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

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.