	UPA1_43	UPA6	N/A	Hybrid
		UPA1_98	UPA6		
		UPA1_100	UPA1		
		UPA1_101	UPA1		
		UPA1_114	UPA6		
		UPA1_293	UPA1		
		UPA1_359	UPA6		
UPA1_6	97078	UPA1_43	UPA6	MBA1	Hybrid
		UPA1_98	UPA6		
		UPA1_100	UPA6		
		UPA1_101	UPA6		
		UPA1_114	UPA6		
		UPA1_293	UPA14?		
		UPA1_359	UPA6		
UPA3_6	10901	UPA3_37	UPA6	MBA3	Hybrid
		UPA3_38	UPA6		
		UPA3_39	UPA6		
		UPA3_98	UPA3		
		UPA3_378	UPA3		
		UPA3_481	UPA6		
		UPA3_512	UPA3		
UPA3_6	8510	UPA3_37	UPA3/6	N/A	Hybrid
		UPA3_38	UPA3* (UUR2/UPA14?)		
		UPA3_39	UPA3		
		UPA3_98	UPA6		
		UPA3_378	UPA3* (UPA14?)		
		UPA3_481	UPA3		
		UPA3_512	UPA3		
UUR9_10	24318	UUR10_0043	UUR10	Serovar markers of both UUR9 and UUR10	Hybrid
		UUR10_0072	UUR10		
		UUR10_0138	UUR10		
		UUR10_0141	UUR10		
		UUR10_0329	UUR10		
		UUR10_0364	UUR10		
		UUR10_0376	UUR10		
		UUR10_0371	UUR10		
		UUR10_0653	UUR10		
		UUR10_0654	UUR10		
		UUR9_ORF01470 (UUR9 only)	UUR9		
		UUR9_ORF01469 (UUR9 only)	UUR9		
		UUR9/10_ORF00475	1 copy of UUR10 and 1 or 2 copies of UUR9		
UUR9_10	25353	UUR10_0043	UUR10	Serovar markers of both UUR9 and UUR10	Hybrid
		UUR10_0072	UUR10		
		UUR10_0138	UUR10		
		UUR10_0141	UUR10		
		UUR10_0329	UUR10		
		UUR10_0364	UUR10		
		UUR10_0376	UUR10		
		UUR10_0371	UUR10		
		UUR10_0653	UUR10		
		UUR10_0654	UUR10		
		UUR9_ORF01470 (UUR9 only)	UUR9		
		UUR9_ORF01469 (UUR9 only)	UUR9		
		UUR9/10_ORF00475	UUR10		

4.4.6 Prevalence of hybrid ureaplasmas

To quickly estimate the prevalence of hybrid ureaplasmas in a large number of clinical isolates containing multiple serovars, quantitative real-time PCR assays were performed using a universal quantification standard, a plasmid pUC19-UU, which contained one copy of each serovar-specific PCR target (except for serovars 4 and 5). The rationale for this test is based on the observation and hypothesis that mixtures of different serovars will still occur as mixtures in different ratios after re-growth because of different growth rates of serovars and different mixing ratio at starting point. On the other hand, a hybrid, as a single organism will contain the same ratio of different serovar markers at the starting point and after re-growth. Thus, by quantification of serovar markers in clinical isolates, we should be able to differentiate the hybrids and mixtures.

Among 271 randomly selected isolates out of 413 containing multiple serovars (except serovars 4 and 5), 28% were hybrids (Figure 4-3). Among the hybrid isolates, 23% were hybrids of two serovars, 4% were hybrids of three serovars, and <1% were hybrid of four serovars. Most of the hybrids contained markers of the same species (Table 4-4). There were 10 UPA hybrids and 59 UUR hybrids. Interspecies hybrids were found in 6 isolates. The most common hybrid types were hybrids of serovars 11 and 12. A clinical subculture could be a pure hybrid strain, a mixture of a hybrid strain and/or non-hybrid strains, or even a mixture of different hybrid strains, like isolate 12599, which was a mixture of two hybrids — a hybrid of serovar 2/11 and another hybrid of serovar 7/12, and a non-hybrid strain of serovar 13.

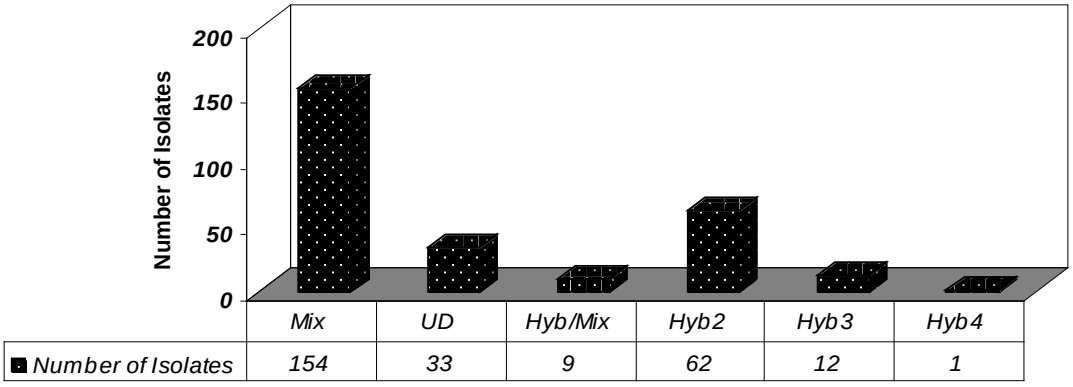


Figure 4-3. Prevalence of hybrid serovars in clinical isolates

Prevalence of hybrid serovars in clinical isolates. A total of 271 randomly selected clinical isolates (except serovars 4 and 5) containing multiple serovars were tested by quantitative real-time PCR, and the quantity of each serovar was calculated. To determine whether an isolate was a hybrid or mixed culture, an arbitrary standard was made based on the largest calculated deviation of the pUC19-UU control between PCR runs. A deviation of ≤5-fold among serovars was considered a hybrid strain (Hyb; the numbers following the term are the number of serovars detected in one); a deviation of 5- to 10-fold was considered a hybrid or mixture (Hyb/Mix); any deviation of >10-fold was considered to represent a mixed culture (Mix); and, finally, if both serovars were in low quantity, it was designated as undetermined (UD).

Table 4-4. Types of *Ureaplasma* hybrids

Hybrid Type	Hybrid serovars	Number of Isolates
Up Hybrids	1,3	1
	1,6	2
	1,14	2
	1,3,6	1
	3,14	1
	3,6	2
	6,14	1
Uu Hybrids	2,7	2
	2,7,12	1
	2,10	2
	2,10,12	2
	2,11	4
	7,9	1
	7,11	4
	7,12	6
	8,9	1
	8,10	1
	8,10,11,12	1
	8,11	7
	9,10	2
	9,11,12	1
	10,11	7
	10,11,12	6
	10,12	1
	11,12	10
Up+Uu Hybrids	1,2	1
	3,9	1
	3,4,11	1
	6,8	1
	6,10	1
	6,11	1
Total		75

4.5 Conclusions

To address the question whether differential pathogenicity exists at the *Ureaplasma* serovar level, an accurate typing method and a large number of clinical isolates are needed. The 14 serovar-specific real-time PCR assays have been proven to separate all 14 serovar type strains without cross-reactions [230]. We applied these assays to type 1061 clinical isolates from different patient populations. Thus far, this is the largest collection of clinical isolates that have been typed to species and serovar level by any method. Results indicated that *U. urealyticum* was significantly increased in men with NGU and women with PID and/or endometritis, which agrees with previous reports [84, 216-220]. However, no agreement in association of particular serovars with diseases was achieved among the different patient groups, including the invasive isolates recovered from usually sterile sites. Previous studies also showed no consistent data correlating individual serovars and pathogenic outcome, even among invasive isolates [4, 180, 196, 205, 222-225]. This suggests that serovar designation is not a reliable sub-species marker for determining differential pathogenicity of *Ureaplasma*.

We utilized the clinical isolates that were available to us for study and we acknowledge there are some potential limitations in the study populations and comparator groups. One such limitation is comparison of vaginal swabs from healthy women to serve as controls for comparison with endometrial tissue in the subjects with disease. A second is the relatively small number of control urine specimens (25) available for analysis from men without NGU. All of the clinical isolates were low

passage organisms, so it is unlikely that there would be selection of individual serovars that grow more rapidly in laboratory culture than others. However, unless one tests the original clinical specimens before any *in vitro* cultivation, it is not possible to know with certainty which serovars were there before the specimen is subjected to laboratory cultivation conditions. Furthermore, some serovars present in mixtures might not have survived prolonged storage and reconstitution. However, even with these potential limitations, our conclusions are not likely to be affected in the view of our finding hybrid serovars as a result of HGT. Thus, we must question the utility of serovar determination in assessment of pathogenicity.

Multiple serovars were detected in about 40% (413/1061) of the clinical isolates, while 6% (67/1061) were not typable and 4% (42/1061) showed conflicting species and serovar results. It has been observed since the earliest *Ureaplasma* typing studies that many clinical isolates contain multiple serovars [222-224, 226, 227]. Although cross-reactive typing reagents and mixed cultures were generally accepted as plausible explanations, it has never been completely clear whether certain strains can contain multiple serovar specificities and how this may occur. Failure to separate multiple serovars from some clinical isolates suggests the occurrence of hybrids. Our studies indicate there are pure organisms carrying multiple serovar markers. DNA sequencing of multiple loci with multiple base-pair polymorphisms provided evidence of HGT, and explains why some isolates were positive in more than one serovar-specific PCR assay. Because serovar-specific markers are exchanged between ureaplasmas, some organisms might acquire multiple markers, and some might completely lose all markers. Therefore, a likely consequence of HGT is the

emergence of hybrid ureaplasmas containing multiple serovar markers from one or more species, and untypable strains. Alternatively, the untypable strains may also represent new serovars that have never been characterized. The 14-serovar classification scheme was expected to be expanded at the time it was established [5] and several later studies using antibody-based or PCR methods have reported a certain number of untypable isolates [205, 222, 238]. To determine whether those untypable isolates represent new serovars or loss of markers, additional analysis such as whole genome sequencing of such isolates would be instructive.

HGT between *U. parvum* and *M. hominis* which co-localize to the mucosal surface of the human urogenital tract was recently reported [235]. Five clusters of genes encoding type I and III restriction /modification systems, transposases, and cell surface proteins were predicted to undergo HGT between the two phylogenetically distinct species.

In this study, we have reported that HGT occurs within *Ureaplasma* spp. Serovar markers and other sequences throughout the genome were exchanged among serovars, including the MBA gene, which was thought to be serovar-specific [202]. This finding questions the definition of serovars, because the phenotypic epitopes on which the serovars are based may change, combine, or be lost after HGT. The mechanisms involved in HGT in ureaplasmas are still not clear. Mobile genetic elements, such as conjugative transposons *Tn916* and *Tn1545*, and plasmids have been identified in ureaplasmas [239-241]. Ureaplasmas may also form biofilms [236], dense structures that enhance gene transfer [242]. Recombination may occur subsequently to transposition of the DNA into the recipient genome. Sequencing

hybrid isolates may help to identify possible hotspots of recombination. Furthermore, two sets of major surface antigen proteins MBA & UU376 and UU171 & UU172 of serovar 3 are phase variable due to a DNA inversion events [99, 184]. One of those phase variants would be undetectable using a serological assay based on the serovar 3 MBA. On the other hand, different serotyping results have been reported to occur in the same strain following subculturing [222, 243].

Genes involved in pathogenicity have not been identified conclusively in *Ureaplasma* spp. and we have shown that individual serovars are unlikely to be associated with differential pathogenicity. Therefore, bacterial load and different host immune responses may be alternative explanations for varied clinical findings. It has been reported that increased bacterial load is associated with NGU [203, 244, 245]. A study using an animal model of urinary tract infection showed that complications associated with *U. parvum* infection are primarily dependent on host-specific factors [246]. Further studies are needed to elucidate mechanisms of specific host response in these and other conditions that may be associated with these organisms.

In conclusion, this study demonstrates that HGT occurs naturally among human *Ureaplasma* species and serovars and *Ureaplasma* pathogenicity is unlikely to be associated with individual serovars. Our data suggest that “serotyping” is impractical and of limited value for assessment of pathogenicity. To clarify the mechanisms related to pathogenicity, future studies should focus on the specific immune response to *Ureaplasma* infections, although there is still the possibility that a gene or group of genes might be present in pathogenic and absent in commensal

ureaplasmas which yet have not been distinguished by examination at the species or serovar level.

Acknowledgements

This study was supported by Federal funds under grants RO1A1072577 and RO1209, contract NO1-AI-30071, from the National Institute of Allergy and Infectious Diseases and grant AI 28279 from the National Institute of Child Health and Human Development. Technical assistance by Donna Crabb and Amy Ratliff is gratefully acknowledged. This work was carried out in collaboration with the Mycoplasma laboratory at the University of Alabama at Birmingham with PI Ken Waites. Li Xiao and I contributed equally to the work described in this chapter. Li Xiao carried out real-time PCR assays and provided the obtained data to me and John Glass for analysis. I analyzed the data and designed new assays for analysis of chimeric strains and designed and assembled the qPCR control plasmid.

Supporting materials

Table S 1. PCR primers for sequencing of polymorphic loci between two serovars.

Compared serovar	Loci	Forward primer sequence (5' to 3')	Reverse primer sequence (5' to 3')	Notes
UPA1_6	UPA1_G0043	ATTAATCAAATGATCATCTTGAATTATGACA	TTTGAACATAAAATTTTATTAACTTGGCCATC	
	UPA1_G0098	CATTAAATAGTGAAATTAAAACCTTTTTTAAACAA	TGCTCATTAAATCATTTATCAATTTAGTGATAAG	
	UPA1_G0100	AAATTAGAAAAAAAACAAAACAACATACAAC	ATCAGAAGTATATAATTCATTATTATTATCACGAC	
	UPA1_G0101	GTATTAGTGAAATAAGTAATCGTTTGGGAA	GTTTGATCATTACAACGTTATGATCATCTA	
	UPA1_G0114	AACTACTAATAGTATTGAGTTATTTTTAGAAGG	TTTGCACTAGTATCATAAATTTCTTTTTCA	
	UPA1_G0293	CAAATTAATATTTGGTGTAAATGAAAACAATG	TTTATAGGATATTCAGGATATTGAATACGAT	
	UPA1_G0359	ATAAATTATTATTAGTAATTCCTAATATGCCAGAT	TACTAATTCTACTTTATGGAATTGGTGTTG	
UPA3_6	UPA3_0037	GATTCTGAATTTATTAGTGAAACTATTAAAACAA	TTTATAGGATATTCAGGATATTGAATACGAT	
	UPA3_0038	GAAAATTTAAATAAATCAGCAATGTATGGG	CAACAGTTGTTGTTTTTAATCAGTTTAAGA	
	UPA3_0039	CAAGACTTGATAGTGATGATAAGATTTACC	GTTTTTTAAGAATTAGATTTCAAGGACGAG	
	UPA3_0098	ATAATTATCAATTTACTAATAATGAATGAGAATCT	GCCTGTAGTTCCTAATAATTTTTTATCATTT	
	UPA3_0378	TCAATCAAGAAATCATTCCAATTTGAGTA	TGTTCTTGTCCACCAATATAAACATC	
	UPA3_0481	TAATACACGTTTTAAATCACCTCTTTTATATAAAT	GTACTAAATATCCTCAAATTTGACTAGGAA	
	UPA3_0512	ACATTCATGTTATTGATTATTTAACTAAACGA	TAATATGAAGTTGTTGGTCTAATAGCTTAC	
UUR9_10	UUR10_0043	GATTTAATCATTGCTGAGGTTAGTGATCTA	ACCCAAGTACATTGAATCTAATTCATATG	
	UUR10_0072	CAAATGATTGATGTAGCAAAACATTCATA	AATAAATCAACTAAGAAGGCAACACATAAC	
	UUR10_0138	TTGGTGATGATGTTTATTTTGATACA	AATTGGTTGTTTTATAATCTGCTTGA	
	UUR10_0141	AACCTAATGATGATAATGATTATCG	TCATCATAAACTACTGATCCTGATG	
	UUR10_0329	ATTATCAGTTAGCATTTTATCTTCTTGTCA	TTCTTGCTTTGTTTCTTGATTTTTAAG	
	UUR10_0364	TTTCTTACAAAGAATGACTAATAATCG	AATATTCACTTCCTACTTTTGTTAAATTAC	
	UUR10_0376	CTGGTGTATTTATTAATTTTATTGCTATTG	TGAAACAGCTGTTAAACAAGATGA	
	UUR10_0371	TTAAAATTTAAAGACACAACAATAGAG	TGATGTTATTTCAGCAATATGATT	
	UUR10_0653	ACATTAAATTATGTAAAATTTGCATC	CTGCAATTTCTAAATCATCAAC	
	UUR10_0654	CAGTAGTTGGATTTATGAAATTATATGC	GCTAATGAATCATAAGCATTACCAG	
	UUR9_ORF01470	GCAAAGAAGACTAAAGAAATAGAGG	CATCTTGTTAAGATCAAATTGATTTAGG	UUR9 only
	UUR9_ORF01469	AATACTAAGGCAGTAAATAGTCTAGC	TTCTTCGTGAGTTTCATCATCG	UUR9 only
	UUR9-10_ORF00475	GTATTTATCCTTGGACTTCATTCATAG	AAGAACATACTCTCTTCTGCATCA	
UP_MBA		TATTTGCAATCTTTATATGTTTTCG		
UP1_6_MBA			GGAATATAAAATAAAAAAAATAGCCCA	
UP3_14_MBA			CAGTTTTAAAATAATAAAAAAATAGCTCA	
UUR10_60.4C_F		TGAATTTAGCGATGTTGATGCA		PCR marker for UUR10
UUR10_60.1C_R			TTATTTGCTACCTGATGCTGATC	
UUR9_61.5_F		AGATACGGGGGGGTTAAGTG		PCR marker for UUR9
UUR9_61.8_R			CGCTGGGAAAATATCAAGACCA	

Chapter 5. Expanding the Genetic Toolbox for Ureaplasmas

5.1 Abstract

Over 50 years of research have yielded little understanding on why UPA and UUR produce invasive diseases in some instances and not in others. Although differences in host factors can lead to increased susceptibility to ureaplasmal infection, evidence suggests that some strains of UPA and UUR may have a greater pathogenic potential than others. Understanding of host-pathogen interactions and factors involved in pathogenicity of these organisms has been hampered by the lack of molecular tools for modification of their genomes. Such tools have been prerequisites for understanding pathogenicity factors and host-pathogen interactions for well-studied pathogens. We attempt to apply traditional molecular microbiology and synthetic biology methods to develop tools for genetic manipulation of *Ureaplasma* species. Adapting the novel synthetic biology tools to UU, like cloning, and engineering bacterial genomes in yeast, and genome transplantation can allow for the creation of targeted single or multiple mutants that will greatly increase the understanding of ureaplasma pathogenicity.

5.2 Introduction

Understanding of the role of ureaplasmas in human diseases has been thwarted thus far due to two major barriers: 1) Lack of suitable diagnostic tests that can identify and characterize the two species and their respective serotypes that could account for differential pathogenicity 2) Lack of genetic manipulation tools for the

creation of deletion mutants and their complementation to study the role of potential pathogenicity factors

A major effort in the field of ureaplasma research over the past 20 years has been focused on overcoming the first barrier. Since 1990, nearly 300 peer-reviewed publications have described basic and clinical investigations involving ureaplasmas using PCR and various other nucleic acid-based techniques. Based on the genome sequences of the 14 ATCC type strains of each ureaplasma serovar, our group has developed species and serovar specific real-time PCR assays to detect ureaplasmas in patient isolates [132], which were applied to a collection of 1,061 clinical isolates [133]. The results of this large scale serotyping study showed 39% of clinical isolates contained more than one serovar, and 28% of the evaluated mixed isolates were chimeric organisms containing markers from up to 4 serovars (Table 5-1and Figure 5-1). This suggested that ureaplasmas are subject to extensive horizontal gene transfer. Therefore it is possible that ureaplasmas do not exist as stable serovars in their hosts, but rather as a dynamic population exchanging genes, most likely not only among themselves, but also with other bacteria in their environment. Based on these findings, the long sought after serotyping assays, are of limited value, and there is no utility in trying to associate individual serovars with specific disease. The differential clinical outcome of ureaplasmal infection may be due primarily to presence or absence of potential pathogenicity factors in the colonizing ureaplasma strain [134] and/or patient to patient differences in terms of autoimmunity and microbiome.

Table 5-1. Distribution of UU Clinical Isolates
The table shows the results of the real-time PCR assay (Chapter 4) applied to 1061 clinical isolates of UU. The results are broken down to number of samples containing one, multiple, or unidentifiable serovar/species

Serovars in isolate	No. Isolates	
1	574 (54.1%)	
2	223	414 (39%)
3	91	
4	41	
5	37	
6-10	22	
No serovar signal	67 (6.3%)	
No species signal	6 (0.6%)	
Total	1061	

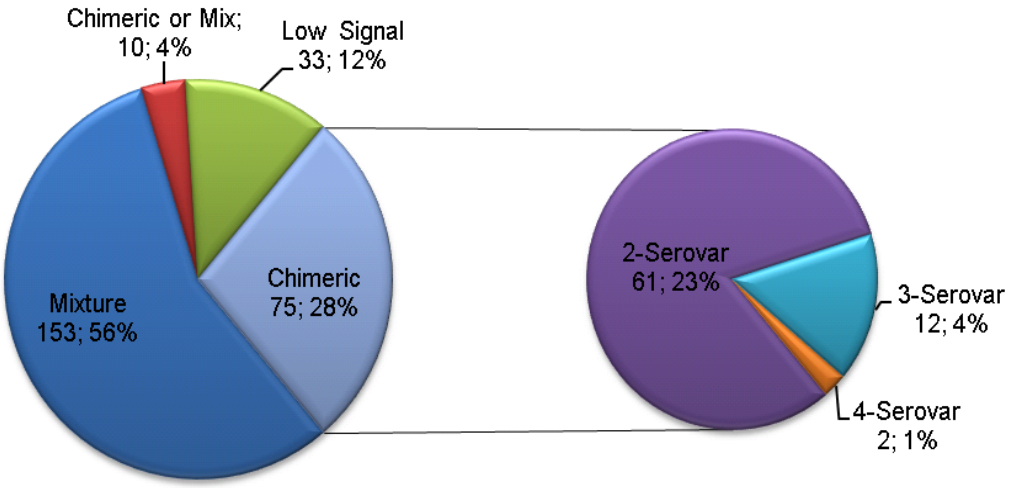


Figure 5-1. Distribution of 271 Mixed Ureaplasma Clinical Isolates.
271 Clinical Isolates with real-time PCR assay singnals for more than one serovar were subjected to quantitative real-time PCR to determine the number of chimeric organisms compared to co-infections. Data label indicates: group type (chimeric or mixture, e.g. co-infection); number of clinical isolates in the group; percentage of 271 tested isolates or of 75 chimeric strains.

The second major barrier to the understanding of ureaplasma pathogenicity is the lack of molecular tools for creation of deletion mutants to study the role of potential pathogenicity factors. This problem has not been explored nearly as exhaustively as the development of serotyping assays. The capacity to create deletion mutants has been a prerequisite in understanding host-pathogen interaction and factors involved in pathogenicity in many well studied pathogens.

Approximately twenty years ago efforts to develop genetic manipulation systems for ureaplasma were unsuccessful (personal communication with Kevin Dybvig). This is one reason little use was made of the first ureaplasma genome sequence, which was published early in the genomic era. New technologies such as TN4001, EZ-tn5 trnasposomes, OriC plasmids, expression of *E. coli* RecA via a suicide plasmid, and the technology developed for synthetic biology offer new possibilities for development of genetic tools for ureaplasma. More recently, Tn4001 was used to randomly mutagenize an UPA3 clinical isolate; however the methods used for that specific ureaplasma clinical isolate have not worked for any other ureaplasmas (Brad Spiller, Cardiff University, personal communication; manuscript in preparation). Nonetheless, transformation conditions for ureaplasmas have not been extensively explored. The development of tools for genome modification of *Ureaplasma* spp. will tremendously facilitate the study and understanding of factors leading to differential pathogenicity of *Ureaplasma* strains. Should transformation with a transposon prove successful, transposon mutagenesis studies can be initiated to screen for mutations of interest. The search of the ureaplasmal IgA protease gene responsible for the reported enzymatic activity [171, 172, 188] could be facilitated by

the use of random transposon mutagenesis in a future study. Mutants can be extremely valuable for study in existing animal models for ureaplasmal infections in humans [27, 56, 121, 247]. Furthermore, the increased understanding of pathogenicity factors can aid in development of better screening methods, drug targets and vaccines to decrease the morbidity associated with ureaplasmal infections.

5.3 Materials and Methods

5.3.1 Ureaplasma cultures

U. parvum and *U. urealyticum* were cultured in 10B growth medium. This growth medium, per 1 L, contains 200 ml Heat-Inactivated horse serum (Gibco Life Technologies) or heat-inactivated porcine serum (Life Technologies), 100 ml 25% yeast extract (Gibco Life Technologies), 688 g Mycoplasma Broth Base (BD BBL), 1 ml 1% phenol red, 5 ml IsoVitaleX Enrichment (Gibco Life Technologies), 1 g urea (Sigma), and 2.5 ml, 400,000 U/ml penicillin (Sigma). Phenol red was used in the 10B growth medium as pH indicator to monitor the growth of ureaplasmas. During ureaplasma growth, urea is hydrolyzed to yield ATP, carbon dioxide, and ammonia. The ammonia changes the growth medium from acidic to basic and the phenol red from yellow to red.

Ureaplasma growth was monitored through a color changing units (CCUs) assay, in which 1:10 serial dilutions of cells up to 10^{-8} were used. The highest dilution to change from yellow to red indicates the approximate number of cells in the starting ureaplasma culture. The organisms can be cultivated on agar plates as well. They will

produce granular brown urease-positive colonies approximately 15–60 μm in diameter on A8 agar after 24–48 hours of incubation at 37°C and 5% CO_2 due to calcium chloride incorporated into the media. Alternatively 10B agar can be prepared and used to grow ureaplasma colonies. Colonies can be observed after 48-72 hours of incubation and appear opaque with light silver tone. Mature colonies exhibit a fried egg shape with irregular, wavy colony edge (Figure 5-2).

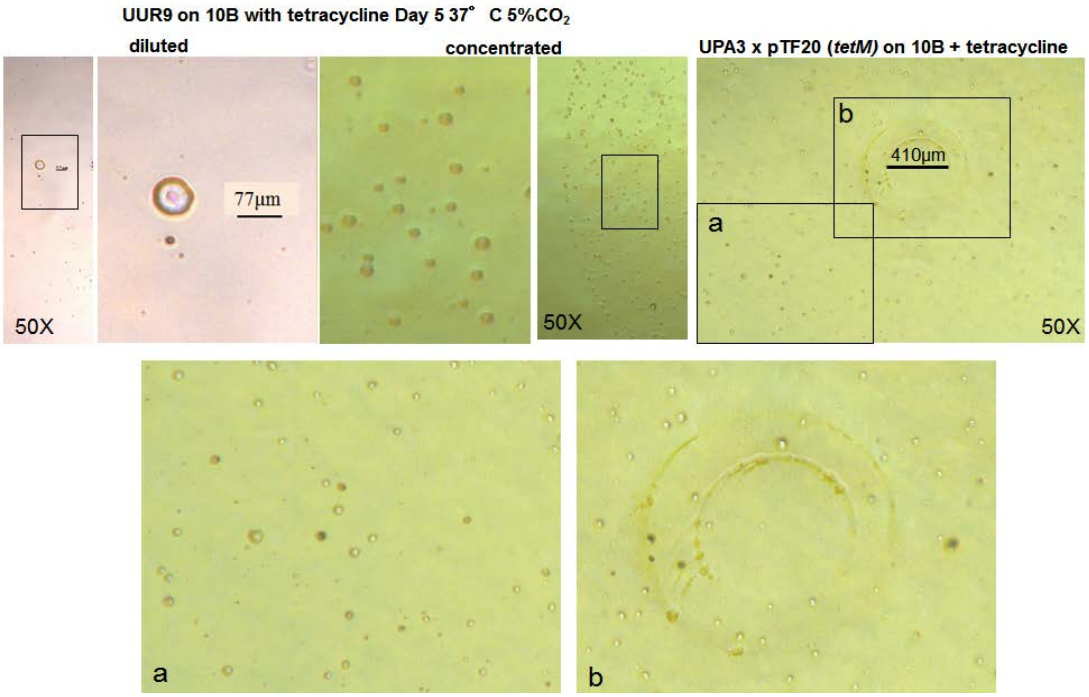


Figure 5-2. *Ureaplasma* colonies on 10B Agar.

The agar plates were incubated at 37°C with 5%CO_2 for 3-7 days. At day 3 colonies can be detected and have a transparent to silverfish gloss and no “fried egg” center. This picture is of 5 day old colonies at 50X magnification. By day 7 the edge of the colony becomes irregular and wavy. When cells are too crowded, the colonies don’t develop the “fried egg” look.

5.3.2 Extracting intact genomic DNA

The buffers, enzymes, and agarose in the following procedure are provided in the CHEF Mammalian Genomic DNA Plug Kit (BioRad catalog number 170-3591). Different cell concentrations per one plug were explored. Finally 20ml of overnight ureaplasma culture (pH 7.5-8.0) per plug was harvested by centrifugation at 10 000 x g for 60 min at 10°C. Media was aspirated, leaving about 1ml of it in the tube. The pellets (invisible) were re-suspended in the remaining media and all pellets were merged together in one 30ml Oakridge Centrifugation tube. Cells were harvested by centrifugation at 27 000 x g for 15 min at 10°C and washed in 20 ml of PBS then centrifuged at 27 000 x g for 15 min at 10°C. A cell suspension was prepared in 50µlper plug of PBS. The 50 well plug mold makes 5 ml of agarose plugs. A 2% low melt agarose solution in sterile water was prepared and melted using a microwave. The cell suspension solution and the 2% low melt agarose solution were equilibrated to 50 °C in a water bath. Next the cell suspension was combined with an equal volume of 2% CleanCut agarose and mixed gently but thoroughly. This results in a final concentration of 1% agarose. The cell/agarose mixture was kept at 50 °C, and transferred to plug molds. The agarose was allowed to solidify. This step can be expedited by placing the molds at 4 °C for 10–15 minutes. This also adds strength to the agarose for removal from the mold. 25 ml of Proteinase K Reaction Buffer was added in a 50 ml conical centrifuge tube. The solidified agarose plugs were pushed into the 50 ml centrifuge tube containing the Proteinase K solution by using the snap off tool provided on the plug mold. The plugs were incubated overnight at 50 °C without agitation. Next the plugs were washed four times in 50 ml of wash buffer for

30 minutes to 1 hour each at room temperature with gentle agitation. 1 mM PMSF was added during the second or third wash to inactivate any residual Proteinase K activity. The plugs were stored at 4 °C for 3 months to 1 year. The recipes of the buffers that come with this kit can be found in the manufacturer's manual.

5.3.3 Restriction Digest and CHEF Gel Analysis of gDNA Plugs

The plugs to be restricted were washed for 30 minutes in 0.1x wash buffer or TE. One plug per digest was placed in a sterile 1.5 ml microcentrifuge tube. The plug was incubated with 1 ml of the appropriate 1x restriction enzyme buffer for about 1 hour with gentle agitation at room temperature. The buffer was removed and 0.3 ml of fresh 1x restriction enzyme buffer was added. 30-50 U per 100 µl plug restriction enzyme, in this case 50U of SmaI or SdaI to linearize the UUR9 genome, and ApaI to cut twice the UUR9 genome was added and incubated overnight at the appropriate temperature. After overnight digestion, the buffer was removed and 1 ml of wash buffer was added. 1⁄4 of a plug per sample was loaded on an agarose gel per well on a 1% agarose gel along with a lambda DNA ladder. The gel was ran in the CHEF Gel III system with the following parameters: Buffer: 0.5x TBE; Temperature: 14 °C; Switch Time: 50-90 seconds; Run Time: 22 hours; Angle: 120°; Voltage Gradient: 6 V/cm. The gel was stained with SybrGold and scanned.

5.3.4 Minimum Inhibitory Concentration Assay:

By using the same methods as described in the literature, we found the MIC for both UPA3 and UUR9 for puromycin, tetracycline and gentamycin [248]. MIC

tests were performed in flat bottom 96-well plates by first adding 360 µl of 10B growth medium with the antibiotic being tested at its highest tested concentration and 180 µl of 10B growth medium without antibiotic to the rest of the wells. Next, the antibiotics were serially diluted in a 1:2 dilution down the rows, leaving the bottom row as a control with no antibiotic. The maximum concentration tested for puromycin was 320 µg/ml, and tetracycline 360 µg/ml. Twenty µl of cells were then added to each well in the first column and serially diluted in a 1:10 dilution across the columns, leaving the right-most column as a control with no cells. The 96-well plates were then incubated at 37°C to grow until no further color changes were observed. Samples were then placed at room temperature for one hour before observations were made as previously described [248].

5.3.5 Testing for Nucleases Present on the cell surface of UU cells

Ureaplasma cells were treated according to the mycoplasma PEG mediated transformation protocol and incubated with plasmid DNA under the transformation conditions in the protocol. Reactions were stopped with phenol/chloroform at 0, 15, 30, and 45 minutes and DNA will be precipitated with ethanol. The samples were analyzed by gel electrophoresis. A sample without cells was used as a negative control. We tested the dependency of the nuclease activity on Ca^{+2} and Mg^{+2} ions by removing either and both ions.

5.3.6 Preparation of Crude Cellular Extracts

To prepare cellular extracts 3L overnight culture of UPA3 recipient ureaplasmas cells were harvested through centrifugation as described above and re-suspended in 0.5 ml Lysis Buffer [20mM Tris pH7.5, 10% Glycerol, 150mM NaCl, 1mM DTT, and 0.1mM EDTA; last two components are added to the buffer immediately before the experiment]. The cell suspension was sonicated on ice until the solution becomes almost clear. The resulting solution was centrifuged to remove cell debris and supernatants were tested for protein concentration [249]. Extracts were stored at -80°C.

5.3.7 Crude Cellular Extract Restriction Enzyme Assay

To test for restriction enzyme activity of *Ureaplasma* we were incubated with a DNA plasmid in all 4 available restriction enzyme buffers by New England Biolabs with the extract at 37°C. The reaction was stopped at different time intervals. The DNA was extracted and precipitated with phenol/chloroform and ethanol and the DNA was analyzed by gel electrophoresis for digestion patterns [249].

5.3.8 Transformation/PEG Concentration Optimization:

Ureaplasma cells grown at 37°C overnight were transformed by chemical mediation of PEG 8,000 (USB) dissolved in 1x phosphate buffered saline solution (PBS, Life Technologies). Different concentrations of PEG solution were explored to determine the optimal PEG concentration. Once optimal PEG concentration was determined 50% PEG was used in subsequent transformation attempts.

UPA3 cells, (5 ml/experiment) were pelleted in bulk at 10,000 g for 20 minutes at 4°C in a Beckman Coulter JA 25.50 rotor. The media was pipetted out leaving 1 ml. the pellet was resuspended and moved to a 1.5 ml microfuge tube. The cell suspension was pelleted at 10000 g for 7 min at 4°C in a Beckman Coulter F301.5 Rotor. Cell pellet was then washed with 1x PBS buffer, and then resuspended and treated on ice with 125 µl of 100 mM calcium chloride/sample for 30 minutes.

During the 30 minute incubation, each experiment was prepared in separate 30 ml centrifuge tubes (Oakridge) by adding 10 µg of yeast tRNA (Sigma), and 1 µl of the plasmid. Then, to one tube at a time, 125 µl of calcium chloride treated cells were added, 1 ml PEG 8,000 (varying concentration for optimization experiments and 50% subsequent experiments) was added, mixed gently by rolling the tube for one minute, and then 5 ml of 10B growth medium without horse serum was added and mixed in the same manner. When all samples were mixed with all components, they were incubated 4 hours to recover at 37°C. During the recovery, a CCU assay in sets of eight tubes per sample was prepared with 900 µl of 10B growth medium with selective antibiotic above the determined MIC for experimental samples. For tetracycline, 10 µg/ml was used.

After the 4 hour recovery, samples were centrifuged at 10,000 g for 20 minutes at 4°C. The supernatant was discarded and samples were re-suspended in 100 µl of 10B growth medium without selective antibiotic and added to the first tube of their corresponding CCU assay set of eight. The sets of eight were then serially diluted from tube to tube in 1:10 dilutions, and incubated at 37°C overnight to grow. The sample with the cells grown to the farthest dilution indicated the most effective

concentration of PEG 8,000 during the PEG optimization. For a successful transformation the transformants were expected to grow in the media containing tetracycline and growth should be sustained in subsequent cultures of the transformant in the presence of tetracycline.

5.3.9 Plasmids Used and Their Construction

The EZ-Tn5 transposon (Epibio) containing a puromycin resistance gene, was used to transform UUR9 and UPA3. Plasmids containing different Tn5 constructs were made. We used the SQFP and SQRP provided by the kit to amplify the backbone of the pMOD plasmid. The promoter and antibiotic resistance genes were PCR amplified with primers containing overhangs complimentary to the sequence of their neighboring assembly DNA pieces. All PCR amplicons were purified and used to assemble to plasmid through Gibson Assembly. The final Tn5 transposon insertion DNA was PCR amplified using the PCRFP and PCRRP primers (primers were provided in the EZ-Tn5™ Custom Transposome™ Construction Kit by Epicenter and sequences of the primers can be found in the kit's manual). The different Tn5 transposon insertion sequences contained genes conferring resistance to either puromycin or tetracycline, and varying promoter regions upstream of the antibiotic resistance. These insertion sequences will be referred to as Tn5relevant antibiotic. i.e. Tn5Puro, Tn5TetM. The resulting transformed cells are selected by CCU assay and/or plated on agar plates for selection and colony analyses. Sanger sequencing with primers SQFP and SQRP (Figure 5-3) can be performed to determine the

transposon insertion location [250]. Figure 5-3 depicts the plasmid map of one variation of the pMOD-2 plasmid.

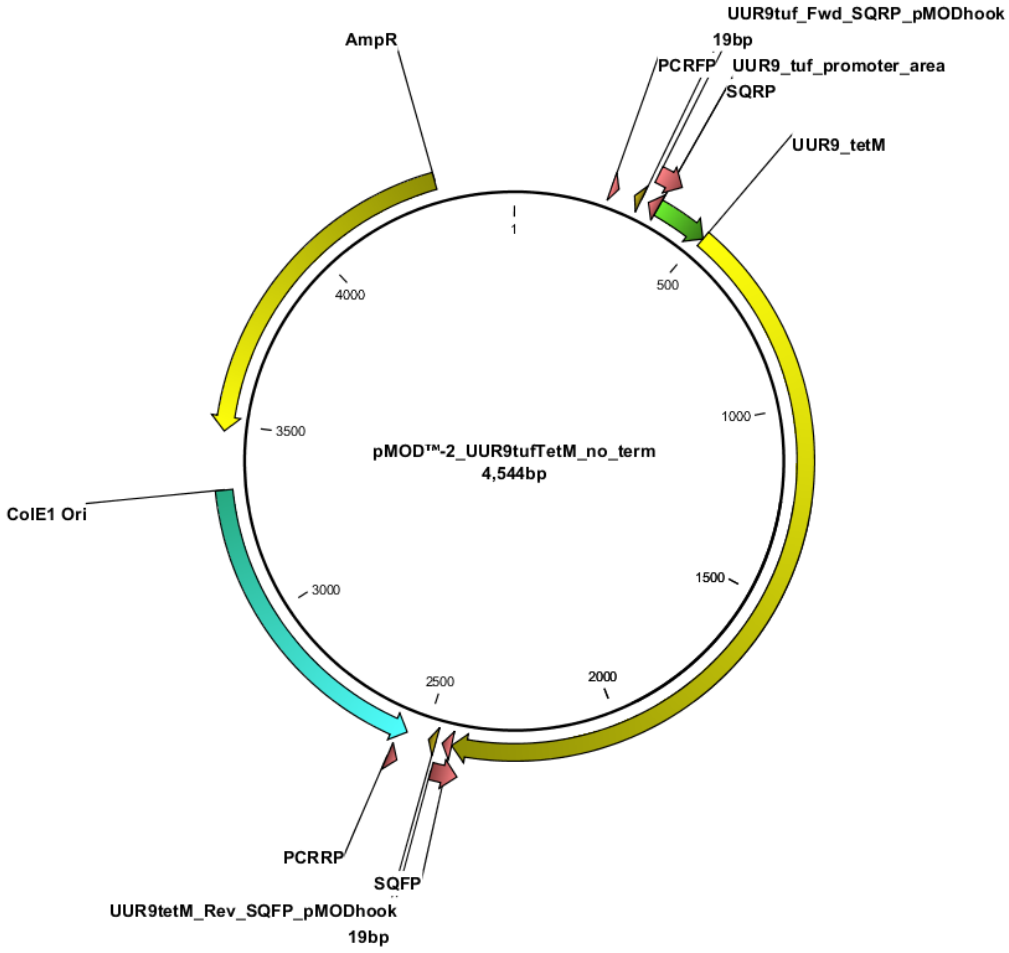


Figure 5-3. pMOD-2 custom EZ-tn5 plasmid.

This is a plasmid map of one variation of the custom design pMOD-2 plasmid provided with the EZ-Tn5™ Custom Transposome™ Construction Kitr by Epicenter. The backbone of the plasmid contains ampicillin (AmpR) resistance gene and *E. coli* ColE1 Ori origin of replication for plasmid propagation and selection in *E. coli.* We inserted PCR amplicons of the UUR9 *tuf* promoter area and the UUR9 *tetM.* To allow for GA into the pMOD-2 plasmid the gDNA areas were amplified using the UUR9tuf_Fwd_SQFR_pMODhook and UUR9tetM_Rev_SQFP_pMODhook primers. The tn5 transposon DNA is PCR amplified using the PCRFP and PCRRP primers as described in the materials and methods section. The 19bp repeats of the tn5 are labeled. The SQFP and SQRP primers can be used in sequencing reactions to determine the insertion site of the transposon.

According to the manufacturer's instructions, we produced transposomes by combining 2 µl of the Tn5 insertion piece, at a concentration of roughly 100 ng/µl, was combined first with 4 µl of EZ-Tn5's transposase, then 2 µl of glycerol. The reaction was then set to incubate at room temperature for 30 minutes. Tn5 transposomes were then used to transform either UUR9 or UPA3 cells, grown overnight with the same transformation procedure described above. Transformation experiments were done using 50% PEG 8,000 as described above. No DNA and no cells controls were used by substituting the transposomes (no DNA) and cells (no cells) with 1 µl and 125 µl of 1x PBS buffer, respectively.

Similar CCU assays in sets of eight tubes per sample were used to monitor ureaplasma growth. However, CCU assay sets also included controls with one set grown in 10B growth medium with selective antibiotic and one set grown in 10B growth medium without selective antibiotic.

For puromycin, 100 µg/ml was used, and for tetracycline, 10 µg/ml was used. Samples were incubated at 37°C overnight to grow. The next morning, all sets were re-titrated and re-titrated one more time 6 hours later. We considered the transformation positive if the samples containing the transposon grew in the tetracycline containing media to a farther dilution than the no transposon control in the tetracycline containing media. Due to the high concentration of the cells used for the transformations to compensate for low transformation efficiency, the no transposon control usually changes the color of the tetracycline containing media in the 10^{-1} dilution.

We used a variety of plasmids to attempt ureaplasma transformation. Most plasmids were mini tn4001 derivatives with either tetracycline, puromycine, or kanamycin resistance genes. We used a pTF20 with *tetM* to perform PEG concentration optimization tests.

We constructed multiple variations of the EZ-tn5 trasnposon plasmid, pMOD2 (Epicenter), by varying the antibiotic gene (tetracycline, puromycin) and its source and the driving promoter. The *tetM* was amplified from pTF20 or UUR9; plasmid; the putative UUR9 *tuf* promoter was amplified spanning the intergenic region upstream of the *tuf* and downstream of the gene before *tuffrom* extracted genomic DNA; the backbone of the pMOD2 was amplified to use as a base for the different tn5 variations. All PCR amplicons were cleaned up using the The Wizard® SV Gel and PCR Clean-Up System (Promega) and used in different combination to produce a variety of tn5 transposons. Primers for all the PCRs were designed to have appropriate overhangs to allow assembly of the different plasmids using Gibson Assembly (GA) [251]. GA reactions were prepared on ice with 25 fmol of each DNA component and 10 µl of 2x GA Master-Mix (Invitrogen), brought to a total volume of 20 µl. Samples were placed in a preheated 50°C PCR machine for 30 minutes.

GA products were placed on ice until ready to electroporate into DH10B or ElectroMax *E. coli*. One µl of product was combined with 40 µl of defrosted *E. coli* cells in a PCR tube and then transferred into a cuvette on ice. Bubbles were minimized by lightly tapping cuvettes on a hard surface. Cuvettes were wiped of any external moisture, and then cells were shocked with a setting of 2000 volts, 25 Farads, and 200 Ohms. Cells were immediately re-suspended in 1 ml of S.O.C

medium (Invitrogen) by pipetting up and down gently a few times then transferred to 14 ml round bottom tubes to recover in a shaker for 1 hour at 37°C. After the recovery, samples were screened by plating 200 µl of each sample on pre-warmed Luria Broth (LB) agar with ampicillin (100 µg/ml) then turned upside-down to grow overnight at 37°C. Individual colonies were picked into 2 ml of LB with tetracycline (10 µg/ml) and placed at 37°C to grow overnight. Cells were spun down and harvested using Qiagen mini-prep kit to extract plasmids. Extracted plasmids were then screened by a PCR that amplified the transposon insertion sequence. Transposon DNA was obtained through PCR using the PCRFP and PCRRP primers and the corresponding pMOD2 customized plasmid as template or through restriction enzyme digest followed by gel purification of the transposon DNA using SYBR Gold stain for visualization (Life Technologies), and then cleaned-up by using Wizard SV Gel and PCR Clean-up System (Promega).

5.4 Results and Discussion

5.4.1 Development of traditional tools for genetic manipulation of UU genomes

Considerable progress has been made in the field of *Mycoplasma* genetics that makes some species in this taxon accessible for genetic studies. Such studies are facilitated by traditional tools, like transformation of cells with transposons and plasmids through electroporation, PEG, and conjugation. Several transposons and their derivatives are used in mycoplasmas to disrupt or insert genes. Autonomous replication plasmids (OriC plasmids) and suicide plasmids have been developed for

some of the mycoplasmas that allow targeted disruption of genes. These tools have led to better understanding of the biology of some mycoplasmas. We believe that these methods and tools can be adapted and used for genetic manipulation of UU species. Through personal communication we know that initial attempts to transform ureaplasma with Tn916 via PEG-mediated transformation approximately twenty years ago were unsuccessful. Tn916 is naturally present in the genome of the ATCC UUR9, and it confers tetracycline resistance through the expression of the *tetM* gene it carries. This suggests that Tn916 is a suitable transposon choice. More recently, Tn4001 was used to randomly mutagenize an UPA3 clinical isolate; however the methods used for that specific ureaplasma clinical isolate have not worked for any other ureaplasmas (Brad Spiller, Cardiff University, personal communication; manuscript in preparation). Artificial transformation of cells usually involves naked DNA. Mycoplasmas produce potent nucleases [252] to scavenge nucleotides from their environment, since they cannot synthesize them. Recent experience at the J. Craig Venter Institute (JCVI) showed that these nucleases can digest the naked donor DNA in genome transplantation experiments. Furthermore, active restriction-modification systems can digest the donor DNA after it entered the cell and before it has been methylated by the recipient cell to make it resistant to those restriction enzymes. The comparative analysis of the genome sequences of the 14 ATCC serovars shows that all serovars have putative restriction-modification systems and putative surface and secreted nucleases [237]. Therefore, initial experiments were completed to characterize the stability of the donor DNA in the presence of UU cells.

After these potential issues are addressed we proceed to develop transformation protocols as well as vectors to genetically manipulate ureaplasmas.

5.4.2 Testing UU cells for the presence of surface and restriction nucleases

Our previous experience with *Mycoplasma pneumoniae* cells shows that they have one or more extracellular nucleases capable of digesting plasmid DNA when it is incubated with the cells under transplantation conditions. Diluting the cells to slow DNA degradation showed that the supercoiled plasmid is cut to a linear molecule before it is destroyed, suggesting that there may be both endo- and exo-nuclease activity present. This nuclease activity is dependent on the presence of Ca^{+2} and Mg^{+2} ions. Removing both these ions abolishes nuclease activity. Pre-incubation of the cells with trypsin was enough to allow PEG-mediated transformation in the presence of $CaCl_2$ of *M. pneumoniae* with plasmids. All 14 sequenced UUs contain 11-14 putative nucleases of different functions. Two of these putative nucleases appear to be membrane bound based on the presence of a transmembrane domain and a signal peptide. One of them has a lipoprotein attachment site and a calcium binding domain. Therefore, before we attempted to transform ureaplasma cells we assayed them for extracellular nuclease activity under the conditions of the transformation protocol. We did not detect surface nuclease activity under the tested conditions. Since there are several formulations of the ureaplasma growth medium in the literature, we tested the presence of nuclease activity using three of these formulations. It is interesting to note that we could not recover the DNA used for the

assay when the growth media was prepared with PPLO or TSB as broth base even in the absence of the cells. We did not investigate the reason the plasmid could not be recovered.

All 14 ureaplasma serovars have 6-12 genes annotated with different predicted restriction-modification functions. Seven of the serovars may have a functional type III RMS, the rest of the serovars appear to have only the type III restriction subunit, which would make the system not functional. Only serovar 9 appears to have both genes of the type II RMS. In most serovars only the type II restriction enzyme is annotated. This system does not require the methylase and nuclease to be in a complex to function, however another methylases present in the genomes may be utilized to methylate and protect the DNA from the type II restriction enzyme. It is an interesting observation that all serovars contain multiple type I specificity proteins, which was also observed in *M. pneumoniae.* Only five serovars (3, 4, 9, 10, 12) have annotated type I methylation and restriction subunits. It is possible that the type I restriction system in these serovars uses the same restriction enzyme in combination with different specificity factors. Because only 3 of the serovars have been completely sequenced and assembled into a single contig, and minimal manual curration has been provided for all the serovars, it is possible that the missing enzymes are in a sequencing gap or have been missed in the automated annotation process. This makes it difficult to predict with confidence the potential restriction-modification activity of these organisms. Therefore, we tested their restriction activity through the preparation of crude cellular extracts. We carried out a restriction nuclease assay using lambda DNA as a substrate and the four available restriction

nuclease buffers available through NEB. We did not detect restriction nuclease activity under these conditions. These results do not indicate that ureaplasmas don't have surface and restriction nucleases. Rather these results show that under the conditions we used the activity of such nucleases was not detectable.

5.4.3 Minimum Inhibitory Concentration (MIC) Assay:

We determined the tetracycline MICs for UUR9 and UPA3 to be over 64 µg/ml and 8 µg/ml, respectively. Our results agreed with previous tetracycline MIC tests [248]. The UUR9 tetracycline MIC is higher because the genome contains a tetracycline resistance gene. Therefore we used 10 µg/ml tetracycline to select transformed cells. Puromycin MIC for both UUR9 and UPA3 were 80 µg/ml, therefore, we used 100 µg/ml puromycin for selection of transformed cells.

5.4.4 Development of a Transformation Protocol for Ureaplasma species

There are three common methods used to transform mycoplasma species: chemical treatment with calcium chloride or polyethylene glycol (PEG), electroporation, and liposome encapsulated DNA delivery [142, 143]. We attempted to transform ureaplasma first by PEG mediated transformation. We used a vector containing a Tn4001 transposon conferring tetracycline resistance through the tetM gene. This transposon and its derivatives have been useful in mycoplasmas and another group was successful in using a Tn4001 derivative transposon to transform several UPA clinical isolates [146-151]. This transposon, however, failed to transform UUR isolates and the ATCC type strains of UPA and UUR.

We carried out multiple optimizations to the PEG mediated transformation procedure used for mycoplasmas. We tested different PEG concentrations, recovery times and conditions, different transposon systems, as well as different culture stages [253, 254]. Our starting point was the transformation protocol developed by Lartigue et al for *Mycoplasma capricolum* [249, 254, 255]

5.4.4.1 Growth Media Composition

Several growth media compositions can be found in the literature for cultivation of Ureaplasmas. In general the media used for culturing ureaplasmas is called 10B. The base of the 10B broth can be Pleuro pneumonia-Like Organisms (PPLO) broth base, Tryptic Soy Broth (TSB) base, or Mycoplasma Broth Base (MBB) [155, 256]. Urea is always present in the broth as ureaplasmas hydrolyze urea for the generation of ATP. Ureaplasmas require serum in the growth media. The serum type can be horse, fetal bovine, or porcine. Different concentrations of serum are found through the literature. Another component of the media is yeast extract. For the initial culturing of clinical isolates fresh yeast extract is preferred (verbal communication with Ken Waites), however, for established laboratory strains commercially prepared yeast extract can be used. The concentration of yeast extract can be varied as well. Optionally, Isovitalex supplements can be added to the media. Since transformation procedures are stressful to the cells being transformed, we chose to add the higher concentration of serum and yeast extract published and add Isovitalex to make the growth media rich in nutrients for the cells to recover after the shock of transformation. We varied the media compositions by using either horse or porcine serum in combination with fresh or commercial yeast extract. We used MBB

as base, since PPLO and TSB interfered with plasmid recovery in the nuclease assay. We determined that ureaplasmas were growing fastest in 10B media with porcine serum and fresh yeast extract. A major difference was seen when porcine serum was used instead of horse serum. Ureaplasma was growing faster in porcine serum. In media with fresh yeast extract the cells were growing slightly faster than in media with commercial yeast extract. Due to the laborious and long process of preparing fresh yeast extract and the minimal effect on growth, we decided to use 10B media prepared with MBB, porcine serum, commercial yeast extract and Isovitalex.

5.4.4.2 PEG toxicity

To test for PEG toxicity ureaplasmas were pelleted and washed wit PBS, then re-suspended in 1ml of PEG 8000 dissolved in PBS in concentrations varying from 5%-60% in 5% increments. The cells were incubated with PEG for 1 minute at room temperature then spun down and washed in PBS. CCU assays were performed to determine the viability of the cells after the treatment. Under these conditions 35% PEG was tolerated best. Cells that were treated with PBS only or 35% PEG had 10^8 CCUs; 5%, 15%, 25%, 30% had 10^7 CCU; 10%, 20%, 40%, 45%, 55%, 70% had 10^6 CCUs; 50% had 10^5 CCUs and 65% had 10^4 CCUs. Next, we tested the PEG toxicity by performing the transformation protocol with different PEG concentrations and no donor DNA (substituted by PBS). Under these conditions 5%-30% PEG concentration were the least toxic to the cells. 35%-40% PEG showed a reduction of viability by about 10^2 CCUs. Poor cell viability was observed in samples treated with 45%-60% PEG, growing only to 10^2 CCU, whereas control grew to 10^7 CCUs.

5.4.4.3 Optimal PEG Concentration

Previous experiments have shown that mycoplasma transformation efficiency is dependent on PEG concentration. The optimal PEG concentration varies among different mycoplasma species. Accordingly, we experimented to find the optimal PEG concentration for UPA3. With the ambition of achieving permanent and more efficient transformation, we tested PEG 8,000 re-suspended in 1X PBS buffer with varying concentrations. We used the pTF20 plasmid with Tn4001 transposon, following the transformation procedures previously described. As a control of the involvement of PEG in the transformation process, we substituted the PEG solution with 1X PBS buffer. All samples were tested in 10B growth medium containing antibiotic. We tested the 0% PEG 8,000 control in both 10B growth medium with antibiotic and in 10B growth medium without antibiotic. The antibiotic control tested whether the PEG aided the insertion of DNA into the cell and potentially into the genome. The no antibiotic control tested whether the cells being used were viable.

After overnight growth, only the samples subjected to 50% PEG solution grew up to the 10^{-2} dilution factor in the presence of tetracycline. The other samples only grew in the 10^{-1} dilution. In the absence of PEG during the transformation procedure no growth was observed in the presence of antibiotic. Thus, subsequent experiments were performed with 50% PEG 8,000. Neither of the transformation samples could be maintained in tetracycline growth media, suggesting transient transformation with the suicide plasmid, due to failed integration of the transposon into the genomic DNA.

5.4.4.4 Ureaplasma Culture pH Optimum before Transformation

We evaluated the best ureaplasma culture growth stage for transformation. We inoculated 50ml media with 200 µl of frozen stock (1000X concentrated overnight culture; pH 7.5-8.0 at harvest time). The inoculated media (1X) was serially diluted to produce 1/2X, 1/4X, and 1/8X concentration of the cells, thus reaching different pHs and culture states at the time of harvest. The pH of the different cultures was as follows: 1X=pH 8.00, 1/2X= pH 7.74, 1/4X=7.54, and 1/8X=7.29. CCUs of these cultures at the time of harvest were: $1X=10^5$, $1/2X= 10^6$, $1/4X=10^8$, and $1/8X=10^8$. Transient tetracycline resistance was observed in the samples with pH 7.74 and 7.54 at the time of harvest. Cultures of these cells could not be maintained in the presence of tetracycline in the growth media. The cells from these samples were harvested and PCR for the presence of the tn5 transposon was performed with positive results for the samples transformed with the tn5 transposomes and negative results for the frozen stock used and the no transposome control sample (Figure 5-4). One explanation is that the transposome is in the ureaplasma cell but did not integrate into the gDNA, thus conferring only a temporary resistance to tetracycline. Alternatively, this may be a false positive PCR result caused by detection of leftover tn5 transposon DNA from the transformation reaction, which did not wash away effectively. If the tn5 transposon integrated into the ureaplasma genomic DNA then an inverse PCR, followed by sequencing should be able to show the tn5 location in the ureaplasma genomic DNA (experiments in progress). The UUR9 *tuf* promoter was amplified using PCR and assembled into the tn5 transposon. Potential errors in the amplicon or primers may be resulting in weak transcription of the *tetM* gene.

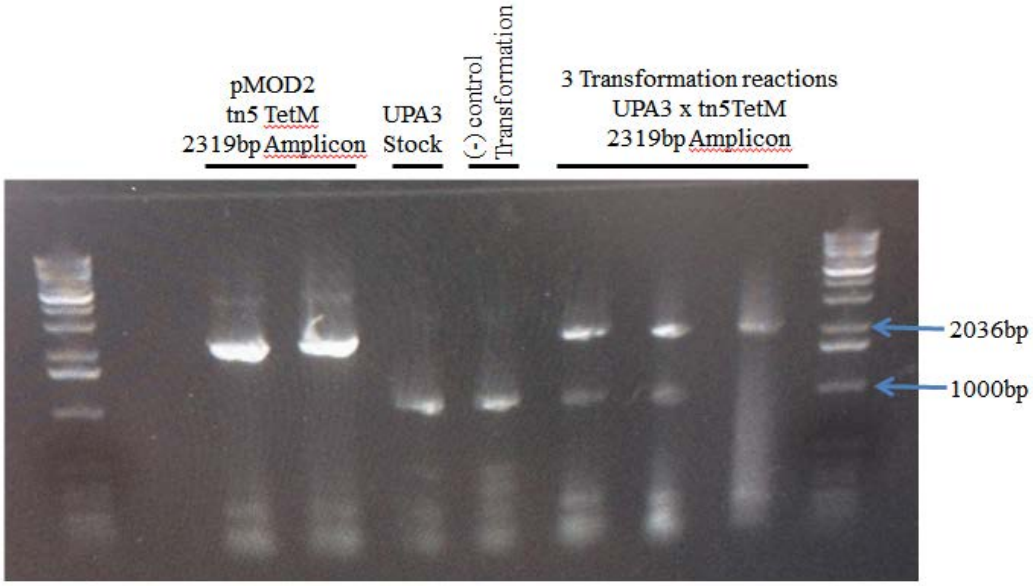


Figure 5-4. PCR for the presence of the Tn5_UUR9*tuf*_TetM after PEG transformation reaction.

The PCRFP and PCRRP primers were used to check for the presence of the custom Tn5_UUR9tuf_TetM transposon in samples that showed transient tetracycline resistance. 1KB DNA ladder by Invitrogen was used on both sides of the gel. Blue arrows point to the 1000 bp and 2036 bp bands in the DNA ladder. The expected amplicon size of the transposon is 2319 bp. As a positive control two samples of the custom pMOD-2 plasmid containing the transposon in question were used with the PCRFP and PCRRP primers. As negative controls a sample of the frozen UPA3 stock used to start the UPA3 culture for transformation was used. The second negative control was the transformation reaction in which no transposomes were added. In both negative control samples an amplicon of about 1000 bp is evident, probably due to the primers amplifying an irrelevant gDNA section in the UPA3. All 3 UPA3 samples transformed with the transposomes show a positive amplicon band of the appropriate size.

5.4.4.5 Transformation results:

Attempts to transform UUR9 and UPA3 with linear Tn5 transposomes containing puromycin resistance gene were unsuccessful. We next attempted to transform using a tn5 containing *tetM* and the mycoplasma *tuf* promoter. To verify that the template plasmid of the tn5 was eliminated and was not a competitor target for the tn5, the linear Tn5TetM PCR product was electroporated into *E. coli* and grown on 100 µg/ml ampicillin plates. No colonies were observed, indicating that there was no plasmid template. The Tn5TetM DNA concentration was 180 ng/µl. A sample of it was diluted to 100 ng/µl and samples of both concentrations were used to prepare transposomes. Transformation of UPA3 cells with these transposomes was unsuccessful. We hypothesized that the reason the transformation failed was due to the use of the mycoplasma *tuf* promoter to operate the *tetM*. Therefore, we are constucting plasmids with the UUR9 *tuf* promoter or the UUR9 promoter in front of the endogenous *tetM* driving the tetM expression in the Tn5TetM transposon. Multiple transformation experiments yielded only transient antibiotic resistance in 10B liquid culture and on 10B agar plates. We have extracted gDNA from one of the samples with transient antibiotic resistance and performed PCR analysis for the presence of the tn5 transposon and *tetM.* Although the PCR was positive we are in the process of determining if the transposon is inserted in the gDNA. We are also constructing a tn5 with UUR9 *tetM* and its upstream promoter as well as a tn5 with UUR9 *tuf* promoter and a kanamycin resistance gene (*aphA*). Alternatively, we will attempt to further modify the transformation protocol to allow a shorter time of exposure of the cells to PEG, to reduce potential toxicity by adding 5ml PBS after the

50% PEG incubation instead of 10B media without serum, and pelleting the cells right away, then recover in 10B media with porcine serum without antibiotic overnight, then select successful transformants in 10B media with antibiotic. The reason for this modification is that when using a transposon that integrates into the gDNA of the recipient, the recipient is not likely to lose the transposon. Furthermore, since the antibiotic gene is under the control of a constitutive promoter, the cells would have enough time to recover in more favorable conditions and produce the resistance protein to be able to survive once transferred to the selective media.

5.4.5 Isolation and analysis of intact genomic DNA

The ultimate goal of the project is to achieve targeted mutations of the ureaplasma genomes. For that goal the ureaplasma genome of interest will be cloned and modified in yeast, and later transplanted into a recipient cell. In order to attempt genome transplantation among ureaplasmas we isolated intact ureaplasma genomic DNA in agarose plugs. Briefly, the cells are suspended in 1% agarose and casted into a mold to produce solid agarose plug, which are treated chemically and enzymatically to remove cell debris and proteins and leave intact, naked, circular, genomic DNA. We determined the proper concentration of cells per agarose plug by analyzing the quality and quantity of genomic DNA in plugs of different cell concentrations. We determined that 20 ml of an overnight culture grown to pH 7.5-8.0 or about 10^7 CCU was optimal. To determine DNA quality we used 3 restriction enzymes. SdaI and SmaI contain only 1 restriction site in the UUR9 genome and ApaI contains two restriction sites. Good genomic DNA quality is considered the highest quantity of

intact, circular, genomic DNA. Intact, circular, genomic DNA has a very faint, if any, band of linear genomic DNA or any other bands. However, when the DNA in the agarose plug is digested with single-cutting restriction enzyme, a single, thick band of linear, full genome size DNA band should appear on the analytical gel.

5.5 Conclusions

The dearth of genetic tools to work with ureaplasmas plays a major role in the lack of understanding of ureaplasmas as an sexually transmitted disease agents and therefore its pathogenicity. Although we are yet to successfully transform and maintain ureaplasma mutants, we have greatly advanced our understanding about the conditions needed to transform ureaplasmas. We suspect that the Tn5 transposons containing the UUR9 endogenous promoter and tetM will be succesful. The use of transposon mutagenesis to create a library of mutants would generate information about the genes present in ureaplasmas and would enable investigation of ureaplasma pathogenicity.

Recent and previous studies have identified potential pathogenicity factors in *Ureaplasma* by sequencing and annotation of the ATCC type strains for each of the 14 serovars of *Ureaplasma*, observing enzymatic activities [154, 155, 171, 172, 188], and comparative genome microarray study of *Ureaplasma* isolates [134]. Such putative pathogenicity factors will remain speculations unless their involvement in pathogenicity is tested by knock-out and subsequent complementation studies. Molecular tools allowing the creation of mutants and their subsequent complementation will eventually lead to the understanding of the factors leading to

differential pathogenicity of *Ureaplasma* strains and to the development of better screening methods, drug targets and vaccines to mitigate the adverse outcomes of ureaplasmal infections.

Our observation of transient antibiotic resistance after transformation with a transposon indicates that UPA3 has the capability to read and express the t*etM* gene on the pTF20 and tn5 constructs. Because the pTF20 does not have an origin of replication compatible with the replication machinery of ureaplasmas and the tn5 is only a linear piece of DNA, if the transposon does not integrate into the genome, the daughter cells will not obtain a copy of the plasmid or linear tn5 transposon, causing death. Therefore, even though we saw an initial resistance to tetracycline in our cultures, the cultures were incapable of sustained growth. We suspect that the transposase gene on the pTF20 was not expressed. Because the Tn4001 transposon was ineffective, we chose to use the transposon Tn5 in a pre-made complex with the Tn5 transposase protein as an alternative, to ensure that the transposase is available for integration of the transposon into the genome. This relieved the cell of some duties and severed reliance on host cell transcriptional and translational machinery.

We worked with five different Tn5 transposon designs with varying promoters and antibiotic resistance genes. As stated above, we have the Tn5 with puromycin and mycoplasma *tuf* promoter, Tn5 with *TetM* and mycoplasma *tuf* promoter , Tn5 containing the UPA3 *tuf* promoter and *tetM* from the pTF20, and Tn5 containing the UUR9 endogenous *tetM* with either the upstream *tetM* promoter or the UUR9 *tuf* promoter. The ineffectiveness of the transformations with Tn5 containing mycoplasma *tuf* promoter may potentially be due to the non-ureaplasma promoter or

differences in codon usage among different bacteria. In pTF20, the *tetM* gene promoter was cloned from *Mycoplasma pulmonis*, and the *tetM* gene was originally from *Enterococcus faecalis* [257]. Although genetically similar to mycoplasmas, ureaplasmas might not be recognizing mycoplasma promoters. Therefore, we designed Tn5 with the endogenous UUR9 *tetM* because we think it may have a higher expression level in ureaplasmas due to the differences in the sequences of the two homologous genes. If needed, we will also create a ureaplasma codon-optimized *tetM* gene.

Alternatively, ureaplasma OriC plasmids are being constructed and tested as well. It has been show in mycoplasmas that after multiple passages, OriC plasmids may integrate into the genome of the host cell [258] or be lost. The capacity of the plasmid to be replicated and segregated may provide enough time for a transposon to integrate into ureaplasma genomic DNA or for the entire plasmid to integrate. I am designing primers for the putative origin replication of several serovars and will clone OriC PCR amplicons to make ureaplasma specific OriC plasmids.

Alternative Transformation Methods

Electroporation will be explored as an alternative method to transform ureaplasma cells. A starting point for electroporation experiments will be the current protocols in use for mycoplasma electroporation in our lab. Although electroporation can be a convenient way to transform ureaplasma, electroporation has not worked to transplant whole genomes. The reason this is not a suitable genome transplantation

method may be the harsh conditions and procedures during the experiment which may fragment the genome, therefore making it not viable.

Although conjugation has been reported [259-261], its mechanism is not well understood and it has worked only in limited number of mycoplasmas. The presence of the conjugal transposon Tn916 in *U. urealyticum* serovar 9, and some genes that are known to facilitate conjugation and integration of conjugal DNA material into the host chromosome (manuscript in preparation) could mean that Ureaplasma may be capable to acquire DNA through conjugation. A possible explanation for the presence of the Tn916 transposon of streptococcal origin [262] in the genome of *U. urealyticum* serovar 9 could be that it was acquired through conjugation from another urogenital bacteria containing this transposon [262], it may have been maintained on a plasmid for a short period of time, based on the isolated report of a *tetM* containing plasmid in a ureaplasma isolate [263] and it was integrated into its genome. Attempts to mate Ureaplasma with an enterococcal donor of Tn916 have failed to produce transconjugants (Dybvig, K., unpublished), however only some ureaplasma serovars have genes that may be involved in conjugation [237], and this information was not available at the time of these experiments. Furthermore, a different donor for mating and conjugal transfer may be required for successful transconjugation.

Viral-based vectors have been used in *Spiroplasma citri* for transformation, and have been shown to achieve higher transformations frequencies [264]. Although some viruses have been identified that infect mycoplasmas, they have not been used extensively as tools for genome modification [265-272]. Our comparative genome analysis of the fourteen ureaplasma serovars shows that only nine serovars contain an

insertion that seems to have occurred due to a chromosomal integration of a phage [237]. This insertion fragment has been shown to be present in some clinical isolates of the same serovar and absent in other clinical isolates of the same serovar. This phage fragment contains genes of unknown function with transmembrane domains, and lipoprotein attachment sites, because of which it is hypothesized that this may be a potential pathogenicity island [134] (manuscript in preparation). Some mycoplasma phages have been shown to be involved in the virulence and pathogenicity of the infected mycoplasma strain [273]. No phages have been isolated yetfrom ureaplasma cultures, the phage present in the 9 of the 14 serovars' genomes might have been trapped in the host chromosome, not be able to excise, and its genome may have been changed by its host.

Other reagents that could be tried for ureaplasma transformation are: Lipofectin, DMRIE-C reagent, and Lipofectamine 2000 by Invitrogen, and FuGENE 6, and DOSPER by Roche.

5.6 Future Directions

5.6.1 Adapt synthetic biology tools to create desired Ureaplasma mutants

After the first reported transformation in 1987 of *Acholeplasma laidlawii* and *Mycoplasma pulmonis* with Tn916 [274], there have been numerous studies to develop tools for genome manipulation [270, 275-281]. Even though tools like transposon mutagenesis and OriC plasmids have been developed for genetic manipulation of some mollicutes, their use for study of virulence factors in these

mollicutes is often difficult, laborious, time consuming and sometimes with limited success [282, 283]. Furthermore, many passages of the transformants are required before chromosome integration [282, 284, 285]. Some mycoplasma species have proven completely recalcitrant to these methods. Although the method developed by Allam *et al*., to enhance targeted homologous recombination in *M. mycoides* [282] holds a great promise as a tool in other mollicutes, the synthetic biology-based approach of cloning genomes of organisms with such limited or no toolset in yeast [286] can open the door to the use of the powerful genetic tools of yeast (including new tools developed at JCVI for manipulation of bacterial genomes cloned as yeast centromeric plasmids [287]) to produce desired mutants. This approach has been used to create a *M. mycoides* strain that lacks its type III restriction gene, which would be impossible by using the previously existing genetic tools for this organism [249]. Developing a protocol to clone ureaplasmal genomes in yeast for genetic manipulation, and transplanting them into a host cell will contribute significantly to the ability of scientists to conduct research that may not be possible by using traditional tools. Examples include the creation of strains with multiple gene deletions, which may be necessary for completely abolishing virulence and better understanding of host-pathogen interactions. Also an attenuated ureaplasma strain could be useful for development of vaccines. Vaccine strains could also be engineered to contain immunogenic proteins of other species to make multivalent vaccines.

5.6.2 Transplantation of UUR9 into UPA or *U. diversum* recipient

We will use UUR9 as donor, because it contains a naturally integrated Tn916 transposon with the *tetM* gene, which will serve as our marker to select transformants. If we achieve transformation, we can attempt to use a different donor serovar by transforming it with a marker of choice (like a puromycin resistance cassette, *lacZ*, or green fluorescent protein (GFP)). In the case of a resistance marker use, we can express the marker at different levels through the use of different promoters as mentioned earlier. The recipient cell needs to be different enough to allow discriminating between the donor and the recipient genome, as well as to minimize the opportunity for homologous recombination. The recipient cell still needs to be similar enough to allow for the recognition and booting up of the donor genome. Therefore, we will use as recipient the farthest phylogenetically removed UPA1. One of the potential problems during transplantation is recombination between the donor and recipient genomes, resulting in a chimeric genome molecule. Such recombination was observed in attempts to transplant *M. genitalium* strain G37 to *M. genitalium* strain 1019. Although we observed recombination among serovars in our serotyping study, chimeras containing markers from a UPA and UUR were less frequent (8%) than intra-species recombination (92%). Furthermore, recombination among UPA serovars appears to be less frequent (13%) than recombination among UUR serovars (79%); see Figure 5-5. This suggests lower recombination capacity of UPA serovars, thus we should observe a lower recombination frequency, if any, when UPA is used as a recipient cell of an UUR genome then if the opposite approach is attempted.

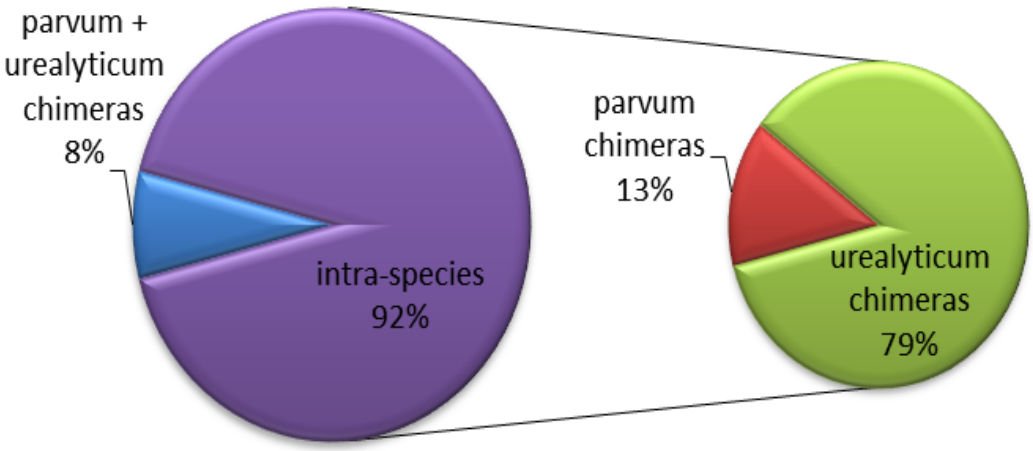


Figure 5-5. Inter- and Intra-species Distribution of 75 Chimeric Ureaplasma Isolates

The majority of recombination is among strains of the same species. Only 8% of the chimeric strains contained markers from both species. Furthermore, of the intra-species chimeras only 13% were chimeras resulting from recombination among UPA strains.

Another option for a recipient cell is the UPA3. Its genome sequence is complete and it appears to potentially have only type I and type III RMS. However, the restriction subunit of the type I and both subunits of the type III RMS contain unverified frameshifts. This may mean that UPA3 does not have a functional RMS, which will have to be tested by our restriction assay. Should this be the case, UPA3 will present a better choice for a recipient cell. Alternatively, we can use *Ureaplasma diversum* as a recipient, in which case we will sequence its genome. Since this ureaplasma is a bovine pathogen, and is more divergent, it may prove to be a better recipient for the transplantation experiments. Due to its greater divergence we expect it to facilitate the differentiation between the donor and recipient genome and minimize the possibility of homologous recombination resulting in the insertion of only the antibiotic resistance marker into the recipient genome. Furthermore, *U. diversum* is a recognized cause of bovine urogenital infections, including granular vulvitis, endometritis, salpingitis, spontaneous abortion, and infertility [88, 201] and may also be involved in bovine respiratory infections [201]. Therefore, sequencing, cloning to engineer, and transplanting *U. diversum* could provide an animal model system for studying ureaplasma induced diseases through observing phenotypic changes of a mutant strain *in vivo.*

For the transplantation procedure we will isolate the donor genome by preparing plugs using CHEF Mammalian Genomic DNA Plug Kit from Bio-Rad [249, 254]. Treating the donor cells briefly with chloramphenicol, prior to cell collection may increase the quantity of whole genomes by synchronizing ongoing rounds of replication and inhibit further rounds [249, 288-290] and compact the

nucleoids [249, 291, 292]. The transplantation protocol will follow the transformation protocol that we will develop for *Ureaplasma*. Since serovar 9 already has a *tetM* marker in its genome, we may attempt transplantation at the same time as transformation, in which case we will follow and adjust the transformation/transplantation protocol used previously described for *M. mycoides* [249, 254]. Since at this step we will be transplanting a natural genome isolated from ureaplasma cells, methylation may not be necessary. Should transplantation be unsuccessful, we can use the crude cellular extracts restriction assay to determine whether restriction enzymes present in the recipient cell are the reason for unsuccessful transplantation. If this is the case we can use the crude cellular extract methylation assay to attempt to protect the donor genome before transplantation.

Ureaplasmas can be visualized using a microscope on agar plates 24-48 hours after plating and incubation at 37°C. Their colonies are of approximately 15–60 μm in diameter. When grown on 10B agar, they form colorless colonies. One option would be to integrate the *lacZ* gene or another fluorescent marker into the genome. We could use an *E. coli* RecA expressing plasmid vector to enhance homologous recombination [282]. The positive transformants now have blue colonies on 10B agar, which should make them easier to visualize.

Transplantation usually requires large quantity of DNA, which may be problematic with ureaplasma cell cultures. Ureaplasmas create an alkaline shift in the growth media, due to their production of ammonia as a result of urea hydrolysis. This alkaline shift prevents the cells to grow to a high cell density, which results in lower DNA yield compared to other mycoplasmas and organisms. One way to overcome

this problem is to pool the DNA from multiple cultures. An alternative is also to use ureaplasma genomes isolated from yeast, which should easily allow higher DNA yield for use in transplantation experiments.

5.6.3 Clone the tetracycline resistant UUR9 genome in yeast

An overview of the process of creating bacterial strains from genomes that have been cloned and engineered in yeast is represented in Figure 5-6 [249]. In order to clone an ureaplasma genome in yeast the integration of a yeast centromere, an autonomously replicating origin sequence and a yeast selection marker for maintenance of the genome in yeast is required. These sequences can be integrated into the ureaplasma genome using a plasmid. The plasmid will also contain a selectable marker like *tetM* (if not present naturally as in serovar 9) and optionally the *lacZ* gene or GFP, which will be used to select for transformants. The latter sequences will also be useful in subsequent genome transplantation from yeast. The advantage of integrating the yeast sequences into the ureaplasma genome prior to yeast cloning is that only viable cells that tolerate the yeast sequences are selected. Transformants will be screened to determine the exact integration sites in the genome by direct genomic sequencing using primers specific for the integrated DNA to sequence into the genome. Clones with insertions at suitable locations will be selected for cloning in yeast. These genomes will be transformed into yeast as described by Lartigue et al.[249]. To verify that the complete ureaplasma genome was cloned in yeast, we will use multiplex PCR with primer pairs that recognize sequences distributed around the genome and CHEF gel analysis.

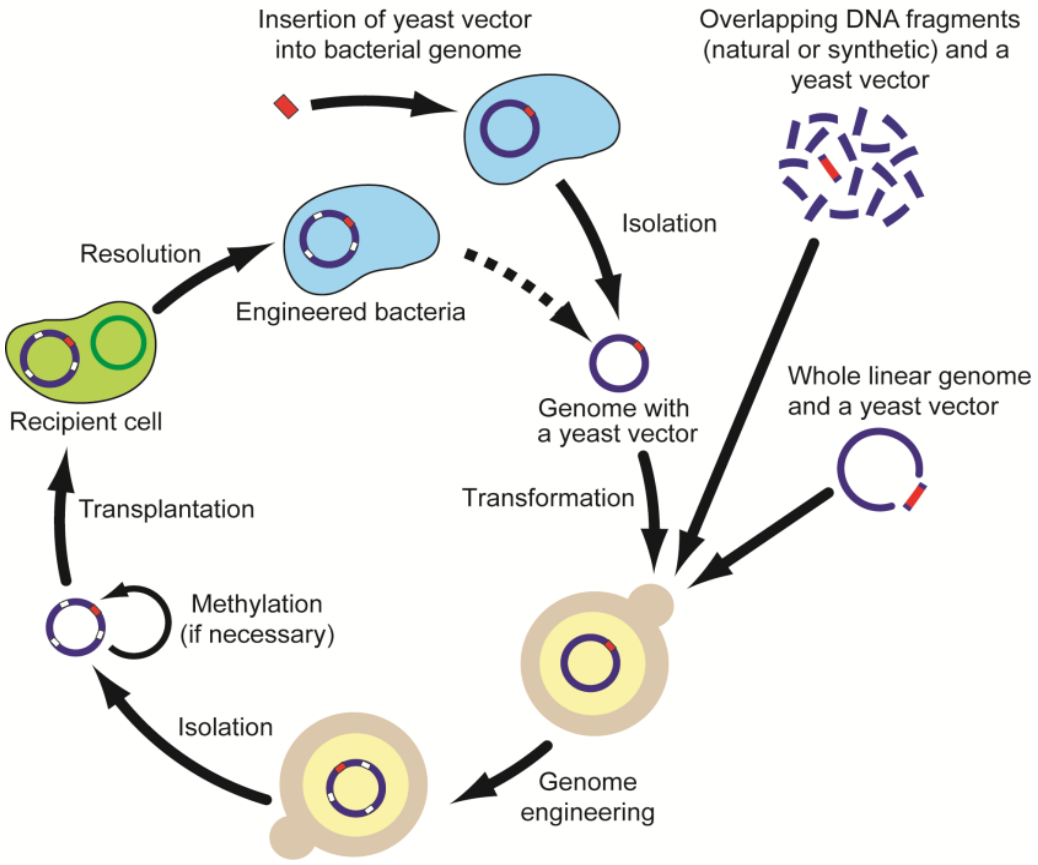


Figure 5-6. Moving a bacterial genome into yeast, engineering it, and transplanting it back into a bacterium.

The figure depicts 3 strategies to clone a bacterial genome in yeast: 1) The yeast centromere vector can be inserted in the bacterial DNA through bacterial transformation (preferred, as only viable insertions will be propagated to be cloned in yeast); 2) overlapping DNA fragments of the bacterial genome are co-transformed into yeast with the yeast centromere vector and the yeast recombination system assembles the bacterial genome as a YAC; 3) the intact bacterial genome is extracted and linearized; the linearized genome is co-transformed into yeast with the yeast centromere vector that has homologous sequences to the bacterial genome on both ends. Once the bacterial genome is in yeast, it can be engineered using yeast tools. The engineered genome is extracted and transplanted into suitable recipient cell to create engineered bacterial strains.
Reused with permission from [249].

In case we were unable to develop a traditional tool for the integration of the yeast maintenance sequences, or in order to clone the genome concurrently with the transformation experiments, we will use the alternative methods for cloning the genome into yeast (Figure 5-5). One such method, transformation associated recombination (TAR) cloning, was used to clone the synthetic and natural *M. genitalium* and *Mycoplasma mycoides* subsp. *capri* (Mmc) genomes into yeast [255, 286, 293, 294]. To use this method we will cut the Ureaplasma genome into several fragments. The fragments can then be transformed into yeast with a yeast centromere plasmid (YCp) that contains homology arms to bridge two of the fragments and other DNA pieces that contain overlapping ends to bridge the remaining fragments. Alternatively, a whole linear genome can be transformed into yeast with an YCp that contains arms homologous to either side of the linear genome. Transformants will be tested as described earlier for clones containing an intact genome. By using these methods there is a possibility that a non-viable genome is created. Should this turn out to be the case, it is possible to use the genetic tools of yeast to change the location of the yeast maintenance sequences. Cloning the ureaplasma genome in yeast gives us an additional opportunity to clone 100-kb genomic fragments as was done to clone the natural 100-kb fragments of Mmc [255]. Using these fragments, we can generate more easily mutant Ureaplasma strains by modifying the fragments and then assembling the genome by combining natural and modified fragments for co-transformation in yeast.

We do not anticipate difficulties in cloning the ureaplasma genome in yeast. Because ureaplasmas use an alternative genetic code (opal stop codon = tryptophan) it

is not likely that ureaplasma genes may cause toxicity to the yeast cell. Furthermore, our experience of cloning genomes into yeast using the standard genetic code, shows that yeast can leave out fragments containing toxic genes. This leads to relatively easy identification of problematic genes, which could be mutated to enable cloning of the full genome. To date in addition to the four mycoplasma species whose genomes we have already reported being cloned in yeast [286], we have now cloned three other bacterial genomes as YACs. This suggests this step will not be difficult.

5.6.4 Delete a gene of interest and transplant the deletion mutant to a recipient cell

Once we have successfully cloned the ureaplasma genome in yeast, we will use the method developed for synthetic biology to delete a gene of interest. Figure 5-6 illustrates the general strategy for knockout construction using the Tandem Repeat Coupled with Endonuclease Cleavage (TREC) method [287]. The first step is to generate a core cassette that contains the endonuclease SceI under the control of a galactose responsive promoter, a *URA3* gene for selection, and a SceI recognition site (asterisk). On one side of these sequences is a 50 bp of DNA identical to DNA adjacent to the 5′ end of the target gene and on the other side is a stretch of DNA homologous to DNA on the 5′ end of the target gene (red line). This will result in a tandem repeat (TR) upon integration. The core cassette is then transformed into yeast containing the ureaplasma genome. After growth on selective medium, colonies are screened for positives that contain the replacement of the core cassette at the target gene locus. The positive clone is then grown on galactose medium to induce

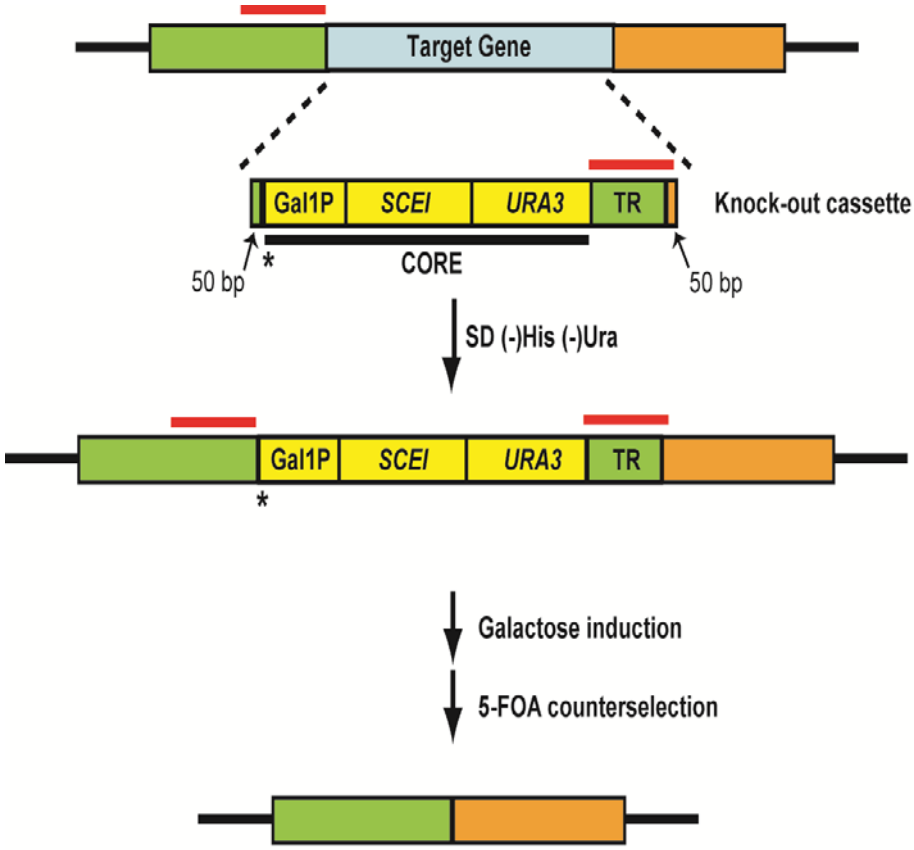


Figure 5-7. TREC method for seamless deletions of target genes in genomes cloned and maintained in yeast

Through homologous recombination, target gene is replaced with a mutagenesis cassette that consists of a knock-out CORE (an18-bp I-SceI recognition site (*), the SCEI gene under the control of a GAL1 promoter, and the URA3 marker) and a DNA fragment (shown as TR in green and red line) identical to a region upstream of the target site. The replacement generates tandem repeat sequences flanking the CORE (red lines). Galactose induces the expression of SceI, which generates a double-strand break (DSB) at the SceI site near the target locus. The DSB promotes an intra-molecular homologous recombination between the repeat sequences, leading to an excision of the CORE. Reused with permission from [249].

expression of SceI which cleaves its recognition site resulting in loss of the core cassette by recombination between the DNA adjacent to the 5′ end of the target gene and the tandem repeat. The cells are plated on medium containing 5-fluoroorotic acid (5-FOA) to select only those cells that have lost the core cassette resulting in the seamless deletion of the target gene. This method was used to remove a *M. mycoides* restriction enzyme gene in an earlier work [249]. The TREC method is needed to manipulate bacterial genomes in yeast because the traditional yeast genetic method of replacement of a gene with only a *URA3* marker and subsequent counter selection with 5-FOA has led to instability of the bacterial genome resulting in unwanted and unpredictable deletions of genomic areas [287]. Once the changes are made to the ureaplasma genome in yeast, the next step is to transplant the altered genomes into the recipient cells to generate the mutant ureaplasma strain. These strains can then be tested for loss of function or gain of function, depending on the nature of the deleted or inserted gene.

5.6.5 Suggested modifications to the ureaplasma genome

As mentioned earlier new experimental ureaplasma strains can be created by the deletion or insertion of genes. We are considering several initial modifications of the ureaplasma genome. One of them is introducing GFP, which can act as an additional marker to aid in the screening for transplants. A fluorescent ureaplasma strain can be useful in following the dissemination of the colonizing strain in tissues. When choosing genes to delete as a proof of concept, one should choose a target that is unlikely to be essential for the survival of the organism. Restriction enzymes are

potential target genes for deletion. RMS are rarely essential for survival in controlled laboratory conditions. Furthermore, the creation of an ureaplasma strain lacking RMS can ease the transplantation process by permanently removing a barrier. Another potential target for deletion is the major surface antigen of ureaplasma – the multiple banded antigen (MBA). Based on our comparative genome analysis of the 14 ATCC serovars we now know that the MBA is part of a large phase variable family. Most serovars contain multiple variable, surface exposed domains which are expressed one at a time by recombining the promoter and the 5` area of the gene (the conserved domain) with a variable domain [237]. One strategy to make a strain that does not express MBA is to delete the currently active *mba* gene (promoter, conserved domain, and variable domain), which will leave the rest of the variable domains silent. Another strategy is to delete the full MBA locus which may be up to 10 Kb long and may include genes that are expressed independently from the *mba* promoter and/or are not part of the MBA phase variable system. Because monoclonal antibodies recognizing different variable domains have been produced, we can confirm the deletion of the active *mba* gene and the silencing of the rest of the variable domains or their deletion by immunoblotting. A strain lacking MBA can be used in a future study to be compared to the wild-type in an animal model system for difference in immunogenicity and development of disease.

Acknowledgements

This work was funded by NIAID project 5R21AI098057. I am very grateful for the experimental support provided by Olivia Chao, our very bright and smart, senior high school intern and mentorship by John Glass, who is the project PI and my dissertation advisor. Many thanks also to Nacyra Assad-Garcia, Sanjay Vashee, Michael Montague, Michael Algire, Vladimir Noskov, Ray-Yuan Chuang for support with troubleshooting, mentoring, and experimental advice. I am thankful also to the rest of the JCVI Synthetic Biology team for useful discussions of the experiments and data.

Chapter 6. Selecting against Recombination Events during Genome Transplantation by Crosslinking the Genome of the Recipient Cell with Mitomycin C

6.1 Introduction

6.1.1 Genome transplantation

In 2007 the first genome transplantation was reported [254]. During genome transplantation a naked bacterial chromosome is installed into a suitable recipient cell, so that the recipient cell is reprogrammed to have only the attributes encoded by the new genome [254, 295]. This process was used to install a chemically synthesized genome version of *Mycoplasma mycoides* into *Mycoplasma capricolum* recipient cells in 2010 [255]. So long as genome transplantation is only possible between different members of the mycoides group of mycoplasmas, much of the potential of the synthetic cell will be unrealized. For instance, the DNA foundry SGI-DNA is advertising that they can synthesize whole bacterial genomes for customers, but in the absence of generalized genome transplantation, there is no use for such synthetic genomes. If we can develop genome transplantation technology, scientists will be able to design, build and test bacteria with synthetic genomes to solve human problems, such as bio-renewable fuels, new pharmaceuticals, and new vaccines. Furthermore, genome transplantation coupled with the modification of bacterial genomes in yeast [286, 287, 294] can expand our understanding of some pathogens for which there are limited or no molecular tools. *Mycobacterium leprae* is one of

many totally intractable pathogens, for which we have no genetic tools. Similarly, there are genetically intractable bacteria with unrealized pharmaceutical or industrial potentials because scientists cannot remodel their genomes.

To realize the potential of this new technology we need to expand it to other pairs of organisms. The first reported genome transplantation used a genome donor and recipient pair of closely related bacteria of the *mycoides* group of mycoplasmas. Carole Lartigue has been able to transplant the genomes of 2 different mycoplasmas within the mycoides group and one mycoplasma species outside of the group, from the next closest related clade into *M. capricolum* recipient cells (manuscript in preparation). However, extending genome transplantation to donor-recipient organism pairs outside of the mycoides group has proven to be a difficult task. During our efforts to develop genome transplantation much of our research was aimed at transplanting genomes from *Mycoplasma genitalium* into either a very closely related species called *Mycoplasma pneumonia* or into different strains of *M. genitalium*. All of these efforts were unsuccessful. Often we obtained false positive transplants as a result of recombination events between the donor and recipient genomes, resulting in the acquisition of selective marker genes by the recipient's genome and thus false positive results. This may potentially hamper the transplantation of ureaplasma genomes as well. We have already reported an extensive horizontal gene transfer among ureaplasmas and even though mosaic strains of UPA and UUR genes were rare, such were observed [104]. As we seek to apply the process to a large variety of organisms in order to realize its full benefits, we need a mechanism to at least

transiently disable the recombination machinery that is virtually ubiquitous in bacteria.

This problem can be solved by making the recipient genome unavailable for recombination. For this purpose we are exploring the use of the DNA alkylating agent Mitomycin C (MMC) to covalently crosslink the recipient genome right before we introduce the donor genomes to the recipient cells. The covalent crosslinking of the recipient genomes will eventually kill the recipient cells. We believe that the cell transcription and translation machinery are still intact shortly after the MMC treatment. Therefore, if a “healthy” intact donor genome is supplied to the still intact recipient cell, its machinery can now read and copy the “healthy” genome allowing the cell to survive.

6.1.2 Mitomycin C

MMC is an antibiotic isolated from *Streptomyces caespitosus* in 1958 [296]. It is an alkylating agent used in antitumor therapy [297, 298]. Three potentially active groups are recognized in MMC: a quinone, a urethane and an aziridine group. MMC requires activation by reduction by host enzymes (discussed later) of the quinone and subsequent loss of the methoxy group[299]. MMC can form different DNA adducts (Figure 6-1; discussed in details elsewhere [300, 301]), however its capacity to crosslink the complementary strands of the DNA molecule is most lethal: a single crosslink per genome can cause the death of a bacterial cell [301].

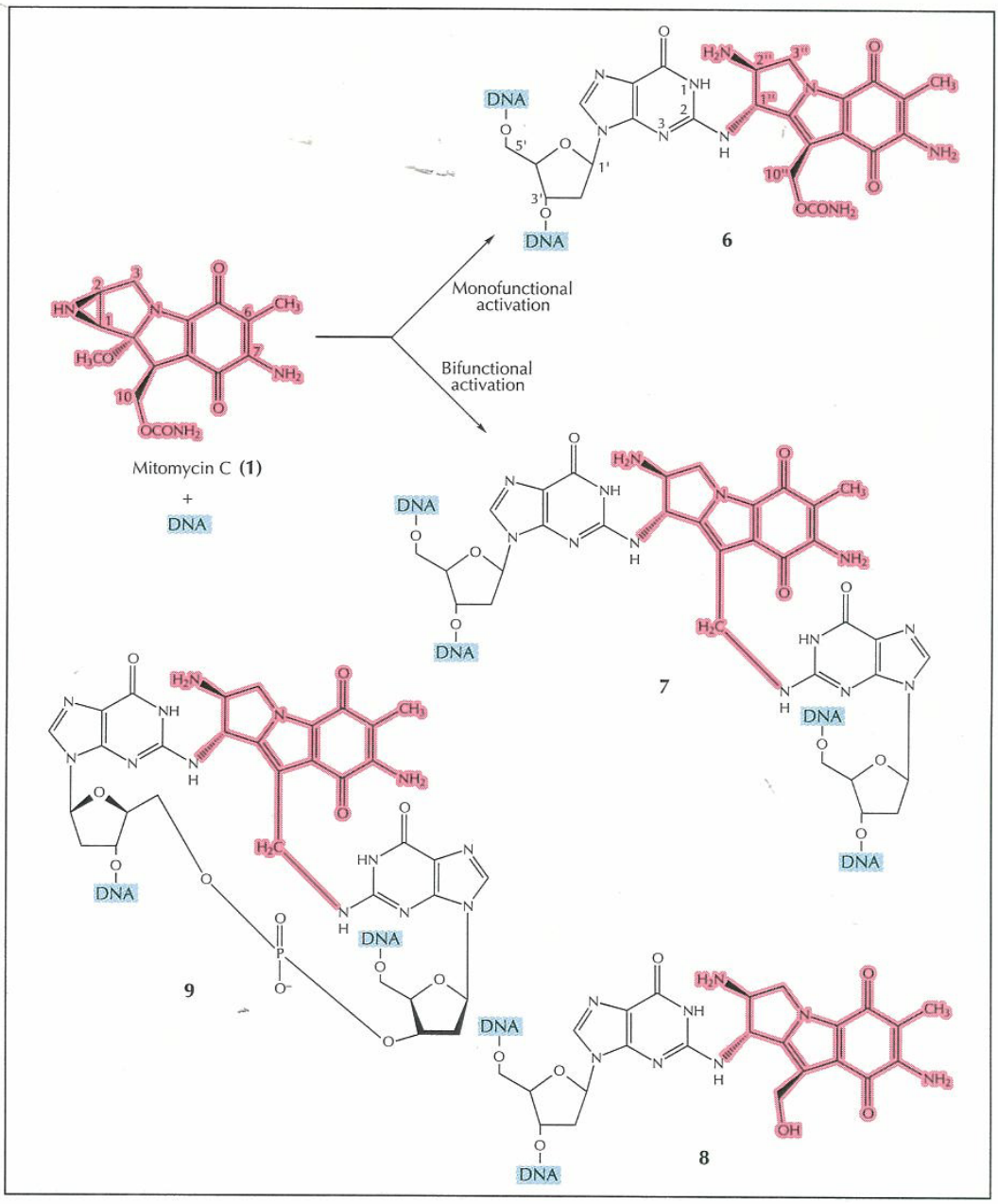


Figure 6-1. Adducts of MMC and DNA formed by reductive activation.

Activation of only one of the alkylating functions gives monoadduct (6), whereas activation of both alkylating functions can give products (7-9). MMC is highlighted in pink. Reused with permission from [301].

MMC can have different effects on the cell depending on concentration, genetic context and environmental conditions. Crosslinking of the host DNA can result in inhibition of DNA synthesis, degradation of DNA, nuclear lysis and formation of giant cells, and mosaicism [299, 302]. At very high concentrations, 30-100 times more that the concentration needed to inhibit DNA synthesis, it has been shown to inhibit RNA synthesis [303].

The action of MMC is relatively specific to double stranded DNA in a sequence dependent manner: 3`-CpG-5`. MMC was shown to have no effect on DNA polymerase isolated from HeLa cells [304]. However, MMC can react with RNA, and can bind to some extent to proteins [305]. DNA repair mechanisms are not inhibited by MMC, as the use of DNA repair inhibitors greatly enhances the sensitivity of the cells to MMC [306]. DNA damage is observed when MMC is administered at any time in the cell cycle; however, mitosis and DNA synthesis are significantly more affected with application during late G_1 and early S phase. G_2 phase is the least sensitive phase, during which the drug inhibits mitosis by reducing the RNA synthesis [299, 307].

The crosslinking efficiency is different for eukaryotic and prokaryotic cells. About a third of the DNA in human tissues shows crosslinking after treatment with large amounts of MMC [308]. In bacteria 10-20% of all MMC has been observed to be associated with the DNA [309] and as high as 100% of the molecules are altered [305].

MMC in its original form is not reactive with DNA. In order for the crosslinking to occur bio-reduction of MMC is required within the cell. At least 5 different enzymes have been described to activate MMC: 1) NADPH:cytochrome C (P-450) reductase; 2) NADH: cytochrome b_5 reductase; 3) Xantine oxidase; 4) Xantine dehydrogenase; 5) DT-diaphorase. The reaction of MMC with these enzymes under different pH conditions results in different reactive MMC species and thus different MMC:DNA adducts are observed. Detailed description of the different reactive MMC species and the conditions under which they occur can be found in the MMC reviews by Sartorelli *et al* from 1994 [310] and Cummings *et al* from 1998 [311].

6.2 Materials and Methods

6.2.1 MMC Solutions and Storage

MMC crystalline powder was dissolved in DMSO to make a stock solution of 1 mg/ml. The MMC stock solution was split in 1ml aliquots and stored at -20°C. Individual tubes were wrapped in aluminum foil to protect the solution from light.

6.2.2 Minimum Inhibitory Concentration

MIC assays were prepared in 96 well plates [248]. MMC was added to SP4 media to make the highest desired concentration for testing in the MIC assay. We started with 16μg/ml as the highest concentration and 360 μlof 16 μg/ml MMC SP4 were added to wells A12-H12. The rest of the wells of the 96 well plate were filled with 180 μl of SP4. Next we serially dilute the antibiotic 1:2 (take 180 μl from A12-

H12 and add in A11-H11, etc, stop at A2-H2, mix and discard 180µl). Wells A1-H1 were left as no antibiotic control. We added 20 µl of 1X overnight culture (about 10^{10} CCUs/ml) in A1-A12. Next we serially diluted 1:10 in each row (take 20 µl from A1-A12 and add to B1-B12 with a multichannel until all rows are done). The plates were incubated at 37°C until no further color change was observed (48-72 hours). The inhibitory concentration was determined by recording the concentration at which no color change was observed in the 10^{-1} dilution (first well of dilution).

6.2.3 Determining the MMC concentration and conditions to inactivate a grown culture

Based on the MICs we tested the MMC concentration needed to inactivate an overnight culture if MMC is added to the culture directly, as well as the optimal incubation time and temperature. We tested 6 concentrations (0 µg/ml – control, 0.25 µg/ml, 1 µg/ml, 2.5 µg/ml, 5 µg/ml, and 10 µg/ml), 3 time points (1, 2, and 4 hours of incubation) and 3 temperatures (4°C, 30°C and 37°C). An overnight culture was split into multiple samples to test all conditions. MMC was added directly to the culture samples to achieve the corresponding concentrations. Samples were incubates at the above temperatures and time periods. To determine the inhibitory effect of MMC in these conditions, CCU assays were performed on the above samples before (control without MMC) and after incubation (all samples) in triplicates in 96 well plates. The wells of the 96 well plates were filled with 180 µl of SP4 without MMC. Twenty µl of the corresponding sample were added to the first row (A1-A12) and 1:10 dilutions were made down the plate. Plates were incubated at 37°C until no further color

change was observed (48-72 hours) and CCU readings were performed to establish the most effective treatment for an overnight culture before transplantation.

6.2.4 Preparation of intact circular genomes in agarose plugs

The buffers, enzymes, and agarose in the following procedure are provided in the CHEF Mammalian Genomic DNA Plug Kit (BioRad catalog number 170-3591). To make 50 plugs with intact circular genomic DNA (gDNA) 1 ml mycoplasma culture (pH 6.5-7.0) per plug was harvested by centrifugation at 4575g for 15 min at 10°C and washed in 20 ml of TRIS-Sucrose Buffer [10 mM Tris (pH 6.5) plus 0.5 M sucrose] then spun as before. A cell suspension was prepared in 2.5 ml of TRIS-Sucrose Buffer. The 50 well plug mold makes 5 ml of agarose plugs. A 2% low melt agarose solution in sterile water was prepared and melted using a microwave. The cell suspension solution and the 2% low melt agarose solution were equilibrated to 50 °C in a water bath. Next the cell suspension was combined with an equal volume of 2% CleanCut agarose and mixed gently but thoroughly. This results in a final concentration of 1% agarose. The cell/agarose mixture was kept at 50 °C, and transferred to plug molds. The agarose was allowed to solidify. This step can be expedited by placing the molds at 4 °C for 10–15 minutes. This also adds strength to the agarose for removal from the mold. 25 ml of Proteinase K Reaction Buffer was added in a 50 ml conical centrifuge tube. The solidified agarose plugs were pushed into the 50 ml centrifuge tube containing the Proteinase K solution by using the snap off tool provided on the plug mold. The plugs were incubated overnight at 50 °C without agitation. Next the plugs were washed four times in 50 ml of wash buffer for

30 minutes to 1 hour each at room temperature with gentle agitation. 1 mM PMSF was added during the second or third wash to inactivate any residual Proteinase K activity. The plugs were stored at 4 °C for 3 months to 1 year. The recipes of the buffers that come with this kit can be found in the manufacturer's manual.

6.2.5 Digests of genomes treated with MMC

The plugs to be restricted were washed for 30 minutes in 0.1x wash buffer or TE. One plug per digest was placed in a sterile 1.5 ml microcentrifuge tube. The plug was incubated with 1 ml of the appropriate 1x restriction enzyme buffer for about 1 hour with gentle agitation at room temperature. The buffer was removed and 0.3 ml of fresh 1x restriction enzyme buffer was added. 30-50 U per 100 µl plug restriction enzyme, in this case 50U of PspXI to linearize the *M. mycoides* genome, was added and incubated overnight at the appropriate temperature. After overnight digestion, the buffer was removed and 1 ml of wash buffer was added. 1⁄4 of a plug per sample was loaded on an agarose gel per well on a 1% agarose gel along with a lambda DNA ladder. The gel was electrophoresed in the CHEF Gel III system with the following parameters: Buffer: 0.5x TBE; Temperature: 14°C; Switch Time: 50-90 seconds; Run Time: 22 hours; Angle: 120°; Voltage Gradient: 6 V/cm. The gel was stained with SYBR Gold (Molecular Probes) and scanned.

6.2.6 Genome Transplantation

Restriction system free *M. capricolum* cells were grown overnight at 30°C to pH 6.5-7.0 in SSG growth media (SOB growth media supplemented with

streptomycin and glucose). If MMC treatment was used, MMC was added to a final concentration of 2.5 μg/ml from a 1 mg/ml stock directly to the culture and the culture was incubated for 2 hours at 30°C. 3 ml of cell culture was used per transplantation reaction. A 100μl sample was obtained from the culture before processing (and before MMC) for CCU and CFU analysis. Remaining culture was centrifuged in bulk for 15 min at 4575 g at 10°C. The media was discarded and the pellet was re-suspended in 1 ml Tris-Sucrose Buffer [10 mM Tris (pH 6.5) plus 0.5 M sucrose], then additional 19 ml of Tris-Sucrose Buffer were added. Cells were pelleted for 15 min at 4575 g at 10°C. The Buffer was discarded and the remains were removed with a Kimwipe. The pellet was re-suspended in 200 μlof 0.1M $CaCl_2$ per transplantation sample and incubated on ice for 30-45 minutes. In a 10ml Falcon tube for each sample was added: 410μlof SP4 growth media without serum; 20μlof melted gDNA plug (using wide bored tips for genomic manipulations); 20 μg of Yeast tRNA (10 μl). After the 30 minutes incubation on ice was done, 200 μl of cells was added in the 10 ml tubes per sample, followed by 650 μl of 10% PEG 6000 dissolved in 2X Tris/$NaCl_2$/$MgCl_2$ and gently mixed well by horizontally inverting and rolling the tubes. The transplantation reactions were incubated for 90 minutes at 30°C, then 5ml of warm SP4 growth media were added to each sample and samples were centrifuged immediately for 15 min at 4575 g. The media was decanted and remaining media was removed by aspiration. Cells were re-suspended in 1.2 ml of warm SP4 media per sample and each sample was platted on two SP4 agar plates with 4μg/ml tetracycline (600 μl of transplantation reaction per plate). Plates were incubated at 37°C for 3-5 days and blue colonies were counted.

6.3 Results and Discussion

6.3.1 Susceptibility of different mycoplasmas to MMC

MMC is a powerful antibiotic. There are organisms that are highly sensitive to MMC and some that are very resistant. We tested the susceptibility to MMC of several mycoplasma species and two ureaplasma species available in our laboratory (Table 6-1). *Bacillus subtilis* growth is inhibited at 0.025 μg/ml, while most Gram positive and Gram negative bacteria are inhibited at 1μg/ml. Many viruses are also sensitive to MMC, however toxoplasma organisms, influenza A, Newcastle disease and Japanese encephalitis viruses are not resistant to MMC [299].

Organism	MMC MIC (μg/ml)
Mycoplasma mycoides subspecies *capri*	1 μg/ml
Mycoplasma capricolum subspecies *capricolum*	1 μg/ml
Mycoplasma pneumoniae	<0.5 μg/ml
Mycoplasma gallisepticum	1 μg/ml
Mycoplasma genitalium	<0.5 μg/ml
Mycoplasma penetrans	4 μg/ml
Mycoplasma imitans	0.063 μg/ml
Mycoplasma alligatoris	2 μg/ml
Mycoplasma crocodyli	<0.25 μg/ml
Ureaplasma urealyticum Serovar 9 ATCC 33175	>80 μg/ml
Ureaplasma parvum serovar 3 ATCC 27815	>80 μg/ml

Table 6-1. Susceptibility of Different Mycoplasmas and Ureaplasmas to MMC

We determined and summarized in this table the lowest concentration of MMC that inhibits the visible growth (color change in the CCU assay) of the microorganisms that were tested.

Several reasons for difference in susceptibility levels and resistance to MMC have been pointed out throughout the literature. The host may be deficient or express different levels of enzymes that can bio-reduce MMC, thus MMC remains non-reactive with DNA or is activated at low levels. None of the bio-reductive enzymes shown to activate MMC seem to be present in the annotated mycoplasma genome sequences, there may be other reductases present in the mycoplasmas capable of activating MMC, as all tested mycoplasma species were sensitive to MMC. The different reduced states of MMC are very unstable and can be re-oxidized to inactivate MMC. In the biosynthetic gene cluster of MMC from *Streptomyces lavendulae*, the *mcrA* gene encodes a MMC oxidase capable of re-oxidizing the activated MMC. This is one of 3 self-protection mechanisms of *Streptomyces caespitosus* [312]. The other two mechanisms are 1:1 binding of MMC to the Mrd resistance protein and an MMC translocase that pumps the antibiotic out of the cell [312]. Effective DNA repair mechanisms have also been suggested to increase resistance to MMC. *Micrococcus radiodurans*, which is extremely effective and accurate in repairing its DNA is four times more resistant to MMC than the *E. coli* B/r [313]. Furthermore, the use of inhibitors of or mutants deficient in the DNA repair pathways, lead to increased susceptibility to MMC [302]. Additionally, four genes in *E. coli* have been linked to MMC resistance when these genes have been overexpressed: the multidrug resistance locus *mdfA; gyrI*, which inhibits the gyrase activity *in vitro*; *rob*, which activates the transcription of efflux pumps, and *sdiA*, which activates expression of genes protecting the cell from UV damage.

We confirmed the stability of MMC in SP4 and SSG media with serum by pre-incubating MMC in the media at a 2.5 μg/ml final concentration for 2 hours or overnight at 30°C and 37°C. The test was done in a 96 well plate format to perform CCUs. Two hundred μl of overnight culture per column were spun down and the pellet was re-suspended in 200 μl per column pre-incubated MMC media. The cultures were deposited in the first well of each column and 1:10 CCU dilutions were performed. Only the cultures in the first row turned the color of the media to yellow. These wells had a large amount cells, which metabolism or death can change the pH of the media and shift the color. The rest of the dilutions did not change color, indicating that MMC was still available for activation and DNA crosslinking. Our experiments show that MMC does not lose its DNA crosslinking potential when incubated at 30°C or 37°C with growth media for at least 18 hours. An extensive review of MMC stability and storage condition is available elsewhere [314].

6.3.2 Inactivation of the recipient cells genomic DNA with MMC

We explored different conditions and concentrations for the inactivation of recipient cell genomes. Because we want to inactivate the recipient genomes just before the transplantation procedure, MMC was added directly to cultures grown overnight. Because an overnight culture is much more cell dense than the cultures in the MIC assays, we explored different concentrations of MMC by using the MIC concentration as a starting point. We explored 1 μg/ml, 2.5 μg/ml, 5 μg/ml, 10 μg/ml, and 20 μg/ml final concentrations when adding MMC to the overnight cultures. We also explored different lengths of incubation, before the recipient cells were washed

and prepared to undergo genome transplantation: 1 hour, 2 hours and 4 hours. The temperature during the incubation period also was varied: 4°C and 30°C (recipient cells are grown at 30°C before transplantation) (Table 6-2 and Table 6-3). We explored these conditions only with *M. capricolum*, the recipient organism for the *M. mycoides* subspecies *capri* donor genomes. This recipient cell—donor genome combination is the only genome transplantation system currently used in our laboratory. As other recipient cell—donor genome transplantation systems are developed the parameters would need to be explored for each recipient species separately. This is due to the difference in metabolism and response to MMC in different species (Table 6-1). We found that if the cells were incubated at 4°C with MMC, there was very slight inhibition of cell growth and in order to achieve moderate inhibition of cell growth much higher concentrations of MMC were needed. Based on the results of these assays and the maximum number of hours that we could extend the transplantation protocol, we chose to use two hours incubation at 30°C for the inactivation of recipient genomes (Table 6-3).

Table 6-2. Efficiency of Genome Transplantation with MMC Treated Cells

M. capricolum recipient cells, which restriction modification system has been inactivated through a transposon insertion, were treated with different concentrations of MMC or not treated with MMC as control. The *M. mycoides* tet^R, *lacZ* genome was used as donor; The pMyc01 plasmid (tet^R, *lacZ*) was used to transform the a separate set of samples in the same conditions. The number of successful transplants resulting from cells that were not treated with MMC was used as 100% efficiency in order to compare the efficiency of transplantation when MMC treated cells are used.The number of viable cells in each sample was determined through a CCU assay performed before the transplantation.

Description *M. capricolum* ΔRe recipient x *M. mycoides*	Average Transplants	Efficiency (compared to untreated cells)	Total Viable cells	Transformation pMyc01 (300ng) Average CFUs
2h 30°C non-treated	23	100%	2.83×10^9	657
2h 30°C 2.5µg/ml	22	96%	4.44×10^2	0
2h 30°C 5µg/ml	10	43%	1.67×10^2	0
1h 30°C non-treated	23	100%	2.58×10^9	230
1h 30°C 10µg/ml	8	35%	5.6×10^1	0

Table 6-3. Efficiency of Genome Transplantation with MMC Treated Cells under different conditions

Comparison of genome transplantation effectiveness between recipient cells treated or not treated with MMC under different conditions (temperature, incubation time, MMC concentration). Condition groups are highlighted for easier comparison. The number of transplants resulting from recipient cells not treated with MMC is used as 100% efficiency in order to compare the efficiency of genome transplantation with MMC treated cell. The total remaining viable recipient cells (*M. capricolum*) after each experiment is given in the last column.

Description	Successful Transplants		*M. capricolum* after experiment
	Average	Efficiency compared to non-treated cells	Total viable *M. capricolum* cells
Original culture before MMC treatment (30°C)			9.8×10^8
2h 30°C Tris-NaCl Buffer No MMC	83	100%	3.0×10^9
2h 30°C Tris-NaCl Buffer 2.5µg/ml MMC	34	41%	1.2×10^2
2h 4°C Tris-NaCl Buffer No MMC	82	100%	1.5×10^9
2h 4°C Tris-NaCl Buffer 2.5µg/ml MMC	36	44%	Not counted
4h 4°C Tris-NaCl Buffer No MMC	110	100%	9.5×10^8
4h 4°C Tris-NaCl Buffer 5µg/ml MMC	58	52%	9.6×10^3
4h 30°C Tris-NaCl Buffer No MMC	157	100%	2.3×10^9
4h 30°C Tris-NaCl Buffer 2.5µg/ml MMC	49	31%	1.2×10^3
2h 30°C Tris-Sucrose Buffer No MMC	15	100%	6.2×10^8
2h 30°C Tris-Sucrose Buffer 2.5µg/ml MMC	7	49%	Not counted
4h 30°C Tris-Sucrose Buffer No MMC	3	100%	7.7×10^8
4h 30°C Tris-Sucrose Buffer 2.5µg/ml MMC	2	50%	Not Counted
Culture before overnight incubation at 4°C			2.7×10^9
18h 4°C No MMC	65	100%	4.0×10^9
18h 4°C 2.5µg/ml MMC	41	63%	3.0×10^3

Next we attempted genome transplantation with MMC treated recipient cells where we varied the final concentration of MMC: 2.5 μg/ml, 5μg/ml, and 10 μg/ml. As a comparison of the efficiency of the transplantation, a transplantation experiment was carried out with recipients that were not treated with MMC. We observed that the higher the MMC concentration and the closer we get to zero surviving recipient cells the lower the efficiency of the transplantation procedure became (Table 6-2). MMC is capable of reacting with RNA and to some extent with proteins [309], although double stranded DNA is the primary and preferred target of MMC. We believe that at higher concentrations, the transcription and translation machinery of the recipient cells may undergo damage and thus be unable to "read" the new intact genome in order to survive, thus reducing the transplantation reaction efficiency to undetectable levels.

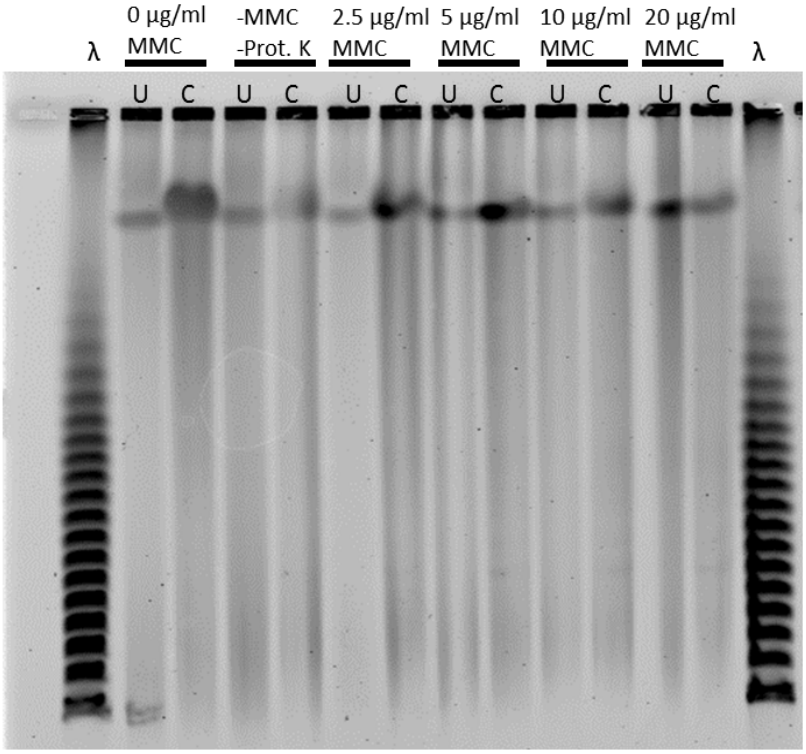
For the *M. capricolum—M. mycoides* subspecies *capri* recipient—donor pair of organisms we found that 2.5 μg/ml was the optimal MMC treatment concentration at which we achieved enough reduction of viable recipient cells and minimum reduction of transplantation efficiency (Table 6-2). Our transplantation efficiency with 2.5 μg/ml MMC treated recipient cells for 2 hours at 30°C, compared to non-treated cells, varied between 40-96% with donor genomes extracted from bacteria. One potential contribution to this variability is the variable quantity and quality of the gDNA in individual plugs within the same batch of prepared plugs, due to the clumping nature of mycoplasma cells. Transplanting bacterial genomes that were cloned and modified in yeast is in general much less efficient than genomes extracted from bacteria [249]. Using MMC treated recipient cells for transplantation of

bacterial genomes extracted from yeast was less efficient than using untreated cells (about 5 colonies on average). We are currently working on improving the transplantation efficiency with MMC treated recipient cells for both sources of gDNA.

To determine the effect of MMC on the gDNA and confirm that the genomes were inactivated we extracted gDNA in agarose plugs from MMC treated *M. mycoides* cells and examined the quality of the DNA by restriction digestion and gel electrophoresis (Figure 6-2). The majority of the gDNA appeared intact. The samples that were not digested with a single cutter restriction enzymehad a similar quantity of linearized gDNA across all MMC concentrations tested compared to the untreated gDNA. This linear fraction of the gDNA results primarily from the mechanical manipulations of the samples. Furthermore, no significant difference was observed in the background smearing in the lanes, suggesting the gDNA is not significantly degraded by MMC treatment. It is interesting to note that the restriction enzyme used in this experiment, PspXI, has a VC^T**CG**A_GB recognition site. It is reasonable to expect that if the complementary strands were cross-linked by MMC, PspXI will produce only nicks in the strands around the cross-linked CpG, which will hold the strands together, not allowing the linearization of the circular genome. We observed that digesting the plug with gDNA that was not treated with MMC produced a much larger linear gDNA band compared to the same undigested plug. However, digesting the samples treated with MMC, did not produce a significantly larger linearized gDNA band compared to the same undigested plugs. This is expected and indicative of effective crosslinking of the *M. mycoides* genomes by MMC (Figure 6-2).

Figure 6-2. Restriction Digest of gDNA from MMC Treated Cells.

Genomic DNA of MMC treated cells was embedded in agarose plugs and digested with PspXI. Legend: λ – Lambda DNA Ladder; U- undigested; C-cut (digested) with PspXI.



MMC treated genomes were also used as donor genomes in transplantation experiments with recipient cells that were not treated with MMC. As expected these donor genomes failed to produce viable transplants. Further studies are needed to explore the potential of these cross-linked genomes to participate in recombination events. Alternatively, recombination may still be possible between the genomes as the primary target of MMC is DNA and the enzymes in the recipient cells are still functional, as evidenced by the capability to obtain transplants. Therefore, it is likely that the enzymes participating in recombination are still active and recombination still occurs. We believe that if this is the case, virtually all recombination events will result in non-viable genomes, because even one gDNA crosslink may be enough to result in the death of the cell.

It is interesting to note that after MMC treatment of the recipient cells, the remaining surviving recipient cells that we observe are a miniscule fraction of the total number of cells that we generally need to use in order to obtain a decent number of transplants. Only approximately 0.000016% of the cells were viable after MMC treatment. Although the surviving recipient cells are about 20 times more than the obtained transplants, we know from previous experiments that on average we need 3 x 10^7 recipient cells to observe one transplantation event. Therefore, we believe that at least some of the transplants we obtained from our MMC treated recipient cells, result from donor genomes, entering recipient cells with cross-linked genomes, thus rescuing them from death, making them so called “zombie cells”. However, further studies will need to be completed to explore the validity of this hypothesis and evaluate alternative explanations of this phenomenon.

6.3.3 Transplanting genomes without antibiotic selection markers

Antibiotic treatment almost always results in decreased growth rates for resistant mutants. This is especially true for protein translation inhibitors like tetracycline. Therefore, antibiotic selection presents an additional challenge for the cells that already need to recover from the PEG treatment during genome transplantation, which may be contributing to the low efficiency of genome transplantation. We explored the use of MMC treated recipient cells to obtain transplants of genomes without antibiotic selection. Because our *M. mycoides* donor genomes contain a *lacZ* gene, we can still use a visual screening method to identify the transplants. Inclusion of Xgal in the SP4 agar results in the *M. mycoides* transplants being blue and easily distinguishable from surviving *M. capricolum* recipient cells, which are white. In the absence of MMC treatment, detecting blue transplants among recipient cells, when the transplantation reaction was plated on agar plates without antibiotic, was impossible, because less than one colony in 100,000 was blue. Due to the high number of recipient cells needed for the transplantation, the plates grow a lawn of recipient cells and the agar turns yellow from red due to the pH shift resulting from the lactic acid produced during cell metabolism detected by the pH indicator phenol red. As a result no blue colonies can be detected. Plating a dilution of the transplantation reaction allowed observing separate recipient cell colonies, but no blue transplants. We used *M. capricolum* restriction free recipient cells treated with 2.5μg/ml MMC for 2 hours at 30°C and transplanted the same *M. mycoides* donor genomes that have been used in previous experiment. These genomes have a tetracycline resistance marker; however, we

plated the transplantation reactions on plates without tetracycline. Using MMC treated recipient cells was the only condition that allowed us to observe blue transplants without selecting on plates with antibiotic (Figure 6-3). Next, we repeated the transplantation experiment with a *M. mycoides* donor genomes with (Figure 6-4) and without tetracycline resistance marker extracted from yeast or bacterial clones. Once again we were able to obtain blue transplants without antibiotic selection for both gDNA sources only when the recipient cells were treated with MMC.

Figure 6-3. Transplanting Genomes Extracted from a Tetracycline Resistant *M. mycoides* strain without tetracycline selection.

Recipient cells: *M. capricolum* restriction system free; gDNA donor: *M. mycoides* tet^R, *lacZ* extracted from the bacterial strain. The first row of plates demonstrates that if the recipient cells are not treated with MMC, transplants (blue colonies) are obtained only if the transplantation reaction is plated on SP4 agar plates containing X-gal and tetracycline (no blue colonies on SP4 agar plates with X-gal only). The second row of plates, demonstrates that when the recipient cells are treated with MMC, transplants (blue colonies) can be obtained on SP4 plates with X-gal and no tetracycline. Next to the plates is a close up of the transplants (blue colonies) and surviving recipient cells (white colonies).

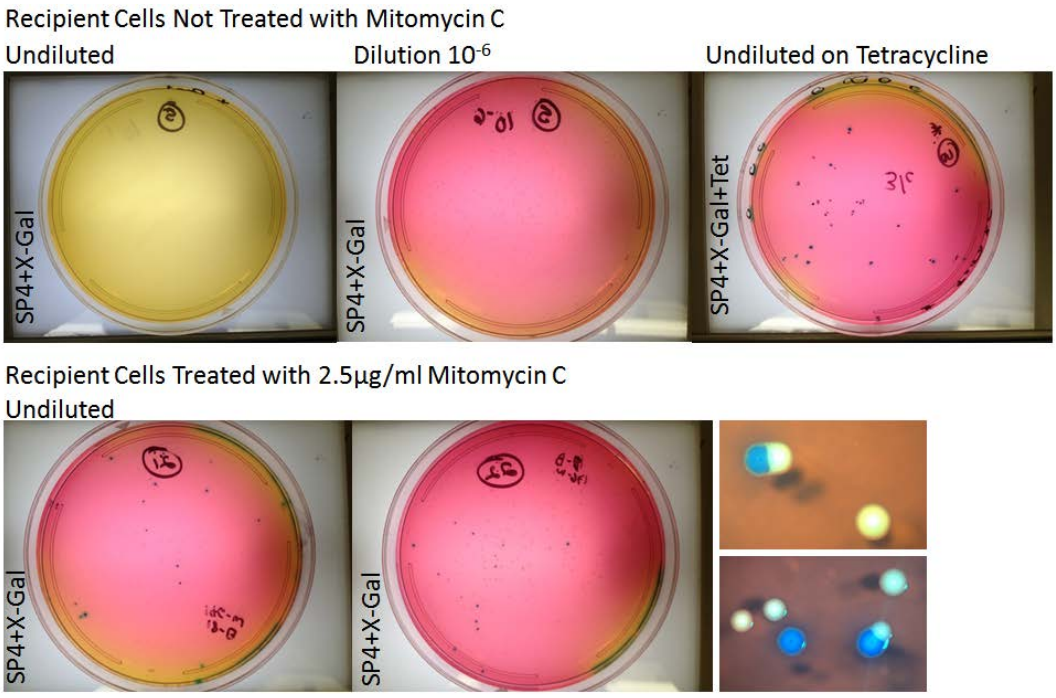
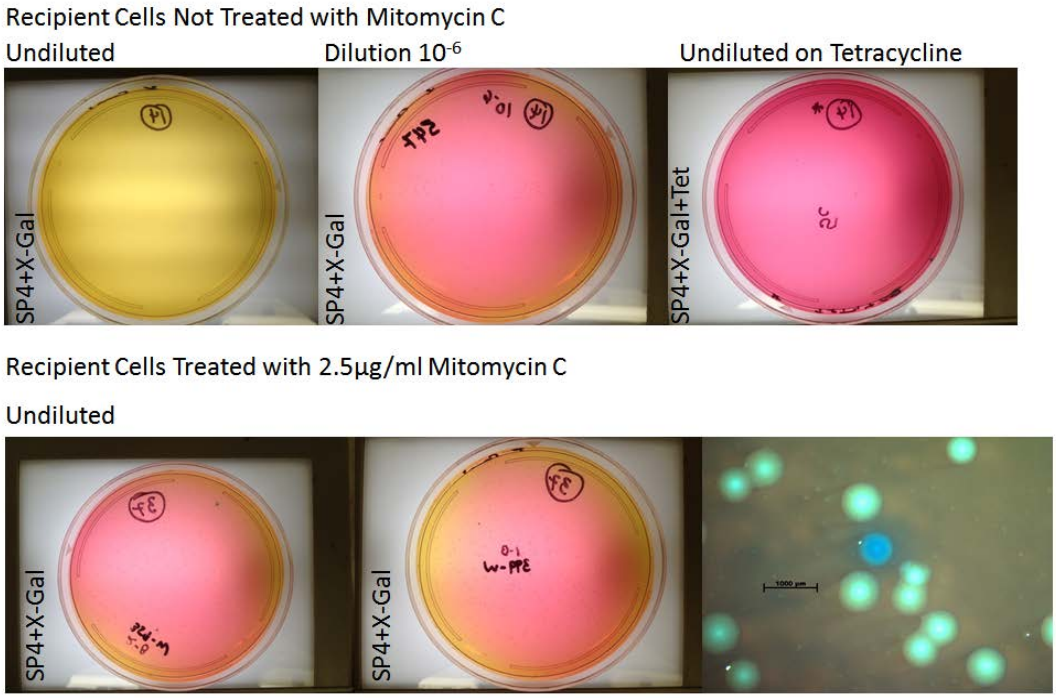


Figure 6-4. Transplanting a Tetracycline Resistant *M. mycoides* Strain Genome Extracted from Yeast without tetracycline selection.

Recipient cells: *M. capricolum* restriction system free; gDNA donor: *M. mycoides* tet^R, *lacZ* extracted from yeast. The first row of plates demonstrates that if the recipient cells are not treated wih MMC, transplants (blue colonies) are obtained only if the transplantation reaction is plated on SP4 agar plates containing X-gal and tetracycline (no blue colonies on SP4 agar plates with X-gal only). The second row of plates, demonstrates that when the recipient cells are treated with MMC, transplants (blue colonies) can be obtained on SP4 plates with X-gal and no tetracycline. Next to the plates is a close up of the transplants (blue colonies) and surviving recipient cells (white colonies).



The selection of antibiotic markers for mycoplasmas is limited to tetracycline puromycin and gentamycin. Tetracycline is the preferred marker in our laboratory as it has the lowest rate of spontaneous resistance. This marker, however, was reserved for genome transplantation in our laboratory. Genome transplantation is being used to iteratively delete segments of the *M. mycoides* genome, in order to find the minimum set of genes needed for growth in the laboratory [295]. Iterative probing of the resulting minimized *M. mycoides* strains with transposons to find out what other segments may be deleted has been difficult due to the need to "recycle" the antibiotic resistance markers. This means that once a change to the genome is made and a clone is established, we need to remove the antibiotic resistance marker from the modified genome, in order to use it along with the next transposon bombardment of the same genome. Currently, there are no methods developed for antibiotic resistance marker recycling in mycoplasmas. The current approach is to grow and passage the desired strain in media without antibiotic selection and wait until the strain becomes susceptible by losing the resistance gene (marker). This process is tedious and time consuming. The use of MMC to treat recipient cells and obtain transplants without antibiotic selection frees the tetracycline resistance marker and allows for iterative transposon probing by using up to tree antibiotic resistance markers, without the need to recycle selective markers. This in itself is an unexpected but useful application of MMC in genome transplantation in our laboratory.

6.4 Discussion

As synthetic biologists, one of our goals is to create technologies that enable biological research by accelerating the rate of and expanding the scope of cycles of design, build and test experimentation as applied to biological systems. Our Venter Institute team demonstrated this by booting up a bacterial cell programmed with a chemically synthesized genome in 2010 [255]. We developed technologies that enabled us to design and build a 1.1 Mb genome and then to test that genome by installing it in a suitable recipient cell using a process called genome transplantation. Four years later, the capacity to design and build whole genomes has greatly expanded. Our Venter Institute team designs and builds new bacterial genomes in weeks instead of months, commercial DNA foundries offer to synthesize 2 Mb genomes, and a team composed mostly of undergraduate students at Johns Hopkins University has redesigned and synthesized an entire yeast chromosome (unpublished data, [255], [315]). However scientists' ability to test synthetic bacterial genomes extends no farther than genome transplantation amongst a few species of bacteria that comprise the mycoides cluster of mycoplasmas. One reason we believe that current genome transplantation methods work in those mycoplasmas is they have very inefficient enzymatic machinery for homologous recombination. Our efforts to transplant genomes between other bacterial species have been thwarted by recombination of the selectable markers in donor genomes into the genomes of recipient cells. This work lays the foundation of an approach that we hope will eliminate that problem.

We have shown that treatment of cells to be used as genome recipients in genome transplantation reactions with the DNA alkylating agent MMC can inactivate the recipient cell genome leaving the cell intact and needing a new, i.e. donor chromosome to remain viable. Using our *M. capricolum—M. mycoides* recipient cell—donor genome system as a test bed, we demonstrated that MMC treatment of the *M. capricolum* recipient cells prior to genome transplantation inactivates the recipient cell genome and actually eliminates the need for antibiotic selection to isolate genome transplants.

We now need to extend this method to enable genome transplantation of genomes not from members of the mycoides cluster of mycoplasmas. Because our data here show that the efficiency of MMC genome inactivation varies among bacterial species, we will choose a recipient cell for our genome transplantation reaction that is efficiently inactivated by MMC. We are now planning to transplant a genome from *Mycoplasma gallisepticum* into a closely related species, *Mycoplasma imitans*. Our experiments (Table 6-1) showed *M. imitans* was more susceptible to MMC treatment than *M. capricolum*. Additionally, in terms of phylogenetic relatedness, which has been demonstrated by Carole Lartigue to be directly correlated to transplantation efficiency in her transplantation studies among members of the mycoides cluster of mycoplasmas (personal communication), *Mycoplasma imitans* is a logical recipient cell species for *M. gallisepticum* donor genomes.

Assuming our efforts to transplant *M. gallisepticum* are successful (or perhaps even if they are not), the next steps after that will be to further expand genome transplantation to other species of mycoplasmas, such as the ureaplasmas, and to

conventional bacteria. For the ureaplasmas we may need to increase the efficiency of the MMC treatment by inclusion of DNA repair inhibitors, or possibly inclusion in the MMC reaction of a suicide plasmid encoding a gene for the expression of one of the 5 different enzymes described to activate MMC. Hopefully we will find that the MMC method for recipient cell genome corruption will be the key to enabling generalized bacterial genome transplantation.

Acknowledgements

This work was funded by NIAID project 5R21AI098057. I am very grateful for the experimental support provided by Nacyra Assad-Garcia in planning and carrying out the experiments. John Glass, who is the project PI and my dissertation advisor provided us with the idea of chemical inactivation of the recipient cell genome. Many thanks also to Sanjay Vashee, Michael Montague, Ray-Yuan Chuang for critical discussions and support with troubleshooting, mentoring, and experimental advice. I am thankful also to the rest of the JCVI Synthetic Biology team for useful discussions of the experiments and data.

Chapter 7. Conclusions and Future Directions

7.1 Ureaplasma clinical importance

Ureaplasmas are considered opportunistic pathogens. As many as 8 in 10 healthy adult women are ureaplasma carriers [3]. While most infections are asymptomatic, there is still a very large burden of ureaplasma disease due to the adverse pregnancy outcomes resulting from ureaplasma infection of the upper urogenital tract. Ureaplasmas have been implicated in infertility, spontaneous abortion, stillbirth, premature birth, low birth weight, and perinatal morbidity and mortality. We believe that our research has added further evidence for the need for new research and changed the discussion in the field away from old ideas about the static nature of serovars and their importance in developing diagnostics and therapeutics.

7.2 Development of diagnostic tools

Over the past 20 years research in the field of ureaplasmas has been focused on developing suitable diagnostic tests that can identify and characterize the two species and their respective serotypes that could account for differential pathogenesis. The development of our real-time PCR assay that can detect each serovar without cross-reactivity and its application to a collection of 1061 clinical isolates changed the trajectory of diagnostic development away from a focus on using DNA approaches to associate serovar with disease outcome. Our hope was to be able to correlate specific serovars with specific clinical outcomes. However, we were surprised by the

complexity of patient isolates, often consisting of multiple serovars or genetic mosaic ureaplasma strains, and even some strains for which the serovar belonging or species could not be determined. Our study revealed extensive horizontal gene transfer among ureaplasmas, which lead us to question the utility of assays seeking to determine ureaplasma clinical isolate serovars.

7.3 What is a serovar?

Ureaplasma serovars were defined based on the reaction of different patient's sera with different ureaplasma patient isolates Fourteen unique reaction groups were established, and thus the ureaplasma isolates were classified as serovars 1 thorough 14 [5, 186]. No additional patient sera have been collected to determine if more serovars exists; although more likely could be found (several of the 1061 we analyzed using or serovar specific real time PCR assays could not be associates with any of the 14 serovars). Some of the reactive antibodies in the sera were defined and determined to bind the MBA. Monoclonal antibodies were later developed for serotyping of ureaplasma isolates. Serotyping assays were used to attempt to determine serovar association with different clinical outcomes. However, cross-reactivity of the antibodies with different serovars hampered these attempts [222]. As the molecular biology technology advanced, multiple PCR assays were developed to address the question of differential pathogenicity of the serovars [3]. The *mba* gene is frequently used in these assays to determine species and/or serovar. The general assumption was that the UU serovar genomes are static and not capable of recombination, therefore serovar/serotype was used interchangeably with genotype.

7.4 Changing how we understand serovars

MBA has been considered to be serovar specific [90, 181, 202, 204], and has been predominantly used in developing assays to determine the ureaplasma serovar of clinical isolates [3, 89]. Our comparative genome analysis (Chapter 2) showed that the MBA was part of a very large phase variable protein family. Furthermore, many serovars contained the MBA TRUs of other serovars (Figure 2-9 and Figure 2-10). We believe that serovars can switch from expressing one MBA TRU to expressing a different MBA TRU. For example UUR4 and UUR12 have an identical set of TRUs (Figure 2-11). A UUR4 strain may switch the expression of the TRU that is believed to be UUR4 specific to the TRU believed to be UUR12 specific. Such a strain will now test positive for UUR12 in assays that utilize the *mba* gene or the MBA protein as a target. It is also interesting to note that these two serovars have virtually identical genomes, with a notable genome rearrangement only in the *mba* locus, resulting in the expression of a different TRU (Figure 2-11).

Our RT-PCR assay described in chapter 3 was applied to 1061 clinical isolates (Chapter 4). We observed ureaplasma clinical isolates with genomes containing RT-PCR targets specific for more than one ATCC serovar. Closer examination of these isolates, through carefully designed molecular sequencing assays, reviled horizontal gene transfer (HGT) among ureaplasmas. Especially interesting was the case of one patient sample that tested positive for UPA1 and UPA6. We determined that this sample contained two hybrid strains. One strain had the MBA sequence of UPA1, but the rest of the genomic sequence was UPA6. The other strain had the MBA sequence

of UPA6, but the rest of the genomic DNA was UPA1. Therefore, HGT represents an additional mechanism that ureaplasmas can utilize to vary their repertoire of MBA TRUs. These observations lead us to believe that due to the plastic nature of the ureaplasma genomes the genotype of an ureaplasma isolate is not important in developing diagnostics and therapeutics. Instead we need to first re-evaluate the relationship between the strain's serovar designation (or serotype) and its genotype. Only then, we can develop effective diagnostic molecular assays for these organisms. Following are some questions demonstrating the need to re-evaluate the serotype to genotype relationship of ureaplasmas:

1. What would be the serovar designation of the clinical strain that contains the UPA6 MBA but its remaining genome is identical to the ATCC UPA1 genome and vice versa?
2. What would be the serovar designation of a UUR4 strain, which MBA TRU was switched to the UUR12's TRU?
3. What would be the serovar designation of a UUR10 strain, which MBA's TRU was switched to a TRU that has not been observed to be active before?
4. How do we determine the serovar designation of ureaplasma isolates that are negative for all serovars of all currently available serotyping and genotyping assays?
5. If serovar designation is described using only the *mba*, is it significant and should we continue to strive to develop assays to differentiate the separate serovars? We found 22 unique TRUs in the 19 sequenced ureaplasma genomes.

Should we expand the serovars to 22 instead of 14, and add more as we find novel TRUs?

There is a need to clarify the definition of the ureaplasma serovars, especially if we want to attribute the differences among the serovars to different clinical outcomes. The extensive HGT we observed in the 1061 clinical samples suggests that ureaplasmas have plastic genomes that are prone to exchanging genetic material most likely during co-infections. HGT and MBA TRU switching can explain previously observed cross-reactivity of serotyping and genotyping assays [5, 86, 222]. MBA is the most widely used target to determine the serovar designation and some of the original sera used to develop the serotyping scheme contained antibodies targeting the MBA. Traditional serotyping techniques are being substituted by molecular techniques like PCR and the genomes of all 14 ATCC serovars have been sequenced. Therefore, even though not all antibody targets have been determined in the original serotyping method, we suggest the adaptation of the MBA as the serovar specific and determining factor. However, it should be made clear that the serovar number describes only the type of actively expressed *mba* at the time of testing. Based on our comparative genome analysis of 19 ureaplasma strains, each stain is capable of expressing only one *mba* gene with a single active TRU. The conserved domain of the *mba* gene contains its lipoprotein anchor and transmembrane domain. All analyzed ureaplasma genomes contain only one conserved domain attached to one TRU and surrounded by alternative TRUs in this locus that do not contain a start codon and thus are not recognized as open reading frames. In chapter 2 we have proposed a hypothetical mechanism for the switching of the TRU attached to the *mba*

conserved domain, involving inverted repeats on both sides of each alternative TRU and a recombinase. It is conceivable that a second *mba* conserved domain may be acquired through HGT. We have not observed this in the tested clinical isolate collection or the sequenced ureaplasma genomes. However, we observed isolates which had exchanged at least the MBA TRU, if not the whole *mba* gene (the *mba* conserved domain sequence is conserved among strains of the same species). If the active MBA determins the serovar designation, then the serovar designation would no longer be strictly associated with a specific genotype of the organism due to the capability of the *mba* TRU to be switched for another TRU in the *mba* locus or to be transferred horizontally as well as the overall high HGT frequency among ureaplasmas. We observed genetic mosaic strains of UPA and UUR, although at a much lower rate. Thus, the serovar designation based entirely on the *mba* would no longer be indicative of the species designation, like it is currently: serovars 1,3,6, and 14 are UPA and serovars 2,4,5,7-13 are UUR. Species would continue to be determined based on the urease PCR assays [316] and genome size [2].

Preserving the serovar designation for ureaplasmas may be beneficial as the *mba* gene gets increasing attention, and accumulates evidence to be recognized as an immunogenic factor leading to the presentation of different severity of pathologies [127, 128, 317]. MBA size variation (the variation of the number of repeats of the TRU) believed to allow the organism to evade the host immune response. Furthermore, the difference in the number of repeats of the same MBA was associated with different pathology. It is conceivable that different TRUs and the difference of the number of their repeats may be associated with differential

pathogenicity. In this case genotyping based on the *mba* gene will be of importance, if the goal is to determine differential pathogenicity outcome based on the MBA variant and size. The MBA target should be carefully selected to account for the phase variability of the gene. Also two formats of the assay can be developed. One assay should determine the currently expressed TRU. This is the TRU attached to the conserved domain of the *mba*. Once specific pathologies have been associated with specific TRUs, a second assay should be developed to detect any TRU of pathogenic potential in the clinical isolate (including those that are not connected to the *mba* conserved domain as ureaplasmas are capable of switching to their use). To determine which TRU is currently active, primers should be designed to anneal to the conserved domain and the TRU. To determine that the TRU is attached to the *mba* conserved domain, and thus it is active, probes spanning the junction between the two should be used in a real-time PCR.

7.5 Shift in literature to host-pathogen studies with animal models and cell lines

We are pleased to see that our comparative genome analysis of all 14 ATCC serovars and 5 clinical isolates, along with our study revealing extensive horizontal gene transfer have provided a pivot point for research in the field of ureaplasma pathogenesis. A clear shift is noticeable in the published literature after our publications. Before our publications, the ureaplasma literature was describing development of assays to genotype the ureaplasmas to the serovar level for the application of these assays to studies aiming to correlate serovars to pathogenic

outcomes was common. A summary of the majority of this literature is represented in a review by Waites *et al* [3]. Based on the observations of our studies, we suggested that differential clinical outcome is based on the presence or absence of potential pathogenicity genes in the ureaplasma strain infecting the patient and/or patient to patient differences in terms of autoimmunity and microbiome [9, 104]. The published studies after our publications and suggestions are dominated more by host-pathogen interaction studies in animal models and cell lines. These studies focus on the role of the host immune response [14, 35, 45, 47, 49, 51, 125, 127, 318-323] and the difference in host alleles for key immune system components [125] in the development of the pathologies associated with ureaplasmas. The interactions of the host immune system and the MBA have also received more attention and studies show that the severity of disease is associated with the size of the MBA (number of repeats of the TRU) [123, 127, 128]. Furthermore, a gene annotated as a putative tyrosine recombinase XerC in ureaplasmas was shown to be functional in a recent study and its suggested involvement in the recombination events responsible for the phase variation of the *mba* locus in UPA3 was confirmed [324]. Although not directly related to our studies and observations, it is worth noting that novel antibiotic treatment plans were developed showing that the administration of certain antibiotics before birth may be beneficial for the infant by preventing the development or lowering the severity of BPD [129-131, 325].

7.6 Need to develop genome manipulation tools to support the understanding of pathogenicity and increase the power of using animal models.

The ureaplasma research community is at a point where their work could be greatly enhanced by the development of genetic tools for the creation of designed mutant ureaplasmas. New molecular diagnostics methods show that ureaplasmas have a large impact on society due to their roles in adverse pregnancy outcomes and infection of neonates. Although we do not understand the nature of ureaplasma pathology, we have a number of hypotheses about specific genes products and antigenic variation that may result in inflammation and disease that cannot be tested at present. We have murine, ovine, and baboon models for ureaplasma pathology [27, 56, 121, 247].

The capacity to create deletion mutants has been a prerequisite in understanding host-pathogen interaction and factors involved in pathogenicity in many well studied pathogens. If we had the ability to manipulate ureaplasma genomes, we would make a series of gene knockouts to evaluate whether any of a set of predicted surface associated proteases contribute to pathogenicity. We would look carefully at the role of the phase variable MBA protein on ureaplasma pathogenesis. MBA is the dominant ureaplasma antigen based on assays using patient serum. Can the disease outcome be affected by eliminating MBA phase variation so that it is always on or always off? There are many similar questions that could be addressed. Furthermore, the increased understanding of pathogenicity factors can aid in

development of better screening methods, drug targets and vaccines to decrease the morbidity associated with ureaplasmal infections.

In chapter 5 we have laid a foundation for the effort to create ureaplasma mutants. We carefully studied the different growth conditions and transformation conditions for ureaplasmas, and believe that we able to insert foreign DNA in the ureaplasma cells, due to the transient antibiotic resistance we observe after transformation reactions. We are on the verge of successfully inserting a transposon into the UPA3 ATCC strain genome. Another group has already been able to transform ureaplasma strains with a tn4001 based transposon and a gentamycin resistance marker using protocols designed by John Glass (manuscript in preparation). However, their transformation success was limited only to UPA clinical isolate strains. We hope that our careful evaluation of transformation conditions and recovery of the transformed cells will allow is to expand the transformation method to all ureaplasma strains, including the ATCC strains. Furthermore, we are utilizing a different transposon system, the tn5, with an already purified transposase, to aid in the insertion of the transposon. We are also using *tetM* as a resistance marker gene, which if successful, will expand the choice of resistance markers to use for the creation of mutants. Once we have successful transformations, we can create a library of transposon mutants for the use of the ureaplasma scientific community. However, our ultimate goal is to make fast and specific alterations of the ureaplasma genome by utilizing the tools developed for synthetic biology at the J. Craig Venter Institute: genome cloning and modification in yeast and subsequent transplantation of the modified genome to recipient ureaplasma cells [249, 286].

To achieve this goal our team is going to continue to work on cloning the UUR9 genome as yeast centromeric plasmids (YCPs). There are several approaches that can be utilized to achieve this goal. The UUR9 intact linearized genome can be co-transformed with the pRS313 plasmid we have constructed. The plasmid contains a YCP vector sequence comprised of yeast centromeric sequences and autonomous replication sequences flanked by two arms of homology to the UUR9 genome. Once the UUR9 linear genome and linear plasmid are in yeast, yeast can recombine the two to create a circular chromosome of UUR9 that will be maintained as one of the yeast chromosomes. Alternatively, once we have achieved successful transformation of ureaplasmas, we can use a transposon to insert the YCP vector sequence in the ureaplasma genome before we attempt to clone it in yeast (Figure 5-5). Yet another alternative is to fuse the ureaplasma cells, that have been transformed and contain the YCP vector sequence, with yeast spheroplasts, which has been used to successfully clone some mycoplasma genomes in yeast [326].

Once we have successfully cloned the ureaplasma genome in yeast, we will use a JCVI method developed to manipulate a bacterial genome cloned in yeast. The Tandem Repeat Coupled with Endonuclease Cleavage (TREC) method [287] allows for knockout construction in yeast. The first step is to generate a core cassette that contains the gene encoding the endonuclease SceI under the control of a galactose responsive promoter, a URA3 gene for selection, and a SceI recognition site. On one side of these sequences is a 50 bp of DNA identical to DNA adjacent to the 5′ end of the target gene and on the other side is a stretch of DNA homologous to DNA on the 5′ end of the target gene. This will result in a tandem repeat (TR) upon integration.

The core cassette is then transformed into yeast containing the ureaplasma genome. After growth on selective medium, colonies are screened for positives that contain the replacement of the core cassette at the target gene locus. The positive clone is then grown on galactose medium to induce expression of SceI, which cleaves its recognition site resulting in loss of the core cassette by recombination between the DNA adjacent to the 5′ end of the target gene and the tandem repeat. The cells are plated on medium containing 5-fluoroorotic acid (5-FOA) to select only those cells that have lost the core cassette resulting in the seamless deletion of the target gene. This method was used to remove a *M. mycoides* restriction enzyme gene in an earlier work [249]. The TREC method is needed to manipulate bacterial genomes in yeast because the traditional yeast genetic method of replacement of a gene with only a URA3 marker and subsequent counter selection with 5-FOA has led to instability of the bacterial genome resulting in unwanted and unpredictable deletions of genomic areas [287].

Once the changes are made to the ureaplasma genome in yeast, the next step is to transplant the altered genomes into recipient cells to generate mutant ureaplasma strains. These strains can then be tested for loss or gain of function or gain of, depending on the nature of the deleted or inserted gene. However, this task is not easy to achieve. It requires that we first optimize the transformation reaction conditions for ureaplasma. Additionally, we need to design assays to detect recombination of the selection markers into the host genome, instead of complete donor genome transplantation. Such recombination has hampered the genome transplantation among other mycoplasma donor-recipient pairs. Considering the extensive HGT we observed

in ureaplasma clinical isolates, it is likely that recombination between the donor and host genome may thwart genome transplantation in ureaplasma donor-recipient pairs, including extending this technology to a wider variety of organisms. We are exploring the use of the DNA cross-linking drug Mitomycin C to prevent false positive genome transplantation results due to recombination between the donor and recipient genomes. We hypothesize that covalently crosslinking the recipient gDNA before the donor genome enters the cells will prevent recombination and thus increase the probability of a true genome transplantation event. Alternatively, recombination between the cross-linked recipient genome and the "healthy" donor genome will result in non-viable offspring cells, providing an additional selection mechanism for cells resulting from true genome transplantation. In chapter 6 we describe the current progress of this project and the use of Mitomycin C to enable marker-less genome transplantation. Should Mitomycin C prove effective in facilitating genome transplantation in recombination prone organisms, the expansion of transplantation to species other than *M. mycoides* & *M. capricolum* could be successful. This can be a springboard to generalized genome transplantation and making use of the capacity to build synthetic genomes in the commercial sector to satisfy human needs, such as development and production of novel drugs, bio-renewable energy, and better understanding of host-pathogen interactions of genetically intractable pathogens, to name a few.

In the short term, the development of traditional molecular microbiology tools like transposon mutagenesis, plasmids for gene knockout and complementation, and adopted synthetic biology technology for making targeted gene deletions and

insertions in ureaplasma genomes will allow better understanding of host-pathogen interaction and mechanisms of pathogenicity for ureaplasmas. We believe these technologies will catalyze the expansion of ureaplasma biology to enable development of better screening methods for pathogenic ureaplasma strains, new drug targets and eventual creation of vaccines to decrease the adverse outcomes of ureaplasmal infections.

Bibliography

1. Shepard, M.C., *The recovery of pleuropneumonia-like organisms from Negro men with and without nongonococcal urethritis.* Am J Syph Gonor Vener Dis, 1954. **38**: p. 113-124.

2. Shepard, M., et al., *Ureaplasma urealyticum gen. nov., sp. nov.:proposed nomenclature for the human T (T-strain) mycoplasmas.* Int J Syst Bacteriol, 1974. **24**: p. 160-171.

3. Waites, K.B., et al., *Ureaplasma,* in *Molecular Detection of Human Bacterial Pathogens*, D. Liu, Editor. 2011, Taylor and Francis, CRC Press: Boca Raton, FL.

4. Piot, P., *Distribution of eight serotypes of Ureaplasma urealyticum in cases of non-gonococcal urethritis and of gonorrhoea, and in healthy persons.* Br J Vener Dis, 1976. **52**(4): p. 266-8.

5. Robertson, J.A. and G.W. Stemke, *Expanded serotyping scheme for Ureaplasma urealyticum strains isolated from humans.* J Clin Microbiol, 1982. **15**(5): p. 873-8.

6. Turunen, H., P. Leinikki, and E. Jansson, *Serological characterisation of Ureaplasma urealyticum strains by enzyme-linked immunosorbent assay (ELISA).* J Clin Pathol, 1982. **35**(4): p. 439-43.

7. Robertson, J.A., et al., *Proposal of Ureaplasma parvum sp. nov. and emended description of Ureaplasma urealyticum (Shepard et al. 1974).* Int J Syst Evol Microbiol, 2002. **52**: p. 587-597.

8. Glass, J.I., et al., *The complete sequence of the mucosal pathogen Ureaplasma urealyticum.* Nature, 2000. **407**(6805): p. 757-62.

9. Paralanov, V., et al., *Comparative genome analysis of 19 Ureaplasma urealyticum and Ureaplasma parvum strains.* BMC Microbiol, 2012. **12**: p. 88.

10. Waites, K.B., et al., eds. *Cumitech 34, Laboratory diagnosis of mycoplasmal infections*. ed. C.e.F.S. Nolte. 2001, American Society for Microbiology: Washington, D.C.

11. Waites, K.B., B. Katz, and R.L. Schelonka, *Mycoplasmas and ureaplasmas as neonatal pathogens.* Clin Microbiol Rev, 2005. **18**(4): p. 757-89.

12. Clegg, A., et al., *High rates of genital mycoplasma infection in the highlands of Papua New Guinea determined both by culture and by a commercial detection kit.* J Clin Microbiol, 1997. **35**(1): p. 197-200.

13. Bayraktar, M.R., et al., *Prevalence and antibiotic susceptibility of Mycoplasma hominis and Ureaplasma urealyticum in pregnant women.* Int J Infect Dis, 2010. **14**(2): p. e90-5.

14. Capoccia, R., G. Greub, and D. Baud, *Ureaplasma urealyticum, Mycoplasma hominis and adverse pregnancy outcomes.* Curr Opin Infect Dis, 2013. **26**(3): p. 231-40.

15. Hunjak, B., et al., *Ureaplasma urealyticum and Ureaplasma parvum in women of reproductive age.* Arch Gynecol Obstet, 2013. **289**(2): p. 407-12.

16. Diaz, L., et al., *Frequency and antimicrobial sensitivity of Ureaplasma urealyticum and Mycoplasma hominis in patients with vaginal discharge.* MEDICC Rev, 2013. **15**(4): p. 45-7.

17. Leli, C., et al., *Microbial and vaginal determinants influencing Mycoplasma hominis and Ureaplasma urealyticum genital colonization in a population of female patients.* Infez Med, 2013. **21**(3): p. 201-6.

18. Eun, H.S., et al., *Serological investigation of Ureaplasma urealyticum in Korean preterm infants.* Korean J Pediatr, 2013. **56**(11): p. 477-81.

19. Jalal, H., et al., *Molecular epidemiology of selected sexually transmitted infections.* Int J Mol Epidemiol Genet, 2013. **4**(3): p. 167-74.

20. Ye, G., et al., *The resistance analysis of Ureaplasma urealyticum and Mycoplasma hominis in female reproductive tract specimens.* Cell Biochem Biophys, 2013. **68**(1): p. 207-10.

21. Iwasaka, T., T. Wada, and H. Sugimori, *Enhancement of colonization of Ureaplasma urealyticum in the mouse genital tract by estrogen treatment.* Am J Obstet Gynecol, 1986. **155**(5): p. 1124-7.

22. Viscardi, R.M., *Ureaplasma species: role in diseases of prematurity.* Clin Perinatol, 2010. **37**(2): p. 393-409.

23. Carroll, S.G., et al., *Lower genital tract swabs in the prediction of intrauterine infection in preterm prelabour rupture of the membranes.* Br J Obstet Gynaecol, 1996. **103**(1): p. 54-9.

24. Romero, R., et al., *Meta-analysis of the relationship between asymptomatic bacteriuria and preterm delivery/low birth weight.* Obstet Gynecol, 1989. **73**(4): p. 576-82.

25. Furr, P.M. and D. Taylor-Robinson, *Prevalence and significance of Mycoplasma hominis and Ureaplasma urealyticum in the urines of a non-venereal disease population.* Epidemiol Infect, 1987. **98**(3): p. 353-9.

26. Takahashi, S., et al., *Detection of Mycoplasma genitalium, Mycoplasma hominis, Ureaplasma urealyticum, and Ureaplasma parvum DNAs in urine from asymptomatic healthy young Japanese men.* J Infect Chemother, 2006. **12**(5): p. 269-71.

27. Novy, M.J., et al., *Ureaplasma parvum or Mycoplasma hominis as sole pathogens cause chorioamnionitis, preterm delivery, and fetal pneumonia in rhesus macaques.* Reprod Sci, 2009. **16**(1): p. 56-70.

28. Kim, M.J., et al., *Widespread microbial invasion of the chorioamniotic membranes is a consequence and not a cause of intra-amniotic infection.* Lab Invest, 2009. **89**(8): p. 924-36.

29. Gerber, S., et al., *Detection of Ureaplasma urealyticum in second-trimester amniotic fluid by polymerase chain reaction correlates with subsequent preterm labor and delivery.* J Infect Dis, 2003. **187**(3): p. 518-21.

30. Abele-Horn, M., et al., *Ureaplasma urealyticum colonization and bronchopulmonary dysplasia: a comparative prospective multicentre study.* Eur J Pediatr, 1998. **157**(12): p. 1004-11.

31. Garland, S.M. and E.D. Bowman, *Role of Ureaplasma urealyticum and Chlamydia trachomatis in lung disease in low birth weight infants.* Pathology, 1996. **28**(3): p. 266-9.

32. Hannaford, K., et al., *Role of ureaplasma urealyticum in lung disease of prematurity.* Arch Dis Child Fetal Neonatal Ed, 1999. **81**(3): p. F162-7.

33. Pacifico, L., et al., *Ureaplasma urealyticum and pulmonary outcome in a neonatal intensive care population.* Pediatr Infect Dis J, 1997. **16**(6): p. 579-86.

34. Perzigian, R.W., et al., *Ureaplasma urealyticum and chronic lung disease in very low birth weight infants during the exogenous surfactant era.* Pediatr Infect Dis J, 1998. **17**(7): p. 620-5.

35. Viscardi, R.M., *Ureaplasma species: role in neonatal morbidities and outcomes.* Arch Dis Child Fetal Neonatal Ed, 2013. **99**(1): p. F87-92.

36. Waites, K.B., et al., *Congenital and opportunistic infections: Ureaplasma species and Mycoplasma hominis.* Semin Fetal Neonatal Med, 2009. **14**(4): p. 190-9.

37. Cassell, G.H., et al., *Isolation of Mycoplasma hominis and Ureaplasma urealyticum from amniotic fluid at 16-20 weeks of gestation: potential effect on outcome of pregnancy.* Sex Transm Dis, 1983. **10**(4 Suppl): p. 294-302.

38. DiGiulio, D.B., et al., *Microbial prevalence, diversity and abundance in amniotic fluid during preterm labor: a molecular and culture-based investigation.* PLoS One, 2008. **3**(8): p. e3056.

39. Okogbule-Wonodi, A.C., et al., *Surfactant protein-A enhances ureaplasmacidal activity in Vitro.* Innate Immun, 2010.

40. Sung, T.J., et al., *Frequency of ureaplasma serovars in respiratory secretions of preterm infants at risk for bronchopulmonary dysplasia.* Pediatr Infect Dis J, 2011. **30**(5): p. 379-83.

41. Da Silva, O., D. Gregson, and O. Hammerberg, *Role of Ureaplasma urealyticum and Chlamydia trachomatis in development of bronchopulmonary*

dysplasia in very low birth weight infants. Pediatr Infect Dis J, 1997. **16**(4): p. 364-9.

42. van Waarde, W.M., et al., *Ureaplasma urealyticum colonization, prematurity and bronchopulmonary dysplasia.* Eur Respir J, 1997. **10**(4): p. 886-90.

43. Couroucli, X.I., et al., *Detection of microorganisms in the tracheal aspirates of preterm infants by polymerase chain reaction: association of adenovirus infection with bronchopulmonary dysplasia.* Pediatr Res, 2000. **47**(2): p. 225-32.

44. Heggie, A.D., et al., *Identification and quantification of ureaplasmas colonizing the respiratory tract and assessment of their role in the development of chronic lung disease in preterm infants.* Pediatr Infect Dis J, 2001. **20**(9): p. 854-9.

45. Kallapur, S.G., B.W. Kramer, and A.H. Jobe, *Ureaplasma and BPD.* Semin Perinatol, 2013. **37**(2): p. 94-101.

46. Wang, E.E., A. Ohlsson, and J.D. Kellner, *Association of Ureaplasma urealyticum colonization with chronic lung disease of prematurity: results of a metaanalysis.* J Pediatr, 1995. **127**(4): p. 640-4.

47. Lowe, J., et al., *Association between Pulmonary Ureaplasma Colonization and Bronchopulmonary Dysplasia in Preterm Infants: Updated Systematic Review and Meta-Analysis.* Pediatr Infect Dis J, 2014.

48. Viscardi, R.M., et al., *Antenatal Ureaplasma urealyticum respiratory tract infection stimulates proinflammatory, profibrotic responses in the preterm baboon lung.* Pediatr Res, 2006. **60**(2): p. 141-6.

49. Collins, J.J., et al., *Repeated intrauterine exposures to inflammatory stimuli attenuated transforming growth factor-beta signaling in the ovine fetal lung.* Neonatology, 2013. **104**(1): p. 49-55.

50. Okogbule-Wonodi, A.C., et al., *Necrotizing enterocolitis is associated with ureaplasma colonization in preterm infants.* Pediatr Res, 2011. **69**(5 Pt 1): p. 442-7.

51. Wolfs, T.G., et al., *Antenatal ureaplasma infection impairs development of the fetal ovine gut in an IL-1-dependent manner.* Mucosal Immunol, 2013. **6**(3): p. 547-56.

52. Viscardi, R.M., et al., *Incidence of invasive ureaplasma in VLBW infants: relationship to severe intraventricular hemorrhage.* J Perinatol, 2008. **28**(11): p. 759-65.

53. Fonseca, L.T., R.C. Silveira, and R.S. Procianoy, *Ureaplasma bacteremia in very low birth weight infants in Brazil.* Pediatr Infect Dis J, 2011. **30**(12): p. 1052-5.

54. Goldenberg, R.L., et al., *The Alabama Preterm Birth Study: umbilical cord blood Ureaplasma urealyticum and Mycoplasma hominis cultures in very preterm newborn infants.* Am J Obstet Gynecol, 2008. **198**(1): p. 43 e1-5.

55. Waites, K.B., et al., *Chronic Ureaplasma urealyticum and Mycoplasma hominis infections of central nervous system in preterm infants.* Lancet, 1988. **1**(8575-6): p. 17-21.

56. Normann, E., et al., *A novel mouse model of Ureaplasma-induced perinatal inflammation: effects on lung and brain injury.* Pediatr Res, 2009. **65**(4): p. 430-6.

57. Berger, A., et al., *Intrauterine infection with Ureaplasma species is associated with adverse neuromotor outcome at 1 and 2 years adjusted age in preterm infants.* J Perinat Med, 2009. **37**(1): p. 72-8.

58. Burstein, G.R. and J.M. Zenilman, *Nongonococcal urethritis--a new paradigm.* Clin Infect Dis, 1999. **28 Suppl 1**: p. S66-73.

59. Gonzalez Pedraza Aviles, A., M.C. Ortiz Zaragoza, and A. Irigoyen Coria, *Bacterial vaginosis a "broad overview".* Rev Latinoam Microbiol, 1999. **41**(1): p. 25-34.

60. Plummer, D.C., S.M. Garland, and G.L. Gilbert, *Bacteraemia and pelvic infection in women due to Ureaplasma urealyticum and Mycoplasma hominis.* Med J Aust, 1987. **146**(3): p. 135-7.

61. Haggerty, C.L., *Evidence for a role of Mycoplasma genitalium in pelvic inflammatory disease.* Curr Opin Infect Dis, 2008. **21**(1): p. 65-9.

62. Lee, J.S., et al., *Concordance of Ureaplasma urealyticum and Mycoplasma hominis in infertile couples: impact on semen parameters.* Urology, 2013. **81**(6): p. 1219-24.

63. Zeighami, H., et al., *Prevalence of Ureaplasma urealyticum and Ureaplasma parvum in semen of infertile and healthy men.* Int J STD AIDS, 2009. **20**(6): p. 387-90.

64. Haggerty, C.L. and R.B. Ness, *Diagnosis and treatment of pelvic inflammatory disease.* Womens Health (Lond Engl), 2008. **4**(4): p. 383-97.

65. Imudia, A.N., et al., *The prevalence of ureaplasma urealyticum, mycoplasma hominis, chlamydia trachomatis and neisseria gonorrhoeae infections, and the rubella status of patients undergoing an initial infertility evaluation.* J Assist Reprod Genet, 2008. **25**(1): p. 43-6.

66. Andrade-Rocha, F.T., *Ureaplasma urealyticum and Mycoplasma hominis in men attending for routine semen analysis. Prevalence, incidence by age and clinical settings, influence on sperm characteristics, relationship with the leukocyte count and clinical value.* Urol Int, 2003. **71**(4): p. 377-81.

67. Witkin, S.S., et al., *Ureaplasma urealyticum and Mycoplasma hominis detected by the polymerase chain reaction in the cervices of women undergoing in vitro fertilization: prevalence and consequences.* J Assist Reprod Genet, 1995. **12**(9): p. 610-4.

68. Cassell, G.H., et al., *Microbiologic study of infertile women at the time of diagnostic laparoscopy. Association of Ureaplasma urealyticum with a defined subpopulation.* N Engl J Med, 1983. **308**(9): p. 502-5.

69. Liu, J., et al., *Prevalence of Ureaplasma Urealyticum, Mycoplasma Hominis, Chlamydia Trachomatis Infections, and Semen Quality in Infertile and Fertile Men in China.* Urology.

70. Salmeri, M., et al., *Prevalence of Ureaplasma urealyticum and Mycoplasma hominis infection in unselected infertile men.* J Chemother. **24**(2): p. 81-6.

71. Eilers, E., et al., *Intrarenal abscesses due to Ureaplasma urealyticum in a transplanted kidney.* J Clin Microbiol, 2007. **45**(3): p. 1066-8.

72. Geissdorfer, W., et al., *Ureaplasma urealyticum meningitis in an adult patient.* J Clin Microbiol, 2008. **46**(3): p. 1141-3.

73. Levi, N., et al., *Aortic graft infection with mycoplasma (Ureaplasma urealyticum).* Eur J Vasc Endovasc Surg, 1995. **10**(3): p. 374-5.

74. Boyle, E.M., Jr., J. Burdine, and R.M. Bolman, 3rd, *Successful treatment of Mycoplasma mediastinitis after heart-lung transplantation.* J Heart Lung Transplant, 1993. **12**(3): p. 508-12.

75. Garcia-de-la-Fuente, C., et al., *Post-operative mediastinitis, pleuritis and pericarditis due to Mycoplasma hominis and Ureaplasma urealyticum with a fatal outcome.* J Med Microbiol, 2008. **57**(Pt 5): p. 656-7.

76. Pigrau, C., et al., *Sternotomy infection due to Mycoplasma hominis and Ureaplasma urealyticum.* Eur J Clin Microbiol Infect Dis, 1995. **14**(7): p. 597-8.

77. Walkty, A., et al., *Ureaplasma parvum as a cause of sternal wound infection.* J Clin Microbiol, 2009. **47**(6): p. 1976-8.

78. Waites, K.B., D.F. Talkington, and C.M. Bebear, *Mycoplasmas*, in *Manual of Commercial Methods in Clinical Microbiology*, A.L. Truant, Editor. 2002, American Society for Microbiology: Washington, D.C. p. 201-224.

79. Waites, K.B., et al., *Mycoplasma and Ureaplasma*, in *Clinical Microbiology Procedure Manual*, H. Isenberg, Editor. 2004, American Society for Microbiology Press: Washington, D.C.

80. Saiki, R.K., et al., *Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase.* Science, 1988. **239**(4839): p. 487-91.

81. Blanchard, A., *Ureaplasma urealyticum urease genes; use of a UGA tryptophan codon.* Mol Microbiol, 1990. **4**(4): p. 669-76.

82. Willoughby, J.J., et al., *Isolation and detection of urease genes in Ureaplasma urealyticum.* Infect Immun, 1991. **59**(7): p. 2463-9.

83. Robertson, J.A., et al., *Polymerase chain reaction using 16S rRNA gene sequences distinguishes the two biovars of Ureaplasma urealyticum.* J Clin Microbiol, 1993. **31**(4): p. 824-30.

84. Deguchi, T., et al., *Association of Ureaplasma urealyticum (biovar 2) with nongonococcal urethritis.* Sex Transm Dis, 2004. **31**(3): p. 192-5.

85. Yoshida, T., et al., *Phylogeny-based rapid identification of mycoplasmas and ureaplasmas from urethritis patients.* J Clin Microbiol, 2002. **40**(1): p. 105-10.

86. Kong, F., et al., *Species identification and subtyping of Ureaplasma parvum and Ureaplasma urealyticum using PCR-based assays.* J Clin Microbiol, 2000. **38**(3): p. 1175-9.

87. Harasawa, R. and Y. Kanamoto, *Differentiation of two biovars of Ureaplasma urealyticum based on the 16S-23S rRNA intergenic spacer region.* J Clin Microbiol, 1999. **37**(12): p. 4135-8.

88. Povlsen, K., J.S. Jensen, and I. Lind, *Detection of Ureaplasma urealyticum by PCR and biovar determination by liquid hybridization.* J Clin Microbiol, 1998. **36**(11): p. 3211-6.

89. Kong, F., et al., *Molecular genotyping of human Ureaplasma species based on multiple-banded antigen (MBA) gene sequences.* Int J Syst Evol Microbiol, 2000. **50 Pt 5**: p. 1921-9.

90. Kong, F., et al., *Comparative analysis and serovar-specific identification of multiple-banded antigen genes of Ureaplasma urealyticum biovar 1.* J Clin Microbiol, 1999. **37**(3): p. 538-43.

91. Teng, K., et al., *Comparison of PCR with culture for detection of Ureaplasma urealyticum in clinical samples from patients with urogenital infections.* J Clin Microbiol, 1994. **32**(9): p. 2232-4.

92. Teng, L.J., et al., *Rapid detection and biovar differentiation of Ureaplasma urealyticum in clinical specimens by PCR.* J Formos Med Assoc, 1995. **94**(7): p. 396-400.

93. Cao, X., et al., *Real-time TaqMan polymerase chain reaction assays for quantitative detection and differentiation of Ureaplasma urealyticum and Ureaplasma parvum.* Diagn Microbiol Infect Dis, 2007. **57**(4): p. 373-8.

94. Yi, J., B.H. Yoon, and E.C. Kim, *Detection and biovar discrimination of Ureaplasma urealyticum by real-time PCR.* Mol Cell Probes, 2005. **19**(4): p. 255-60.

95. Mallard, K., K. Schopfer, and T. Bodmer, *Development of real-time PCR for the differential detection and quantification of Ureaplasma urealyticum and Ureaplasma parvum.* J Microbiol Methods, 2005. **60**(1): p. 13-9.

96. Cao, X., et al., *Two multiplex real-time TaqMan polymerase chain reaction systems for simultaneous detecting and serotyping of Ureaplasma parvum.* Diagn Microbiol Infect Dis, 2007. **59**(1): p. 109-11.

97. Xiao, L., et al., *Detection and characterization of human Ureaplasma species and serovars by real-time PCR.* submitted for publication, 2009.

98. Knox, C.L., P. Giffard, and P. Timms, *The phylogeny of Ureaplasma urealyticum based on the mba gene fragment.* Int J Syst Bacteriol, 1998. **48 Pt 4**: p. 1323-31.

99. Zimmerman, C.U., et al., *Alternate phase variation in expression of two major surface membrane proteins (MBA and UU376) of Ureaplasma parvum serovar 3.* FEMS Microbiol Lett, 2009. **292**(2): p. 187-93.

100. Chun, J.Y., et al., *Dual priming oligonucleotide system for the multiplex detection of respiratory viruses and SNP genotyping of CYP2C19 gene.* Nucleic Acids Res, 2007. **35**(6): p. e40.

101. Kim, T.-H., et al., *Detection of cryptic microoorgganisms in patients with chronic prostatitis by multoplex polymerase chain reaction.* The Korean Journal of Urology, 2007. **3**: p. 304-9.

102. Choe, H.S., et al., *Performance of Anyplex II multiplex real-time PCR for the diagnosis of seven sexually transmitted infections: comparison with currently available methods.* Int J Infect Dis, 2013. **17**(12): p. e1134-40.

103. Payne, M.S., et al., *High-Resolution Melt PCR Analysis for Genotyping of Ureaplasma parvum Isolates Directly from Clinical Samples.* J Clin Microbiol, 2013. **52**(2): p. 599-606.

104. Xiao, L., et al., *Extensive horizontal gene transfer in ureaplasmas from humans questions the utility of serotyping for diagnostic purposes.* J Clin Microbiol, 2011. **49**(8): p. 2818-26.

105. Yao, F., et al., *Establishment of multiplex PCR for simultaneous detection of four venereal pathogens.* Biomed Environ Sci, 2013. **26**(7): p. 622-4.

106. McKechnie, M.L., et al., *Simultaneous identification of 14 genital microorganisms in urine by use of a multiplex PCR-based reverse line blot assay.* J Clin Microbiol, 2009. **47**(6): p. 1871-7.

107. Cao, X., et al., *Visual DNA microarrays for simultaneous detection of Ureaplasma urealyticum and Chlamydia trachomatis coupled with multiplex asymmetrical PCR.* Biosens Bioelectron, 2006. **22**(3): p. 393-8.

108. Kong, F. and G.L. Gilbert, *Multiplex PCR-based reverse line blot hybridization assay (mPCR/RLB)--a practical epidemiological and diagnostic tool.* Nat Protoc, 2006. **1**(6): p. 2668-80.

109. Mori, Y. and T. Notomi, *Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases.* J Infect Chemother, 2009. **15**(2): p. 62-9.

110. Wiedmann, M., et al., *Ligase chain reaction (LCR)--overview and applications.* PCR Methods Appl, 1994. **3**(4): p. S51-64.

111. Deiman, B., P. van Aarle, and P. Sillekens, *Characteristics and applications of nucleic acid sequence-based amplification (NASBA).* Mol Biotechnol, 2002. **20**(2): p. 163-79.

112. Loens, K., et al., *Development of real-time multiplex nucleic acid sequence-based amplification for detection of Mycoplasma pneumoniae, Chlamydophila pneumoniae, and Legionella spp. in respiratory specimens.* J Clin Microbiol, 2008. **46**(1): p. 185-91.

113. Loens, K., et al., *Detection of Mycoplasma pneumoniae by real-time nucleic acid sequence-based amplification.* J Clin Microbiol, 2003. **41**(9): p. 4448-50.

114. Barken, K.B., J.A. Haagensen, and T. Tolker-Nielsen, *Advances in nucleic acid-based diagnostics of bacterial infections.* Clin Chim Acta, 2007. **384**(1-2): p. 1-11.

115. Benstein, B.D., et al., *Ureaplasma in lung. 1. Localization by in situ hybridization in a mouse model.* Exp Mol Pathol, 2003. **75**(2): p. 165-70.

116. Notomi, T., et al., *Loop-mediated isothermal amplification of DNA.* Nucleic Acids Res, 2000. **28**(12): p. E63.

117. Jean, J., D.H. D'Souza, and L.A. Jaykus, *Multiplex nucleic acid sequence-based amplification for simultaneous detection of several enteric viruses in model ready-to-eat foods.* Appl Environ Microbiol, 2004. **70**(11): p. 6603-10.

118. Weusten, J.J., et al., *Principles of quantitation of viral loads using nucleic acid sequence-based amplification in combination with homogeneous detection using molecular beacons.* Nucleic Acids Res, 2002. **30**(6): p. e26.

119. Chen, Q., et al., *Real-time monitoring of the strand displacement amplification (SDA) of human cytomegalovirus by a new SDA-piezoelectric DNA sensor system.* Biosens Bioelectron, 2009. **24**(12): p. 3412-8.

120. Walker, G.T., et al., *Multiplex strand displacement amplification (SDA) and detection of DNA sequences from Mycobacterium tuberculosis and other mycobacteria.* Nucleic Acids Res, 1994. **22**(13): p. 2670-7.

121. Famuyide, M.E., et al., *Surfactant protein-A limits Ureaplasma-mediated lung inflammation in a murine pneumonia model.* Pediatr Res, 2009. **66**(2): p. 162-7.

122. Viscardi, R.M. and J.D. Hasday, *Role of Ureaplasma species in neonatal chronic lung disease: epidemiologic and experimental evidence.* Pediatr Res, 2009. **65**(5 Pt 2): p. 84R-90R.

123. Shimizu, T., Y. Kida, and K. Kuwano, *Ureaplasma parvum lipoproteins, including MB antigen, activate NF-{kappa}B through TLR1, TLR2 and TLR6.* Microbiology, 2008. **154**(Pt 5): p. 1318-25.

124. Peltier, M.R., et al., *Characterization of the macrophage-stimulating activity from Ureaplasma urealyticum.* Am J Reprod Immunol, 2007. **57**(3): p. 186-92.

125. Winters, A.H., et al., *Single nucleotide polymorphism in toll-like receptor 6 is associated with a decreased risk for ureaplasma respiratory tract colonization and bronchopulmonary dysplasia in preterm infants.* Pediatr Infect Dis J, 2013. **32**(8): p. 898-904.

126. Kallapur, S.G., et al., *Chronic fetal exposure to Ureaplasma parvum suppresses innate immune responses in sheep.* J Immunol, 2011. **187**(5): p. 2688-95.

127. Dando, S.J., et al., *The role of the multiple banded antigen of Ureaplasma parvum in intra-amniotic infection: major virulence factor or decoy?* PLoS One, 2012. **7**(1): p. e29856.

128. Knox, C.L., et al., *The severity of chorioamnionitis in pregnant sheep is associated with in vivo variation of the surface-exposed multiple-banded antigen/gene of Ureaplasma parvum.* Biol Reprod. **83**(3): p. 415-26.

129. Grigsby, P.L., et al., *Maternal azithromycin therapy for Ureaplasma intraamniotic infection delays preterm delivery and reduces fetal lung injury in a primate model.* Am J Obstet Gynecol, 2012. **207**(6): p. 475 e1-475 e14.

130. Acosta, E.P., et al., *Transplacental Transfer of Azithromycin and Its Use for Eradicating Intra-amniotic Ureaplasma Infection in a Primate Model.* J Infect Dis, 2013.

131. Keelan, J.A., et al., *Maternal administration of solithromycin, a new, potent, broad-spectrum fluoroketolide antibiotic, achieves fetal and intra-amniotic*

antimicrobial protection in a pregnant sheep model. Antimicrob Agents Chemother, 2014. **58**(1): p. 447-54.

132. Xiao, L., et al., *Detection and characterization of human Ureaplasma species and serovars by real-time PCR.* submitted for publication, 2009.

133. Xiao, L., et al., *Extensive horizontal gene transfer in human ureaplasmas questions the utility of serotyping for diagnostic purposes [abstract]*, in *18th Congress of the International Organization for Mycoplasmology*. 2010: Chianciano Terme, Italy.

134. Momynaliev, K., et al., *Comparative genome analysis of Ureaplasma parvum clinical isolates.* Res Microbiol, 2007. **158**(4): p. 371-8.

135. Kong, F., et al., *Phylogenetic analysis of Ureaplasma urealyticum--support for the establishment of a new species, Ureaplasma parvum.* Int J Syst Bacteriol, 1999. **49 Pt 4**: p. 1879-89.

136. Robertson, J.A., et al., *Comparison of 16S rRNA genes within the T960 and parvo biovars of ureaplasmas isolated from humans.* Int J Syst Bacteriol, 1994. **44**(4): p. 836-8.

137. Delcher, A.L., et al., *Improved microbial gene identification with GLIMMER.* Nucleic Acids Res, 1999. **27**(23): p. 4636-41.

138. Griffiths-Jones, S., et al., *Rfam: an RNA family database.* Nucleic Acids Res, 2003. **31**(1): p. 439-41.

139. Lowe, T.M. and S.R. Eddy, *tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.* Nucleic Acids Res, 1997. **25**(5): p. 955-64.

140. Laslett, D. and B. Canback, *ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences.* Nucleic Acids Res, 2004. **32**(1): p. 11-6.

141. Selengut, J.D., et al., *TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes.* Nucleic Acids Res, 2007. **35**(Database issue): p. D260-4.

142. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

143. Haft, D.H., et al., *Genome Properties: a system for the investigation of prokaryotic genetic content for microbiology, genome annotation and comparative genomics.* Bioinformatics, 2005. **21**(3): p. 293-306.

144. Crabtree, J., et al., *Sybil: methods and software for multiple genome comparison and visualization.* Methods Mol Biol, 2007. **408**: p. 93-108.

145. Laing, C., et al., *Pan-genome sequence analysis using Panseq: an online tool for the rapid analysis of core and accessory genomic regions.* BMC Bioinformatics. **11**: p. 461.

146. manatee.sourceforge.net

147. Gelfand, Y., A. Rodriguez, and G. Benson, *TRDB--the Tandem Repeats Database.* Nucleic Acids Res, 2007. **35**(Database issue): p. D80-7.

148. Zhang, Z., et al., *A greedy algorithm for aligning DNA sequences.* J Comput Biol, 2000. **7**(1-2): p. 203-14.

149. Sonnhammer, E.L. and R. Durbin, *A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis.* Gene, 1995. **167**(1-2): p. GC1-10.

150. Altschul, S.F., et al., *Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.* Nucleic Acids Res, 1997. **25**(17): p. 3389-402.

151. Wu, J. and J. Xie, *Hidden Markov model and its applications in motif findings.* Methods Mol Biol, 2010. **620**(2010): p. 405-16.

152. Finn, R.D., et al., *The Pfam protein families database.* Nucleic Acids Res, 2010. **38**(Database issue): p. D211-22.

153. DeSilva, N.S. and P.A. Quinn, *Characterization of phospholipase A1, A2, C activity in Ureaplasma urealyticum membranes.* Mol Cell Biochem, 1999. **201**(1-2): p. 159-67.

154. De Silva, N.S. and P.A. Quinn, *Localization of endogenous activity of phospholipases A and C in Ureaplasma urealyticum.* J Clin Microbiol, 1991. **29**(7): p. 1498-503.

155. De Silva, N.S. and P.A. Quinn, *Endogenous activity of phospholipases A and C in Ureaplasma urealyticum.* J Clin Microbiol, 1986. **23**(2): p. 354-9.

156. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.

157. Huson, D.H. and C. Scornavacca, *Dendroscope 3 - An interactive viewer for rooted phylogenetic trees and networks.* In preparation.

158. Waterhouse, A.M., et al., *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-91.

159. Harasawa, R. and G.H. Cassell, *Phylogenetic analysis of genes coding for 16S rRNA in mammalian ureaplasmas.* Int J Syst Bacteriol, 1996. **46**(3): p. 827-9.

160. Maniloff, J., *Phylogeny and Evolution*, in *Molecular Biology and Pathogenicity of Mycoplasmas*, S. Razin and R. Herrmann, Editors. 2002, Kluwer Academic: New York. p. 41.

161. Wang, H. and P. Mullany, *The large resolvase TndX is required and sufficient for integration and excision of derivatives of the novel conjugative transposon Tn5397.* J Bacteriol, 2000. **182**(23): p. 6577-83.

162. Dougherty, B.A., et al., *Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasmid pMRC01 from Lactococcus lactis DPC3147.* Mol Microbiol, 1998. **29**(4): p. 1029-38.

163. Schroder, G., et al., *TraG-like proteins of DNA transfer systems and of the Helicobacter pylori type IV secretion system: inner membrane gate for exported substrates?* J Bacteriol, 2002. **184**(10): p. 2767-79.

164. Byrd, D.R. and S.W. Matson, *Nicking by transesterification: the reaction catalysed by a relaxase.* Mol Microbiol, 1997. **25**(6): p. 1011-22.

165. Read, T.D., et al., *Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.* Nucleic acids research, 2000. **28**(6): p. 1397-406.

166. Kater, L.A., E.J. Goetzl, and K.F. Austen, *Isolation of human eosinophil phospholipase D.* J Clin Invest, 1976. **57**(5): p. 1173-80.

167. Marques, L.M., et al., *Invasion of Ureaplasma diversum in Hep-2 cells.* BMC Microbiol. **10**: p. 83.

168. Fliegera, A., et al., *Critical evaluation of p-nitrophenylphosphorylcholine (p-NPPC) as artificial substrate for the detection of phospholipase C*.* Enzyme Microb Technol, 2000. **26**(5-6): p. 451-458.

169. Park, D.W., et al., *Regulation of cyclooxygenase-2 expression by phospholipase D in human amnion-derived WISH cells.* Mol Pharmacol, 2002. **61**(3): p. 614-9.

170. Lin, L., et al., *The Neisseria type 2 IgA1 protease cleaves LAMP1 and promotes survival of bacteria within epithelial cells.* Mol Microbiol, 1997. **24**(5): p. 1083-94.

171. Kilian, M., et al., *Immunoglobulin A1 protease activity in strains of Ureaplasma urealyticum.* Acta Pathol Microbiol Immunol Scand B, 1984. **92**(1): p. 61-4.

172. Kilian, M. and E.A. Freundt, *Exclusive occurrence of an extracellular protease capable of cleaving the hinge region of human immunoglobulin A1 in strains of Ureaplasma urealyticum.* Isr J Med Sci, 1984. **20**(10): p. 938-41.

173. Punta, M., et al., *The Pfam protein families database.* Nucleic Acids Res, 2012. **40**(Database issue): p. D290-301.

174. Somarajan, S.R., T.R. Kannan, and J.B. Baseman, *Mycoplasma pneumoniae Mpn133 is a cytotoxic nuclease with a glutamic acid-, lysine- and serine-rich*

region essential for binding and internalization but not enzymatic activity. Cell Microbiol. **12**(12): p. 1821-31.

175. Abdullah, K.M., et al., *A neutral glycoprotease of Pasteurella haemolytica A1 specifically cleaves O-sialoglycoproteins.* Infect Immun, 1992. **60**(1): p. 56-62.

176. Mehta, P.K., et al., *Identification of Mycobacterium marinum macrophage infection mutants.* Microb Pathog, 2006. **40**(4): p. 139-51.

177. Falagas, M.E., G.I. Betsi, and S. Athanasiou, *Probiotics for the treatment of women with bacterial vaginosis.* Clin Microbiol Infect, 2007. **13**(7): p. 657-64.

178. Tsang, P., et al., *Identification of genes associated with mutacin I production in Streptococcus mutans using random insertional mutagenesis.* Microbiology, 2005. **151**(Pt 12): p. 3947-55.

179. Ben-Menachem, G., et al., *The thioredoxin reductase system of mycoplasmas.* Microbiology, 1997. **143 (Pt 6)**: p. 1933-40.

180. Zheng, X., et al., *Serotype diversity and antigen variation among invasive isolates of Ureaplasma urealyticum from neonates.* Infect Immun, 1992. **60**(8): p. 3472-4.

181. Zheng, X., et al., *Small repeating units within the Ureaplasma urealyticum MB antigen gene encode serovar specificity and are associated with antigen size variation.* Infect Immun, 1995. **63**(3): p. 891-8.

182. Zheng, X., et al., *Epitope mapping of the variable repetitive region with the MB antigen of Ureaplasma urealyticum.* Clin Diagn Lab Immunol, 1996. **3**(6): p. 774-8.

183. Monecke, S., J.H. Helbig, and E. Jacobs, *Phase variation of the multiple banded protein in Ureaplasma urealyticum and Ureaplasma parvum.* Int J Med Microbiol, 2003. **293**(2-3): p. 203-11.

184. Zimmerman, C.U., R. Rosengarten, and J. Spergser, *Ureaplasma antigenic variation beyond MBA phase variation: DNA inversions generating chimeric*

structures and switching in expression of the MBA N-terminal paralogue UU172. Mol Microbiol, 2011. **79**(3): p. 663-76.

185. Ron, Y., et al., *Identification and characterization of a site-specific tyrosine recombinase within the variable loci of Mycoplasma bovis, Mycoplasma pulmonis and Mycoplasma agalactiae.* Gene, 2002. **292**(1-2): p. 205-11.

186. Robertson, J.A. and G.W. Stemke, *Modified metabolic inhibition test for serotyping strains of Ureaplasma urealyticum (T-strain Mycoplasma).* J Clin Microbiol, 1979. **9**(6): p. 673-6.

187. Smith, D.G., W.C. Russell, and D. Thirkell, *Adherence of Ureaplasma urealyticum to human epithelial cells.* Microbiology, 1994. **140 (Pt 10)**: p. 2893-8.

188. Robertson, J.A., M.E. Stemler, and G.W. Stemke, *Immunoglobulin A protease activity of Ureaplasma urealyticum.* J Clin Microbiol, 1984. **19**(2): p. 255-8.

189. Deguchi, T., et al., *Association of Ureaplasma urealyticum (biovar 2) with nongonococcal urethritis.* Sexually transmitted diseases, 2004. **31**(3): p. 192-5.

190. Povlsen, K., et al., *Relationship of Ureaplasma urealyticum biovar 2 to nongonococcal urethritis.* European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2002. **21**(2): p. 97-101.

191. Maeda, S., et al., *Detection of Mycoplasma genitalium, Mycoplasma hominis, Ureaplasma parvum (biovar 1) and Ureaplasma urealyticum (biovar 2) in patients with non-gonococcal urethritis using polymerase chain reaction-microtiter plate hybridization.* International journal of urology : official journal of the Japanese Urological Association, 2004. **11**(9): p. 750-4.

192. Ondondo, R.O., et al., *Differential association of ureaplasma species with non-gonococcal urethritis in heterosexual men.* Sexually transmitted infections, 2010. **86**(4): p. 271-5.

193. Abele-Horn, M., et al., *Association of Ureaplasma urealyticum biovars with clinical outcome for neonates, obstetric patients, and gynecological patients*

with pelvic inflammatory disease. Journal of clinical microbiology, 1997. **35**(5): p. 1199-202.

194. Povlsen, K., P. Thorsen, and I. Lind, *Relationship of Ureaplasma urealyticum biovars to the presence or absence of bacterial vaginosis in pregnant women and to the time of delivery.* European journal of clinical microbiology & infectious diseases : official publication of the European Society of Clinical Microbiology, 2001. **20**(1): p. 65-7.

195. Shepard, M.C., et al., *Ureaplasma urealyticum gen. nov. sp. nov.: proposed nomenclature for the human T (T-strain) mycoplasmas.* International Journal of Systematic Bacteriology, 1974. **24**: p. 160-171.

196. Shepard, M.C. and C.D. Lunceford, *Serological typing of Ureaplasma urealyticum isolates from urethritis patients by an agar growth inhibition method.* J Clin Microbiol, 1978. **8**(5): p. 566-74.

197. Black, F.T., *Modifications of the growth inhibition test and its application to human T-mycoplasmas.* Appl Microbiol, 1973. **25**(4): p. 528-33.

198. Rosendal, S. and F.T. Black, *Direct and indirect immunofluorescence of unfixed and fixed Mycoplasma colonies.* Acta Pathol Microbiol Scand 1972. **80**: p. 615-622.

199. Stemke, G.W. and J.A. Robertson, *Modified colony indirect epifluorescence test for serotyping Ureaplasma urealyticum and an adaptation to detect common antigenic specificity.* J Clin Microbiol, 1981. **14**(5): p. 582-4.

200. Echahidi, F., et al., *Development of monoclonal antibodies against Ureaplasma urealyticum serotypes and their use for serotyping clinical isolates.* Clin Diagn Lab Immunol, 2000. **7**(4): p. 563-7.

201. Quinn, P.A., L.U. Arshoff, and H.C. Li, *Serotyping of Ureaplasma urealyticum by immunoperoxidase assay.* J Clin Microbiol, 1981. **13**(4): p. 670-6.

202. Watson, H.L., D.K. Blalock, and G.H. Cassell, *Variable antigens of Ureaplasma urealyticum containing both serovar-specific and serovar-cross-reactive epitopes.* Infect Immun, 1990. **58**(11): p. 3679-88.

203. Yoshida, T., et al., *Quantitative detection of Ureaplasma parvum (biovar 1) and Ureaplasma urealyticum (biovar 2) in urine specimens from men with and without urethritis by real-time polymerase chain reaction.* Sex Transm Dis, 2007. **34**(6): p. 416-9.

204. Teng, L.J., et al., *Ureaplasma urealyticum biovar specificity and diversity are encoded in multiple-banded antigen gene.* J Clin Microbiol, 1994. **32**(6): p. 1464-9.

205. Yoshida, T., et al., *Polymerase chain reaction-based subtyping of ureaplasma parvum and ureaplasma urealyticum in first-pass urine samples from men with or without urethritis.* Sex Transm Dis, 2005. **32**(7): p. 454-7.

206. Waites, K.B., et al., *Mycoplasma and Ureaplasma,* in *Clinical Microbiology Procedure Handbook*, H. Isenberg, Editor. 2004, American Society for Microbiology Press: Washington, D.C.

207. Blanchard, A., et al., *Detection of Ureaplasma urealyticum by polymerase chain reaction in the urogenital tract of adults, in amniotic fluid, and in the respiratory tract of newborns.* Clin Infect Dis, 1993. **17 Suppl 1**: p. S148-53.

208. Katz, B., et al., *Characterization of ureaplasmas isolated from preterm infants with and without bronchopulmonary dysplasia.* J Clin Microbiol, 2005. **43**(9): p. 4852-4.

209. Abele-Horn, M., et al., *Polymerase chain reaction versus culture for detection of Ureaplasma urealyticum and Mycoplasma hominis in the urogenital tract of adults and the respiratory tract of newborns.* Eur J Clin Microbiol Infect Dis, 1996. **15**(7): p. 595-8.

210. Luki, N., et al., *Comparison of polymerase chain reaction assay with culture for detection of genital mycoplasmas in perinatal infections.* Eur J Clin Microbiol Infect Dis, 1998. **17**(4): p. 255-63.

211. Yoon, B.H., et al., *The clinical significance of detecting Ureaplasma urealyticum by the polymerase chain reaction in the amniotic fluid of patients with preterm labor.* Am J Obstet Gynecol, 2003. **189**(4): p. 919-24.

212. Knox, C.L. and P. Timms, *Comparison of PCR, nested PCR, and random amplified polymorphic DNA PCR for detection and typing of Ureaplasma*

urealyticum in specimens from pregnant women. J Clin Microbiol, 1998. **36**(10): p. 3032-9.

213. Pitcher, D., M. Sillis, and J.A. Robertson, *Simple method for determining biovar and serovar types of Ureaplasma urealyticum clinical isolates using PCR-single-strand conformation polymorphism analysis.* J Clin Microbiol, 2001. **39**(5): p. 1840-4.

214. *Invasive pneumococcal disease in children 5 years after conjugate vaccine introduction--eight states, 1998-2005.* MMWR Morb Mortal Wkly Rep, 2008. **57**(6): p. 144-8.

215. Dworkin, M.S., L. Park, and S.M. Borchardt, *The changing epidemiology of invasive Haemophilus influenzae disease, especially in persons > or = 65 years old.* Clin Infect Dis, 2007. **44**(6): p. 810-6.

216. Povlsen, K., et al., *Relationship of Ureaplasma urealyticum biovar 2 to nongonococcal urethritis.* Eur J Clin Microbiol Infect Dis, 2002. **21**(2): p. 97-101.

217. Maeda, S., et al., *Detection of Mycoplasma genitalium, Mycoplasma hominis, Ureaplasma parvum (biovar 1) and Ureaplasma urealyticum (biovar 2) in patients with non-gonococcal urethritis using polymerase chain reaction-microtiter plate hybridization.* Int J Urol, 2004. **11**(9): p. 750-4.

218. Ondondo, R.O., et al., *Differential association of ureaplasma species with non-gonococcal urethritis in heterosexual men.* Sex Transm Infect. **86**(4): p. 271-5.

219. Abele-Horn, M., et al., *Association of Ureaplasma urealyticum biovars with clinical outcome for neonates, obstetric patients, and gynecological patients with pelvic inflammatory disease.* J Clin Microbiol, 1997. **35**(5): p. 1199-202.

220. Povlsen, K., P. Thorsen, and I. Lind, *Relationship of Ureaplasma urealyticum biovars to the presence or absence of bacterial vaginosis in pregnant women and to the time of delivery.* Eur J Clin Microbiol Infect Dis, 2001. **20**(1): p. 65-7.

221. Bradshaw, C.S., et al., *Etiologies of nongonococcal urethritis: bacteria, viruses, and the association with orogenital exposure.* J Infect Dis, 2006. **193**(3): p. 336-45.

222. Stemke, G.W. and J.A. Robertson, *Problems associated with serotyping strains of Ureaplasma urealyticum.* Diagn Microbiol Infect Dis, 1985. **3**(4): p. 311-20.

223. Cracea, E., S. Constantinescu, and M. Lazar, *Serotypes of Ureaplasma urealyticum isolated from patients with nongonococcal urethritis and gonorrhea and from asymptomatic urethral carriers.* Sex Transm Dis, 1985. **12**(4): p. 219-23.

224. Naessens, A., et al., *Serotypes of Ureaplasma urealyticum isolated from normal pregnant women and patients with pregnancy complications.* J Clin Microbiol, 1988. **26**(2): p. 319-22.

225. De Francesco, M.A., et al., *Detection of Ureaplasma biovars and polymerase chain reaction-based subtyping of Ureaplasma parvum in women with or without symptoms of genital infections.* Eur J Clin Microbiol Infect Dis, 2009. **28**(6): p. 641-6.

226. Piot, P., *Comparison of growth inhibition and immunofluorescence tests in serotyping clinical isolates of Ureaplasma urealyticum.* Br J Vener Dis, 1977. **53**(3): p. 186-9.

227. Horowitz, S.A., et al., *Can group- and serovar-specific proteins be detected in Ureaplasma urealyticum?* Pediatr Infect Dis, 1986. **5**(6 Suppl): p. S325-31.

228. Echahidi, F., et al., *Development of an enzyme-linked immunosorbent assay for serotyping ureaplasma urealyticum strains using monoclonal antibodies.* Clin Diagn Lab Immunol, 2001. **8**(1): p. 52-7.

229. Rosendal, S. and F.T. Black, *Direct and indirect immunofluorescence of unfixed and fixed Mycoplasma colonies.* Acta Pathol Microbiol Scand B Microbiol Immunol, 1972. **80**(4): p. 615-22.

230. Xiao, L., et al., *Detection and characterization of human Ureaplasma species and serovars by real-time PCR.* J Clin Microbiol, 2010. **48**(8): p. 2715-23.

231. Smets, B.F. and T. Barkay, *Horizontal gene transfer: perspectives at a crossroads of scientific disciplines.* Nat Rev Microbiol, 2005. **3**(9): p. 675-8.

232. Woese, C.R., J. Maniloff, and L.B. Zablen, *Phylogenetic analysis of the mycoplasmas.* Proc Natl Acad Sci U S A, 1980. **77**(1): p. 494-8.

233. Sirand-Pugnet, P., et al., *Being pathogenic, plastic, and sexual while living with a nearly minimal bacterial genome.* PLoS Genet, 2007. **3**(5): p. e75.

234. Teachman, A.M., et al., *Gene transfer in Mycoplasma pulmonis.* J Bacteriol, 2002. **184**(4): p. 947-51.

235. Pereyre, S., et al., *Life on arginine for Mycoplasma hominis: clues from its minimal genome and comparison with other human urogenital mycoplasmas.* PLoS Genet, 2009. **5**(10): p. e1000677.

236. Garcia-Castillo, M., et al., *Differences in biofilm development and antibiotic susceptibility among clinical Ureaplasma urealyticum and Ureaplasma parvum isolates.* J Antimicrob Chemother, 2008. **62**(5): p. 1027-30.

237. Paralanov, V., et al., *Comparative Genome Analysis of all 14 Ureaplasma parvum and Ureaplasma urealyticum Serovars [abstract]*, in *18th Congress of the International Organization for Mycoplasmology*. 2010: Chianciano Terme, Italy.

238. Knox, C.L., et al., *Ureaplasma parvum and Ureaplasma urealyticum are detected in semen after washing before assisted reproductive technology procedures.* Fertil Steril, 2003. **80**(4): p. 921-9.

239. de Barbeyrac, B., et al., *A Tn1545-like transposon carries the tet(M) gene in tetracycline resistant strains of Bacteroides ureolyticus as well as Ureaplasma urealyticum but not Neisseria gonorrhoeae.* J Antimicrob Chemother, 1996. **37**(2): p. 223-32.

240. Roberts, M.C. and G.E. Kenny, *TetM tetracycline-resistant determinants in Ureaplasma urealyticum.* Pediatr Infect Dis, 1986. **5**(6 Suppl): p. S338-40.

241. Lu, C., et al., *Phenotypic and Genetic Characteristics of Macrolide and Lincosamide Resistant Ureaplasma urealyticum Isolated in Guangzhou, China.* Curr Microbiol, 2009.

242. Molin, S. and T. Tolker-Nielsen, *Gene transfer occurs with enhanced efficiency in biofilms and induces enhanced stabilisation of the biofilm structure.* Curr Opin Biotechnol, 2003. **14**(3): p. 255-61.

243. Lin, J.S. and E.H. Kass, *Serotypic heterogeneity in isolates of human genital T-mycoplasmas.* Infect Immun, 1973. **7**(3): p. 499-500.

244. Bowie, W.R., et al., *Etiology of nongonococcal urethritis. Evidence for Chlamydia trachomatis and Ureaplasma urealyticum.* J Clin Invest, 1977. **59**(5): p. 735-42.

245. Brunner, H., W. Weidner, and H.G. Schiefer, *Quantitative studies on the role of Ureaplasma urealyticum in non-gonococcal urethritis and chronic prostatitis.* Yale J Biol Med, 1983. **56**(5-6): p. 545-50.

246. Reyes, L., M. Reinhard, and M.B. Brown, *Different inflammatory responses are associated with Ureaplasma parvum-induced UTI and urolith formation.* BMC Infect Dis, 2009. **9**: p. 9.

247. Yoder, B.A., et al., *Effects of antenatal colonization with ureaplasma urealyticum on pulmonary disease in the immature baboon.* Pediatr Res, 2003. **54**(6): p. 797-807.

248. Beeton, M.L., et al., *Concurrent titration and determination of antibiotic resistance in ureaplasma species with identification of novel point mutations in genes associated with resistance.* Antimicrob Agents Chemother, 2009. **53**(5): p. 2020-7.

249. Lartigue, C., et al., *Creating bacterial strains from genomes that have been cloned and engineered in yeast.* Science, 2009. **325**(5948): p. 1693-6.

250. Glass, J.I., et al., *Essential genes of a minimal bacterium.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(2): p. 425-30.

251. Gibson, D.G., et al., *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nature methods, 2009. **6**(5): p. 343-5.

252. Minion, F.C., et al., *Membrane-associated nuclease activities in mycoplasmas.* J Bacteriol, 1993. **175**(24): p. 7842-7.

253. Dybvig, K., G.E. Gasparich, and K.W. King, *Artificial Transformation of Mollicutes via Polyethylene Glycol- and Electroporation-Mediated Methods,* in *Molecular and Diagnostic Procedures in Mycoplasmology, Volume 1: Molecular Characterization* S. Razin and J.G. Tully, Editors. 1995, Academic Press, Inc.: San Diego, CA.

254. Lartigue, C., et al., *Genome transplantation in bacteria: changing one species to another.* Science, 2007. **317**(5838): p. 632-8.

255. Gibson, D.G., et al., *Creation of a bacterial cell controlled by a chemically synthesized genome.* Science, 2010. **329**(5987): p. 52-6.

256. Atlas, R.M., *Hadbook of Microbiological Media*. 2 ed, ed. L.C. Parks. 1997, Boca Raton: CRC Press.

257. Dybvig, K., C.T. French, and L.L. Voelker, *Construction and use of derivatives of transposon Tn4001 that function in Mycoplasma pulmonis and Mycoplasma arthritidis.* J Bacteriol, 2000. **182**(15): p. 4343-7.

258. Lartigue, C., et al., *Host specificity of mollicutes oriC plasmids: functional analysis of replication origin.* Nucleic Acids Res, 2003. **31**(22): p. 6610-8.

259. Barroso, G. and J. Labarere, *Chromosomal gene transfer in Spiroplasma citri.* Science, 1988. **241**(4868): p. 959-61.

260. Mahairas, G.G., J.A. Cao, and F.C. Minion, *Genetic exchange of transposon and integrative plasmid markers in Mycoplasma pulmonis.* J Bacteriol, 1990. **172**(5): p. 2267-72.

261. Roberts, M.C. and G.E. Kenny, *Conjugal transfer of transposon Tn916 from Streptococcus faecalis to Mycoplasma hominis.* J Bacteriol, 1987. **169**(8): p. 3836-9.

262. Roberts, M.C. and G.E. Kenny, *Dissemination of the tetM tetracycline resistance determinant to Ureaplasma urealyticum.* Antimicrob Agents Chemother, 1986. **29**(2): p. 350-2.

263. Jahn, G., A.A. Bialasiewicz, and H. Blenk, *Evaluation of plasmids in tetracycline resistant strains of Neisseria gonorrhoeae and Ureaplasma urealyticum in a case of severe urethritis.* Eur J Epidemiol, 1985. **1**(4): p. 294-300.

264. Gasparich, G.E., et al., *Optimization of methods for transfecting Spiroplasma citri strain R8A2 HP with the spiroplasma virus SpV1 replicative form.* Plasmid, 1993. **29**(3): p. 193-205.

265. Clapper, B., et al., *Bacteriophage MAV1 is not associated with virulence of Mycoplasma arthritidis.* Infect Immun, 2004. **72**(12): p. 7322-5.

266. Dybvig, K., et al., *Isolation of a virus from Mycoplasma pulmonis.* Isr J Med Sci, 1987. **23**(5): p. 418-22.

267. Dybvig, K. and J. Maniloff, *Integration and lysogeny by an enveloped mycoplasma virus.* J Gen Virol, 1983. **64 (Pt 8)**: p. 1781-5.

268. Dybvig, K., et al., *Identification of an enveloped phage, mycoplasma virus L172, that contains a 14-kilobase single-stranded DNA genome.* J Virol, 1985. **53**(2): p. 384-90.

269. Dybvig, K., T.L. Sladek, and J. Maniloff, *Isolation of mycoplasma virus L2 insertion variants and miniviruses.* J Virol, 1986. **59**(3): p. 584-90.

270. Marais, A., J.M. Bove, and J. Renaudin, *Spiroplasma citri virus SpV1-derived cloning vector: deletion formation by illegitimate and homologous recombination in a spiroplasmal host strain which probably lacks a functional recA gene.* J Bacteriol, 1996. **178**(3): p. 862-70.

271. Voelker, L.L. and K. Dybvig, *Characterization of the lysogenic bacteriophage MAV1 from Mycoplasma arthritidis.* J Bacteriol, 1998. **180**(22): p. 5928-31.

272. Zou, N., K. Park, and K. Dybvig, *Mycoplasma virus P1 has a linear, double-stranded DNA genome with inverted terminal repeats.* Plasmid, 1995. **33**(1): p. 41-9.

273. Wagner, P.L. and M.K. Waldor, *Bactriophages in Bacterial Pathogenesis*, in *The bacteriophages*, R. Calendar, Editor. 2006, Oxford University Press: Oxford ; New York. p. xiii, 746 p.

274. Dybvig, K. and G.H. Cassell, *Transposition of gram-positive transposon Tn916 in Acholeplasma laidlawii and Mycoplasma pulmonis.* Science, 1987. **235**(4794): p. 1392-4.

275. Dybvig, K. and A. Woodard, *Construction of recA mutants of Acholeplasma laidlawii by insertional inactivation with a homologous DNA fragment.* Plasmid, 1992. **28**(3): p. 262-6.

276. Hahn, T.W., K.A. Krebes, and D.C. Krause, *Expression in Mycoplasma pneumoniae of the recombinant gene encoding the cytadherence-associated protein HMW1 and identification of HMW4 as a product.* Mol Microbiol, 1996. **19**(5): p. 1085-93.

277. Hedreyda, C.T., K.K. Lee, and D.C. Krause, *Transformation of Mycoplasma pneumoniae with Tn4001 by electroporation.* Plasmid, 1993. **30**(2): p. 170-5.

278. Knudtson, K.L. and F.C. Minion, *Construction of Tn4001lac derivatives to be used as promoter probe vectors in mycoplasmas.* Gene, 1993. **137**(2): p. 217-22.

279. Mahairas, G.G. and F.C. Minion, *Random insertion of the gentamicin resistance transposon Tn4001 in Mycoplasma pulmonis.* Plasmid, 1989. **21**(1): p. 43-7.

280. Minion, F.C. and P.A. Kapke, *Transformation of mycoplasmas.* Methods Mol Biol, 1998. **104**: p. 227-34.

281. Renaudin, J., et al., *Integrative and free Spiroplasma citri oriC plasmids: expression of the Spiroplasma phoeniceum spiralin in Spiroplasma citri.* J Bacteriol, 1995. **177**(10): p. 2870-7.

282. Allam, A.B., et al., *Enhancement of targeted homologous recombination in Mycoplasma mycoides subsp. capri by inclusion of heterologous recA.* Appl Environ Microbiol, 2010. **76**(20): p. 6951-4.

283. Halbedel, S. and J. Stulke, *Tools for the genetic analysis of Mycoplasma.* Int J Med Microbiol, 2007. **297**(1): p. 37-44.

284. Cordova, C.M., et al., *Identification of the origin of replication of the Mycoplasma pulmonis chromosome and its use in oriC replicative plasmids.* J Bacteriol, 2002. **184**(19): p. 5426-35.

285. Lee, S.W., G.F. Browning, and P.F. Markham, *Development of a replicable oriC plasmid for Mycoplasma gallisepticum and Mycoplasma imitans, and gene disruption through homologous recombination in M. gallisepticum.* Microbiology, 2008. **154**(Pt 9): p. 2571-80.

286. Benders, G.A., et al., *Cloning whole bacterial genomes in yeast.* Nucleic Acids Res, 2010. **38**(8): p. 2558-69.

287. Noskov, V.N., T.H. Segall-Shapiro, and R.Y. Chuang, *Tandem repeat coupled with endonuclease cleavage (TREC): a seamless modification tool for genome engineering in yeast.* Nucleic Acids Res, 2010. **38**(8): p. 2570-6.

288. Bernander, R., T. Akerlund, and K. Nordstrom, *Inhibition and restart of initiation of chromosome replication: effects on exponentially growing Escherichia coli cells.* J Bacteriol, 1995. **177**(7): p. 1670-82.

289. Drakulic, M. and M. Errera, *Chloramphenicol-sensitive DNA synthesis in normal and irradiated bacteria.* Biochim Biophys Acta, 1959. **31**(2): p. 459-63.

290. Skarstad, K., E. Boye, and H.B. Steen, *Timing of initiation of chromosome replication in individual Escherichia coli cells.* EMBO J, 1986. **5**(7): p. 1711-7.

291. Murphy, L.D. and S.B. Zimmerman, *A limited loss of DNA compaction accompanying the release of cytoplasm from cells of Escherichia coli.* J Struct Biol, 2001. **133**(1): p. 75-86.

292. Seto, S. and M. Miyata, *Partitioning, movement, and positioning of nucleoids in Mycoplasma capricolum.* J Bacteriol, 1999. **181**(19): p. 6073-80.

293. Gibson, D.G., et al., *Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome.* Science, 2008. **319**(5867): p. 1215-20.

294. Gibson, D.G., et al., *One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20404-9.

295. Glass, J.I., *Synthetic genomics and the construction of a synthetic bacterial cell.* Perspect Biol Med, 2012. **55**(4): p. 473-89.

296. Wakaki, S., et al., *Isolation of new fractions of antitumor mitomycins.* Antibiot Chemother (Northfield Ill), 1958. **8**(5): p. 228-40.

297. Philips, F.S., H.S. Schwartz, and S.S. Sternberg, *Pharmacology of mitomycin C. I. Toxicity and pathologic effects.* Cancer Res, 1960. **20**: p. 1354-61.

298. Schwartz, H.S., S.S. Sternberg, and F.S. Philips, *Pharmacology of Mitomycin C. Iv. Effects in Vivo on Nucleic Acid Synthesis; Comparison with Actinomycin D.* Cancer Res, 1963. **23**: p. 1125-36.

299. Crooke, S.T. and W.T. Bradner, *Mitomycin C: a review.* Cancer Treat Rev, 1976. **3**(3): p. 121-39.

300. Spanswick, V.J., J. Cummings, and J.F. Smyth, *Current issues in the enzymology of mitomycin C metabolic activation.* Gen Pharmacol, 1998. **31**(4): p. 539-44.

301. Tomasz, M., *Mitomycin C: small, fast and deadly (but very selective).* Chem Biol, 1995. **2**(9): p. 575-9.

302. Vig, B.K., *Genetic toxicology of mitomycin C, actinomycins, daunomycin and adriamycin.* Mutat Res, 1977. **39**(2): p. 189-238.

303. Shatkin, A.J., *Actinomycin inhibition of ribonucleic acid synthesis and poliovirus infection of HeLa cells.* Biochim Biophys Acta, 1962. **61**: p. 310-3.

304. Bach, M.K. and W.E. Magee, *Biochemical effects of isatin beta-thiosemicarbazone on development of vaccinia virus.* Proc Soc Exp Biol Med, 1962. **110**: p. 565-7.

305. Waring, M.J., *Drugs which affect the structure and function of DNA.* Nature, 1968. **219**(5161): p. 1320-5.

306. Rauth, A.M., B. Barton, and C.P. Lee, *Effects of caffeine on L-cells exposed to mitomycin C.* Cancer Res, 1970. **30**(11): p. 2724-9.

307. Stein, G.S. and H. Rothstein, *Mitomycin C may inhibit mitosis by reducing "G2" RNA synthesis.* Curr Mod Biol, 1968. **2**(5): p. 254-63.

308. Szybalski, W., *Chemical Reactivity of Chromosomal DNA as Related to Mutagenicity: Studies with Human Cell Lines.* Cold Spring Harb Symp Quant Biol, 1964. **29**: p. 151-9.

309. Szybalski, W. and V.N. Iyer, *Crosslinking of DNA by Enzymatically or Chemically Activated Mitomycins and Porfiromycins, Bifunctionally "Alkylating" Antibiotics.* Fed Proc, 1964. **23**: p. 946-57.

310. Sartorelli, A.C., et al., *Mitomycin C: a prototype bioreductive agent.* Oncol Res, 1994. **6**(10-11): p. 501-8.

311. Cummings, J., et al., *Enzymology of mitomycin C metabolic activation in tumour tissue: implications for enzyme-directed bioreductive drug development.* Biochem Pharmacol, 1998. **56**(4): p. 405-14.

312. Mao, Y., M. Varoglu, and D.H. Sherman, *Molecular characterization and analysis of the biosynthetic gene cluster for the antitumor antibiotic mitomycin C from Streptomyces lavendulae NRRL 2564.* Chem Biol, 1999. **6**(4): p. 251-63.

313. Sweet, D.M. and B.E. Moseley, *The resistance of Micrococcus radiodurans to killing and mutation by agents which damage DNA.* Mutat Res, 1976. **34**(2): p. 175-86.

314. Velpandian, T., et al., *Evaluation of the stability of extemporaneously prepared ophthalmic formulation of mitomycin C.* J Ocul Pharmacol Ther, 2005. **21**(3): p. 217-22.

315. Annaluru, N., et al., *Total Synthesis of a Functional Designer Eukaryotic Chromosome.* Science, 2014.

316. Ruifu, Y., et al., *Biovar diversity is reflected by variations of genes encoding urease of Ureaplasma urealyticum.* Microbiol Immunol, 1997. **41**(8): p. 625-7.

317. Robinson, J.W., et al., *Ureaplasma parvum serovar 3 multiple banded antigen size variation after chronic intra-amniotic infection/colonization.* PLoS One. **8**(4): p. e62746.

318. Xiao, L., et al., *Suppression of Antimicrobial Peptide Expression by Ureaplasma species.* Infect Immun, 2014.

319. Pandelidis, K., et al., *Role of biofilm formation in Ureaplasma antibiotic susceptibility and development of bronchopulmonary dysplasia in preterm neonates.* Pediatr Infect Dis J, 2012. **32**(4): p. 394-8.

320. Kwak, D.W., et al., *Co-infection with vaginal Ureaplasma urealyticum and Mycoplasma hominis increases adverse pregnancy outcomes in patients with preterm labor or preterm premature rupture of membranes.* J Matern Fetal Neonatal Med, 2014. **27**(4): p. 333-7.

321. Allam, A.B., et al., *Immune profiling of BALB/C and C57BL/6 mice reveals a correlation between Ureaplasma parvum-Induced fetal inflammatory response syndrome-like pathology and increased placental expression of TLR2 and CD14.* Am J Reprod Immunol, 2014. **71**(3): p. 241-51.

322. Fu, M., et al., *Effects of Ureaplasma urealyticum lipid-associated membrane proteins on rheumatoid arthritis synovial fibroblasts.* J Int Med Res, 2013. **41**(5): p. 1655-70.

323. Triantafilou, M., et al., *Synergic activation of toll-like receptor (TLR) 2/6 and 9 in response to Ureaplasma parvum & urealyticum in human amniotic epithelial cells.* PLoS One, 2013. **8**(4): p. e61199.

324. Zimmerman, C.U., R. Rosengarten, and J. Spergser, *Interaction of the putative tyrosine recombinases RipX (UU145), XerC (UU222), and CodV (UU529) of Ureaplasma parvum serovar 3 with specific DNA.* FEMS Microbiol Lett, 2013. **340**(1): p. 55-64.

325. Dando, S.J., et al., *Maternal administration of erythromycin fails to eradicate intrauterine ureaplasma infection in an ovine model.* Biol Reprod, 2010. **83**(4): p. 616-22.

326. Karas, B.J., et al., *Direct transfer of whole genomes from bacteria to yeast.* Nat Methods, 2013. **10**(5): p. 410-2.