



Genetic and Functional Studies of the Mip Protein of *Legionella*

Rodney Mark Ratcliff, BSc (Hons), MASM

Infectious Diseases Laboratories
Institute of Medical and Veterinary Science

and

Department of Microbiology and Immunology
University of Adelaide.

Adelaide, South Australia.

A thesis submitted to the University of Adelaide for the degree of

Doctor of Philosophy

15th March 2000

amended 14th June 2000
archival copy

Table of Contents

Chapter One: Introduction	1
Background	1
Morphology and Taxonomy	2
Legionellosis	3
Mode of transmission	5
Environmental habitat	7
Interactions between <i>Legionella</i> and phagocytic hosts	8
Attachment	9
Engulfment and internalisation	11
Intracellular processing	13
Intracellular replication of legionellae	13
Host cell death and bacterial release	15
Virulence	18
Genetic factors involved with intracellular multiplication and host cell killing (the <i>icm/dot</i> system)	20
Legiolysin	25
Msp (Zn ⁺⁺ metalloprotease)	25
Lipopolysaccharide	28
The association of flagella with disease	29
Type IV fimbriae	30
Major outer membrane proteins	31
Heat shock proteins	33
Macrophage infectivity potentiator (Mip) protein	34
Virulence traits of <i>Legionella</i> species other than <i>L. pneumophila</i>	39
Chapter One (continued): Introduction to bacterial classification and phylogeny	41
Identification of <i>Legionella</i>	41
Phylogeny	46
Methods of phylogenetic analysis	52
Parsimony methods	53
Distance methods	55
UPGMA cluster analysis	56
Neighbour-joining method	56
Maximum likelihood methods	57
Rooting trees	58
Reliability of a tree	58
Ecological diversity	60
Aims and objectives	64
Chapter Two: Materials and methods	71
Growth media	71
Chemicals and reagents	72
Antibody	73
Enzymes	73
Synthesis of synthetic oligonucleotides	73
Culture and maintenance of bacterial strains	74
Genetic transfer methods	74
Transformation	74
Electroporation	75
Screening of transformants	76
DNA extraction protocols	76

Chromosomal DNA extraction methods.....	76
Method 1	76
Method 2 (rapid)	77
Method 3 (heat lysis)	77
Plasmid DNA extraction methods	77
Method 1 (large scale)	77
Method 2 (small scale).....	78
PEG precipitation purification	79
Analysis and manipulation of DNA	79
DNA quantitation.....	79
Restriction endonuclease digestion of DNA	80
Analytical and preparative separation of restriction fragments	80
Dephosphorylation of DNA using alkaline phosphatase	81
In vitro cloning.....	81
Plasmid nested deletions	81
Labelling of DNA fragments with Digoxigenin-dUTP	82
Southern blot analysis	82
Colony hybridisation probing method	83
Hybridisation and development	84
DNA sequencing and analysis	84
Polymerase chain reaction	85
Protein Analysis	85
Whole cell method	85
SDS polyacrylamide gel electrophoresis (PAGE)	86
Western blot analysis	86
Colony transfer and immuno-blotting.....	87
Phylogenetic analysis of sequence data	87
Analysis of the <i>Legionella</i> genus.....	87
Analysis of the <i>anisa-worsleiensis</i> clade	88
Chapter Three: Functional implications of the inter-species gene variation	90
Introduction	90
Method and sequencing strategy	91
Sequencing strategy for other species	93
Results	94
The open reading frame	94
Flanking sequences	96
Discussion	97
Chapter Four: Phylogeny of the genus <i>Legionella</i>.....	105
Introduction	105
Phylogenetic analysis	105
Results	106
Maximum parsimony analysis	108
Distance analysis.....	108
Congruence among distance and MP trees	109
Non-congruence among distance and MP trees.....	110
Discussion	110
Chapter Five: Further phylogenetic analysis of <i>Legionella</i> species within the (<i>anisa-worsleiensis</i>) clade	123
Introduction	123
Sequence strategy.....	124
Results	125
<i>m</i> sp <i>A</i> sequence analysis	125
Maximum parsimony (MP) analysis.....	127

Maximum likelihood (ML) and distance analysis	127
Congruence and non-congruence among MP, ML and NJ trees	128
Discussion	130
Chapter Six: A <i>mip</i> sequence-based classification scheme	137
Introduction.....	137
Summary of methodology and sequence strategy.....	139
Results.....	141
Discussion	143
Chapter Seven: Ecological diversity of <i>Legionella</i> strains	155
Introduction.....	155
Sequence strategy.....	158
Results.....	158
<i>mip</i> data.....	159
<i>mspA</i> data.....	160
Discussion	162
Chapter Eight: Summary discussion	173
Functional analysis.....	173
Phylogeny.....	174
Further analysis of the <i>anisa</i> - <i>worsleiensis</i> clade.....	176
Classification scheme.....	176
Ecological diversity.....	178
Bibliography	182
Appendices	240
Appendix 1: Distance matrices used in Chapter Four.....	241
Appendix 2: Distance matrices used in Chapter Five	242
Appendix 3: Copies of published papers	243
Appendix 4: Alignment files used in Chapters Four and Five.....	244

Summary

The genus *Legionella* comprises a diverse group of bacterial endosymbionts of some freshwater protozoa. They are also a significant cause of atypical pneumonia in humans, which involves many of the same endocytic and intracellular processes required for endosymbiotic life. Of the approximately 40 species currently recognised, nearly half have been associated with human disease. The macrophage infectivity potentiator (Mip) protein, an immunophilin of the FKBP class which exhibit peptidyl-prolyl *cis/trans* isomerase (PPIase) activity, was the first to be reported as a virulence factor for legionellae, and the sequence of the *mip* gene published for *L. pneumophila*, *L. micdadei* and *L. longbeachae*. This study reports the nucleotide sequence, and the predicted amino acid sequence of the *mip* gene for an additional 35 *Legionella* species, and compares all of the *mip* sequences both functionally and phylogenetically.

The sequences were 69-97% conserved at the nucleotide level and 82-99% at the amino acid level, with total conservation of the amino acids in the seven sites determined to be associated with PPIase activity. No apparent difference could be determined in the arrangement of amino acids which would predict a functional difference in Mip from species associated with disease, and Mip in species isolated only from the environment.

Additionally, a phylogenetic comparison of the *mip* gene sequences with published *16S rRNA* sequences, using both genetic distance and maximum parsimony methods was performed. Few well supported relationships were apparent from both data sets, the most robust being a clade comprising (((*cincinnatiensis*, *longbeachae*, *sainthelensi*, *santicrucis*))

gratiana) (*moravica*, *quateirensis*, *shakespearei*, *worsleiensis*) *anisa*, *bozemanii*, *cherrii*, *dumoffii*, *gormanii*, *jordanis*, *parisiensis*, *pneumophila*, *steigerwaltii*, *tucsonensis*, and *wadsworthii*). These clades were phylogenetically analysed further using approximately 460 bp of the nucleotide sequence from the *mspA/proA* gene.

Further, a species-specific identification scheme for *Legionella* was developed, targeting approximately 700 bp of the *mip* gene, utilising gene amplification with universal primers and direct amplicon sequencing. All species could be identified with the exception of *L. geestiana*, but serotypes could not always be differentiated. Additionally, the genotypic classification of 350 wild strains from several continents was consistent with their phenotypic classification, with the exception of a few strains where serological cross reactivity was complex, potentially confusing phenotypic classification. Strains thought to represent currently uncharacterised novel species were also found to be genetically unique with regard to the *mip* gene.

Among the wild strains examined to validate the classification scheme, *mip* sequence identity was observed for some species, despite the strains being isolated in diverse geographical locations. Wild-strain *mip* sequence identity within seven species was compared with that from a 450 bp segment from the *mspA/proA* gene, and a remarkable level of intraspecies strain relationship identity was observed, even among the strains from diverse geographical locations. The ecological implications of the intraspecies strain relationships is discussed, and a model of global legionellae dispersal within amoebic cysts is proposed to account for the ecological observations.