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

Colonies of several *Legionella* strains on charcoal yeast extract agar (CYE) after 4 days incubation at 37°C in air. Various magnifications show typical ground-glass opalescent appearance. Some pure strains exhibit pleomorphic growth or colour. The top two photographs demonstrate typical red (LH) and blue-white (RH) fluorescence exhibited by some species when illuminated by a Woods (UV) Lamp.

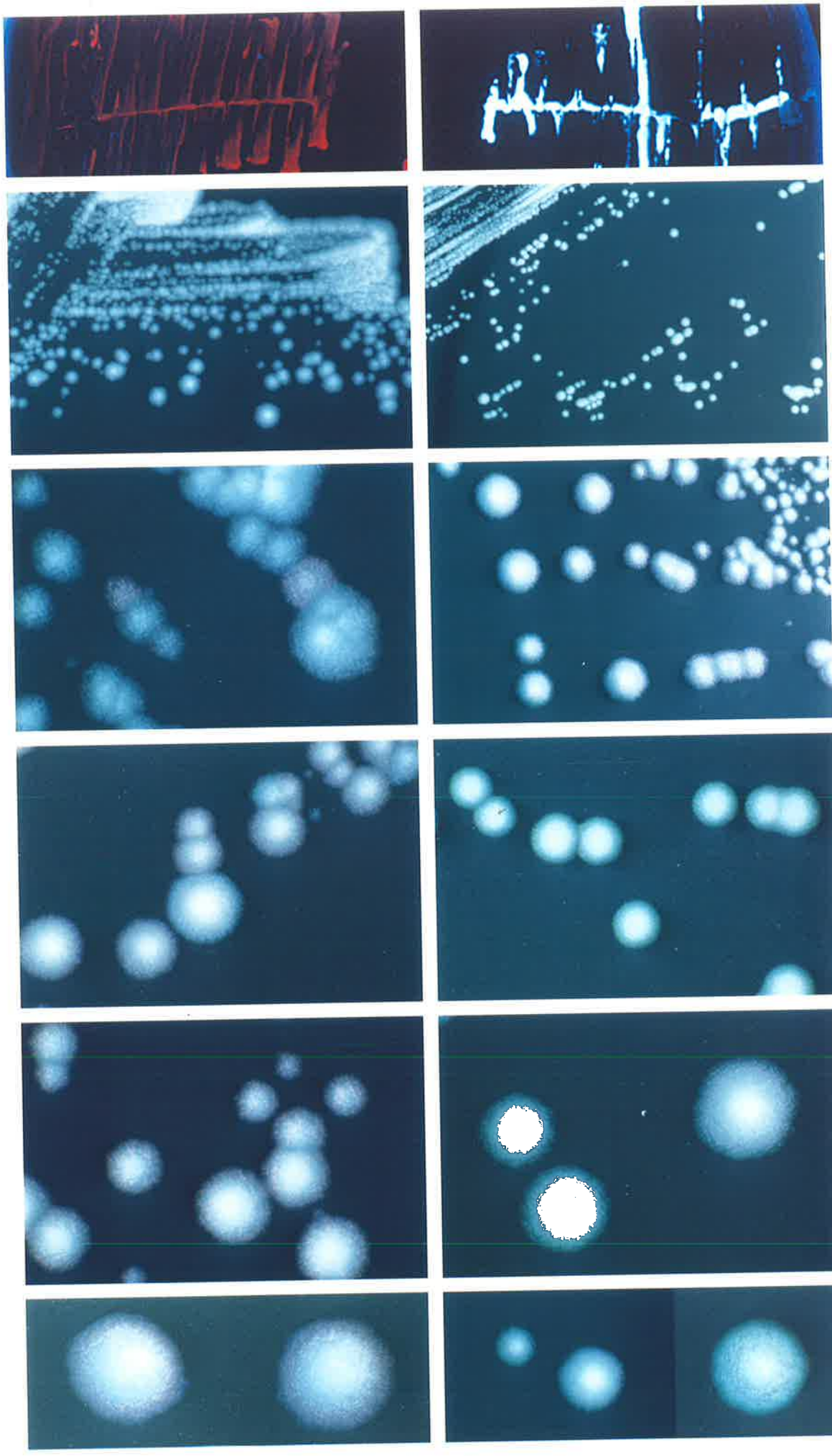


Table of Contents

Table of Abbreviations

The following abbreviations and gene designations have been used throughout this thesis.

ATP	adenosine-5'-triphosphate
C-terminal	carboxyl-terminal of a protein sequence
CYE	Charcoal Yeast Extract Agar
dot	defective in organelle trafficking
dNTP	deoxy-nucleotide tri-phosphate
EDTA	ethylenediaminetetra acetic acid
ER	endoplasmic reticulum
FKBP	FK506 binding protein
GTP	guanosine-5'-triphosphate
GTR+I+G	general time reversible method, I is the proportion of invariant sites, and G is the γ-shape parameter
icm	intracellular multiplication
IPTG	isopropyl-α-D-thiogalactopyranoside
kDa	kilodalton
Kb	kilobase
LLAPs	*Legionella*-like amoebic pathogens
LPS	lipopolysaccharide
MEE	multilocus enzyme electrophoesis
mip or Mip	macrophage infectivity potentiator gene or protein
ML	maximum liklihood method
MOMP	major outer membrane protein
MP	maximum parsimony method
mspA	major secretory protein gene (also termed *proA*)
Msp	major secretory protein
MW	molecular weight
MLST	multilocus sequence typing
NBT	nitroblue tetrazolium chloride
NJ	neighbour joining method
N-terminal	amino-terminal of a protein sequence
nt	nucleotide
OMP	outer membrane protein
O/N	over night
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PLF	phagosome-lysosome fusion
PPIase	peptidyl-prolyl *cis/trans* isomerase
PFGE	pulsed-field gel electrophoresis
QP	Quartet puzzling method
RBS	ribosomal binding site
RFLP	restriction fragment-length polymorphism
rpoB	RNA polymerase B-subunit gene
sg	serogroup
SDS	sodium dodecyl sulphate
TAE	Tris-Acetate-EDTA
TE	Tris-EDTA
TEMED	N,N,N,N,-tetramethylethylenediamine
TBS	Tris buffered saline
Tris	Tris([hydroxymethyl] aminomethane)

TTBS	Tween 20 / Tris buffered saline
UPGMA	unweighted pair group method with arithmetic averages
v/v	volume per volume
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-α-D-galactopyranoside
X-P	5-bromo-4-chloro-3-indolyl-phosphate toluidine salt

Declaration

I declare that the work described herein contains no material that has been previously submitted for the award of any degree or diploma in any university and to the best of my knowledge and belief contains no material previously published or written by another person, except where due reference is made in the text.

I give consent to this copy of my thesis, when deposited in the University Library, being available for loan and photocopying.

Date: 22ND JUNE 2000 Signature:

Acknowledgements

I would like to acknowledge the support of the following people.

- Drs. Jan Lanser and Michael Heuzenroeder, IMVS, Adelaide, and Professor Paul Manning, University of Adelaide, for their supervision, encouragement and support. Without their support, it just would not have happened.

- Ms. Robyn Doyle, and Ms. Rina Dalcin, IMVS, Adelaide, for technical help and for being such a pleasure to work with. Robyn especially inducted me into the mysteries of molecular biology, as she has with many others, before and since. Often to her own detriment, Robyn was always generous with her time and reagents, making difficult tasks so much easier.

- Ms. Norma Sangster, IMVS, Adelaide, for the supply of the Australian *Legionella* strains and assistance with the maintenance and storage of the remaining strains. Norma's safeguard of the collection was a significant source of confidence in its integrity.

- Dr. Stephen Donnellan, Evolutionary Biology Unit, South Australian Museum, Adelaide, for assistance with the phylogenetic analyses. His lucid skill with critical thought was a constant challenge to my muddle-headed equivalent.

- Richard Lumb, IMVS, Adelaide, for assistance in the preparation of the manuscript. I now know that infinitives are not to be split, and am confident that all i's are dotted, and t's are crossed.

- The following persons for the supply of *Legionella* stains and DNA. I am proud to now count many of these colleagues among my friends:

  - Dr. Timothy Harrison, Public Health Laboratory Services, Collindale, London, England. Also Dr. Norman Fry from the same laboratory for collaborative assistance.

  - Dr. Vladimir Drasar, OHS – National *Legionella* Reference Laboratory, District Hygiene Station, Vyškov, Czech Republic.

  - Dr. Jürgen Helbig, Institut für Mikrobiologie und Hygiene, Universitätsklinikum Dresden, Dresden, Germany.

  - Mr. Robert Benson, National Centre for Infectious Diseases, Centre for Disease Control and Pevention, USA. Robert also supplied some DNA samples including several LLAP DNA samples.

  - Professor Atsushi Saito, and Dr. Michio Koide, First Department of Internal Medicine, Nishihara, Okinawa, Japan.

  - Dr. Maddalena Castellani-Pastoris, Dept. of Bacteriology, Instituto Superiore di Sanita, Rome, Italy.

  - Professor Jerome Etienne, Laboratoire de Microbiologie, Hopital Edouard Herriot, Lyon, France.

  - Dr. Richard Birtles, Unite des Rickettsies, Faculte de Medecine, Universite de la Mediterrannee, Marseille, France, for the supply of some of the LLAP DNA.

- My family; Julie, my wife, and my children Leah, Dale and Kyle. Only those who have similarly lived with one involved with a long-term, time consuming project will truly understand and appreciate the cost and sacrifice experienced by them. I thank them for their support, even when the writing dragged on past one completion date after another, and wish there were special words potent enough to convey the intensity of my gratitude to them all.

- And lastly I wish to thank science itself. I still remember the profound sense of wonder with which I viewed my first sequence result. Insignificant as it was, except I guess to the *Legionella santicrucis* strain from which I obtained it, that sequence was still raw, unadulterated knowledge, never before known. In reality, such an experience occurs only to a privileged few. That I can be counted among that few is beyond words.

Be for ever a student. He and he alone is an old man who feels that he has learnt enough and has need for no more knowledge.

Sivananda (1887-1963) Indian physician, sage.

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.

Publications

Ratcliff, R. M., S. C. Donnellan, J. A. Lanser, P. A. Manning, and M. W. Heuzenroeder. 1997. Interspecies sequence differences in the Mip protein from the genus *Legionella*: implications for function and evolutionary relatedness. Mol Microbiol. **25:**1149-1158.

Ratcliff, R. M., J. A. Lanser, P. A. Manning, and M. W. Heuzenroeder. 1998. Sequence-based classification scheme for the genus *Legionella* targeting the *mip* gene. J Clin Microbiol. **36:**1560-1567.

Adeleke, A. A., B. S. Fields, R. F. Benson, M. I. Daneshvar, J. M. Pruckler, R. M. Ratcliff, T. G. Harrison, R. S. Weyant, R. J. Birtles, D. Raoult, and M. A. Halablab. 2000. *Legionella drosanskii* sp. nov., *Legionella rowbothamensis* sp. nov. and *Legionella fallonii* sp. nov: Three unusual new *Legionella* species. Int J Syst Evol Microbiol. Manuscript submitted for publication.

Manuscripts in preparation

Ratcliff, R. M., S. C. Donnellan, J. A. Lanser, P. A. Manning, and M. W. Heuzenroeder. 2000. Further phylogenic analysis of the *Legionella* genus.

Ratcliff, R. M., S. C. Donnellan, J. A. Lanser, P. A. Manning, and M. W. Heuzenroeder. 2000. Gene sequence diversity purging in *Legionella* suggests strains migrate globally.

Chapter One

Introduction

Background

In 1976, a mysterious epidemic of atypical pneumonia occurred among 4,500 delegates attending a Convention of the American Legion in Philadelphia (Fraser and McDade, 1979). A total of 221 delegates contracted the disease, and 34 died (Fraser, *et al.*, 1977, Yu, 1995). Although the exact source of this outbreak was never established conclusively, the aetiological agent was determined to be a previously unrecognised pathogen, *Legionella pneumophila* (Brenner, *et al.*, 1979, McDade, *et al.*, 1977). Once recognised, sero-epidemiologic studies implicated *L. pneumophila* in previous outbreaks of disease (Yu, 1995). Two years previously, in the same hotel, 11 members of an Oddfellows Convention had contracted the disease, but it had gone unnoticed (Terranova, *et al.*, 1978). Even earlier, in 1965, at a psychiatric hospital in Washington, DC, 81 patients had contracted a similar respiratory disease, with 15 deaths. Of affected patients, 85% revealed antibody seroconversion for *L. pneumophila* when stored sera were tested in 1977 (Thacker, 1978). The earliest published evidence of infection in fact dates from as early as 1947, when an uncharacterised organism isolated using guinea pig inoculation from the blood of a patient with a febrile respiratory disease, was subsequently proven to be *L. pneumophila* (McDade, *et al.*, 1979).

Within two years of the recognition and naming of *L. pneumophila*, a second species, *L. micdadei*, was identified as the cause of pneumonia in immuno-compromised patients in Pittsburgh, PA. and Charlottesville, VA., USA. As the fastidious growth requirements of *Legionella* became understood, further species were characterised, either associated with human disease, or isolated from the environment during the course of outbreak investigations (Brenner, *et al.*, 1980, Cherry, *et al.*, 1982, Hebert, *et al.*, 1980). A non-pneumonic, flu-like form of the disease, called Pontiac Fever, has also been described (Kaufmann, *et al.*, 1981), the earliest recognised being in a county health department facility in Pontiac, Michigan, where at least 144 people were affected (Glick, *et al.*, 1978,). To date, at least 40 species are known, and more exist (Ratcliff, *et al.*, 1998). At least six species have more than one serotype (Benson, *et al.*, 1996, Harrison and Saunders, 1994). Nearly half of the species have been associated with human disease (Table 1.1). Serological evidence of disease in numerous species of both domestic and wild animals has been sought, but the results are inconclusive, with the exception of a fatal pneumonia in a bovine calf caused by *L. pneumophila* (Collins, 1986, Fabbi, *et al.*, 1998).

Morphology and Taxonomy

Legionella organisms are non-spore-forming, Gram negative cocco-bacilli, 0.5μm in diameter, and 1 to 2 μm in length. Most species are motile by means of one or more polar flagella (Fallon, 1990). They replicate by "pinching" binary fission, and are nutritionally fastidious to such an extent that it is thought that intracellular replication within select host cells is the primary and perhaps sole means of proliferation in the environment (Fields, 1993). *In vitro*, many species can be cultured on Charcoal Yeast Extract (CYE) Agar, a specialised medium supplemented with cysteine, iron salts, α-ketogluterate, and activated charcoal as a detoxifier (Feeley, *et al.*, 1979). Colonies 1-2mm in diameter form after 2-3 days incubation at 37°C at high humidity, and longer (7-10 days) for some species. The growth of most species is enhanced by the presence of carbon dioxide (CO_2) at 3% in air, and there is an indication

that some species prefer a lower incubation temperature such as 30°C. Colonies are usually round, smooth-edged, and raised, and exhibit a "cut glass' or opalescent appearance that may be white, grey, green, or pale pink to rose in colour. Colonial variation is exhibited by some species, in both size and colour (Fallon, 1990). Colonies of some species fluoresce blue-white, or red-brown when examined under ultra-violet light.

Other species, designated *Legionella*-like Amoebic Pathogens (LLAPs), because of their inability to be cultivated on CYE are only recoverable in co-culture with amoebae or similar host cells (Rowbotham, 1983). Some have been isolated on CYE after prolonged incubation and repeated subculture (Adeleke, *et al.*, 2000). Using amoebic co-culture, LLAPs have been isolated from the sputa of patients with legionellosis, implicating them as the causative agent (Adeleke, *et al.*, 1996, Hookey, *et al.*, 1996).

L. pneumophila is also able to enter a non-replicative viable but nonculturable (VBNC) state, where the bacteria are no longer culturable on routine culture media (Steinert, *et al.*, 1997). Reactivation of dormant legionellae occurs in the presence of amoebae, and the reactivated bacteria are just as virulent as those cultured on artificial media, as measured by animal invasion models. However, reactivation of dormant legionellae has not been observed in animal models.

Legionellosis

Legionellosis usually presents as an atypical pneumonia, with lobar-segmental or patchy pulmonary infiltration, described histologically as acute fibrinopurulent pneumonia involving the alveoli (Winn and Myerowitz, 1981). The alveolar infiltrate consists of macrophages, polymorphonuclear leucocytes and erythrocytes. There is little or no inflammation of the upper airways and bronchioles. Clinical symptoms include non-productive cough, pulse-temperature dissociation, abnormalities in liver function tests,

diarrhoea, hyponatremia, hypophosphatemia, myalgia, confusion and multiple rigors (Edelstein, 1993). The spectrum of symptoms can be quite variable, and reliance on classic symptoms would result in missed diagnosis, except perhaps when identifying cases during outbreaks. A non-pneumonic "flu-like" form called Pontiac Fever is also recognised. Risk factors such as cigarette smoking, chronic lung disease and immuno-suppression, especially that caused by corticosteroid therapy have been consistently implicated. Surgery is a major predisposing factor in nosocomial infections, especially among transplant recipients. The incidence is low among patients with acquired immuno-deficiency, but the clinical manifestations, such as lung abscesses and bacteraemia are more severe. Infection in children is rare, but has been diagnosed in the immunocompromised. Extra-pulmonary disease is recognised, and reported in many organs, most likely the result of bacteraemia seeding during the course of pneumonia. Some wound infections have been attributed to direct inoculation from *Legionella*-contaminated water, such as spa water and wound irrigation fluid (Edelstein, 1993).

The pathogenic sequence in mammalian hosts is presumed to be as follows. *L. pneumophila* is inhaled into the alveoli in the lower airways, where they are taken up into alveolar macrophages and infiltrated blood monocytes. By evading the usual lytic processes within the phagocytic cells, they multiply, drawing nutrients from the host cell. Eventually the host cell dies, and the bacterial progeny are released into the alveoli again, to be taken up by other phagocytic cells, and probably alveolar epithelial cells (Cianciotto, *et al.*, 1995a), thereby establishing and amplifying an intracellular infection. Lung tissue damage occurs, either as a direct result of the effects of toxic bacterial products, or as a response to the inflammatory response triggered by the infection (Cianciotto, *et al.*, 1989a).

Mode of Transmission

Legionellosis has been attributed to the inhalation of viable organisms in fine aerosols (up to five microns in diameter) into the lung, where they invade the alveolar macrophages, other phagocytic cells, and probably alveolar epithelial cells (Abu Kwaik, *et al.*, 1998a, Cianciotto, *et al.*, 1989a, Cianciotto, *et al.*, 1995a, Horwitz, 1992, O'Connell, *et al.*, 1996). Aerosol generating systems which have been linked with disease are generally man-made, and include cooling towers, spa and whirlpool baths, bathroom showers and respiratory-therapy equipment (Baron and Willeke, 1986, Bollin, *et al.*, 1985, Bornstein, *et al.*, 1989, Breiman, *et al.*, 1990, Dennis, *et al.*, 1984, Groothuis, *et al.*, 1985, Mangione, *et al.*, 1985, Spitalny, *et al.*, 1984a, Spitalny, *et al.*, 1984b, Vogt, *et al.*, 1987, Woo, *et al.*, 1992). An outbreak among 28 shoppers who visited a grocery store was attributed to a misting machine used to spray produce on display (Mahoney, *et al.*, 1992). Recently, in February 1999, a very large outbreak occurred, in which 242 people who visited a large flower show near Amsterdam in the Netherlands became ill and 28 subsequently died. The cause has been attributed to a whirlpool spa that was on display at the show. The *Legionella* strain isolated from the spa was identical to that found in some of the patients (http://www.hcinfo.com/outbreaks-news.htm). Potable water distribution systems in hospitals, nursing homes, workplaces and homes have also been implicated, and indeed there is some controversy whether these are a more significant reservoir than cooling towers (Lin, *et al.*, 1998), with aspiration a significant mode of transmission (Muder, *et al.*, 1986, Stout and Yu, 1997). The evidence cited is the high incidence of disease amongst patients who have undergone surgical procedures where there is a high propensity for aspiration. Person to person transmission has never been documented (Abu Kwaik, *et al.*, 1998b). An explanation for this may be that the "invasion unit" is a protozoan infected with *Legionella* (Shuman, *et al.*, 1998). While most protozoal trophozoites would be >5 μm in diameter, and hence greater than the aerosol particle size thought optimal for inhalation into lung alveolae, this may not necessarily preclude legionellae-infected trophozoites as an "invasion unit". Alternatively, viable legionellae have been observed within

protozoan cysts (Adeleke, *et al.*, 1996), and cysts of protozoa known to host legionellae have been detected in air currents at levels which suggest they would be frequently inhaled from contaminated air (Rogerson and Detwiler, 1999). In support for this hypothesis, Cirillo, *et al.* (1994) determined that prior growth of *L. pneumophila* in *Acanthamoeba castellanii* increased their invasiveness at least 100-fold for epithelial cells and 10-fold for macrophages, and postulated that the intracellular environment stimulates invasion-related protein expression. An additional possibility arises from the observations of Rowbotham (1983, 1986) and Berk, *et al.* (1998) that amoebae such as *A. castellanii* and *Acanthamoeba polyphaga*, which may be present in cooling tower water, are known to expel vesicles of 1 to 5 μm in size, just prior to encystment. If the amoebae are infected with *L. pneumophila*, such vesicles may contain viable bacteria in high numbers (theoretical estimates vary from 20-200 bacteria (Berk, *et al.*, 1998) to 365-1483 bacteria (Rowbotham, 1986) depending on size) and are resistant to biocides and some environmental conditions. Such vesicles are of the correct size to be inhaled, and even a single vesicle could contain an infectious dose (Rowbotham, 1986). Such vesicles may explain the infectious-dose paradox (O'Brien and Bhopal, 1993) of how cooling towers with relatively low numbers of legionellae can be implicated in outbreaks, and how legionellae-containing aerosols can be carried long distances from the source without the bacteria dying from desiccation (Berk, *et al.*, 1998). In contrast, large numbers of *in vitro*-grown organisms (100,000 for guinea pigs) are required to initiate infection in experimental animals (Doyle, *et al.*, 1998). However, the minimum infectious dose for human infection is not known. Support for amoebic involvement in human disease is the finding that patients infected with *Legionella* also have antibodies directed to free-living amoebae isolated from the water to which the patients had been exposed (Winiecka-Krusnell, *et al.*, 1998, in Winiecka-Krusnell and Lindner, 1999). Interestingly, these observations also raise the additional paradox that biocide use may actually promote the infectious risk from a cooling tower by precipitating amoebic encystment, and expulsion of vesicles packed with viable bacteria (Winiecka-Krusnell and Linder, 1999).

Legionellosis caused by *Legionella longbeachae*, which has been relatively rarely reported in the USA and Europe, but is significant in Australia and New Zealand, has a strong association with gardening. While not proven, contaminated compost and potting soil, in which the organism is widely distributed, is suspected as the organism reservoir (Steele, *et al.*, 1990a, Steele, *et al.*, 1990b). The disease process is similar for legionellosis caused by *L. pneumophila*, but the risk factors are not as clear, with patients less likely to have underlying chronic health problems, or to be smokers (Walker and Weinstein, 1992).

Environmental Habitat

Legionella are ubiquitous inhabitants of aquatic environments and damp soil. Their ability to multiply intracellularly within protozoa was first described in 1980 (Rowbotham, 1980). It is now understood that they may be unable to multiply extracellularly in the environment. Thirteen species of amoebae and two species of ciliated protozoa have been shown to support intracellular bacterial replication and so act as potential environmental hosts (Abu Kwaik, *et al.*, 1998b, Cameron, *et al.*, 1991, Fields, *et al.*, 1993, Hughes and Steele, 1994, Lee and West, 1991, Steele and McLennan, 1996, Steele, *et al.*, 1990a, Steele, *et al.*, 1990b). Their ability to replicate intracellularly within amoebae is significantly greater than that demonstrated by other intracellular parasites such as *Chlamydia pneumonia* and *Mycobacterium avium*. This sophisticated association of legionellae with environmental protozoa accounts for their continuous presence in the environment, rendering them tolerant to harsh environmental extremes of pH, temperature, osmolarity, oxidising agents, and chemical disinfectants and biocides (Barker and Brown, 1994, Brown and Barker, 1999, Barker, *et al.*, 1986, Barker, *et al.*, 1993, Winiecka-Krusnell and Linder, 1999). The increased resistance to environmental extremes has been shown to be intrinsic when compared with *in vitro*-grown bacteria, suggesting that such resistance is induced at least in part, by the intracellular environment (Byrne and Swanson, 1998). Resistance is also afforded by direct protection of

the bacteria within vesicles expelled by the protozoan host, and within the highly resistant protozoan cyst (Berk, *et al.*, 1998).

Interactions between *Legionella* and phagocytic hosts

Endocytosis, the uptake of extracellular material within plasma membrane-derived vacuoles, is one of the basic traits of eukaryotic cells (Rittig, *et al.*, 1999). This can be either as receptor-independent uptake of extracellular fluid (pinocytosis) or as receptor-mediated uptake of macromolecules and particles, including bacteria (phagocytosis). Many protozoa phagocytose bacteria as a major nutrient source. Almost all mammalian cell types are able to facultatively or constitutively perform phagocytosis, provided there is an adequate receptor-ligand binding system. Cell lines where this uptake mechanism is most evident, such as neutrophils, monocytes and macrophages, are designated professional phagocytes. However, other cell types, such as alveolar epithelial cells can be recruited to perform phagocytosis.

An increasing number of receptors known to promote phagocytosis include Fc receptors, complement receptors, scavenger receptors, mannose receptors and those for extracellular matrix components. Phagocytosis-promoting ligands are either direct structural patterns on the surface of the macromolecule, particle or bacteria (non-opsonic), or host-derived opsonins coating the surface (opsonic) (Rittig, *et al.*, 1999).

The classic phagocytic "zipper" mechanism (Fig. 1.1) was first established 25 years ago (Griffin, *et al.*, 1975) to include:

1. **Attachment** of the particle to the phagocytic receptor
2. **Engulfment** of the adherent particle by protrusions extending from the cell surface
3. **Internalisation**, combining the formation of a sealed vacuole (phagosome) and its budding off from the restored cell membrane

4. **Intracellular processing** of the phagosome, eventually turning into a phagolysosome upon participation in the general endocytic trafficking of the host cell.

More recently, variations of this classic mechanism, and also unconventional mechanisms of phagocytisis, have been recognised. The cellular processes occurring between legionellae and their phagocytic hosts have been extensively studied as part of this process, in order to understand the sophisticated interactions required to enable intracellular parasitism, an ability that is central to the ecology and pathogenesis of *Legionella*. Both mammalian phagocytic cells and protozoan hosts, two evolutionary disparate hosts, have been studied. Most work has been focused on *L. pneumophila*, since it is the most common pathogenic species.

Attachment

The initial interaction between an intracellular pathogen and the host cell is mediated by the attachment of a bacterial ligand or adhesin to a surface receptor on the host cell. This ligand-receptor binding precipitates signal transduction processes which in turn initiate polymerisation of actin at the receptor site, leading to phagocytic uptake (Petty and Todd, 1993, Wang, *et al.*, 1993). For *L. pneumophila*, two attachment mechanisms have been described, one opsonic, used by human-derived phagocytic cells, and the second non-opsonic, used by protozoan hosts.

In the first, the complement component C3 binds selectively to the major outer membrane protein (MOMP), mediating phagocytosis of liposome-MOMP complexes by human monocytes and macrophages (Bellinger-Kawahara and Horwitz, 1990). Coxon *et. al.* (1998) determined that opsonin-dependent phagocytosis of *L. pneumophila* by human monocytes involved attachment to the CR1 and CR3 integrin receptors on the host cell

surface, which triggered both tyrosine protein kinase (TPK) and protein kinase C (PKC) phosphorylation signals, initiating the localised accumulation of F-actin at the cell periphery near the site of bacterial entry. They also determined that TPK phosphorylation of six major cellular host derived proteins occurred, three of which were associated with the cytoskeleton. Avirulent *L. pneumophila*, *Escherichia coli* and zymosan entry into human monocytes induced similar TPK phosphorylation events, indicating that CR3-mediated uptake into human monocytes is unrelated to the ultimate virulence mechanism associated with intracellular replication.

The second attachment mechanism is both opsonin and microfilament independent, and involves type IV pili (Stone and Abu Kwaik, 1998). *L. pneumophila* defective in the expression of the $pilE_L$ gene manifested reduced attachment to both epithelial and protozoan cells. Homologs of $pilE_L$ were detected using Southern hybridization in all serogroups of *L. pneumophila* but in less than half of 16 other *Legionella* species (Stone and Abu Kwaik, 1998). There was no correlation between disease association and the presence of $pilE_L$ among the 16 non-*L. pneumophila* species. Although the attachment receptor is not known, Abu Kwaik *et. al.* (1998b) have postulated that for *Hartmannella vermiformis*, the receptor may be a galactose/*N*-acetylgalactosamine (Gal/GalNAc) lectin similar to the β2 intergrin-like Gal/GalNAc lectin of the pathogenic protozoan *Entamoeba histolytica*. Integrins are heterodimeric protein tyrosine kinase receptors, which upon ligand binding undergo tyrosine phosphorylation, resulting in recruitment and rearrangement of the cytoskeleton. For both *L. pneumophila* and *L. micdadei*, attachment to the Gal/GalNAc lectin of *H. vermiformis* has been shown to trigger signal transduction events that result in dramatic tyrosine dephosphorylation of the lectin receptor and other cytoskeletal proteins (Abu Kwaik, *et al.*, 1998c, Venkataraman, *et al.*, 1997, Venkataraman, *et al.*, 1998). However, the same workers have proposed that the mechanisms of attachment and subsequent uptake may vary between protozoan hosts, based upon the different outcomes for 89 invasion-defective insertion

mutants of *L. pneumophila* infecting *H. vermiformis* and *Acanthamoeba polyphaga* (Harb, *et al.*, 1998). Interestingly, the same mutants displayed similar defective phenotypes in their cytotoxicity and intracellular replication within U937 macrophage-like cells and *A. polyphaga*, two very distant hosts in evolutionary terms.

Engulfment and Internalisation

Following the initial host-parasite interaction, *Legionella* are internalised by the host cell, either as a single bacterial cell, or in pairs (Fields, 1996). This mechanism is dependent both on the host cell and *Legionella* species involved. For human monocytes, and several protozoan hosts, phagocytosis of *L. pneumophila* occurs by at least two mechanisms. The first is the conventional, receptor-mediated, cytoskeleton-independent, endocytosis in which phagocyte pseudopods or micropseudopods circumferentially engulf the particle as the result of progressive interactions between receptors on the surface of the pseudopod and ligands on the particle surface (classic "zipper" mechanism). This zipper-like interaction continues until they meet and fuse at the distal side of the particle, enclosing it within a membrane-bound vacuole (Horwitz, 1984). Fusion of the membranes is not a spontaneous event, but depends on fusiogenic factors adjacently present on the approaching membrane surfaces (Rittig, *et al.*, 1998b), and possibly the absence of antifusiogenic factors (Rittig, *et al.*, 1999). In examining *L. pneumophila* uptake into human monocytes, polymorphonuclear leucocytes and alveolar macrophages, Horwitz (1984) observed a second novel mechanism that seemed to predominate, which was designated "coiling" phagocytosis (Fig 1.1). This involved engulfment in a single multi-turn unilateral pseudopod with coiling whorls giving rise to largely self-apposed pseudopodal surfaces. The coiling phagocytosis mechanism occurred not just with live *L. pneumophila*, but also with formalin-, gluteraldehyde- or heat-killed *L. pneumophila* cells. Similar coiling phagocytosis has also been observed for protozoan hosts, but at a lower frequency (Bozue and Johnson, 1996, Venkataraman, *et al.*, 1998), and during the phagocytosis by human monocytes of other intracellular parasites, such as *Borrelia*

burgdorferi, *Leishmania*, *Leptospira* and fungal cells (Rittig, *et al.*, 1998b). The events which trigger coiling phagocytosis rather than the conventional form are not known. Rittig, *et al.* (1998a) argue that it is actively induced, rather than a passive event, a reaction of the phagocyte to the attachment of certain types of particles, and triggered by specific, but as yet unknown, factors. They postulate that it is the result of a disturbance of conventional "zipper" phagocytosis, caused not by specific "coiling"-initiating receptors, but by asymmetrical receptor clustering on the surface of the phagocytic cell to one side of the initial ligand-receptor interaction. A unilateral pseudopod is initiated by the receptor cluster, which fails to fuse with its stem after circumventing the particle, continuing to extend and rolling into itself rather than creating a phagosome. The presence of a continuously extending pseudopod which fails to fuse suggests that the engulfment and internalisation steps of endocytosis are separate steps which while normally linked are dependent on distinct sets of control and effector factors. During coiling phagocytosis, these two steps somehow become uncoupled. Additionally, coiling phagocytosis suggests that the fusion of the pseudopod tips has a negative feedback upon further pseudopod extension (Rittig, *et al.*, 1999). The significance of coiling phagocytosis in the pathogenesis of *Legionella* remains unclear. Killed *L. pneumophila* are taken up by coiling phagocytosis, whereas many clinical isolates of *L. pneumophila* are taken up exclusively by conventional phagocytosis (Elliott and Winn, 1986, Rechnitzer and Blom, 1989). Uptake of *L. micdadei* by human monocytes and *H. vermiformis* does not utilise microfilament-dependent coiling, although the uptake mechanism utilised by the two hosts does differ (Abu Kwaik, *et al.*, 1998c, Joshi and Swanson, 1999). In fact, uptake of *L. micdadei* by coiling phagocytosis has never been reported (Weinbaum, *et al.*, 1984). Similarly, *L. pneumophila* Knoxville-1 is internalised by human macrophages using classic phagocytosis (Rechnitzer and Blom, 1989), as is *L. dumoffii* by the non-professional phagocytic epithelial Vero cell line. (Maruta, *et al.*, 1998). These results suggest that the mode of uptake may be independent of the subsequent fate of the bacteria.

Intracellular processing

Following coiling phagocytosis, the many layers of the coil resolve to form a single layered membrane-bound compartment, similar to that formed following conventional phagocytosis (Shuman, *et al.*, 1998). The mechanism of resolution is not understood, but Clemens and Horwitz (1992) determined that the bacteria do influence the composition of the phagosomal membrane as it is being formed (also see section dealing with the *icm/do*t locus).

In the classical understanding of endocytosis, following internalization of a particle such as a bacterium within a phagosome, the resultant phagosome matures. F-actin is depolymerised from the phagosome, and the newly denuded membrane becomes accessible to a series of complex fusion and fission events firstly with early and then late endosomes, and ultimately lysosome(s) to form a phagolysosome. During the maturation process, the phagosome moves into the cell on microtubules, and it is this trafficking which facilitates and controls maturation by progressively exposing the phagosome to the various components of the endosomal system (Aderem and Underhill, 1999). The phagolysosome contents ultimately acidify and are flooded with digestive enzymes, killing the bacteria within. The rates of phagosome-lysosome fusion (PLF) vary dramatically depending on the nature of the ingested particle. As an example, within 30 minutes of co-cultivation of *Proteus mirabilis* with *A. castellanii*, 91% of phagosomes containing *P. mirabilis* had fused with and were acidified by the amoebic lysosome (Abu Kwaik, 1996). However, phagosomes containing latex may not undergo PLF for several hours (Aderem and Underhill, 1999). The rate at which phagosomes mature is thought to be influenced by the nature of the interaction between the phagosomal membrane and the particle surface, but the mechanisms involved are not understood.

Intracellular replication of legionellae

As a means of escaping the potentially harsh and nutrient poor aquatic environment, legionellae have evolved the ability to avoid the usual outcome of particles internalised within

the phagosome, and so survive and multiply within the phagocytic cell. Two crucial events in this survival process for the bacteria are the blocking of the PLF process and preventing the acidification of the vacuole. Further, the phagosome becomes surrounded by mitochondria and host cell vesicles within the first 60 minutes post-infection (Abu Kwaik, 1996). By four hours post-infection, the cytosolic face of the phagosome is surrounded by a multilayered membrane that may be derived from, or correspond to rough endoplasmic reticulum (ER) (Abu Kwaik, 1996). The detection of an ER-specific chaperone protein, BiP, in the membrane supports this association (Swanson and Isberg, 1993). At some point following the formation of the ER-associated phagosome, the internalised bacteria begin to replicate. Typically, an increase in bacterial numbers is not detected until 6-8 hours post-infection. Whether, this lag is due to the usual "lag phase" of bacterial growth, or due to the time required for the development of nutritional conditions within the phagosome is not known (Shuman, *et al.*, 1998). Given that the generation time of bacteria within the phagosome, after the lag phase, is approximately two hours, similar to that found *in vitro*, nutritional conditions within the phagosome must be adequate for sustained multiplication. The mechanism whereby nutrients reach the bacteria within the phagosomes is not known, and the role of the recruited vesicles and ER associated with the cytosolic face of the phagosome is also unclear (Shuman, *et al.*, 1998).

For example, iron is an essential nutrient for all living organisms. Within the intracellular environment, iron is not freely available, being bound to transferrin, complexed with ferritin or sequestered in labile iron and haem pools (Abu Kwaik, *et al.*, 1998a). It has been postulated, based on the nutrient requirements on artificial media, that *Legionella* require high concentrations of iron salts (1 mM) as well as cysteine (400 mg/l), and has lead workers to study how such levels may be achieved within the phagosome. Hickey and Cianciotto (1997), studying iron requirements, have described a *fur* gene and Fur-regulated genes for *L. pneumophila*, and postulate that *L. pneumophila* utilises siderophores to acquire iron. They

have determined that a mutant defective in the expression of the siderophore aerobactin synthetase, an important enzyme in iron scavenging, is also defective in intracellular replication in macrophages. Alternatively, other workers have suggested the high levels of cysteine and iron in artificial media are unnecessary (Shuman, *et al.*, 1998). They showed that iron and cysteine react to form redox products, including cystine, and as a consequence *Legionella* will grow in much lower levels if the proportions of iron and cysteine are kept in balance. Consequently, it may not be necessary to infer extraordinary high levels of either cysteine or iron within the phagosome. Bird and Horwitz (1989) showed that a minimal amount of iron is essential by demonstrating that intracellular replication of *L. pneumophila* within blood monocytes could be prevented in the presence of iron chelators or the cytokine IFN-γ, to reduce the intracellular iron concentration. Supplementation of iron reversed these effects.

In addition, it is thought that the concentration of NaCl within the phagosome is very low, based on the observation that continuous passage of *L. pneumophila* can produce spontaneous mutants which are both NaCl tolerant and attenuated in intracellular survival in macrophages and guinea pigs (McDade and Shepard, 1979). Also, many of the *icm/dot* and *pmi* mutants (see later section) exhibit the same NaCl tolerant phenotype (Gao, *et al.*, 1997, Jacob, *et al.*, 1994, Vogel, *et al.*, 1996, Vogel, *et al.*, 1998). The association of high NaCl levels with attenuation of virulence remains to be elucidated.

Host cell death and bacterial release

At the end of the intracellular growth phase, as nutrients within the phagosome are depleted by the bacteria which now pack the phagosome, the bacteria exit the exponential growth phase, and concurrently undergo a dramatic phenotypic switch, characterised by an increase in sodium sensitivity, osmotic resistance, motility, cytotoxicity, and *icm/dot* loci upregulation (see later) (Byrne and Swanson, 1998). Compared with bacteria grown on

artificial media, these bacteria are short, thick, and highly motile, and possess a thick, smooth cell wall, a higher β-hydroxybutyrate content, and express a different array of proteins. Approximately 18 hours post-infection, there is a dramatic cytopathic effect on the host, causing lysis and death of the host, with the release of the bacteria into the environment. These observations are consistent with a model in which the bacteria, in response to nutrient deprivation within the phagosome, switch from metabolic functions favouring maximum replication, to a state triggering release from the host cell, surviving increased osmotic stress, promoting dispersal into the environment, and the re-establishment of a new intracellular niche protected from lysosomal degradation. If successful in re-establishing the nutrient rich intracellular environment, the bacteria return to their replication favoured phenotype (Byrne and Swanson, 1998).

The cytopathic mechanism remains unclear, although various explanations have been postulated. Initially, the major secretory protein (Msp) protease was thought to be responsible, but was discounted by Blander, *et al.* (1990) who demonstrated that protease-null mutants still achieve the same cytopathic effect of wild-type *L. pneumophila*. Moffat, *et al.* (1994a, 1994b) determined that similar null mutants, while not demonstrating any phenotypic differences from the wild-type in protozoa and explanted guinea pig macrophages, did attenuate infection in a guinea pig model, and argue that although not observed, it is likely Msp is active at some time during intracellular life in protozoa. Others have postulated the existence of cytotoxins, but they have not been clearly demonstrated. An alternate view is that the bacteria induce apoptosis, or programmed cell death, in the host, similar to that determined for *Shigella flexneri* and *Bordetella pertussis* (Muller, *et al.*, 1996). During apoptosis, a strictly regulated suicide program is initiated within the dying cell, involving the activation of a family of cysteine proteases (caspases) which subsequently leads to cellular and organelle dismantling, DNA cleavage into nucleosome-sized fragments, and caspase-mediated cleavage of proteins within the host cell (Salvesen and Dixit, 1998, Zychlinsky and Sansonetti, 1997). The ability

of *L. pneumophila* to precipitate apoptosis has been challenged (Zhu and Loh, 1996) but recently Hägele, *et al.* (1998) demonstrated that *L. pneumophila* is indeed able to induce apoptosis in human monocytes but not in *A. castellanii*. They observed that the apoptotic death of human monocytes by *L. pneumophila* is dependent on factors such as multiplicity of infection, time post-infection and intracellular location of the bacteria. Noting that the TNF-α-mediated signal-transduction pathway is not involved in transducing the signal for apoptosis, they postulate a currently undetermined factor interacting directly with the caspase cascade similar to the apoptosis induction mechanism of *S. flexneri*.

A third explanation of the cytopathic effects is that the by-products of bacterial metabolism, such as NH_3, accumulate to a level which effectively poisons the host cell, precipitating cell death (Shuman, *et al.*, 1998). A fourth possibility is that the bacterial numbers themselves overwhelm the host cell by rupturing the membranes that are no longer capable of containing the expanding bacterial bulk. Recently, a fifth model has been proposed, based on the observed cytopathic effect that can occur from *icm/dot* loci-encoded pore formation in the host cell membrane (see later for a detailed explanation) (Byrne and Swanson, 1998). When the numbers of bacteria invading each host cell is very large, so many pores are formed in the host membrane that integrity and function are destroyed, and cell lysis results. Similarly, as a consequence of the large numbers of bacteria packed into the phagosome, *icm/dot* loci-encoded pore formation may cause lysis of the phagosome membrane. It has been noted however that this type of cytotoxicity requires the intimate membrane contact resulting from ligand-receptor binding and attachment (Kirby and Isberg, 1998). There is no evidence that this is achieved during phagosome release.

Recently, Gao and Abu Kwaik (1999a, 1999b) detected caspase 3 induction of apoptosis in both macrophages, alveolar epithelial cells and peripheral blood monocytes. However, they noted caspase 3 is maximally expressed 3 hours post-infection, and that even

extracellular *L. pneumophila* induced caspase 3-mediated apoptosis by a mechanism that is dependent on the Icm/Dot secretion apparatus (Gao and Abu Kwaik, 1999a). They postulate that apoptosis may play an important role in blocking phagosome maturation and endocytic fusion through the endosomal pathway. In addition they postulate that there may be two stages to human infection, with apoptosis playing a different role in each. The first occurs early in infection when bacterial numbers are low. Host cells are induced to undergo apoptosis when bacterial numbers reach a certain threshold, regardless of the growth phase, to maximise the spread of bacterial uptake. Further, host cell death by apoptosis rather than necrosis may reduce the inflammatory response at the site of infection, enhancing bacterial proliferation. During the second phase, late in infection, when there are large numbers of post-exponential growth phase bacteria, they propose that bacterial cytotoxicity and host cell necrosis is more significant. They propose that Icm/Dot pore formation plays a significant role during this phase, to precipitate host cell necrosis. They also suggest that the large numbers of bacteria being released during this phase similarly lyse neighbouring cells by the same mechanism (Gao and Abu Kwaik, 1999a, Gao and Abu Kwaik, 1999b).

To date, until further evidence can be supplied, none of these models can be discounted, and it is possible that different mechanisms may be used by various legionellae for different hosts.

Virulence

The association with human disease varies both between *Legionella* species, and among strains of the same species (Doyle, *et al.*, 1998, Joshi and Swanson, 1999, Maruta, *et al.*, 1998). In the USA, ninety percent of all cases of legionellosis are caused by *L. pneumophila*, followed by *L. micdadei* (Marston, *et al.*, 1994). Of the cases caused by *L. pneumophila*, 80% are caused by serogroup 1, even though there are 13 additional serogroups. Approximately half of all species have never been associated with disease (Table 1.1). While

it remains unclear whether this distribution is due to the greater inherent virulence of the strains isolated with higher frequency, or due to the prevalence of strains in the environment, it has been shown in both cell culture and animal models that there are species and strain differences which do influence virulence potential (Doyle, *et al.*, 1998, Maruta, *et al.*, 1998). At the same time, it has been demonstrated that species never associated with disease may be capable of doing so (O'Connell, *et al.*, 1996). Since it seems highly likely that all *Legionella*, irrespective of species, survive in the environment as intracellular parasites of amoebae and protozoa, then they must express a sufficient metabolic repertoire to facilitate intracellular life. As noted previously, there are small, but measurable differences between strains and species in the mechanisms by which this is achieved (Maruta, *et al.*, 1998). It has also been noted that at least some of the factors which promote the capacity of legionellae to invade and multiply intracellularly are inducible, as has the possibility that symptoms seen in human disease, and presumably the infection outcome as well, are directly related to the metabolic products produced during the process of invasion and intracellular survival. Consequently, it may be completely accidental that legionellae are capable of causing mammalian disease, with the advantage conferred by evolution of being able to invade a wide spectrum of amoebal and protozoal hosts fortuitously also conferring competence in mammalian cells as well. This issue of the origin of mammalian infectivity, may be clarified by the elucidation of the newly discovered *mil* loci, mutations within which confer on the mutant various degrees of defects within macrophages but still exhibit the wild-type phenotype within protozoa (Abu Kwaik, *et al.*, 1998b, Gao, *et al.*, 1998). Whatever the origin, purpose, and diversity of infectious capability within mammals, legionellae do produce various virulence determinants which affect the disease outcome (Dowling, *et al.*, 1992). A discussion of some specific determinants follows.

Genetic factors involved with intracellular multiplication and host cell killing (the icm/dot system)

Using both spontaneous and transposon-induced mutants to understand the genetics of intracellular multiplication and host cell death, two groups of workers simultaneously but independently recently defined a gene locus which may account for most of the genes that enable *L. pneumophila* to grow within and kill human macrophages (Berger and Isberg, 1993, Marra, *et al.*, 1992, Segal and Shuman, 1998). Named *icm* (for "intracellular multiplication") (Marra, *et al.*, 1992) or *dot* (for "defective in organelle trafficking") (Berger and Isberg, 1993), mutants in these linked genes (termed region I) were found to be completely defective in preventing PLF, and were also unable to cause lethal pneumonia in guinea pigs. A second unlinked region (region II) has subsequently been recognised (Brand, *et al.*, 1994, Purcell and Shuman, 1998, Segal, *et al.*, 1998, Segal and Shuman, 1997). The complete *icm/dot* locus comprises a total of 23 *icm/dot* genes in the two unlinked chromosomal clusters of approximately 22 kb each (see Fig. 1.2a). No further *icm/dot* genes are required for intracellular growth (Vogel, *et al.*, 1998). These genes may represent a pathogenicity island(s) acquired to prevent PLF within the host cell, thus enabling intracellular multiplication, as well as to mediate host cytotoxicity by pore formation (Andrews, *et al.*, 1998, Berger, *et al.*, 1994, Brand, *et al.*, 1994, Kirby, *et al.*, 1998, Purcell and Shuman, 1998, Segal, *et al.*, 1998, Segal and Shuman, 1997, Vogel, *et al.*, 1998). The evidence for gene acquisition is the striking similarity of this gene family to the DNA conjugal transfer systems of an integrated plasmid, like the IncN plasmid pkM101, and the presence of type IV secretion systems such as those found in *Agrobacterium tumefaciens* (*virB* operon) and *B. pertussis* (*ptl* operon) (Burns, 1999, Vogel, *et al.*, 1998, Winans, *et al.*, 1996). Mutations in these genes render the bacterium either totally or partially incapable of killing human macrophages. Further analysis reveals many mutants are unable to prevent PLF, and as a consequence are unable to multiply intracellularly, do not kill host cells, and do not cause disease in animals (Wiater, *et al.*, 1998). Since PLF occurs rapidly in these mutants, within 30 minutes of bacterial uptake (Wiater, *et*

al., 1998), and *de novo* protein synthesis is not required to prevent PLF in wild type *L. pneumophila* (Horwitz and Silverstein, 1983), expression of the *icm/dot* genes must occur before the bacteria are phagocytosed. Heterogeneity in the level of expression of the *icm/dot* genes by the bacterial population prior to phagocytosis, could explain the observation for wild-type *L. pneumophila* that nearly one third fail to prevent PLF (Wiater, *et al.*, 1998). It could also explain why *L. pneumophila* grown in co-culture with various protozoa are more invasive for mammalian cells (Brieland, *et al.*, 1996, Cirillo, *et al.*, 1994). Prior co-culture would promote expression of *icm/dot* genes, increasing the proportion of bacteria in the population that would be able to prevent PLF, and display increased intracellular multiplication. To date, 13 of the 23 genes have been proven to be absolutely required for intracellular growth and killing in human macrophages. The same 13 are similarly absolutely required for intracellular growth and killing in *A. castellanii* (Segal and Shuman, 1999), indicating that intracellular growth in the two evolutionary distinct hosts utilises many of the same genes.

Sequence analysis of the *icm/dot* gene products suggests that 14 of the 23 proteins are located in the bacterial inner membrane, and four contain an ATP/GTP-binding site. This information, when linked to the fact that four of these 14 genes also exhibit striking sequence similarity to genes involved in conjugal DNA transfer, tumorogenic DNA transfer in *Agrobacterium tumefaciens*, and in the case of *B. pertussis*, the export of a toxin, has lead to the postulate that the *icm/dot* genes encode a type IV secretory system (see Fig 1.2b). This hypothesis has been supported by the observation that wild-type *L. pneumophila* can mobilise derivatives of the non-self-transmissible, heterologous IncQ plasmid RSF1010, and that this conjugation is dependent on several *icm/dot* genes (Segal, *et al.*, 1998, Vogel, *et al.*, 1998). Because of the rapidity with which viable expression of the *icm/dot* system renders bacteria able to avoid PLF, the molecule exported by this system is unlikely to be DNA, as there is insufficient time to allow DNA to be transferred and expressed (Wiater, *et al.*, 1998). It is

likely that the *icm/dot* system transfers an effector protein or proteins into the host cell, to modify the endocytic pathway in some way which prevents subsequent PLF (Segal and Shuman, 1998). Pore formation requires intimate membrane contact resulting from ligand-receptor binding and attachment (Kirby and Isberg, 1998).

Kirby, *et al.* (1998) observed that when wild-type *L. pneumophila* were mixed with macrophages in ratios >10:1, rapid cytotoxicity of the macrophages results, the result of pore formation in the eukaryotic membrane precipitating an osmotic lysis termed contact lysis (also Husmann and Johnson, 1994). Using osmoprotection of red blood cells, the pore size was estimated at 3nm in diameter (Kirby, *et al.*, 1998). It was also observed that bacteria containing mutations in several of the *icm/dot* genes were defective for this activity, and postulated that pore formation and cytotoxicity, and the *icm/dot* export system are linked, since both are dependent on the same gene expression. Thus it is possible that the effector protein/s exported by the *icm/dot* system is/are involved in pore formation in the host. Since the kinetics of high multiplicity infection-induced cytotoxicity are such that the effect occurs within minutes of contact, pore formation is possibly the first event initiating the process that eventually prevents PLF (Kirby and Isberg, 1998). During a more typical infection process, where there are similar numbers of bacteria and phagocytic cells, fewer pores are formed, sufficient to initiate the signalling which prevents PLF but insufficient to cause host cell lysis, clearly an inefficient strategy for establishing productive infection. Such a system shares some functional resemblance to YopB/YopD from *Yersinia pseudotuberculosis*, IpaB/IpaC from *Shigella flexneri*, and SipB/SipC from *Salmonella typhimurium* (Kirby, *et al.*, 1998). Which, if any, of the Icm/Dot proteins are effector proteins transported by the Icm/Dot apparatus is not known. However, Segal and Shuman (1998) have noted that one of the *icm/dot* genes (*icmL/dotI*) encodes for a product containing two amphipathic β-sheet regions, structures found in pore-forming toxins. They also note that another *icm/dot*-encoded effector candidate (*icmG/dotF*) contains a coil/coil domain, known to be important in recognition of the vesicle

and its target in the SNARE system (Hay and Scheller, 1997). The term SNARE (SNAP receptor) defines a class of vesicle trafficking proteins which act as a receptor for α-SNAP (α-soluble NSF (*N*-ethylmaleimide-sensitive factor) attachment factor) and thought to represent a form of lock-and key mechanism which allows only correctly targeted transport vesicles to dock and fuse with target membranes (Hay and Scheller, 1997). As the SNARE system is involved in a variety of fusion events, this product may interact with the SNARE system to prevent PLF. It is possible pore formation within the phagosome membrane may act as a conduit for nutrients to enter the phagosome from the host cytoplasm (Kirby, *et al.*, 1998).

Zuckman, *et al.* (1999) reported the existence of a *L. pneumophila icmW* mutant which retains the pore-forming activity found in virulent bacteria, but is defective in the inhibition of PLF and intracellular growth. They note that IcmW is one of the few potentially soluble proteins encoded within the *icm/dot* loci, and that it is not required for pore formation, but is required for intracellular growth and phagosome trafficking. However, they determined that IcmW is neither secreted not translocated, and postulate that *icm/dot* expression results in at least two separate events. Firstly, expression results in pore formation, which is essential but not on its own sufficient to enable evasion of PLF. The second event occurs either co-ordinately or subsequent to pore formation, in the form of a signal transduction event. It is this second event which involves IcmW (Zuckman, *et al.*, 1999).

High multiplicity infection-induced cytotoxicity may play a significant role in human disease. It is likely that high local concentrations of bacteria would occur within the lung, and as a consequence be cytotoxic to alveolar and bronchiolar epithelial cells, monocytes and macrophages. Rapid necrotic lysis of professional phagocytic cells *in vivo* may additionally release destructive enzymes, increasing inflammation and local tissue destruction (Kirby, *et al.*, 1998). Such an effect may account for the leukocytoclastic process characteristic of this form of pneumonia (Winn and Myerowitz, 1981).

In a recent report, Coers, *et al.* (1999) found that *L. pneumophila dot* and *icm* mutants not normally capable of intracellular replication, were able to do so within a phagosome established by a wild-type *L. pneumophila* which retained its ability to alter phagosome trafficking. Co-infection of wild-type *L. pneumophila* with mutants in the *dotA*, *dotI*, *icmW* or *icmX* genes was able to restore intracellular replication of the mutants. They concluded that *L. pneumophila* tailor the phagosome into a specialised organelle permissive for bacterial growth through the specific actions of the *icm/dot* genes. Further, they determined that in macrophages pre-infected with virulent *L. pneumophila*, subsequent infection (more than five minutes) with either heat-killed *Saccharomyces cerevisiae* or a *dotA* mutant did not alter the phagosome trafficking of *S. cerevisiae* or the *dotA* mutant. Such phagosomes containing the avirulent cells were delivered by the normal endocytic pathway to lysosomes for PLF. The authors concluded that the Icm/Dot transport apparatus does not generate a signal that accumulates and affects the trafficking of newly formed phagocytic compartments in the host cell. Using a *thyA* mutant *L. pneumophila* which behaved like a virulent wild type in the presence of thymidine, but ceased to replicate when thymidine was removed, they determined that the coordinate growth of the wild type was not required for the growth of a *dotA* mutant. They propose a model in which factors transported by the *L. pneumophila* Icm/Dot apparatus act to remodel the phagosome into a specialised organelle, and that, once this compartment has been established, the continued transport of the factors by the Dot/Icm apparatus is not required. PLF is prevented, and nutrients and additional membrane which support bacterial growth inside the vacuole are delivered by the host cell (Coers, *et al.*, 1999).

A second type IV secretion system has also been recently reported (Segal, *et al.*, 1999). This system is distinct from that encoded by the *icm/dot* loci, comprises 11 genes (termed *lvh*) and resides on a DNA island with a GC content higher than the *L. pneumophila* chromosome. It is dispensable for intracellular growth in both human macrophages and *A.*

castellanii, but is partially required for RSF1010 conjugation. Although some of the *lvh* genes can substitute for some components of the *icm/dot* system for RSF1010 conjugation, they can not substitute for components of the same system when involved with intracellular growth. The involvement of this locus in intracellular growth, and the exact nature of the inter-relationship with the *icm/dot* system, remains to be elucidated.

Legiolysin

Haemolysis has long been associated with virulence in bacteria. In 1991, researchers in Germany isolated a 39-kDa protein from *L. pneumophila*, termed Lly (legiolysin), which conferred the phenotypic characteristics of culture medium browning, probably the result of pigment production, and fluorescence (Bender, *et al.*, 1991, Wintermeyer, *et al.*, 1991). Notably, haemolysis was conferred on *E. coli* K-12 clones carrying the *lly* determinate. Although a specific *lly* probe hybridised to DNA from other *Legionella* species, at low stringency, no hybridisation was observed for other genera. The gene sequence was determined, and found to exhibit some homology to genes coding for enzymes involved with the degradation of aromatic amino acids (Wintermeyer, *et al.*, 1994). A *lly*-negative mutant lost the ability to produce brown pigment, and confer fluorescence, but retained haemolysis, and was unaffected in intracellular survival in both U937 macrophage-like cells and *A. castellanii* (Wintermeyer, *et al.*, 1994). No longer thought to be a virulence determinant, the exact role of legiolysin in the intracellular life of legionellae remains to be elucidated.

Msp (Zn^{++} metaloprotease)

All wild-type strains of *L. pneumophila* produce and secrete large amounts of a 38-kDa zinc-metalloprotease, also called the major secretory protease (Msp) (Moffat, *et al.*, 1994b). The sequence of the *proA* gene which encodes the protease was determined by Quinn and Tomkins (1989), and exhibits extensive homology with a class of bacterial neutral proteases which includes thermolysin from *Bacillus thermoproteolyticus* and elastase from

Pseudomonas aeruginosa (Black, *et al.*, 1990). The same gene was independently reported, and termed *mspA* (Szeto and Shuman, 1990). The *mspA/proA* gene contains a single 1,629-bp open reading frame (ORF), giving a predicted polypeptide of 543 amino acids with a computed molecular mass of approximately 60-kDa. A signal sequence was identified within the first 16 N-terminal amino acids. The mature protease sequence was determined to begin at residue Glu-208. Tn*phoA* fusions indicated that the entire *mspA/proA* ORF is transcribed and translated, including the 5' leader sequence, suggesting the entire polypeptide is transported to the periplasm before cleavage at Glu-208 to produce the mature protease (Moffat, *et al.*, 1994a). The disparity between ORF size and mature protein length is typical of this class of proteases. Site directed mutation determined that Glu-378 is located in the putative active site and is critical for proteolytic activity and cytotoxicity of CHO cell monolayers (Moffat, *et al.*, 1994a). The *proA* gene from *L. longbeachae* has also been sequenced (GeneBank X83035), giving a predicted polypeptide of 529 amino acids, 71% homologous to that for *L. pneumophila*. The equivalent of Glu-208 and Glu-378, namely Glu-205 and Glu-366 respectively, are conserved, as are large regions in the C-terminal region of the protein in the proximity of Glu-366.

Consistent with this evidence of the periplasmic maturation of Msp before export across the outer membrane is the recent evidence of Liles, *et al.* (1999) that export of Msp is dependent on prepilin peptidase. Prepilin peptidases cleave the leader sequences from prepilin-like proteins as part of the type II secretion pathway. A *pilD*-deficient *L. pneumophila* mutant was unable to export Msp or exhibit haemolytic activity. Functionality was restored with the reintroduction of *pilD*. Interestingly, the *pilD*-deficient *L. pneumophila* mutant was also impaired in its ability to grow in *H. vermiformis* and macrophage-like U937 cells. In addition, the mutant was 100-fold less pathogenic in a guinea pig virulence model. As at least two additional proteins were unable to be exported by the mutant, the reduced virulence of the *pilD*-deficient *L. pneumophila* mutant can not be attributed directly to the effect of Msp.

However, the results are the first evidence of a significant role for type II secretion in intracellular parasitism. Similar findings of inhibition of Msp excretion and impaired ability to multiply in *A. castellanii* have also been more recently reported for a strain with mutations in two of the *lsp* genes (*lspGH*), members of the family of genes encoding the type II secretion pathway (Hales and Shuman, 1999). They also attribute the involvement of the type II secretion system in intracellular growth to the export of proteins other than Msp.

In 1989, Blander and Horowitz (1989) demonstrated that guinea pigs sub-lethally infected with *L. pneumophila* develop a strong cell-mediated immune response to Msp and when immunised with Msp developed both humoral and cell-mediated immune responses to Msp. These protected against lethal aerosol challenges with *L. pneumophila* and limited *L. pneumophila* multiplication in their lungs. However, two subsequent studies with protease-deficient *L. pneumophila* mutants failed to demonstrate an effect on intracellular replication or disease measured in a guinea pig model of infection (Blander, *et al.*, 1990, Szeto and Shuman, 1990). Expression of Msp during infection was demonstrated in human alveolar macrophages with immuno-gold studies, with the enzyme being detected both within phagosomes and distributed throughout the cell (Rechnitzer, *et al.*, 1992). Although the authors did not comment on observing any cytopathic effects, they subsequently reported a reduction of chemotactic activity of human neutrophils, but not human monocytes (Rechnitzer and Kharazmi, 1992). Moffat, *et al.* (1994b) were also unable to demonstrate an effect on the uptake, survival or multiplication of a similar mutant *L. pneumophila* in *A. castellanii*, or explanted guinea-pig macrophages. However, they were able to demonstrate attenuation of virulence in a guinea-pig model employing intratracheal inoculation, and postulated a role in the acute necrotizing inflammatory response associated with the disease.

As has been already mentioned, Msp has alternately been postulated to have a role in the process of cell death and bacterial release. This hypothesis has been discounted by

Blander, *et al.* (1990) who determined the null mutations in the protease structural gene do not effect the cytopathic effect of the wild-type *L. pneumophila*. Similarly, a *mspA/proA* mutant strain was also determined to be able to induce apoptosis in HL-60 macrophage-like cells, coincidentally also demonstrating that the pleiotropic cytokine TNF-α, which is cleaved by Msp, is not the apoptotic death trigger (Hell, *et al.*, 1993, Muller, *et al.*, 1996).

Recently, James, *et al.* (1997) demonstrated that the virulence of a clinical isolate of *L. pneumophila* was significantly attenuated when cultured in an iron-limited environment and noted that expression of Msp was reduced five-fold in response to iron depletion. They also demonstrated that iron-loaded transferrin enhanced the growth of steady-state, iron-limited cultures, and postulated that transferrin represents a potentially important iron source for *L. pneumophila in vivo*. Although unable to detect cell surface transferrin receptors, they were able to demonstrate *in vitro* that transferrin was digested by Msp activity in the culture supernatants. Siderophores were not produced under these culture conditions.

To date, no clear function for Msp in intracellular life has been determined, and its exact role in *L. pneumophila* infection remains to be elucidated.

Lipopolysaccharide

Lipopolysaccharide (LPS) is an important virulence determinant for many Gram negative bacteria (Zhang, *et al.*, 1998). Gabay and Horwitz (1985) extracted LPS from *L. pneumophila* and found that it was the major antigen recognised by patient sera in the indirect fluorescent-antibody assay. LPS has been reported in other *Legionella* species (Ciesielski, *et al.*, 1986, Sonesson, *et al.*, 1993, Sonesson, *et al.*, 1994a, Sonesson, *et al.*, 1994b, Sonesson, *et al.*, 1994c, Sonesson, *et al.*, 1994d). However, *Legionella* LPS lacks the typical endotoxicity associated with the LPS of many Gram negative bacilli. It is deficient in the hydroxy-fatty acids generally associated with the endotoxic structural components of lipid A,

and as a consequence it is only a weak inducer of the pyrogenic response in rabbits, and exhibits a relatively low toxicity in mice (Wong, *et al.*, 1979). In addition, the patterns of LPS from virulent and avirulent strains of the same serogroups on SDS-PAGE are similar (Conlan and Ashworth, 1986). As a consequence, it is unlikely that LPS contributes significantly to the virulence of *L. pneumophila* (Dowling, *et al.*, 1992).

The association of flagella with disease

The ability of bacteria such as *Vibrio cholerae* and *Pseudomonas* spp. to possess enhanced colonising ability and pathogenicity as a consequence of the expression of flagella and the resultant motility has been demonstrated (Attridge and Rowley, 1983, De Weger, *et al.*, 1987). *L. pneumophila* have been shown to possess one or more polar or subpolar flagella with a 47-kDa filament sub-unit (Chandler, *et al.*, 1980, Elliott and Johnson, 1981, Fallon, 1990, Rodgers, *et al.*, 1980), and that a common flagellum antigen is present in most if not all *Legionella* species (Bornstein, *et al.*, 1991). The *flaA* gene from both *L. pneumophila* and *L. micdadei* has been sequenced and the polypeptide structure of the flagellin subunit predicted (Bangsborg, *et al.*, 1995, Heuner, *et al.*, 1995). Considerable similarity with the flagellin subunits of other bacteria is observed especially in the *C*-terminal and *N*-terminal regions.

To determine the role of flagella expression in infection, Pruckler, *et al.* (1995) produced flagella-deficient *L. pneumophila* mutants by both insertion mutation or selection of spontaneous mutants, and studied their ability to infect and multiply in amoebae and a human monocyte-like cell line U937. The ability of the mutants to do so was varied, and indicated that while flagella expression was not essential for virulence, some mutants, and especially the insertion mutants, were attenuated in their ability to multiply in co-culture with amoebae, and unable to do so in U937 cells. The authors suggest that the regulatory systems for flagella expression are more profoundly affected in the insertion mutants than those occurring naturally, and postulate a coordinated regulation of flagella production with other genes

expressed during infection. They cite the bi-phasic nature of intracellular growth and flagella production as evidence, and note that the switch from the non-motile "multiplicative phase" to the highly motile "active infective phase" occurs at a well defined point while inside the host cell, just before release. Highly motile bacteria would theoretically be more capable of becoming associated with, and possibly attaching to, a new host cell, than non-motile ones, and be more effective in sustaining the infectious process. Since the synthesis and assembly of flagella and the associated motility systems in *E. coli* and *Salmonella typhimurium* are complex and controlled by approximately 40 genes grouped into three transcriptional levels, each with its own regulatory system (Jones and Aizawa, 1991), coordinated expression of *L. pneumophila* flagella, which is likely to be similarly complex, and other genes associated with infection is highly possible.

Recently, Heuner, *et al.* (1999) reported that *flaA* expression is not only temperature regulated, but is also influenced by growth phase, the viscosity and osmolarity of the medium and down-regulated by the presence of the amino acids serine and threonine. They also conclude that motility may be controlled to enable the legionellae to escape the spent host and locate another, in order to survive in an aquatic environment.

Type IV fimbriae

Various pathogenic bacteria have been shown to possess short fimbrial structures on their cell surface, which facilitate adhesion to eukaryotic cells and are required for or enhance virulence (Hultgren, *et al.*, 1993, Strom and Lory, 1993). *L. pneumophila* contains a type II secretion system that encodes the synthesis of a type IV pilus, possessing homologues of both the type IV pilus biosynthetic genes *pilBCD*, and the type IV pilin gene *pilE* (Liles, *et al.*, 1998, Stone and Abu Kwaik, 1998). A number of other *Legionella* species have been shown by Southern hybridisation to possess similar homologues (Stone and Abu Kwaik, 1998). Electron microscopy has shown that *L. pneumophila* express pili of variable lengths, either

long (0.8 to 1.5 μm) or short (0.1 to 0.6 μm) (Rodgers, *et al.*, 1980, Stone and Abu Kwaik, 1998). A mutation in the *pilE* gene of *L. pneumophila* resulted in the loss of expression of the long pilus, and a two-fold decrease in adhesion to amoebae and macrophages (Stone and Abu Kwaik, 1998). The mutant was still able to replicate at rates equivalent to the wild-type, indicating that the type IV pilus is not required for intracellular growth. Since the presence of type IV pili may provide an advantage for colonisation of lung tissue, the variable expression among *Legionella* species (present in all *L. pneumophila* serogroups but variable among other species) may account for the different frequencies of association with human disease. However, there appears to be no correlation between the species determined to possess type IV pili and those associated with human disease.

Major outer membrane proteins

A major outer membrane protein (MOMP) has been reported for *L. pneumophila*, but the specific details of its structure are unclear. A 29-kDa protein, common to the genus, was reported by Ehret and Ruckdeschel (1985a, 1985b, 1985c) and shown to be a component of an aggregate of 24- to 29-kDa subunits stabilised by disulphide bonds to form a 95-kDa protein complex (Butler, *et al.*, 1985). The complex was reported to be structurally similar to the MOMP found in *Chlamydia* (Newhall and Jones, 1983). Hoffman, *et al.* (1992b) reported an alternate finding of a MOMP complex comprising 70- and 120-kDa proteins, which behaves similar to an *E. coli*-like porin, based on its folding behaviour and ability to form membrane channels (Gabay, *et al.*, 1985). The nucleotide and predicted amino acid sequences of two outer membrane proteins have been reported, a 25-kDa protein coded for by the *ompM* gene (High, *et al.*, 1993), and a 28- to 31-kDa *ompS*-encoded protein (Hoffman, *et al.*, 1992a). The latter protein is reported to be an anchor protein involving attachment of peptidoglycan. The *ompM*-encoded 25-kDa protein was confirmed to be a component of MOMP by similarities in its silver staining characteristics, poor migration through polyacrylamide in the absence of a strong reducing agent such as β-mercaptoethanol, and immunoblotting studies using a

MOMP-specific monoclonal antibody (High, *et al.*, 1993). The poor migration through polyacrylamide without the presence of a powerful reducing agent such as β-mercaptoethanol was attributed to nine cysteine residues present in the deduced amino acid sequence. The authors reported that the *ompM* and *ompS* sequences showed no sequence similarity to each other, or to any other prokaryotic genes. However, a careful examination of the sequences does show short regions of reasonable similarity, which may be coincidental, but may indicate a degree of functional relatedness (R. Doyle, 2000, personal communication). The authors suggest that the size discrepancies of the proteins comprising the MOMP described by Ehret and Ruckdeschel (1985) and Butler, *et al.*(1985) could be due to amounts of LPS and peptidoglycan still complexed to the MOMP under the different extraction conditions, citing that *L. pneumophila* MOMP is known to be tightly bound to LPS (Gabay, *et al.*, 1985, Hindahl and Iglewski, 1986). MOMP-like structures could be detected in other serotypes of *L. pneumophila* (Hindahl and Iglewski, 1986) and other *Legionella* species (Butler, *et al.*, 1985), using immunologic studies. However, primers specific for the published *ompM* gene sequence failed to amplify similar genes when tested on DNA extracted from a number of other *L. pneumophila* and *L. longbeachae* strains (R. Doyle, personal communication, 1999). Similarly, whole cell protein extracts failed to demonstrate a MOMP-like homologue in *L. longbeachae*, even in the presence of β-mercaptoethanol. Primers specific for the published *ompS* gene did produce an amplicon of the expected size in *L. pneumophila* sg 1 (Philadelphia-1 strain) but not other species. However, homologues of the *ompS* gene could be detected by Southern hybridisation in *L. micdadei* and several *L. longbeachae* strains using the above amplicon as a probe at low stringency (approximately 30% base pair mismatch) (R. Doyle, personal communication, 1999).

Krinos *et al.* (1999) recently reported that electroporation of the *ompM* gene into a low virulence derivative of *L. pneumophila* which demonstrated reduced expression of MOMP but enhanced expression of the 31-kDa protein, increased expression of MOMP and eliminated

the expression of the 31-kDa protein. No explanation was given for loss of expression of the 31-kDa protein, and although the authors assumed it was the same *ompS*-encoded protein reported by Hoffman *et al.,* (1992a) this was not determined experimentally.

The association of MOMP with virulence is based on two observations. Firstly, an *E. coli* clone expressing the 25-kDa *ompM*-encoded protein showed increased virulence in the fertile chicken egg virulence assay (High, *et al.*, 1993). Secondly, Krinos *et al.* (1999) showed that following the electroporation of the *ompM* gene into the low virulence derivative of *L. pneumophila* mentioned above, expression of MOMP increased the virulence of the strain in a chick embryo virulence assay. The increased virulence and cell adhesion could be eliminated in the presence of MOMP-specific monoclonal antibody. It has been postulated that *L. pneumophila* MOMP binds the complement component C3, mediating opsonin-dependent phagocytosis by human monocytes and macrophages (Bellinger-Kawahara and Horwitz, 1990). The findings of Krinos *et al.* (1999) suggest that the 25-kDa *ompM*-encoded protein serves as an adhesive molecule for host cells. Given the potential importance of MOMP in contributing to the intracellular virulence of *L. pneumophila* and perhaps other *Legionella* species, it is surprising that so little has been reported clarifying its role in the last five years.

Heat shock proteins

Among the few surface exposed proteins identified for *L. pneumophila* is the GroEL homologue Hsp60, an essential heat shock protein (Garduño, *et al.*, 1998a). Hsp60 synthesis in virulent *L. pneumophila* is up-regulated following association with host cells, resulting in an increase in both the amount of Hsp60 which is surface exposed, and also released into newly formed and mature phagosomes (Fernandez, *et al.*, 1996, Garduño, *et al.*, 1998a, Hoffman, *et al.*, 1990). In contrast, Hsp60 up-regulation or release in avirulent *L. pneumophila* does not occur. In addition, *L. pneumophila* Hsp60 has been shown to induce synthesis of interleukin-1β in macrophages through a mechanism that involves ligand-

receptor interactions in the absence of Hsp60 internalization (Retzlaff, *et al.*, 1996). Recently, the same researchers demonstrated that surface exposed Hsp60 in virulent *L. pneumophila* acted as an adhesin-invasin and mediated the internalization and unique trafficing of latex particles in HeLa cells (Garduño, *et al.*, 1998b). They also observed that endosomes containing Hsp60-coated beads did not fuse with secondary lysosomes similar to endosomes containing virulent *L. pneumophila*. In contrast, endosomes containing virulent *L. pneumophila*, and endosomes containing Hsp60-coated beads, did not become associated with the ER, implying that such association is determined by other factors. The authors determined that basal levels of Hsp60 in both virulent and avirulent *L. pneumophila* are similar, and that avirulent strains appear to be defective not in Hsp60 synthesis, but in the ability to transport Hsp60 across the outer membrane and thus can not display the protein on the bacterial surface, or release it from the cell. They postulate that Hsp60 may be the effector protein transported by the *icm/dot* system, even though it is not encoded within either region I or II of the icm/*dot* system. This postulate is not supported by experimental evidence to date.

Macrophage Infectivity Potentiator (Mip) Protein

The existence of a 24-kDa protein surface antigen of *L. pneumophila* was reported as early as 1984, as researchers began to investigate the genetic basis of virulence (Engleberg, *et al.*, 1984a, Engleberg, *et al.*, 1984b, Pearlman, *et al.*, 1985). Cianciotto, *et al.* (1989b) later reported that a site-directed mutation in the gene produced a mutant 80-fold less infective for macrophages compared to the wild-type. Consequently, the gene responsible was designated *mip* (*m*acrophage *i*nfectivity *p*otentiator). Engleberg *et al.* (1989) reported that the *mip* gene was 699 base pairs in length, corresponding to an inferred protein of 233 amino acids. A putative transcriptional start site was determined, and probable promoter consensus sequences upstream. Downstream, a region of dyad symmetry typical of a factor-independent transcriptional terminator was recognised. The inferred polypeptide sequence suggested the presence of a secretory signal sequence of 20 residues, which was confirmed by *N*-terminal

sequencing of the mature protein. Putative secondary structure and hydropathy analysis predicted a 60 residue, mildly hydrophilic, alpha-helix in the *N*-terminal half of the mature protein, with beta sheets and turns, and increased hyrophobicity, for the remaining *C*-terminal portion. They proposed the protein could act as a potent polycation, with a predicted pI of 9.8, and postulated a surface located polycationic protein with the 60 residue alpha-helix forming an elongated arm projecting from the cell surface (Engleberg, *et al.*, 1989). By Southern hybridisation, it was determined that the *mip* gene was present in all serogroups of *L. pneumophila*, and *mip*-like genes were present in the 29 other *Legionella* species tested (Cianciotto, *et al.*, 1990a), but reduced stringency conditions were required to detect the gene in species other than *L. pneumophila*. Immunoblot studies also revealed a Mip-related protein with sizes in the range of 24- to 31-kDa among the other species.

The *mip* gene sequences for *L. micdadei* and *L. longbeachae* were also determined (Bangsborg, *et al.*, 1991, Doyle, *et al.*, 1998). Significant homology with respect to *L. pneumophila mip* existed in the sequences; 71% for *L. micdadei* and 79% for *L. longbeachae.* Interestingly, the inferred amino acid sequence for *L. micdadei* Mip was larger (25.93-kDa) and contained 10 additional residues following the signal sequence, consistent with the larger size predicted from the immunoblot studies (Cianciotto, *et al.*, 1990a). Additionally, Mip from *L. longbeachae* and *L. pneumophila* demonstrate aberrant migration on SDS-PAGE gels (Doyle, 1999, personal communication). *L. longbeachae* Mip is a lower molecular weight (one reside smaller) but migrates more slowly than *L. pneumophila*. The inferred amino acid sequence of *L. longbeachae* Mip is slightly more negatively charged, which may account for the mobility differences.

Studies using knockout-mutations in *mip* in *L. pneumophila*, *L. micdadei* and *L. longbeachae* have shown Mip$^-$ mutants to be partly defective in facilitating intracellular infection in monocytes, macrophages, alveolar epithelial cells and protozoa, and are

attenuated in an animal model of virulence (Cianciotto, *et al.*, 1989b, Cianciotto, *et al.*, 1990b, Cianciotto, *et al.*, 1995a, Doyle, *et al.*, 1998, O'Connell, *et al.*, 1995). To elucidate the role of Mip in intracellular infection, Cianciotto and Fields determined that *mip* mutants of *L. pneumophila* associate with *H. vermiformis* cells in comparable numbers to that for wild-type bacteria, suggesting that Mip was not involved in the attachment to and uptake into the host cell (Cianciotto and Fields, 1992). Similarly, Mip does not appear to have a role in uptake of *L. pneumophila* into macrophages (Cianciotto, *et al.*, 1989b). Despite a 20-1000-fold reduction in viable cells following uptake, depending on the bacterial species and type of host cell, and a prolonged lag phase, *L. pneumophila, L. longbeachae* and *L. micdadei mip* mutants multiply at rates often comparable to that of the wild-type (Cianciotto, *et al.*, 1989b, Doyle, *et al.*, 1989, Doyle, 2000, personal communication, O'Connell, *et al.*, 1995). Thus, the common phenotype of *mip* mutants is the reduced ability to survive immediately after entry into the host cell. More recently, a mutant strain of *Salmonella enterica* serovar Typhimurium, with a defective *fkpA* gene, which encodes a homologue of *mip*, similarly show reduced ability to survive intracellularly in a macrophage-like cell line (Horne, *et al.*, 1997). In apparent conflict with these findings is the demonstration that Mip is one of three antigens upregulated during intracellular infection, being maximally expressed four to eight hours after uptake (Susa, *et al.*, 1996). To be consistent with the findings from the mutational studies, and in agreement with the observation that intracellular-grown *L. pneumophila* are more invasive in establishing a new intracellular infection, upregulation of *mip* expression could be part of the increased invasiveness which promotes the re-establishment of a new intracellular infection.

Mip exhibits peptidyl-prolyl *cis/trans* isomerase (PPIase) activity of the FK506 Binding Protein (FKBP) class (Fischer, *et al.*, 1992). FKBPs are one of two protein PPIase super families, the other being Cyclophilins, that constitute the immunophilin class (Hacker and Fischer, 1993), so named because of their ability to bind immunosuppressive peptide derivatives. FKPBs bind an experimental compound FK506, and cyclophilins are capable of

binding to the cyclic undecapeptide cyclosporin A. Recently, a third class of PPIases termed parvulins, has been described (Rahfeld, *et al.*, 1994a, Rahfeld, *et al.*, 1994b, Rahfeld, *et al.*, 1995, Rudd, *et al.*, 1996). PPIase activity catalyses the slow conformational interconversion around the angle ω of the imidic bond of proline in oligopeptides, bringing about conformational change to the secondary structure of the target protein. Although sharing some functional similarity in acting as PPIases, all three classes are structurally dissimilar, and their amino acid sequences do not exhibit any noticeable homology to each other (Hacker and Fischer, 1993, Rahfeld, *et al.*, 1994a).

Immunophilins are housekeeping proteins widely distributed in both procaryotes and eucaryotes, and FKBPs and cyclophilins seem to assist the folding and maturation of proteins; for example, transferrin and collagen in the mammalian endoplasmic reticulum, the translocation of mitochondrial precursor proteins, and rhodopsin in the *Drosophila* eye (Wintermeyer, *et al.*, 1995). In the ligand-bound form, complexed with their respective immunosuppressive peptide, and also to a significantly reduced extent in the unbound form, these proteins are additionally known to interfere with signal transduction in T cells, by interacting with the protein phosphatase calcineurin, thereby inhibiting the nuclear transport of the transcriptional factor NF-AT, a critical step in T-cell activation (Schreiber, 1991). Gel-permeation chromatography, SDS-PAGE, small angle X-ray solution scattering and enzyme activity studies have revealed that *L. pneumophila* Mip is principally active as a homodimer, with the contact regions between the two monomers probably located between the *N*-terminal regions (Schmidt, *et al.*, 1994, Schmidt, *et al.*, 1995).

Mip-like analogues, some of which have been demonstrated to exhibit PPIase activity of the FKBP class, have been detected in many procaryotes, such as *Neisseria meningitidis*, *Chlamydia trachomatis*, *Coxiella burnetii*, *Bartonella* (*Rochalimaea*) *quintana*, *Erwinia chrysanthemi*, *Pseudomonas aeruginosa*, *Escherichia coli*, *Aeromonas hydrophila*,

Salmonella enterica serovar Typhimurium as well as other *Enterobacteriaceae*, and also in eucaryotic cells such as *Neurospora crassa*, *Saccharomyces cerevisiae*, *Trypanosoma cruzei*, bovine calf thymus cells, human Jurkat T-cells and a human cancer cell line (Cianciotto, *et al.*, 1995b, Horne and Young, 1995, Horne, *et al.*, 1997, Lundemose, *et al.*, 1992, Mo, *et al.*, 1995, Moro, *et al.*, 1995, Pahl and Keller, 1992, Pissavin and Hugouvieux-Cotte-Pattat, 1997, Rahfeld, *et al.*, 1994b, Rahfeld, *et al.*, 1996, Rockey, *et al.*, 1996, Wong, *et al.*, 1997).

The C-terminal region of Mip shows significant homology to FKBPs from other procaryotes and eukaryotes. In addition, 10 residues have been determined to be involved with the interaction of human FKBP and FK506. Many of these are conserved in microbial FKBPs; six for *L. pneumophila* Mip (Hacker and Fischer, 1993) (Note: the reference states erroneously that eight out of ten are conserved, but the alignment included reveals only six. The error has been confirmed by the authors (personal communication, 1997)). By using site directed mutagenasis, Ludwig, *et al.* (1994) were able to show that three of these residues conserved between human and *L. pneumophila* FKBPs, namely Asp-162, Tyr-205 and Phe-222, as well as an additional residue, Phe-161 (also conserved in human FKBP) are involved in the enzymic-activity and probably binding of FK506 by *L. pneumophila* Mip. Since some of these residues are separated from each other by as many as 22 residues, it is postulated that this region of the protein, which contains a number of β sheets, is substantially folded to bring the residues into proximity with each other to form the active PPIase site (Hacker and Fischer, 1993). These similarities confirm that there are strong structural and functional similarities for all FKBPs. Hacker and Fischer have postulated that *L. pneumophila* Mip is anchored to the outer surface membrane by the *N*-terminal region, with the *C*-terminal active region projected from the surface by the large α-helical domain, to bridge the distance between the bacteria and putative partner molecules (Hacker and Fischer, 1993). While monoclonal antibodies specific for epitopes in close proximity to the PPIase site will react with whole cells, suggesting that the active site is surface exposed (Helbig, *et al.*, 1995), the more traditional

model proposed by Engleberg, *et al.* (1989), where the hydrophobic β-sheet regions form a series of trans-membrane domains, and the α-helical section projects from the cell surface, is more likely. Wintermeyer, *et al.* (1995) assessed the ability of *L. pneumophila* mutants, expressing only altered Mip proteins with little residual PPIase function, to survive intracellularly in epithelial cells, monocytes and *A. castellanii*. These studies suggested that the mutants behaved like the wild-type, with no reduction in the ability to survive intracellularly, further implying that PPIase activity is not associated with Mip-dependent intracellular survival. One possible explanation is that PPIase activity is being supplied by a second protein or protein complex. An 18-kDa cyclophilin exhibiting PPIase activity has been reported for *L. pneumophila* (Schmidt, *et al.*, 1996). However, if the cyclophilin was acting as a helper, the ability of *mip* mutants to survive would be similarly preserved, but this does not occur.

Thus the role of Mip in intracellular life, its mode of action, and host target has not been elucidated to date. Mip may be acting as a foldase in the conformational maturation of other *Legionella* derived proteins involved with uptake by the host and intracellular survival mechanisms. Other possible functions include the denaturation of host proteins or structures involved with host defence, perhaps in the cell wall or lysosome vacuole membrane, or in targeting host receptors or polypeptides, which trigger events inhibiting normal host defence mechanisms (Hacker and Fischer, 1993). Conversely, since it is not even known if *L. pneumophila* Mip is acting in a manner similar to that of other PPIases in this ubiquitous, important class of proteins, it may have an entirely novel role in the intracellular life of legionellae.

Virulence traits of Legionella *species other than* L. pneumophila.

While *L. pneumophila* is the most reported species causing Legionnaires' disease, other species, predominantly *L. micdadei* and *L. longbeachae*, have been associated with

disease (Table 1.1). Only a few studies of the virulence mechanisms of these species compared to those of *L. pneumophila* have been performed.

Joshi and Swanson (1999) compared four clinical isolates of *L micdadei* with a well characterised strain of *L. pneumophila*, assessing differences among the strains for seven virulence traits, namely infection and replication in cultured macrophages, evasion of PLF, contact-dependent cytopathicity, sodium sensitivity, osmotic resistance and conjugal DNA transfer. Three of the four *L. micdadei* strains were less virulent by nearly every measure. However, one *L. micdadei* strain was a surprising exception, evading lysosomes and replicating in macrophages as efficiently as *L. pneumophila*, despite lacking both contact-dependent cytopathicity and regulated sodium sensitivity. The authors note that *L. micdadei* lacks phospholipase C, and Msp, although the latter, based on the work of Dowling, *et al.* (1992) may be erroneous. Sequence showing significant similarity to the *mspA/proA* gene has been detected in *L. micdadei* although the expression and functionality of its product is not known (see Chapter Five). They also note that internalisation of *L. micdadei* via coiling phagocytosis and association with the host ER has never been observed, including the exceptional *L. micdadei* strain. The authors conclude that the differential virulence between *L. micdadei* and *L. pneumophila* is not the result of a single phenotypic or genetic trait, and postulate that *Legionella* is a diverse genus and different species including *L. micdadei* utilise different strategies to parasitise host cells.

Chapter One (continued)

Introduction to bacterial classification and phylogeny

Because of the broad nature of the issues dealt with in this study in relation to *Legionella*, it is necessary to extend the introduction to cover additional topics. Some of the material included is a more basic introduction than in the previous sections, as these subjects are not as widely understood.

Identification of *Legionella*

The identification of most clinically significant bacteria utilizes their ability to exhibit various biochemical abilities as a result of their genetic capability (Berdal and Olsvik, 1983, Mauchline and Keevil, 1991, O'Brien, *et al.*, 1983). Commonly this is based on the presence of enzymes enabling the degradation of various substrates to form a detectable product such as the fermentation of various sugars producing a reduction in pH. It may also include the ability of an organism to grow in the presence or absence of certain compounds, such as the presence of high salt, citrate as the sole source of carbon, or aerobic or anaerobic atmospheres. However, apart from a requirement for cysteine, which is used as a presumptive identification test, *Legionella* are considered biochemically inert by such identification systems commonly found in most laboratories (Hookey, *et al.*, 1996, Maiwald, *et al.*, 1998, Winn, 1995). Commercial identification systems are similarly of little use, although the expensive BIOLOG substrate utilisation system (BIOLOG Inc., Hayward, USA) was used successfully after extensive modification (Mauchline and Keevil, 1991).

By far the most successful identification technique involves the serological characterisation of isolated strains (Maiwald, *et al.*, 1998). Monoclonal antibodies (Mabs) have been used to identify strains to genus level, or as *L. pneumophila* in the case of a commercially available Mab recognising an outer membrane protein (Gosting, *et al.*, 1984). The genus-specific Mabs recognise the 58- to 60-kDa heat-shock protein (Steinmetz, *et al.*, 1991) or Mip (Helbig, *et al.*, 1995). Polyclonal serogroup specificity for *Legionella* is primarily based on LPS epitopes (Conlan and Ashworth, 1986). The reactivity of such polyclonal antisera, usually extensively cross-absorbed to increase specificity, has enabled legionellae to be sero-grouped. *L. pneumophila* is the most complex species serologically, (currently 15 serotypes), and several other species, including *L. longbeachae*, *L. feeleii* and *L. bozemanii* have at least two serogroups. The anti-sera is used in agglutination or ELISA reactions, or fluorescein-conjugated in fluorescent microscopy. However, polyclonal antisera has been reported to cross-react with a number of other bacteria, such as *Bordetella* (Benson, *et al.*, 1987), *Francisella tularensis* (Roy, *et al.*, 1989), *Pseudomonas* (Edelstein and Edelstein, 1989), *Bacteroides fragilis* (Edelstein and Edelstein, 1989), and *Corynebacterium* (Summersgill and Snyder, 1990). In addition, many species and sero-groups show considerable cross-reactivity even after substantial cross-absorption, making the definitive identification of some wild strains impossible (Harrison and Taylor, 1988). Further, a number of currently unclassified new species are known to exist (Wilkinson, *et al.*, 1990). Until the sero-specificity of such strains is known, wild strains from these species are likely to be miss-identified, and even then cross-reactivity may prevent speciation. As an example, red-pigmented strains of *L. taurinensis* are serologically indistinguishable from *L. spiritensis* (V. Drasar, 1998, personal communication) whereas DNA homology studies indicate they are related to *L. rubrilucens* (Lo Presti, *et al.*, 1999).

As a result of such difficulties, other workers have investigated the use of chromatographic finger-printing, and developed profiles of fatty acids and ubiquinones for each species (Lambert and Moss, 1989, Wilkinson, *et al.*, 1990). Originally promising when approximately 20 species were classified, the apparently unique profile for some species was determined to be no longer clearly distinguishable from the newly characterised species, especially when wild-strain variation was included (Diogo, *et al.*, 1999, Wilkinson, *et al.*, 1990). The methods, and especially fatty acid analysis using gas chromatography, were also shown to be affected by the preparative methods and equipment used (Wilkinson, *et al.*, 1990), necessitating each laboratory to prepare its own profiles. Such a task required access to all type strains and a comprehensive range of wild strains, was expensive and very time consuming. The inclusion of hydroxylated fatty acids improved the discrimination, but as this required the analysis of both mono-and di-hydroxylated fatty acids, the extraction methods and chromatography are even more time consuming, and the resulting profiles so complex that computer-assisted analysis is advised (Diogo, *et al.*, 1999, Jantzen, *et al.*, 1993).

As molecular techniques became more feasible, such techniques have been evaluated for *Legionella*. Typing using allozyme profiles, ribotyping, RFLP and PFGE were all found to be very useful, but primarily at a sub-species level, when determining the relationships between strains during epidemiological studies, and such techniques remain extremely valuable (Haertl and Bandlow, 1991, Harrison, *et al.*, 1990, Lanser, *et al.*, 1992, Riffard, *et al.*, 1998, Schoonmaker, *et al.*, 1992). Recently, a method utilising random amplified polymorphic DNA (RAPD) analysis was reported (Bansal and McDonell, 1997, Lo Presti, *et al.*, 1998b). Also used predominantly in intraspecies strain discrimination, its successful use in species identification has been reported for such taxa as *Leptospira*, *Candida* and *Enterococcus* (Descheemaeker, *et al.*, 1997, Lehman, *et al.*, 1992, Ralph, *et al.*, 1993). However, Lo Presti, *et al.* (1998) reported that the method of Bansal and McDonnell (1997) required optimisation to obtain reproducibility with the previously published profiles,

necessitating the availability of all the type strains and many wild strains. RAPD analysis is quick and economical to perform, but its robustness in other laboratories for the identification of legionellae is uncertain. Recent reviews question both the reproducibility and robustness of RAPD compared with other methods (Jones, *et al.*, 1997, Koeleman, *et al.*, 1997). A more reliable method, termed amplified-fragment length polymorphism (AFLP) has been reported for quite a variety of life forms, including bacteria, fungi, parasites, plants and animals (Savelkoul, *et al.*, 1999). The method is analogous to RFLP, as it involves digestion of genomic DNA with restriction endonucleases (RE). However, following digestion, oligonucleotide adapters are ligated to each end of the restriction fragments, and the fragments amplified by stringent amplification before the ladder of different-sized fragments are visualised by staining after polyacrylamide gel electrophoresis. This method has been used for typing *L. pneumophila* strains (Valsangiacomo, *et al.*, 1995) but its usfulness for typing other species, or to speciate strains is unreported.

Gene-sequencing methods have become more widely available, economical and simple, following the introduction of improved sequencing methods such as PCR amplicon sequencing with dye terminator chemistry. In turn, this has increased the amount of published sequence, making sequence-based identification the method of choice for some taxa which are difficult to identify by traditional biochemical methods (Amann, *et al.*, 1988, Collins, *et al.*, 1994, Dewhirst, *et al.*, 1992, Lehman, *et al.*, 1992, Ludwig and Schleifer, 1994, Pascual, *et al.*, 1995, Ruimy, *et al.*, 1994, Sallen, *et al.*, 1996). Genotypic schemes have the great advantage of being unaffected by colony age and growth conditions, and in contrast to chromatographic methods, also unaffected by preparative methods, analysis conditions and type of equipment used. Additionally, because a gene sequence is essentially a long digital string, with each digit being one of only 4 nucleotides, genotypic schemes are less ambiguous and can utilise significantly more discriminating data than phenotypic ones, and in a form that lends itself to widely available computer analysis software. Finally, the access to increasingly

comprehensive sequence data bases such as GeneBank via the internet, means that the maintenance of personal databases and analysis software is no longer essential.

Many genotypic schemes utilise variation in the *rRNA* genes, targeting the conserved regions with universal primers, to amplify and sequence regions of variability between the conserved regions (Collins, *et al.*, 1994, Dewhirst, *et al.*, 1992, Ludwig and Schleifer, 1994, Ruimy, *et al.*, 1994, Sallen, *et al.*, 1996). Such regions of high conservation and increased variation reflect regions of the ribosome affected by different rates of functional constraint on mutation events. However, the regions of high conservation predispose *rRNA* genes to homologous recombination events leading to horizontal gene transfer (Strätz, *et al.*, 1996). In addition, ribosomal RNAs are encoded by tandem genes, a rare phenomenon in bacteria. As a consequence, recombination events which produce non- or less-functional gene products that are fatal within other genes can be tolerated within the *rRNA* genes, increasing the frequency of recombinational events (Ambler, 1996). Thus more than one species within a genus may possess the same *rRNA* gene sequence, and very recent gene transfer events would cause erroneous identification of such recipient strains. The practical reality of the latter erroneous identification is unknown, but remains a theoretical possibility.

In contrast, protein-encoding genes are less constrained from mutational events as a consequence of the redundancy in the genetic code, especially at the third codon position. Thus sequences from protein-encoding genes are likely to contain more genetically informative sites. In addition, they are less likely to be involved in homologous recombination events compared to *rRNA* genes as they usually lack the characteristics which predispose *rRNA* genes to recombination events (Achtman, 2000). Genes such as those encoding antigenic determinants should be avoided as they may contain mechanisms which promote mutational events to a point where the rate of substitutions masks earlier events (see the later discussion on substitution saturation in Phylogeny). Genes coding for "house keeping"

proteins make more suitable targets. However, the increased sequence variability in protein-encoding genes may limit the availability of universal primers able to generate amplicons for all species within the genus of interest. The practical outcome of these considerations, if universal primers can be determined, is that genotyping schemes targeting protein-encoding genes are usually able to resolve taxa to species level, and often show intra-species sequence variation which may be additionally informative.

Phylogeny

Phylogeny is the study of relationships, and more importantly the grouping of various life forms into an ordered arrangement, based on shared, derived similarities. Such groupings or classifications have been attempted for centuries, using phenotypic characteristics and features to infer relationship and order among living organisms. Understanding of the evolutionary processes of mutation and natural selection, the rules of inheritance, and especially the nature of genetic material, the genetic code and how it is expressed, has greatly increased over the last 50-100 years. In turn, this has produced a burgeoning of interest in phylogenetic relationship, and an increased sophistication and accuracy in the methods used to infer such relationships. Nowhere has this increase been more marked than with the availability of molecular markers. Beginning in the 1950s and 1960s with protein electrophoretic mobility, it was followed by the analysis of restriction fragment-length polymorphisms (RFLPs) in the 1970s and 1980s. RFLP methods utilised the newly described existence of specific restriction endonucleases and improved laboratory techniques such as Southern hybridisation for observing molecular alterations. Most recently, and perhaps most significantly, interest in phylogenetic relationships has been fostered by the availability and use of the gene sequence itself. This has been made more accessible by polymerase chain reaction (PCR) and sophisticated, economical and widely used automated gene sequencing techniques. A fundamental assumption is made when applying sequence data to phylogeny, that the relatedness of life forms is directly proportional to the similarity of their gene

sequences. Mutations in the sequence progressively accumulate over time from random errors made during DNA replication, and the number of mutations and hence evolutionary relatedness is quantifiable by comparing the level of sequence similarity. While there are flaws in these basic assumptions, nevertheless the availability of gene sequence data has revolutionised phylogenetics.

The usefulness of molecular data is several fold (Avise, 1994b).

1. Molecular data are genetic, and thus involve the most basic information available on phylogenetic relationships.
2. Molecular data are universal to all life. Consequently, relationships can be analysed based on essentially any gene or gene product present in very diverse genera.
3. Molecular markers, and especially gene sequence, contain orders of magnitude more information, and in a less ambiguous form than phenotypic data. For a phenotypic trait such as the ability of an organism to express a biochemical phenotype such as glucose fermentation, only two states are possible, positive or negative. Often the result is not clearly differentiated, and all shades of expression from negative, through equivocal to positive may be found among the population being studied. By comparison, the utilisation of the gene sequence or sequences coding for the phenotypic trait, may contain several thousand base pairs, each and every one can be unambiguously determined, as only four distinct character states are possible for each base site, represented by one of the four nucleotide bases. The difference can be likened to a single "analogue" phenotypic character, compared with many "digital" genetic sequence characters.
4. Molecular markers can distinguish one of the central problems of phenotypic phylogeny, in that they can distinguish the component of biological similarity due to descent from common ancestry (homology) from that due to convergence from different ancestors (analogy).

5. Molecular markers allow direct comparisons of relative levels of genetic differentiation among essentially any groups of organisms. Consequently, it provides a common "ruler" for measuring divergence, which has lead to the concept of the molecular chronometer.
6. Sequence-based molecular data contain previously inaccessible information about the fundamental basis of evolutionary change. Since it is the evolutionary substitution of nucleotide bases which eventually flows through to phenotypic diversity, sequence data allows mechanistic appraisals of the evolutionary process.

The recognition that evolutionary distance between two life-forms is rarely in exact direct proportion to the accumulated nucleotide substitutions has lead to a plethora of mathematical models which attempt to correct for bias. However, as the actual evolutionary history of a life form is not known, the effectiveness of the correction methods can only be assessed using mathematical inference, and the testing of hypothetical data, to determine when the assumptions are valid, and more importantly, under what conditions they fail. The issues which need to be considered, and which can introduce bias, are as follows.

1. **Alignment.** All sequence-based phylogenetic models rely on an alignment of the gene sequences. Each alignment position or site is considered to be a character, with the four possible nucleotides (or a gap in the alignment, the result of an insertion or deletion) being character states. Only characters (alignment position) containing variable character states (nucleotides or gap) are phylogenetically informative. Similar sequences can be aligned with little or no ambiguity, so the phylogenetically informative alignment positions are easily apparent. However the alignment of disparate sequences (either the whole of a sequence or a particular region) is often unclear, with the consequence that bias is introduced into the number and content of informative alignment positions. For this reason sequence alignments, which are usually performed with the assistance of a computer-based alignment program, must be manually verified, and alignment ambiguities excluded

from further analysis. How a phylogenetic model deals with gaps (considered a character state, or excluded either pair-wise or globally) also needs to be considered.

2. **Gene trees versus organism trees**. An assumption is made that the evolutionary tree obtained from a gene sequence alignment reflects the evolutionary history of the organism. However, this is not necessarily true. Genes which experience artificially high rates of mutation (e.g. genes coding for antigenic structures), or are involved in recombinational events leading to insertion of heterologous genetic material, or have experienced convergent evolution, are all examples of influences which cause the evolution of a gene to diverge from and not reflect the evolution of the organism. The presence of this type of bias is often not apparent. The use of sequence from multiple genes of types not known for high rates of mutation or recombination will increase the likelihood that the recovered phylogenetic tree does reflect the evolution of the organism.
3. **Bias in the rate of substitution of each type of nucleotide base.** Based solely on statistical chance, substitutions should be spread equally among the four nucleotide bases. This should ultimately lead to the ratio of A:T:G:C at substitution sites to be 25:25:25:25. In reality, this rarely occurs, and the substitution rate will be biased toward one or several bases. In such cases a method must be chosen which does not assume equal rates of substitution for each base.
4. **Transitional and transversional substitutions**. Substitutions can be defined as a **transition** (ts) when one nucleotide base type ie purine (A and G) or pyrimidine (C and T) is replaced by a base of the same type ie a purine with a purine, or a pyrimidine with a pyrimidine. **Transversions** (tv) occur when a purine is replaced with a pyrimidine, or vice versa. Again, based solely on statistical chance, the rate of each should be equal i.e. ts/tv ratio is one. If the ts/tv ratio does not equal one for a given data set, a method which allows for the bias must be chosen.
5. **Synonymous and non-synonymous substitutions.** Substitutions which do not change the overlying amino acid , as is often the case for substitutions in the third codon position, are

termed **synonymous** substitutions. Those which do change the overlying amino acid are termed **non-synonymous** substitutions. Frequently, the newly encoded amino acid resulting from a non-synonymous substitution has a deleterious affect on the function of the protein encoded by the gene, and the cell does not survive. As a result, the rate of non-synonymous substitutions is very much less than the rate for synonymous substitutions. Substitution rates within the three codon positions are almost always least in the second codon position, and highest in the third codon position (Yang, 1996). Substitution rates can also vary in different regions of the gene as a consequence of the degree with which a non-synonymous change is tolerated. Mutations are very restricted in critical regions of a protein sequence such as those coding for the protein active site, but can be tolerated to a much greater extent in non-critical regions. Substitution rate variation along a sequence is a major cause of bias in phylogenetic inference, and measures such as the **gamma (γ) distribution** are used to accommodate the measured substitution distribution bias (Yang, 1993).

6. **Different substitution rates along branches.** Most models assume an equal rate of evolution for each taxon. Taxa which have evolved more rapidly than "sister" taxa introduce bias which is very difficult to compensate for, and usually affects the resolution of the tree.
7. **Very divergent taxa.** Sequence data from very divergent taxa are very difficult to analyse with confidence. Firstly, the data are "saturated" for substitutions, in that substitutions are being themselves substituted, with the loss of the relationship information contained in the replaced substitutions. Secondly, as mentioned above, the sequence alignment can only be approximated, introducing statistical bias. Both information loss and alignment bias are difficult to compensate for with any degree of confidence. Fundamentally, alternative gene sequence which is less divergent, and less saturated needs to be determined for such taxa. Often, meaningful information can be obtained by excluding third codon positions, where

saturation is potentially most evident, or targeting rRNA sequences where functional constraints on the substitution rate is higher.

Phylogenetic studies in procaryotes have targeted *rRNA* (Woese, 1991). Since it is abundantly expressed, it was amenable to sequencing by early chemical cleavage techniques, or later by reverse transcription techniques. In addition, it is universally present and relatively stable genetically, comparable to mitochondrial sequences in animals. PCR and improved sequencing techniques such as dye-terminator sequencing have enabled direct sequencing of the *rRNA* genes from the genome, making reverse transcription unnecessary. Additionally, rRNA contains regions of high conservation constrained from evolutionary change by ribosome function, but interspersed with more variable regions less constrained by function, and as a consequence rich in phylogenetic information. Thus the utilisation of "universal" primers specific to the conserved regions, enables the amplification and sequencing of the genetically informative variable regions (Ludwig and Schleifer, 1994).

However, when only data from a single gene are available, it is difficult to discriminate between phylogenetic history and the effects of selection and/or horizontal transmission (Woese, 1991). Recent comparisons using nucleotide sequences of protein encoding genes reinforce this concern (Amann, *et al.*, 1988, Kamla, *et al.*, 1996). Bäumler has argued that the evolution of virulence in *Salmonella* is the direct result of horizontal transfer, and that genetic exchange is still occurring at high frequency in *Salmonella enterica* (Bäumler, 1997). Echeita and Usera (1998) recently reviewed eight different outbreaks of typhoid fever and concluded that genetic rearrangements by homologous recombination in the *rrn* gene can occur during the emergence of an outbreak. Of particular concern with the widespread use of *rRNA* genes is that they are present in multiple copies in prokaryotes and are highly conserved at the sequence level. These characteristics may predispose *rRNA* operons to a higher likelihood of involvement in homologous recombination especially as the

degree of homology of sequences is an important determinant of the frequency of recombination (Shen and Huang, 1986). Indeed Strätz, *et al.* (1996) have used these properties to transform *rRNA* operons with orthologous sequences from distantly related bacteria. Additionally, Cilia, *et al.* (1996) determined that point mutations occurring within the *rRNA* genes for *Escherichia* and *Salmonella* tended to be propagated to other *rRNA* genes within the bacterium, by conversions involving short domains. However, they determined that the sequence homogenisation produced by such conversions did not necessarily involve the whole gene. As a consequence, when such homogenising events were rare, some of the nodes within a phylogenetic tree may represent the conversion events rather than the evolution of the organism; a *rRNA* gene tree rather than a "species" tree. There is a need to test the many available *16S rRNA* bacterial phylogenies with relationships based on other loci (Woese, 1991).

Methods of phylogenetic analysis

Phylogenetic relationships are commonly visualised as a phylogenetic tree, with the topology (branching pattern) inferring the nature of the relationships between the member taxa. Since the true topology is unknown, the construction of a phylogenetic tree is a statistical inference of the true tree. This involves two processes, namely an estimation of the topology, and an estimation of the branch lengths (Nei, 1996). There is a dilemma when estimating a topology where the number of taxa is large, since the number of possible topologies is very large, and it becomes very difficult to select the correct one. It is usual to use a phylogenetic algorithm to search among the possible topologies to choose the most likely one based on the presence of desirable properties according to some specified optimisation criteria e.g. fewest character state changes under parsimony. While the statistical foundation for such optimisation is not well established, computer simulations show that they generally work well under biologically realistic conditions (Nei, 1996). There are three main groups of methods used in phyolgenetic inference; **parsimony** methods, **distance** methods

and **likelihood** methods. Each method employs different algorithms based on various optimisation criteria to infer relationship, and each inherently includes assumptions about the nature of the data it is analysing. If the data set does not conform to the assumptions on which a method is based, the topology of the phylogenetic tree deduced will be false. However, as the true topology is not known, the error in the inferred one may not be immediately obvious.

Parsimony Methods

Parsimony is a common notion in science, in that simpler hypotheses to explain observations are preferred over more complicated ones. Maximum parsimony involves the construction of all possible topologies for a given set of taxa, and then determining the tree or trees which minimise the total tree length, which is a measure of the number of evolutionary steps (transformations from one character state to another) required to explain a given set of data (Sober, 1983, Sober, 1989, Swofford, *et al.*, 1996). **Exhaustive** enumeration of all possible trees is only feasible for a few taxa (up to 10-15). Given that there will be 2×10^{20} possible trees for 20 taxa, parsimony methods require a shortened search algorithm for >10-15 taxa. In the **branch-and-bound** search method, the tree length is scored from a root, extending out along each possible branch in series. As the total tree length is being scored for each possible topology, only the shortest is retained. The analysis for any new topology can be discontinued as soon as the total tree length exceeds the current shortest length, as can any other topology which shares the same branching to that point. When the data set is very large, a **heuristic** method of optimisation must be used, which while sacrificing the guarantee of always determining the most optimal topology, shortens computation time to realistic limits. An awareness of data bias which increases the likelihood of optimisation error is essential.

These methods choose a topology according to some criteria, and proceed to rearrange it to minimise the total tree length (Swofford, *et al.*, 1996). **Stepwise addition** starts with three taxa only, perhaps the triplet in the data set possessing the shortest possible total branch

length, followed by testing the addition of each taxon in the data set to select which additional taxon produces the tree with the shortest total branch length, and is to be used in the next step. **Star decomposition** starts with all taxa connected to a central node as a star tree. All possible trees resulting from paring of taxa are assessed by the criteria determining which is optimum (see Fig. 1.3). In the optimal tree, the pair is collapsed to a single branch, and paring of the remaining taxa or branches is reassessed for the optimal tree. The criteria used to assess the optimal tree for branches which contain multiple taxa merged as a result of preceding step/s, is averaged, so that the composite branch behaves as if it were a single taxon. The second pair, which may now include the merged taxa branch, is collapsed to a single branch within this tree, and the process repeats itself, stepwise collapsing the pair which produces the optimal tree to a single branch. Stepwise addition and star decomposition often do not find the optimal tree unless the number of taxa is small or the data contains few ambiguities. Rearrangement of the initial estimated optimal tree by swapping branches (subtree pruning and regrafting, or tree bisection and reconnection) in predefined rearrangements may allow selection of the true optimal tree.

When more than one tree is determined, a strict consensus tree can be constructed where the bifurcations of branches, which are conflicting within the trees, are collapsed to a common node. A significant advantage in parsimonious methods, and perhaps why they are so widely used, is that the number of trees produced is a measure of the phylogenetic informativeness of the data set being analysed. However, while parsimony methods do not require explicit models of evolutionary change, they do make assumptions, the violation of which can lead to problems. The difficulty is knowing precisely what those assumptions are. One example where parsimony is an inconsistent estimator of the true topology, occurs when grossly unequal rates of change occur along different branches, termed "long branch attraction" (Swofford, *et al.*, 1996). It is also difficult to treat the phylogenetic inference in a statistical framework, as there is no natural way to compute the means and variances of

minimum numbers of base substitutions obtained by the parsimony procedure (Nei, 1996). It is, however, the only method that can easily handle insertions and deletions of nucleotides, which sometimes give important information.

Distance methods

In distance methods, an evolutionary distance is calculated for all pairs of taxa to produce a matrix of estimated pairwise distances between taxa, and a tree is constructed using optimisation criteria. The evolutionary distance used for this purpose is usually an estimate of the number of substitutions at each site. There are many distance measures available, each one attempting to compensate for bias such as occurs when the rate of substitution for each nucleotide is unequal, or sufficiently large that sites may be saturated for change i.e. substitutions are occurring over the top of, and thus masking, previous substitutions. The simplest distance measure is the p-distance, and is a measure of the number of sites at which two sequences differ. A p-distance of 0.01 is an average of one substitution every 100 bases. In all but very similar sequences, this is usually a poor measure of the actual number of evolutionary changes, and usually underestimates the actual amount of evolutionary change due to substitution saturation (Page and Holmes, 1998). Other distance measures attempt to correct the observed distance by estimating the amount of substitutions which have been overprinted. For instance, the Jukes-Cantor model (Jukes and Cantor, 1969) assumes all four bases are present in equal amounts, and that the ts/tv ratio is one. If the base frequencies are substantially unequal, but the ts/tv ratio is close to one, the Tajima-Nei model gives a better correction (Tajima and Nei, 1984). Where bias exists in both the base frequencies and the ts/tv ratio, the Tamura or Tamura-Nei correction models give a better prediction of the actual evolutionary distance (Kumar, *et al.*, 1993). Thus, it is imperative that the data be analysed statistically so that the most appropriate distance measure can be chosen. As a general rule, if two distance measures give similar distance values, the simpler one is preferred, because it has a smaller variance (Kumar, *et al.*, 1993).

UPGMA cluster analysis

The simplest distance method employs the "Unweighted Pair Group Method with Arithmetic Averages" (UPGMA) cluster analysis (Sneath and Sokal, 1973). The tree is constructed (see Fig. 1.4) in a manner similar to **star decomposition**, by linking the least distant pairs of taxa to form a single cluster which is then treated as if it were a single taxon, with the distances to other taxa being the average of the distances to the taxa it contains. This is repeated stepwise with the next least distant pair of taxa, one of which may now be a cluster of taxa, to form a second cluster, and so on, linking each pair of least distant taxa or clusters, until the last two clusters are merged into a single cluster containing all the original taxa (Swofford, *et al.*, 1996). The tree is displayed as a dendrogram with the location of the node of each cluster scaled to designate the average distance between the taxa it contains. The scale may be expressed as distance or percentage similarity. The tree is rooted implicitly at the point where the last two clusters (ie the deepest) are joined. The terminal taxa are right justified to zero distance or 100% similarity, and the branch lengths from a node are not scaled (Avise, 1994a).

Neighbour-joining method

The neighbour-joining method is related conceptually to cluster analysis, eg UPGMA, but also allows for unequal rates of molecular change among branches. In contrast to cluster analysis, neighbour-joining keeps track of nodes on a tree rather than taxa or clusters of taxa (Saitou and Nei, 1987, Swofford, *et al.*, 1996). It does so by constructing, at each step of the analysis, a modified distance matrix, that has the net effect of adjusting the branch lengths between each pair of nodes on the basis of the mean divergence from all other nodes. The tree is constructed by linking the least distant pair of nodes as defined by the modified matrix at each step similar to that for UPGMA, but modifying the distance matrix at each step. Unlike UPGMA, the branch lengths can differ (Avise, 1994a).

Maximum likelihood methods

Maximum likelihood (ML) methods attempt to evaluate possible models of evolutionary history which would give rise to the observed data, and selects the one with the higher probability as the preferred "most likely" model using a heuristic tree search procedure (Cavalli-Sforza and Edwards, 1967, Felsenstein, 1981, Felsenstein, 1993). Each possible tree topology is determined, and branch-lengths assigned to determine the maximum-likelihood estimate for each. The branch-length optimisation procedure is very tedious, and because the number of possible trees grows exponentially with the number of taxa, computation time required to perform the analysis on even moderate numbers of taxa has been prohibitive, even with heuristic search techniques (Olsen, *et al.*, 1994). However, ML methods have become much more popular with the increased availability of powerful computers.

The value of the methods are debated. Swofford, *et al.* (1996) argue that ML predicts with more consistency, yields estimates which have a lower variance and often outperforms distance or parsimony methods. On the other hand, Nei argues that the ML models are complicated, may not give a higher probability of obtaining the true tree compared with simpler models, require enormous computational time, and for protein-encoding DNA sequences which are distantly related, have problems due to transition/transversion bias, and higher rates of synonymous to non-synonymous change (Nei, 1996). He does accept that the latter criticisms have largely been overcome by the protein-likelihood method, which hypothetical modelling suggests will produce reasonably true trees for distantly related protein sequences.

A recent innovation of ML is quartet puzzling (QP), in which the ML tree of every possible quartet of sequences is constructed, then the quartet trees are repeatedly combined to an overall tree using a form of Bayesian probability weighting (puzzling step), followed by a

final computation of the majority rule consensus of all intermediate steps to give the QP tree (Strimmer, *et al.*, 1997, Strimmer and von Haesler, 1996). The authors determined using computer simulation that QP was always equal to or better than NJ in reconstructing the true tree, out-performing NJ 10-fold when the ts/tv bias is high. Cao, *et al.* (1998) determined that repeated local rearrangements of the quartet trees sometimes became trapped in local optima, and did not always guarantee the highest possible ML tree. They recommended that trees inferred by NJ or other methods should be used as starting trees, or the tree with the highest likelihood from several analysis runs should be used to avoid this problem.

Rooting trees

While most methods do not require a tree to be rooted, it is often desired, in which case the root must be located using extrinsic information. This is most commonly inferred by the inclusion in the data set of one or more assumed outgroup taxa. The location at which the outgroup joins the unrooted tree containing the ingroup taxa implies a root with respect to the ingroup taxa. This assumes that the ingroup taxa are monophyletic, a state which also needs to be extrinsically justified. Rooting is frequently the most precarious step in any phylogenetic analysis. If a too distantly related outgroup is chosen, there may be so many changes in the branch connecting the outgroup to the ingroup, that the sequences become effectively randomised, which can lead to "long branch effects" causing artificial rooting along longer ingroup branches (Swofford, *et al.*, 1996). It is preferable to leave a tree unrooted if only highly divergent outgroups are available. Alternatively, if rooting is important, the analysis can be performed on the ingroup taxa first, followed by the connection of the outgroup taxon to the unrooted tree.

Reliability of a tree

Virtually the only way to perform a phylogenetic analysis on a data set is with one of the many computer software phylogenetic packages available, such as MEGA, PAUP, or

PHYLIP. However, it is often the case that the user has no appreciation for the validity of the tree produced. Maximum parsimony methods will produce multiple trees if implied, each equally valid for the data set, and as a consequence does inherently contain an indication of the likely reliability of the predicted tree to the real tree. However, distance methods will produce only one tree, but the data may not support such an exact result. One way of testing the reliability of a tested tree is to examine the reliability of each interior branch (Nei, 1996), commonly done by performing a bootstrap test (Efron and Gong, 1983, Efron and Tibshirani, 1993, Felsenstein, 1985). It is important to note that the bootstrap test tests for ambiguities within the data, i.e. how consistently a node is supported. Nodes which are the result of only a small number of substitutions will not receive high boot strap values, even though the data is unambiguous for those nodes. The test is performed as follows. From an alignment of the nucleotide sequences making up the data set that produced the phylogenetic tree, a pseudo-replicate set of the same size is constructed, by randomly choosing sites from the original set for inclusion in the pseudo-replicate set. Sites chosen for inclusion in the pseudo-replicate set are not removed, so can be randomly chosen again. As a consequence, when the pseudo-replicate set contains the same number of sites as the real set, some sites will be included more than once, and others will be absent. A tree is constructed from the pseudo-replicate set, which may vary from the original as a consequence of the randomised inclusion of sites. This process of the creation of a pseudo-replicate set from which a tree is constructed, is repeated many times e.g. 500-1000. The trees so produced are then analysed to see how many times each branch bifurcation, or node, is present. If the frequency is high, then the node is unambiguously supported, and is consequently considered to be statistically significant. Significance is usually expressed as a percentage, and while disagreement occurs between researchers, 70% and above is considered to imply the node is well supported by the data. Lower values indicate that other topological arrangements involving the node are also supported, and that the selected node is the result of stochastic effects within the data. One such effect is as a consequence of the saturation of individual base sites for change, where the

intermediate substitutions have themselves been substituted. If this has happened for a significant number of sites, the true relationship/s may no longer be apparent, and the inferred relationship/s erroneous. This can commonly occur for synonymous substitutions which occurred early in the evolution of the genus. Lack of functional constraints on the rate of substitution means these site are likely to have undergone additional substitutions, resulting in the loss of information about the more deeply resolved nodes. For a heavily saturated set of sequences, only the most recent "tip" nodes may be supported by high boot strap values, with the topology of the more deeply resolved nodes no longer being clear.

Ecological diversity

For close to 20 years, a bacterial species has been somewhat arbitrarily defined as a phenotypically distinct group of strains with at least 70% whole chromosome homology, as measured by DNA-DNA hybridisation (Johnson, 1986, Wayne, *et al.*, 1987). The definition has been universally applied to all bacteria, and frequently the groupings of bacteria based on 70% DNA-DNA homology are the same as those based on phenotypic characteristics. However, the arbitrariness of the cut-off value is increasingly being questioned for its ability to always group bacteria into true ecological units (Vandamme, *et al.*, 1996). Such a cut-off value is not predicted by any genetic evolutionary theory or model, and it is not clear what determines the fraction of genomic segments that anneal.

The rapid increase in availability of gene sequences, and their use in systematics, has lead to the concept of "sequence space" to clarify the true nature of a species. "Sequence space" is a multidimensional visualisation of the clusters of sequence variation within a gene, group of genes and ultimately within the whole genome, expressed by individual strains within the population of strains comprising the species (Ambler, 1996). The question then is whether the space between the species cluster in "sequence space" is void, or does it contain a background of intermediate strains? Is the boundary surrounding each strain cluster defined,

or diffuse and arbitrary, as the edges of the species cluster gradually merge with the background of sequence variation? Ambler presents data from *Rhodospirillaceae* (purple photosynthetic bacteria) and the pseudomonad group which suggests that the sequence clusters are entire, the boundaries defined and the intervening spaces devoid of intermediate forms, although he did find some species are "tight" in that the sequence variation is limited, whereas other species possessed more intraspecies variation ie they are "loose". Species of *Salmonella,* the pathogenic *Neisseria*, and *E. coli* have similarly been reported to exhibit sequence clusters within sequence space and to be devoid of intermediate forms (Maynard Smith, 1970, Spratt, *et al.*, 1995). Irrespective of the "tightness" of the sequence divergence among strains of the same cluster, the average divergence among these strains is much less than the average sequence divergence between strains from different clusters. Recent theory by Cohan (1994a, 1994b, 1996, 1998) suggests that each sequence similarity cluster corresponds to an ecologically distinct population and is a better basis from which to define "species", and on which to construct a classification scheme (Palys, *et al.*, 1997). It is possible that the absence of intermediate forms between species is an artefactual consequence of examining too few strains (Ambler, 1996). While only the determination of more sequences from more strains will truly clarify this issue, the evidence to date suggests that even if there are intermediate forms, they are sufficiently rare that the concept of species clusters in "sequence space" is likely to remain valid.

Cohan has refined his theories into a Coalescence Model (Fig. 1.5) of neutral sequence divergence within and between ecological populations of bacteria, to explain the presence of distinct "sequence space" strain clusters (Cohan, 1994a, Cohan, 1995, Palys, *et al.*, 1997). The model proposes that adaptive mutation or genetic acquisition events can confer a competitive advantage on a cell, resulting in a selective advantage for that cell and its descendants, such that they out-compete all other cells within the population. Such natural selection events consequently purge genetic diversity within the population, at all loci, as only the genome

sequence from the dominant strain remains within that population. However, the natural selection events resulting from the selective advantage do not effect members of other populations outside the ecological niche, although they too are subject to similar natural selection and diversity purging events within their populations. Repetitive rounds of natural selection within the same population, termed periodic selection, as well as within related but distinct populations, promotes the distinctness of ecological populations as separate sequence clusters, at all loci, whether or not the particular gene of interest is responsible for the selective advantage driving the diversity purging events.

Concurrent to the increased diversity occurring between populations as a result of periodic selection, is the possibility of recombination events between the populations in the loci being examined, to produce sequence homogenisation and reduction of between-population diversity. The effect of such events can be substantial even when rare, as there is no fitness penalty when incorporating another species' gene sequence, as the encoded protein (whole or part) is already functionally suitable (Palys, *et al.*, 1997). It is the interplay of the diversifying effect of periodic selection, and the homogenising effect of interpopulation recombination which will eventually define the observed sequence cluster size and distinctness.

The natural outcome of this model is that ecologically distinct populations should eventually diverge into distinct sequence clusters at virtually every locus. Based on current sequence diversity data, groups known to be ecologically diverse have shown a distinctness ratio (between/within) of 2 or greater. Thus future sequence cluster divergence analysis where the distinctness ratio is ≥ 2 should similarly indicate distinct populations (Palys, *et al.*, 1997).

The model predicts the definition of a species to be a distinct population occupying the same ecological niche, and that the distinctiveness should be evident from analysing most

loci, especially those coding for house-keeping functions. However, because of the effects of inter-population recombination, the formal classification of bacteria should consider the sequence variation at more than one unlinked protein-encoding loci, to ensure that such horizontal recombination events do not confuse the classification (Achtman, 2000, Enright and Spratt, 1999, Palys, *et al.*, 1997). Classification based on such criteria does not need to be at the exclusion of *16S rRNA* gene sequences, DNA hybridisation, and phenotypic characteristics. Rather it is probable that classification based on each will reveal significant similarities, and with the combination of methods, the nature of discrepancies will be better understood, and the true classification and inter-species evolutionary relationships clearer.

Aims and objectives

The *mip* gene sequences from *L. pneumophila*, *L. micdadei* and *L. longbeachae* have already been determined. Southern hybridisation studies suggest *mip*-like homologues exist in other species, but that differences between the homologues are likely. Using the known sequences as a basis, the first objective is to determine the nucleotide sequence for the *mip* gene from as many other *Legionella* species as is possible, either by cloning and dye primer sequencing, or by gene amplification and dye terminator sequencing.

Using these *mip* gene sequences , the following objectives were determined:

- From the inferred amino acid sequence, assess the structural and potential functional differences between species, to determine if such differences account for the variation in disease association among *Legionella* species.
- Phylogenetically analyse the *Legionella* genus utilising variation present in the *mip* gene nucleotide sequences.
- Compare the phylogenetic topologies determined for *Legionella* from the *mip* gene with other published analyses based on different genes.
- Develop a genotyping scheme for the *Legionella* genus utilising variation present in the *mip* gene nucleotide sequences, and compare wild-type strain genotypes with identifications determined from other classification schemes.
- Compare the clonality of wild-type strains from different *Legionella* species.

Table 1.1. *Legionella* species currently formally characterised, and published association with human disease. Note that the association of *L. cherrii* with disease is based on serological evidence only.

Organism	Association with human disease	Reference(s)
L. adelaidensis	No	Benson, *et al.*, 1991
L. anisa	Yes	Gorman, *et al.*, 1985, Thacker, *et al.*, 1990
L. birminghamensis	Yes	Wilkinson, *et al.*, 1987, Fang, *et al.*, 1989
L. bozemanii sg 1 & 2	Yes	Brenner, *et al.*, 1980, Fang, *et al.*, 1989
L. brunensis	No	Wilkinson, *et al.*, 1988
L. cherrii	Yes	Brenner, *et al.*, 1985, Fang, *et al.*, 1989
L. cincinnatiensis	Yes	Jernigan, *et al.*, 1994, Thacker, *et al.*, 1988
L. dumoffii	Yes	Brenner, *et al.*, 1980, Fang, *et al.*, 1989
L. erythra	No	Brenner, *et al.*, 1985
L. fairfieldensis	No	Thacker, *et al.*, 1991
L. feeleii sg 1 & 2	Yes	Herwaldt, *et al.*, 1984, Thacker, *et al.*, 1985
L. geestiana	No	Dennis, *et al.*, 1993
L. gormanii	Yes	Morris, *et al.*, 1980, Fang, *et al.*, 1989
L. gratiana	No	Bornstein, *et al.*, 1989
L. hackeliae sg 1 & 2	Yes	Brenner, *et al.*, 1985, Wilkinson, *et al.*, 1985b
L. israelensis	No	Bercovier, *et al.*, 1986
L. jamestowniensis	No	Brenner, *et al.*, 1985
L. jordanis	Yes	Cherry, *et al.*, 1982
L. lansingensis	Yes	Thacker, *et al.*, 1992
L. londiniensis	No	Dennis, *et al.*, 1993
L. longbeachae sg 1 & 2	Yes	Lim, *et al.*, 1989, McKinney, *et al.*, 1981
L. lytica	Yes	Hookey, *et al.*, 1996
L. maceachernii	Yes	Brenner, *et al.*, 1985, Wilkinson, *et al.*, 1985a
L. micdadei	Yes	Hebert, *et al.*, 1980, Fang, *et al.*, 1989
L. moravica	No	Wilkinson, *et al.*, 1988
L. nautarum	No	Dennis, *et al.*, 1993
L. oakridgensis	Yes	Orrison, *et al.*, 1983, Lo Presti, *et al.*, 1998a
L. parisiensis	Yes	Lo Presti, *et al.*, 1997
L. pneumophila sg 1 -16	Yes	Brenner, *et al.*, 1979
L. quateirensis	No	Dennis, *et al.*, 1993
L. quinlivanii sg 1 & 2	No	Benson, *et al.*, 1989, Birtles, *et al.*, 1991
L. rubrilucens	No	Brenner, *et al.*, 1985
L. sainthelensi sg 1 & 2	Yes	Benson, *et al.*, 1990, Campbell, *et al.*, 1984
L. santicrucis	No	Brenner, *et al.*, 1985
L. shakespearei	No	Verma, *et al.*, 1992
L. spiritensis	No	Brenner, *et al.*, 1985
L. steigerwaltii	No	Brenner, *et al.*, 1985
L. taurinensis	No	Lo Presti, *et al.*, 1999
L. tucsonensis	Yes	Thacker, *et al.*, 1989
L. wadsworthii	Yes	Edelstein, *et al.*, 1982, Fang, *et al.*, 1989
L. waltersii	No	Benson, *et al.*, 1996
L. worsleiensis	No	Dennis, *et al.*, 1993

Figure 1.1. Microbial uptake by multilayered pseudopods compared with zipper-type phagocytosis. **(a)** According to the classic model of zipper phagocytosis (Griffin, *et al.*, 1975), particles are engulfed by a circumferential, cup-like pseudopod. The phagocytic cup is sealed by fusion of the approaching membranes, which enables a closed phagosome to bud off from the plasma membrane. **(b)** During uptake of yeast or zymosan particles (Rittig, *et al.*, 1998b), ~2% of the cups are not sealed, and the approaching rims overlap (overlapping phagocytosis). Depending on the experimental settings, the overlapping pseudopods either fuse to a confluent phagosome wall or slide past each other and build up pseudopod stacks before transforming into a phagosome wall. This discontinuous transformation is manifested by intracellular remnants of the former membrane gap. **(c)** Up to 80% of all phagocytosed *Leishmania* spp. promastigotes (Rittig, *et al.*, 1998b), 60% of *Borrelia burgdorferi* (Rittig, *et al.*, 1992) and 100% of *Legionella pneumophila* (Horwitz, 1984) are enveloped by a unilateral pseudopod that rolls into itself rather than fusing with its stem (coiling phagocytosis). With *Leishmania* promastigotes, the resulting pseudopod whorls fuse to a confluent phagosome wall, manifested by the presence of remnants of the former membrane gap. With *B. burgdorferi*, the dissipation of the internalized membranes leaves the bacteria in the cytosol. *L. pneumophila* somehow ends up in a ribosome-studded replicative vacuole, but this transformation has not been observed (indicated by the dotted arrow) Reproduced from Rittig, *et al.* (1998a).

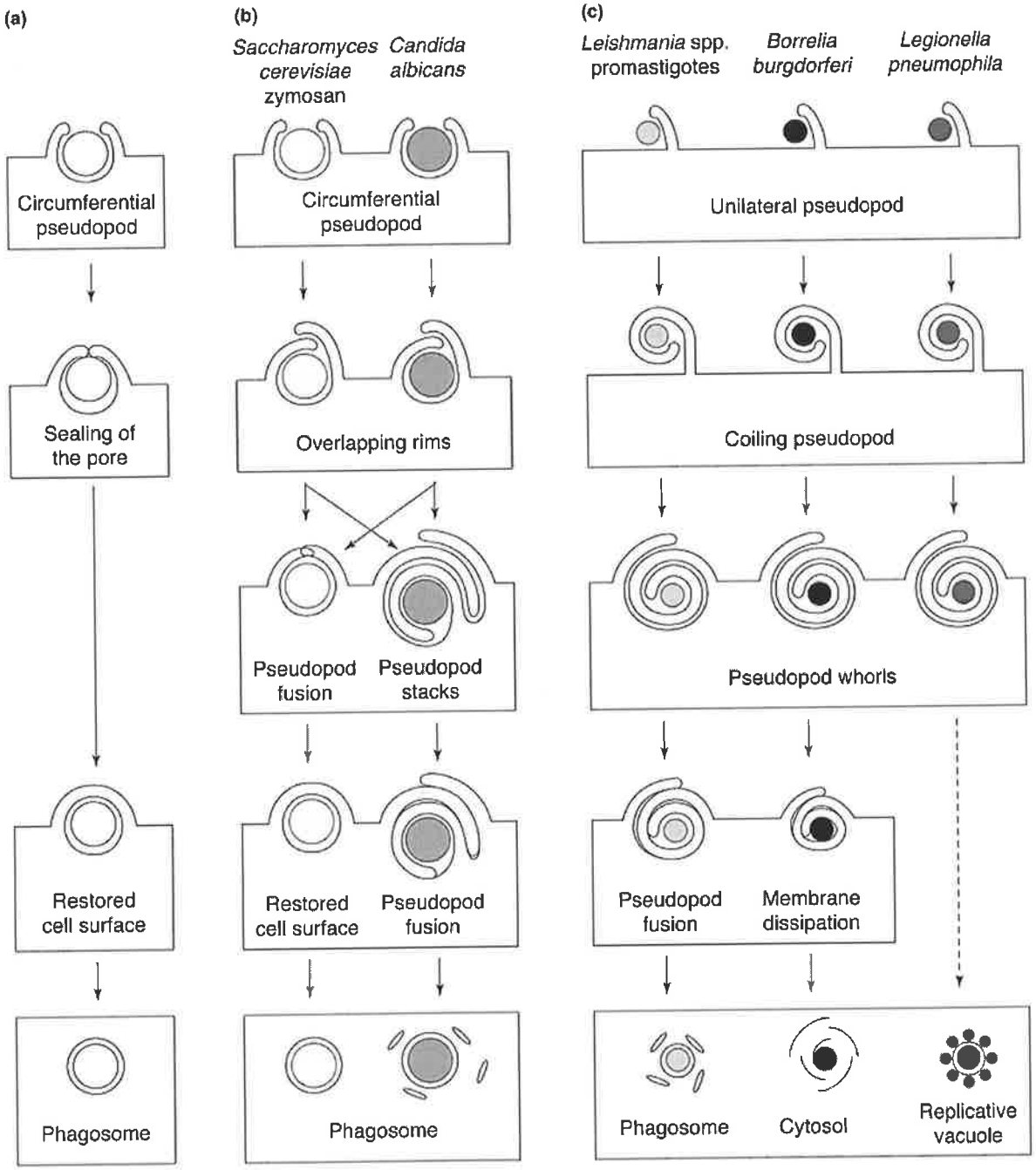
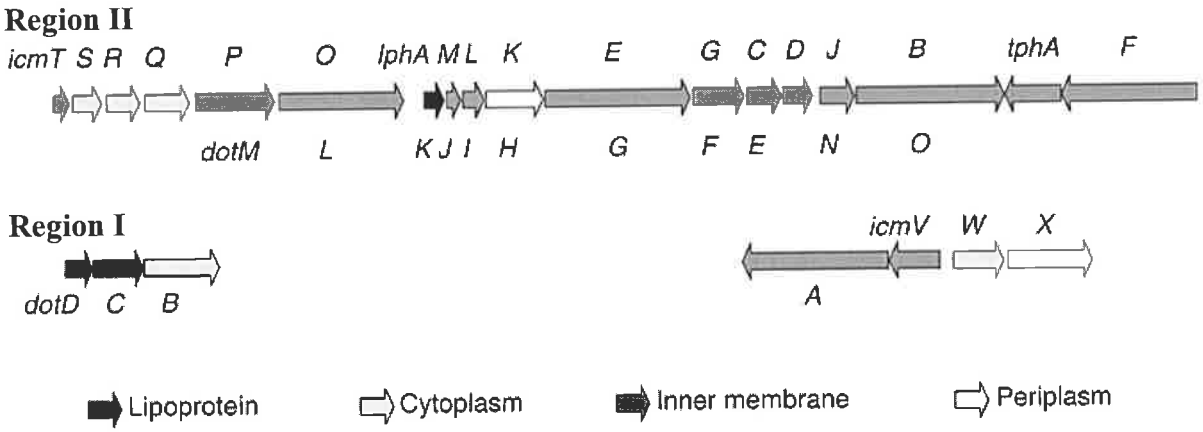
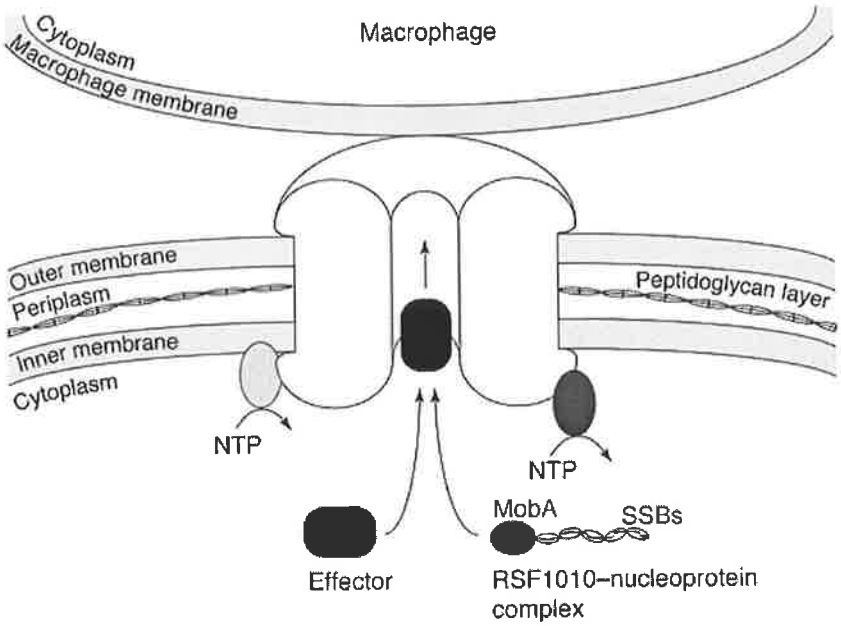

(a)
Circumferential pseudopod
Sealing of the pore
Restored cell surface
Phagosome
(b)
Saccharomyces cerevisiae zymosan
Candida albicans
Circumferential pseudopod
Overlapping rims
Pseudopod fusion
Pseudopod stacks
Restored cell surface
Pseudopod fusion
Phagosome
(c)
Leishmania spp. promastigotes
Borrelia burgdorferi
Legionella pneumophila
Unilateral pseudopod
Coiling pseudopod
Pseudopod whorls
Pseudopod fusion
Membrane dissipation
Phagosome
Cytosol
Replicative vacuole

Figure 1.2. (a) Linkage map of the two *icm/dot* regions. Region I contains the *icmV, W, X – dot A, B, C, D* genes, and region II contains the *icmT, S, R, Q, P, O - lphA – icmM, L, K, E, G, C, D, J, B - tphA - icmF* or the *dot M, L, K, J, I, H, G, F, E, N, O* genes. Coding regions are indicated by arrows, and the different tints indicate the predicted location of the protein in the bacterial cell. The *icm* gene designations are marked above the gene map, and the *dot* gene designations are marked under the gene map. **(b)** A model for the *L. pneumophila* virulence transfer system. Some of the *icm/dot* gene products are expected to interact with one another to form a protein complex, which will form a channel through the bacterial inner and outer membrane and will constitute the the Icm/Dot transfer system or transferosome. Other *icm/dot* gene product(s) are expected to be effector molecules. Some of the *icm/dot* gene products may be involved in assembly of the complex, and the energy required for assembly and/or transfer of the effector molecule may be provided by nucleotide triphosphatase activities of several ATP/GTP-binding proteins. Another substrate that is able to be transferred by the Icm/Dot transfer system is the plasmid RSF1010-nucleoprotein complex, which has been shown to conjugate between bacteria in a *icm/dot*-dependent manner. **(c)** There are three possibilities for the location of the effector protein(s): in the phagosome membrane, in the macrophage cytoplasm or in the phagosome space. Coers, *et al.* (1999) suggests the first or last possibility are more likely. Reproduced from Segal and Shuman (1998).

(a)


Region II
icmT S R Q P O lphA M L K E G C D J B tphA F
dotM L K J I H G F E N O
Region I
dotD C B
icmV W X
A
Lipoprotein
Cytoplasm
Inner membrane
Periplasm

(b)


Macrophage
Cytoplasm
Macrophage membrane
Outer membrane
Periplasm
Inner membrane
Cytoplasm
Peptidoglycan layer
NTP
NTP
MobA
SSBs
Effector
RSF1010–nucleoprotein complex

(c)

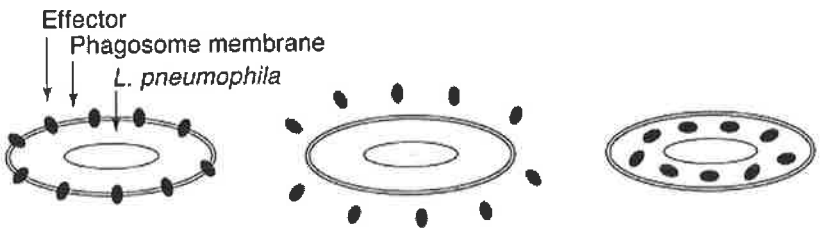

Effector
Phagosome membrane
L. pneumophila

Figure 1.3. An hypothetical analysis of 5 taxa, using the maximum parsimony method for phylogenetic inference, with heuristic tree selection using the star decomposition method. At each step, the criteria for the optimal tree is evaluated for each possible joining of a pair of lineages leading away from a central node. The best tree found during each step (shaded) becomes the starting point for the next step. Adapted from Swofford, *et al.* (1996).

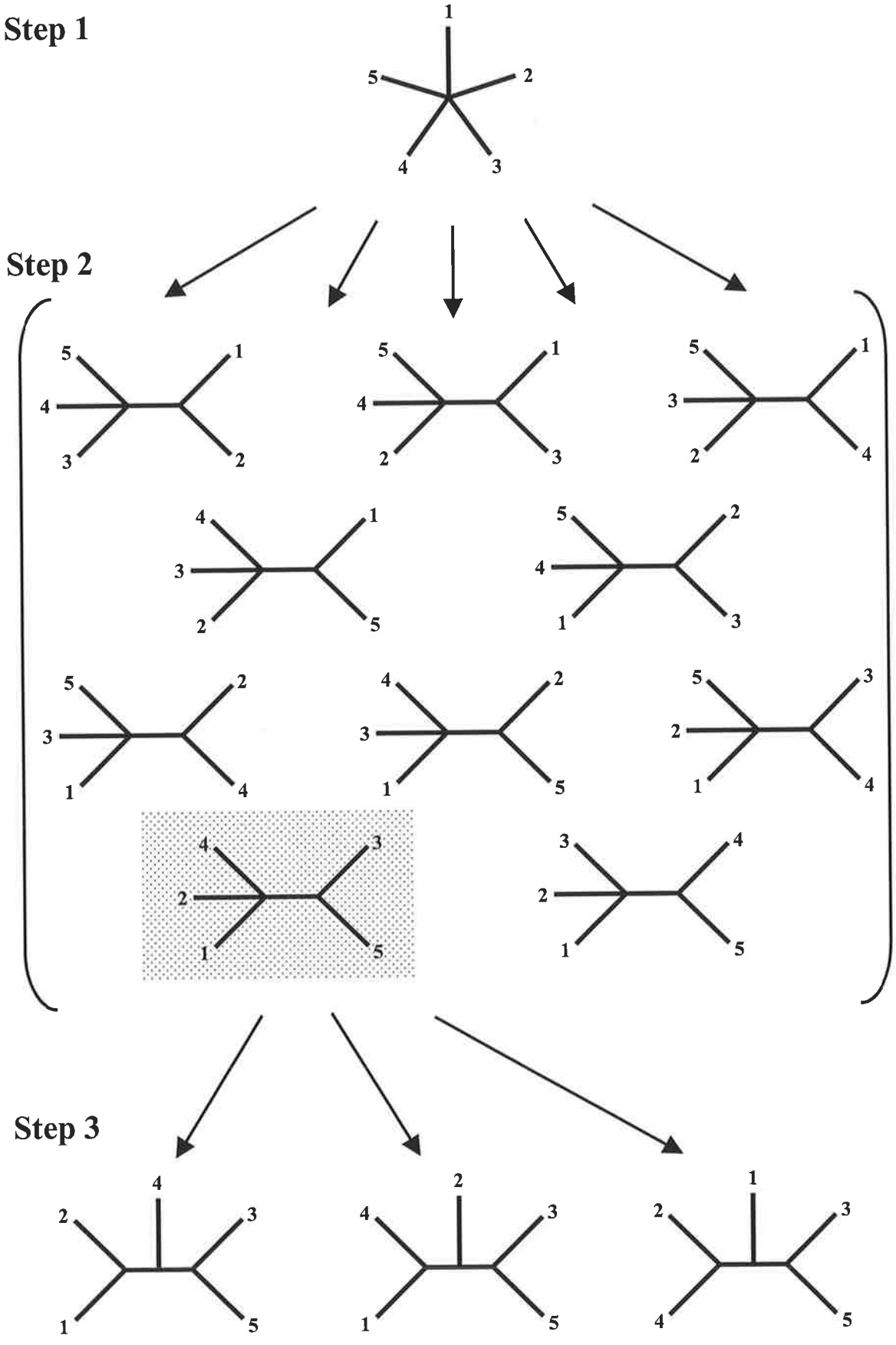

Step 1
Step 2
Step 3

Figure 1.4. Comparison of 5 hypothetical *rRNA* sequences by distance methods. Table A contains the UPGMA cluster analysis of the evolutionary distance estimates (Sneath and Sokal, 1973). The distance metric used is the average number of nucleotide substitutions per 100 alignment positions, expressed as a fraction of one. At each step, the smallest distance between pairs (bold) determines which pair of taxa are collapsed to a single branch, and the distance estimates recalculated (similar to star decomposition). The resulting tree is drawn in B. The vertical bar joining two taxa or clusters indicates the evolutionary distance on the scale. For comparison, tree C indicates the phylogenetic tree determined by neighbour-joining (NJ) (Saitou and Nei, 1987). The length of each branch is scaled to the evolutionary distance estimate between taxa. In this example, the distance metric (weighted least squares) is different from that used in UPGMA, weighted to adjust for the perceived evolutionary substitution mechanism. Adapted from Swofford, *et al.* (1996).

A UPGMA analysis

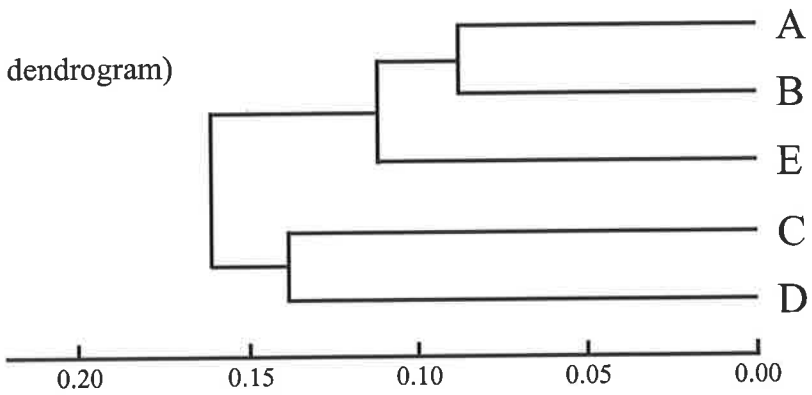
	A	B	C	D	E
A	-	**0.1715**	0.2147	0.3091	0.2326
B		-	0.2991	0.3399	0.2058
C			-	0.2795	0.3943
D				-	0.4289
E					-

	A-B	C	D	E
A-B	-	0.2569	0.3245	**0.2192**
C		-	0.2795	0.3943
D		-	-	0.4289
E				-

	A-B-E	C	D
A-B-E	-	0.3027	0.3593
C		-	**0.2795**
D			-

	A-B-E	C-D
A-B-E	-	**0.3310**
C-D		-

B (UPGMA dendrogram)



C (Neighbour-joining dendrogram)

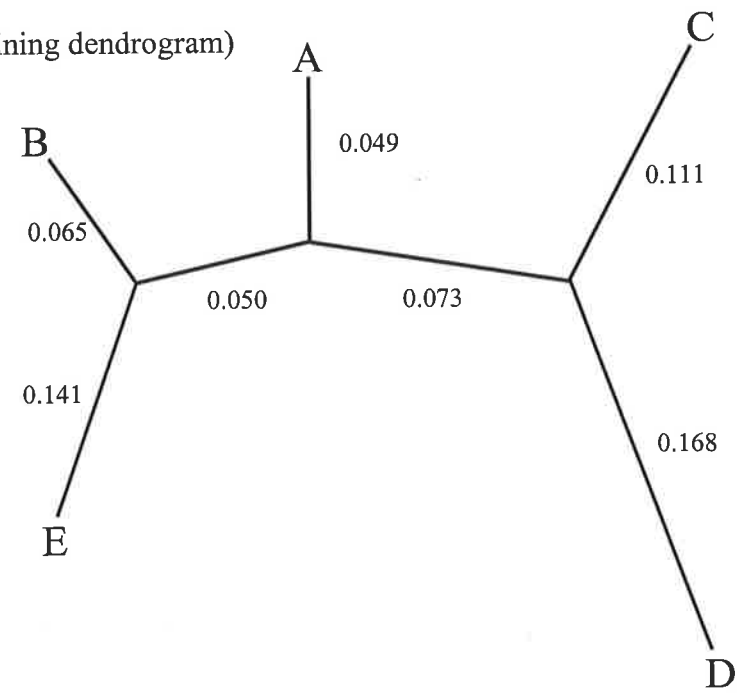


Fig 1.5. Cohan's Coalescence Model of periodic selection and diversity purging (Cohan, 1994a, Cohan, 1995, Palys, *et al.*, 1997). The figure demonstrates that two populations will become distinct sequence similarity clusters where the between-population divergence is much greater than the within-population divergence.

A The derived population consists of the descendants of a mutant (X) capable of utilising a new ecological niche. The adaptive mutant in the derived population (*) is capable of outcompeting all other members of the derived population. Note that at this point, the two populations do not appear as distinct clusters. Moreover, the ancestral population is not even a monophyletic group.

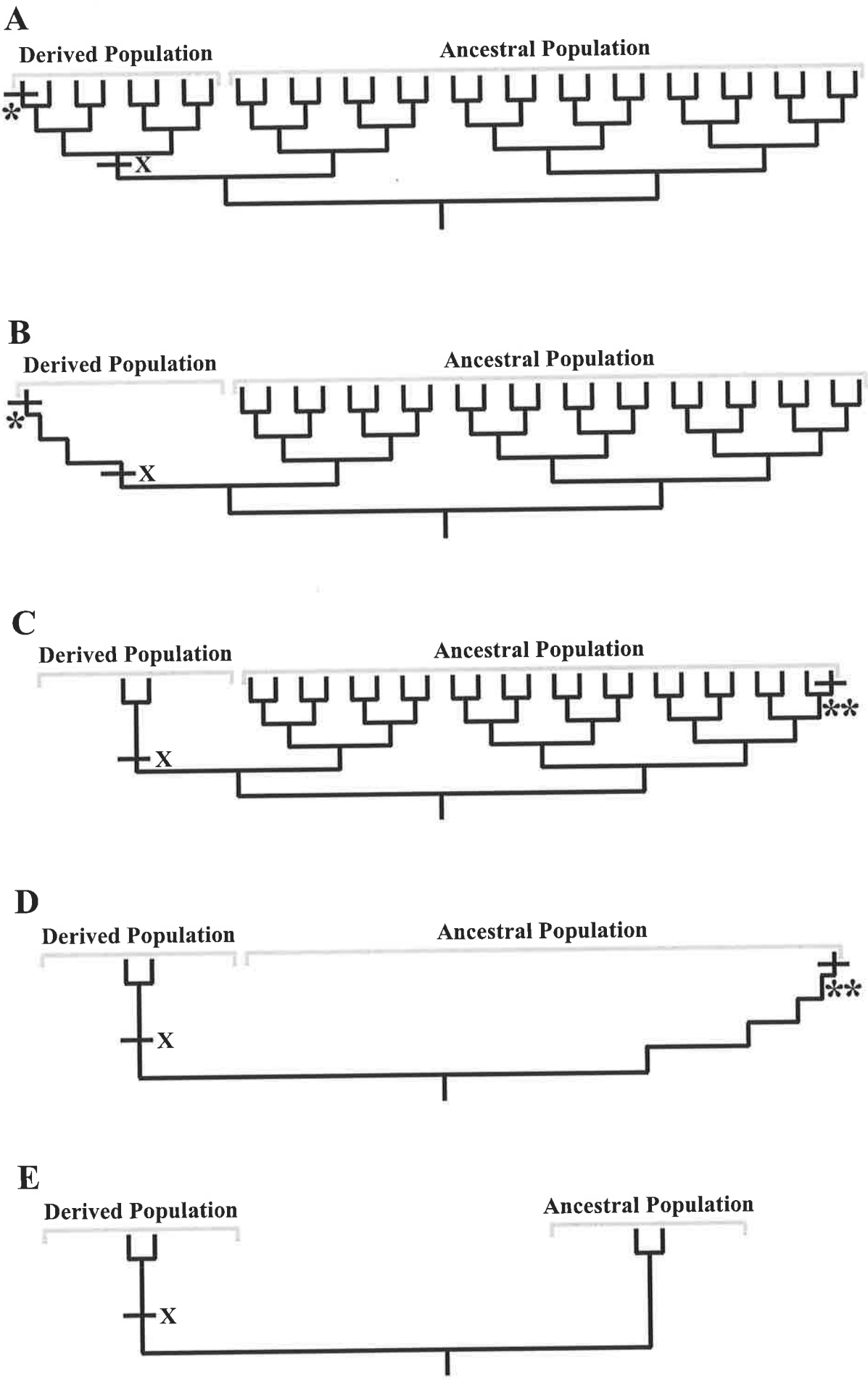
B The adaptive mutant (*) has driven all other lineages within the derived population to extinction.

C With time, the derived population becomes more genetically diverse. One cell in the ancestral population (**) has developed a mutation that allows it to outcompete other members of the population.

D The adaptive mutant (**) has outcompeted other members of the ancestral population

E The ancestral population becomes more genetically diverse. At this point, each population has a distinct sequence cluster as well as a monophyletic group.

N.B. A similar outcome would occur if the events X, * and ** were the acquisition of selectively advantageous genetic material by horizontal gene transfer at a different loci, and which produced the derived population as for X, and/or the adaptive mutant effects * and/or **. Reproduced from Palys, *et al.* (1997).


A
Derived Population
Ancestral Population
*
X
B
Derived Population
Ancestral Population
*
X
C
Derived Population
Ancestral Population
**
X
D
Derived Population
Ancestral Population
**
X
E
Derived Population
Ancestral Population
X

Chapter Two

Materials and Methods

Growth media

The following solid media were used for bacterial cultivation.

Legionellae strains were grown on charcoal yeast extract (CYE) (CM665; Oxoid, Basingstoke, Hampshire, United Kingdom) with α-ketoglutarate and L-cysteine supplement (code SR110; Oxoid) and 1% bovine serum albumin. *E. coli* K-12 strains used in cloning experiments were grown on Columbia blood agar base (CA) (1% Bacto pantone, 1% Bacto Bitone, 0.3% tryptic digest of beef heart, 0.1% corn starch, 0.5% NaCl and 1.5% Bacto agar). Both media were purchased from Med Vet Science, South Australia. Xgal/IPTG/Amp plates (20 ml) were poured from melted 100 ml CA supplemented with 400 µl 25 mg/ml ampicillin, 160µl Xgal (20 mg/µl in dimethyl formamide) and 160 µl IPTG (24 mg/ml).

The following liquid media were used for bacterial cultivation.

For whole cell protein analysis, legionellae strains were grown in *Legionella* broth (Buffered Yeast Extract – BYE) containing 1% yeast extract, 1% N-(-2 acetamido)-2-aminoethanesulfonic acid buffer, 0.1% α-ketoglutarate, 0.4% L-cysteine-HCl and 0.025% ferric pyrophosphate (Steinmetz, *et al.*, 1991). Luria-broth (L-broth) (1% Oxoid tryptone L42, 0.5% Oxoid yeast L21, and 0.5% NaCl) was the general liquid growth medium for *E. coli* K-12 strains used in cloning experiments (Sambrook, *et al.*, 1989a). Bacteria for plasmid enrichment were grown in Terrific Broth, which consisted of 1.2% Bacto tryptone, 2.4%

Bacto yeast extract, 0.5 glycerol, 1.15% KH_2PO_4 and 6.25% K_2HPO_4 (Sambrook, *et al.*, 1989a). SOC medium (2% Bacto tryptone, 0.5% Bacto yeast, 10 mM NaCl, 2.5 mM KCl, 10mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) was used to maximise the recovery of transformants following electroporation (Sambrook, *et al.*, 1989a). Trypticase soy broth with glycerol (TSB glycerol) (1.7% Bacto tryptone, 0.3% Bacto soyatone, 1.5% NaCl, 0.25% K_2HPO_4 and 24% glycerol) was used as freezing medium for maintenance of bacterial stocks (Atlas, 1993).

Where appropriate, ampicillin at a final concentration of 100 µg/ml was added to broth and solid media during cloning experiments to select for, or maintain, transformants.

Chemicals and reagents

All chemicals were Analar grade. Butanol, caesium chloride (CsCl), 4-chloro-1-naphthol, di-potassium hydrogen orthophosphate (K_2HPO_4), di-sodium hydrogen orthophosphate (Na_2HPO_4), ethidium bromide, ethylene-diamine-tetra-acetic-acid (EDTA), glacial acetic acid, formamide, glucose, glycerol, hydrogen peroxide, iso-amyl alcohol, isopropanol, methanol, magnesium chloride ($MgCl_2$), magnesium sulphate (Mg_2SO_4), β-mercaptoethanol, phenol, polyvinyl-pyrrolidone, potassium chloride (KCl), di-potassium hydrogen orthophosphate (K_2HPO_4), potassium di-hydrogen orthophosphate (KH_2PO_4), Sigma 7-9 (Tris; hydroxymethane), sodium chloride (NaCl), sodium citrate, sodium dodecyl sulphate (SDS), sodium hydroxide (NaOH), sodium lauroylsarcosine (sarkosyl), sucrose and Tris (hydroxymethane) amino methane were obtained from BDH Chemicals, England or the Sigma Chemical Company, USA. Guanidinium isothiocyanate (enzyme grade) was purchased from Gibco BRL, Gaithersburg, MD, USA, ampicillin from Boehringer Mannheim, GmbH, Germany, and electrophoresis grade reagents were obtained from the sources indicated: acrylamide (Bio-Rad Laboratories, Richmond, CA or National Diagnostics, Atlanta, Georgia,

USA), ammonium persulphate (Bio-Rad) and N,N,N,N,-tetramethylethylenediamine (TEMED) (Sigma).

The four deoxyribonucleotide triphosphates (dATP, dTTP, dGTP, dCTP) were obtained from Pharmacia (Uppsala, Sweden). Adenosine-5'-triphosphate sodium salt (ATP), 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), and isopropyl-β-D-thiogalacto-pyranoside (IPTG), and the Digoxigenin labelling reagents, nitroblue tetrazolium chloride (NBT) and 5-brom-4-chlor-3-indolyl-phosphate toluidine salt (X-P) were purchased from Boehringer Mannheim. Agarose was from Progen Industries, Qld., Australia, or Bio-Rad Laboratories. The later also supplied preparative grade low melt point agarose.

Antibody

Absorbed polyvalent *L. longbeachae* anti-Mip serum was generously supplied by Robyn Doyle, IMVS, Adelaide, for use in Mip immuno-blotting experiments.

Enzymes

AmpliTaq and Amplitaq Gold DNA polymerases and buffers used in the amplification reactions (Buffer II, containing 100 mM Tris and 100 mM KCl; and 25mM $MgCl_2$) were obtained from Perkin Elmer Norwalk, CO, USA, lysozyme from the Sigma Chemical Co., and Klenow enzyme and proteinase K from Boehringer Mannheim. Restriction endonucleases were purchased from either Amersham, Boehringher Mannheim, New England Biolabs or Pharmacia, and used with the appropriate buffer from the same supplier.

Synthesis of synthetic oligonucleotides

Synthetic oligonucleotides (primers) were synthesised using reagents purchased from Applied Biosystems or Ajax Chemicals (acetonitrile). Synthesis was performed on an Applied Biosystems 381A DNA synthesizer. Oligonucleotides (100 µl) were extracted once with

butanol (1 ml) to remove residual salts, dried *in vacuo* and reconstituted with 100 µl of purified water (Sawadogo and Van Dyke, 1991).

Culture and maintenance of bacterial strains

The strains of *Legionella* used in this thesis are given in the relevant chapters; type strains in Table 3.1, and wild strains, including source, in Table 6.1. For long term storage, all strains were maintained as a thick suspension in TSB glycerol at -70°C. Fresh cultures of *Legionella* strains were grown for 3-4 days on CYE plates at 37°C in air in a humidified incubator, with the exception of strains from Dr. Vladimir Drasar, which were incubated for up to 7 days at 30°C in a candle jar. *E. coli* K-12 strains used in cloning experiments were grown for 18 hours on CA or L-broth at 37°C. Ampicillin was added when appropriate to maintain transformants.

Genetic transfer methods

Transformation

Transformation-competent *E. coli* K12 strain DH5α cells were prepared as follows: an overnight shaken culture (in L-broth) was diluted 1:20 into pre-warmed L-broth and incubated with shaking for 2.5 hr at 37°C (A_{600} of 0.6, *ca.* 4 x 10^8 cells/ml). The cells were chilled on ice for 30 mins, pelleted at 4,354 x *g* at 4°C, resuspended in 10 ml of cold 100 mM $MgCl_2$, centrifuged again and resuspended in 2 ml of cold 100 mM $CaCl_2$. This was allowed to stand for 60 mins on ice before the addition of DNA. Alternatively, 200 µl of sterile glycerol was added to the suspension, mixed gently and stored at -70°C in 100 µl aliquots for further use.

Transformation was performed according to the method described by Brown, *et al.* (1979). Competent cells (200 µl) were mixed with DNA (volume made to 100 µl with 1x TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and left on ice for a further 1 hr. The cell-

DNA mixture was heated at 42°C for 2 mins and then 1 ml of L-broth was added followed by incubation with shaking at 37°C for 1 hr. The culture (100 µl) was then plated onto Xgal/IPTG/Amp selection plates and incubated over night (O/N) at 37°C and transformants selected.

Electroporation

Electrocompetent cells were prepared using the method of Dower, *et al.* (1988). *E. coli* DH5α was made competent for electroporation with plasmid DNA as follows: an overnight shaken culture (in L-broth) was diluted 1:20 into L-broth and incubated for 3 h at 37°C with gentle agitation at 150 rpm (A_{600} of 0.6, *ca.* 4 x 10^8 cells/ml). The log-phase culture was chilled on ice for one hour and then harvested at 4,354 x *g* for 20 mins at 4°C. The pellet was washed twice in 250 ml of ice-cold purified water and once in 5 ml of ice-cold 10% glycerol. The final pellet was resuspended in 800 to 1000 µl of 10% glycerol and stored at -70°C in 40 µl aliquots for further use.

Electroporation was performed according to the method described by Dower, *et al.* (1988). Purified ligated DNA (approximately 250 µg in 5 µl) was added to 40 µl of ice-thawed electrocompetent cells and held in an ice slurry until used. The entire cell-DNA suspension was loaded into the bottom of a chilled, dried 0.2 cm Bio-Rad electroporation cuvette (Cat. No. 165-2086) ensuring no air bubbles are trapped in the suspension (Load suspension to one side of cuvette bottom, and then gently tap cuvette on bench to evenly distribute the suspension across the cuvette). Note that air bubbles can cause the cuvette to arc during electroporation. Electroporation of the cells-DNA suspension was performed using the Bio-Rad Gene Pulser with capacitance extender (Model No. 1652078) at the following settings: 200 Ω resistance, 25 µF capacitance, 2.5 kV pulse strength. Immediately following electroporation (less than 10 seconds), 1.5 ml of SOC medium was immediately added to the bacterial suspension and incubated for 45 mins at 37°C with very gentle shaking. A 100 µl

sample of the culture was then plated onto Xgal/IPTG/Amp selection plates selection plates neat, and diluted ten- and 100-fold, and incubated O/N at 37°C, to determine the optimum dilution to recover transformants. The remainder of the transformed culture was then plated at the optimum dilution determined, incubated, and transformants recovered.

Screening of transformants

Two methods were used. In early experiments, transformants were examined using colony transfer and immuno-blotting. In later experiments, transformants were examined using colony hybridization probing. The method details are given in the DNA and protein manipulation sections.

DNA extraction protocols

Chromosomal DNA extraction methods

Chromosomal DNA was extracted with one of the three following procedures:

Method 1:

This method of genomic DNA extraction (Manning, *et al.*, 1986) was used to recover high quality DNA for cloning purposes. Bacterial culture (50 ml) was pelleted at 2,831 x *g* for 15 mins. The bacterial pellet was resuspended in 2 ml of Sucrose-Tris solution (50 mM Tris, pH 8.0, 25% sucrose) and 1 ml of 10 mg/ml of lysozyme in 250 mM EDTA, pH 8.0, and incubated for 60 mins in an ice bath. Then 750 µl of TE solution (100 mM Tris, pH 8.0, 10 mM EDTA), 250 µl of lysis solution (1 M Tris, pH 8.0, 10% sarkosyl, 250 mM EDTA) and 40 mg of pronase was added and incubated for 2 hr at 56°C. The lysed suspension was extracted three times with Tris-saturated phenol to remove proteins associated with DNA and then with chloroform-isoamylalcohol (24:1) to remove any residual phenol. The supernatant from the last extraction was dialysed against 1 x TE solution overnight at 4°C. Dialysed DNA was stored at 4°C.

Method 2:

This rapid genomic DNA extraction method (Saunders, *et al.*, 1990) was used to recover chromosomal DNA from many of the wild *Legionella* strains. Approximately 100-200 mg of cells from a culture plate were resuspended in 400 µl of 2 mg/ml of lysozyme and incubated for 10 mins at 25°C. The suspension was lysed with 400 µl of 5 M guanidinium isothiocyanate in 100 mM EDTA, pH 7.0, and emulsified by vortexing vigorously for 30 secs. DNA and cytosolic proteins were removed from the lysate by performing a single extraction (17,000 x *g* for 5 mins) with equal volumes of phenol, chloroform, isoamyl-alcohol mixture (25:24:1). Chromosomal DNA in the supernatant was precipitated by adding an 0.8 vol of isopropanol. The precipitated DNA was washed in 80% ethanol and pelleted at 17,000 x *g* for 5 mins. The DNA pellet was dried *in vacuo*, resuspended in 500 µl of TE solution and stored at 4°C.

Method 3 (heat lysis)

Crude DNA suitable for PCR amplification was also used for many of the wild strains used. A moderately turbid suspension of organisms was made in 500 µl sterile water (approximately 10^9 cfu/ml), and either microwaved on full power (650W) for two minutes, or boiled or steamed for 10 - 15 mins. 2µl of this crude extract was used in the amplification reaction.

Plasmid DNA extraction methods

Plasmid DNA was isolated by one of the two following procedures:

Method 1:

Large scale plasmid purification was performed by the three step alkali lysis method (Garger, *et al.*, 1983). Cells from a 500 ml culture (Terrific Broth) were harvested at 4,400 x *g* for 15 mins at 4°C and resuspended in 12 ml of solution-1 (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM EDTA). Freshly prepared lysozyme (4 ml of 20 mg/ml in solution-1) was

mixed with the cell suspension and incubated at room temperature for 10 mins. Addition of 27.6 ml of solution-2 (0.2 M NaOH, 1% (w/v) SDS), followed by a 5 mins incubation on ice resulted in total cell lysis. After the addition of 14 ml solution 3 (5 M potassium acetate, pH 4.8) and incubation on ice for 15 mins, protein, chromosomal DNA and high MW RNA were removed by centrifugation (4,400 x *g* for 15 mins at 4°C). The supernatant was then extracted with an equal volume of a TE saturated phenol, chloroform, isoamyl-alcohol mixture (25:24:1) at 4,400 x *g* for 15 mins at 4°C. Plasmid DNA from the aqueous phase was precipitated with 0.8 volume of 100% isopropanol at room temperature for 20 mins and collected by centrifugation (11,000 x *g* for 30 mins at 4°C). After washing in 70% (v/v) ethanol (14,500 x *g* for 20 mins at 4°C), the pellet was dried *in vacuo* and resuspended in 800 µl of 1 x TE. Plasmid DNA was purified from contaminating protein and RNA by centrifugation on a two step CsCl ethidium bromide gradient according to Garger, *et al.* (1983). The DNA, CsCl, ethidium bromide mixture was prepared by mixing the 800 µl DNA in TE with 1.455 g of solid CsCl and 150 µl of 10 mg/ml of ethidium bromide. Final volume of the mixture was made up to 1.6 ml with TE solution. The mixture was then added to the bottom of a 4.2 ml polyallomer tube filled with 3.2 ml of less dense CsCl in TE solution (1.470 g/ml, Refractive index 1.3780). The contents of the tube were centrifuged at 372,000 x *g* for 3 hr at 25°C. The DNA band was removed by side puncture of the tube with a 19 gauge needle attached to a 1 ml syringe. The ethidium bromide was extracted using isoamylalcohol. CsCl was then removed by dialysis overnight against three changes of 5 litres 1x TE at 4°C. DNA was stored at 4°C.

Method 2:

Small scale plasmid purification was performed by the three step alkali lysis method using a modification of Garger, *et al.* (1983). Overnight bacterial cultures (1.5 ml L-broth) were transferred to a microfuge tube, harvested by centrifugation (17,000 x *g* for 5 mins), and resuspended in 200 µl of solution-1 (50 mM glucose, 25 mM Tris-HCl, pH 8.0, 10 mM

EDTA). The addition of 400 µl of solution 2 (0.2 M NaOH, 1% (w/v) SDS) followed by a 5 mins incubation on ice resulted in cell lysis. After the addition of 300 µl of solution-3 (5 M potassium acetate, pH 4.8) and a 5 mins incubation on ice, protein, chromosomal DNA and high MW RNA were collected by centrifugation (17,000 x *g* for 5 mins). The supernatant was transferred to a fresh tube and extracted once with an equal volume of a TE saturated phenol, chloroform, isoamylalcohol mixture (25:24:1). Plasmid DNA from the aqueous phase was precipitated with an 0.8 volume of 100% isopropanol at room temperature for 20 mins and collected by centrifugation (25,000 x *g* for 30 mins at 4°C). After washing in 70% (v/v) ethanol (17,000 x *g* for 5 mins), the pellet was dried *in vacuo* and resuspended in 40 µl of 1 x TE.

PEG precipitation purification

Small scale plasmid preparations which were used in sequencing reactions were further purified, following the instructions in the Perkin Elmer Dye Terminator Sequencing Cycle Sequencing Ready Reaction Kit Instructions (PM 402078). Briefly, 32µl of the small scale plasmid preparation was mixed with 8 µl 4 M NaCl, followed by 40 µl autoclave 13% PEG_{8000}. After thorough mixing, the samples were refrigerated at -20°C O/N, and the precipitated plasmid DNA pelleted by centrifugation (25,000 x *g* for 30 mins at 4°C). After washing in 500 µl 70% (v/v) ethanol (17,000 x *g* for 5 mins), the pellet was dried *in vacuo*, resuspended in 20 µl of water, and stored at -20°C.

Analysis and manipulation of DNA

DNA quantitation

The concentration of DNA in solutions was determined by measurement of absorption at 260 nm and assuming an A_{260} of 1.0 is equal to 50 mg of DNA/ml (Sambrook, *et al.*, 1989b).

Restriction endonuclease digestion of DNA

Cleavage reactions with the restriction enzymes indicated in the text were performed in the buffer indicated by the manufacturer. 0.1-0.5 mg of DNA or purified restriction fragments were incubated with the appropriate enzyme buffer and 5 U of each restriction enzyme in a final volume of 20 µl, at 37°C O/N. The reactions were terminated by heating at 85°C for 10 mins. Prior to loading onto a gel, a one tenth volume of tracking dye (15% (w/v) Ficoll, 0.1% (w/v) bromophenol blue, 0.1 mg/ml RNase A) was added.

The sizes of restriction enzyme fragments were estimated by comparing their relative mobility with that of *Eco*RI digested *Bacillus subtilis* bacteriophage SPP-1 DNA. The sizes (kilobase, kb) of the fragments were: 8.51, 7.35, 6.11, 4.84, 3.59, 2.81, 1.95, 1.86, 1.51, 1.39, 1.16, 0.98, 0.72, 0.48, 0.36 and 0.09 (Geneworks, Adelaide, Australia).

Analytical and preparative separation of restriction fragments

Electrophoresis of digested DNA was carried out at room temperature on horizontal, 1%, or 1.5% or 2% (w/v) agarose gels, 13 cm long, 13 cm wide and 0.7 cm thick. Gels were run at 64.5 mA for 4-5 hr in 1x TAE buffer (40 mM sodium acetate, 40 mM Tris and 2 mM EDTA). After electrophoresis, the gels were stained in distilled water containing 2 mg/ml ethidium bromide. DNA bands were visualised by trans-illumination with ultraviolet light (300 ηm) and photographed using Polaroid 667 positive film.

For preparative gels, Bio-Rad low-gelling-temperature agarose at a concentration of 1.0% (w/v) was used for separation of restriction fragments, which were recovered by the following method: DNA bands were excised and the agarose melted at 65°C and then vortexed vigorously with equal volumes of Tris saturated phenol. Residual phenol was removed with chloroform and the DNA precipitated with two volumes of ethanol and one tenth volume of 3 M sodium acetate, pH 5.2. DNA was collected by centrifugation (17,000 x

g for 5 mins), washed once with 70% (v/v) ethanol and dried *in vacuo* before being resuspended in 1x TE buffer.

Dephosphorylation of DNA using alkaline phosphatase

Digested plasmid DNA (0.1-0.5 mg) was incubated with 1 unit of alkaline phosphatase for 30 mins at 37°C. The reaction was terminated by the addition of 5 M EDTA, pH 8.0, to a final concentration of 3 mM followed by heating at 65°C for 10 mins. The reaction mix was then extracted twice with hot (56°C) TE saturated phenol and once with chloroform, isoamylalcohol (24:1). DNA was precipitated overnight at -20°C with 2.5 volumes of ethanol and 1/10 volume of 3 M Na acetate, pH 4.8. The precipitate was collected by centrifugation (25,000 x *g* for 30 mins at 4°C), washed once with 1 ml of 70% (v/v) ethanol, dried *in vacuo* and dissolved in 1x TE buffer.

In vitro cloning

DNA to be subcloned (3 μg) was cleaved in either single or double restriction enzyme digests. This was combined with 1 μg of similarly cleaved vector DNA, then ligated with 8 U of T_4 DNA ligase in a volume of 20 μl in a final buffer concentration of 20 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 10 mM dithiothreitol, 0.6 mM ATP for 16 hr at 4°C. The ligated DNA was then used directly for transformation of *E. coli* strains. Transformants were screened for insertional inactivation of the appropriate drug resistance, wherever possible, prior to plasmid DNA isolation.

Plasmid nested deletion

Plasmid DNA (10μg) was digested with a restriction endonuclease enzyme cutting only at the plasmid multiple cloning site (MCS), and not within the DNA insert Suitable enzymes were chosen by detecting a single fragment of the correct size using gel electrophoresis. *Hind*III, followed by *Sma*I, was used for all clones in this thesis. After

*Hind*III digestion, the DNA was precipitated with 2.5 volumes of ethanol and 1/10 vol. 3 M Na acetate pH 4.8, the pellet washed in 70% ethanol, dried *in vacuo*, and resuspended in 50 µl 1 x Klenow buffer. Following end-fill protection with Klenow enzyme (1 U) and α phospho-dNTPs, the DNA was similarly reprecipitated, and digested with *Sma*I ready for deletion digestion, and reprecipitated a third time to purify the pellet. The exonuclease III deletion, ligation and transformation were performed according to the manufacturer's recommendation (Promega Corp, Madison, WI, USA), using 12 time points of 1 min. The deleted plasmids from each time point were ligated and transformed as before. From 15 colonies for each time point, a set of plasmids each sequentially 200bp smaller were selected for sequencing.

Labelling DNA fragments with Digoxigenin-dUTP

Labelling DNA fragments with Digoxigenin-dUTP was performed according to the manufacturer's instructions. PCR products or DNA digests (18 µl) were heated at 100°C with 2 µl of 10 x hexanucleotide. The mixture was cooled on ice for 1 min after which 2 µl of 10 x Digoxigenin labelling mix and 1 µl of 2 U/µl of Klenow was added. Mineral oil was layered over the reaction mix and left to incubate for 18 h at 37°C. The labelled DNA was then added to 30 ml of hybridisation fluid (375 mM NaCl, 37.5 mM tri-sodium citrate, 0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin, 0.1% SDS, 1% skim milk) and stored at -20°C.

Southern blot analysis

Southern blot analysis was performed according to the method of Southern (1975). Briefly, the ethidium bromide stained agarose gel was soaked for 8 to 10 mins in 250 mM HCl. The gel was then washed twice in 500 ml of denaturing solution (1.5 M NaCl, 0.5 M NaOH) and twice in 500 ml of neutralizing solution (0.5 M Tris, pH 7.4: 1.5 M NaCl). DNA bands from the gel were allowed to transfer by capillary action onto a reinforced nylon filter (Amersham) for 18 hr at 25°C in a 10 x SSC solution (7.5 M NaCl, 750 mM tri-sodium

citrate, pH 7.0). The transfer apparatus was assembled to achieve the capillary action by laying the gel on a pre-soaked wick of Whatman filter-paper suspended over the SSC solution with two opposing edges of the wick in contact with the SSC solution. On top of the gel, and cut exactly to size to ensure none came in contact with the wick, were laid the pre-soaked nylon filter, followed by three layers of pre-soaked Whatman filter-paper, and a 100 mm thick block of paper towel. During assembly air bubbles between the pre-soaked filters and gel were removed to ensure even transfer, and the paper towel block was lightly and evenly weighted to maintain even capillarity.

Colony hybridisation probing method

Colony hybridisation probing was performed using a modification of Paton, *et al.* (1996). L-broth (150 µl) was added to each well of a labelled U-bottomed microtitre plate and one transformant was noculated per well to the same location and orientation on duplicate trays. Trays were incubated O/N at 37°C. The transformants in one set of trays were lysed with alkali and DNA transferred to nylon membrane for hybridisation probing to detect transformants containing the required DNA insert, as follows:

Plates were centrifuged in a plate centrifuge (1,500 rpm for 15 mins) to pellet cells, and the supernatant tipped off and drained upside down on absorbent towel. The cells were resuspended in 10 µl TE buffer using a plate shaker, and 5 µl 10% SDS and 50 µl 0.5 M NaOH, 1.5 M NaOH added, mixing thoroughly after each addition. A nylon membrane cut to the same size as the trays was labelled appropriately, and approximately 3 µl from each well was spotted onto the filter using a replicator, marking a grid on the membrane to locate the samples before they dried. After air-drying for approximately 10 mins, the membrane/s were carefully wrapped in plastic cling-film, and the DNA UV-fixed to the membrane using UV light (UV trans-illuminator, 300ηm) for 2 mins. The membrane/s were then hybridised and developed as for Southern hybridisation.

Selected transformants were recovered from the duplicate plate or tray, checked for purity on Xgal/IPTG/Amp plates, and cultured onto CA/Amp plates for -70°C storage.

Hybridisation and development

The nylon filter was then soaked in 6 x SSC solution (3.75 M NaCl, 375 mM tri-sodium citrate, pH 7.0) for 5 mins and the DNA cross-linked onto nylon under ultraviolet illumination for 2 mins. The nylon filter was then probed with hybridisation fluid containing Digoxigenin-labelled DNA for 18 hr at 42°C. Any unbound Digoxigenin-labelled DNA was removed by washing the nylon filter four times in 5 x SSC containing 0.1% SDS, twice at RT for 10 mins and twice at 50°C for 20 mins (Stringency of washing is approximately 10-20% nucleotide base pair mismatch). The nylon filter was blocked with 5% skim milk in buffer-1 (100 mM Tris HCl, pH 7.5, 150 mM NaCl) for 90 mins at 25°C. The filter was then incubated for 30 mins at 25°C with an anti-Digoxigenin-alkaline phosphatase antibody conjugate (1:5000). Following two 15 mins washes in buffer-1 to remove unbound antibody-conjugates, the filter was developed with NBT and X-P in buffer-3 (buffer-1 with 50 mM $MgCl_2$, pH 9.5) in the dark from 2 hr to 24 hr. Development was terminated in 1 x TE buffer.

DNA sequencing and analysis

The nucleotide sequence of clones, including deleted nests of clones, was determined by dideoxy chain-termination using the PRISM™ Ready Reaction Dye Primer Cycle Sequencing Ready Reaction M13 Kit (Perkin Elmer). Nucleotide primers flanking the cloned insert (M13 forward and reverse primers) were used to sequence the insert. The nucleotide sequence of PCR amplified products was determined by dideoxy chain-termination chemistry using the PRISM™ Ready Reaction DyeDeoxy™ Terminator Cycle Sequencing Kit (FS or Big Dye) (Perkin Elmer), after the amplicons were purified using QIAquick PCR Purification Kit (Qiagen GmbH, Hilden, Germany). Sequencing reactions were purified using ethanol/Na acetate precipitation according to the manufacturer's recommendations, and DNA sequencing

was performed on an ABI 373A or 377 DNA sequence analyser (Applied Biosystems, USA). The assembly of contiguous nucleotide sequence, alignment of *mip* sequence from different species and strains, determination of open reading frames, and comparison of amino acid sequence data were performed using Gene Compar 2.0 (Applied Maths, Kotrijk, Belgium).

Polymerase chain reaction

Each PCR reaction mix contained 200 µM of each nucleotide dATP, dGTP, dTTP, and dCTP, 22 pmol of each primers, 50 mM KCl, 10 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 1 ng of genomic DNA as template and 2 U of AmpliTaq DNA polymerase per 25 µl solution. The amplification protocol varied especially with the annealing temperature, depending on the primer set being used, and specific details are given in the relevant chapters. However, a hot start at 94°C was used for all amplifications. Temperature cycling was performed in a Corbett Research (Sydney, Australia). Thermal Sequencer Model FTS-960. All reactions were analysed by agarose gel electrophoresis. The size of the PCR products were estimated by comparing with lanes containing either SPP-1 DNA markers or Amplisize™ DNA size standards (50 to 2000 bp ladder) (Bio-Rad).

Protein analysis

Whole cell method

Legionellae or transformants were grown in 5 ml of the appropriate broth at 37°C, O/N for transformants and 3 days for legionellae, and harvested at 4,000 rpm for 20 mins. The pellet was resuspended thoroughly in 500 µl of sterile purified water, followed by 500µl of double strength Lugtenberg's solution, to give a final concentration of 6.25 mM Tris, pH 6.8, 2% SDS, 10% glycerol and 5% β-mercaptoethanol (Lugtenberg, *et al.*, 1975). The thoroughly mixed suspensions were refrigerated at -20°C until needed. Just before loading the suspensions were steamed for 6 minutes. Some legionellae suspensions were very viscous so were diluted a further 10-fold in buffer.

SDS polyacrylamide gel electrophoresis (PAGE)

The protocol for SDS PAGE was modified from the method of Laemmli (1970). The stacker gel contained 4% acrylamide in 2 mM Tris, pH 6.8 and 0.001% SDS. The separating gel contained 15% acrylamide in a solution containing 11 mM Tris, pH 8.8 and 0.6% SDS. After loading (10 µl for transformants, 40 µl for diluted legionellae suspensions), the gel was run at 20 mA for 18 to 24 hr. All samples were heated at 100°C for 6 mins prior to loading.

Prestained size markers (Bio-Rad) were phosphorylase B (104-kDa), bovine serum albumin (80-kDa), ovalbumin (50-kDa), carbonic anhydrase (32-kDa), soyabean trypsin inhibitor (27-kDa) and lysozyme (18-kDa) were used for molecular size determinations.

Western blot analysis

The method for western blot analysis was according to Towbin, *et al.* (1979). Protein bands from SDS PAGE gel were electrotransferred onto reinforced nitrocellulose membrane (NCM) (Amersham) at 10 to 20 V for 18 hr in electroblot buffer (25 mM Tris, 192 mM glycine, 5% methanol). The NCM was blocked with 5% skim milk in Tris-buffered-saline buffer (TBS) (100 mM Tris, 500 mM NaCl) for 30 mins at 37°C. The NCM was then reacted with primary antibody in TBS containing 0.5% Tween-20 (TTBS) for 18 hr at 25°C. Any unbound antibody was washed off with two 15 mins washes in TTBS. The NCM was then reacted with an anti-species immunoglobulin conjugated with horseradish peroxidase for 4 to 6 hr at 25°C. Any unbound antibody was washed off with two 15 mins washes in TBS. To detect the presence of the antigen-antibody complexes peroxidase substrate (30 mg 4-chloro-1-naphthol dissolved in 10 ml -20°C methanol was added to 49.5 ml TBS containing 50 µl 30% hydrogen peroxide) and allowed to incubate for 10 to 15 mins (RT) as described by Hawkes, *et al.* (1982).

Colony transfer and immuno-blotting

Colony transfer and immunoblotting was followed according to the method of Henning, *et al.* (1979). Duplicate plates were prepared containing transformants spot inoculated to the same location in each with 50 transformants per plate. After O/N incubation at 37°C, one set of plates were refrigerated at 4°C, and the other processed as follows. Nylon filter discs (8.5 cm diameter) were cut to fit, and placed on the surface of each plate, completely covering the colonies, allowed to wet completely and the colonies to adhere (approximately 3 mins). The filters were then carefully removed and placed colony-side up on Whatman filter paper soaked in 0.5 M HCl for 30 mins. The filters were then washed with a stream of saline vigorously expelled from a Pasteur pipette to remove the cell debris. Non-specific binding was blocked in 5% skim milk, agitating for 30 mins. Primary and secondary antibody were added as for western blot described above.

Phylogenetic analysis of sequence data

Analysis of the Legionella *genus*

Nucleotide sequences of the putative reading frame of the *mip* gene from 38 species of *Legionella* and the outgroup *Coxiella burnetii* were aligned with GeneCompar 2.0 using the default settings. Subsequently, the alignment of the hyper-variable region (nucleotide 55-132; codons 19 - 44) was adjusted manually giving a total of 753 aligned sites. Because of the high degree of length variation in the hyper-variable region, homology of sites was uncertain and the region was excluded from all further analyses. The additional 8 nucleotides present at the carboxy terminal end of *L. israelensis* were also excluded as their unique presence in this species is phylogenetically uninformative. The *16S rRNA* sequences for the same taxa that were available for the *mip* data set were also aligned with GeneCompar using the default settings, giving a total of 1405 aligned sites. As there were very few putative indels (insertion/deletions) required in the ribosomal RNA alignment, the entire alignment was used

in the phylogenetic analysis. The alignment of the *16S rRNA* sequences used by Hookey, *et al.*, (1996) was not available for comparison and may differ from the alignment used in the present study because Hookey, *et al.* (1996) included taxa additional to those considered here: *L. bozemanii* (two strains), *L. geestiana*, "Glasgow" 86/35/84, LLAP-3, *Sarcobium lyticum*, and *L. pneumophila* subsp. *fraseri*.

Transition/transversion (ts/tv) ratios were estimated by maximum likelihood analysis using DNAML option in PHYLIP version 3.5 (Felsenstein, 1993). Optimal likelihood values were derived independently for each of the two nucleotide data sets by using a range of ts/tv ratios (1, 2, 5, 10, 15 or 20). Nucleotide frequencies, codon usage and evolutionary distances between sequences were calculated using MEGA (Kumar, *et al.*, 1993). Phylogenetic analyses using the maximum parsimony criterion of optimality were performed with PAUP version 3.1 (Swofford, 1993) and distance analyses were performed with MEGA. Bootstrap analysis was used to estimate the reliability of nodes. The phylogenetic analyses for the two nucleotide data sets were performed independently to assess the degree of conflict between the resultant trees. The number of transitions and transversions were compared with increasing evolutionary distance to determine if the substitutions were accumulating in a linear manner, or were saturated.

Analysis of the L. anisa – L. worsleiensis *clade*

Nucleotide sequences from the putative reading frame of the *mspA/proA* gene from the 19 species of *Legionella* within the *anisa – worsleiensis* clade and the outgroup *L. micdadei* were aligned with GeneCompar using the default settings to give a total of 438 aligned sites. The three nucleotides (positions 46-48, forming one inframe codon) uniquely present in the sequence from *L. micdadei* were excluded, as they are phylogenetically uninformative. Nucleotide frequencies, evolutionary distance p and ts and tv frequencies and ratio were calculated using MEGA. Additional specific information is given in Chapter Five.

Table 2.1. General strains and plasmids used in this thesis.

Strains	Relevant characteristics	Source or reference
E. coli DH5α	F^-, ø80d*lacZ*ΔM15, *recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17* (r_K^-, m_K^+), *supE44*, *relA1*, *deoR*, Δ(*lacZYA-argF*)U169	(Hanahan, 1985)
DH5α [pIMVS26]	*E. coli* DH5α containing pIMVS26 plasmid	(Doyle, *et al.*, 1998)
DH5α [pRMR-anisa]	*E. coli* DH5α containing pRMR-anisa plasmid	This study
DH5α [pRMR-cin]	*E. coli* DH5α containing pRMR-cin plasmid	This study
DH5α [pRMR-dum]	*E. coli* DH5α containing pRMR-dum plasmid	This study
DH5α [pRMR-sant]	*E. coli* DH5α containing pRMR-sant plasmid	This study

Plasmids	Relevant characteristics	Source or reference
pGEM®-7Zf(-)	cloning vector, ampicillin resistant	Promega
pIMVS26	pGEM carrying a 8 kb fragment containing *mip* gene from *L. longbeachae* sg 1 ATCC 33462	(Doyle, *et al.*, 1998)
pRMR-anisa	pGEM carrying a 2.1kb *Cla*I-*Sac*I fragment containing *mip* gene from *L. anisa* ATCC 35292	This study
pRMR-cin	pGEM carrying 1.7 kb *Sph*I fragment containing *mip* gene from *L. cincinnatiensis* ATCC 43753	This study
pRMR-dum	pGEM carrying a 1.5 kb *Sph*I fragment containing *mip* gene from *L. dumoffii* ATCC 33279	This study
pRMR-sant	pGEM carrying 1.7 kb *Sph*I fragment containing *mip* gene from *L. santicrucis* ATCC 35301	This study

Chapter Three

Functional implications of the inter-species gene variation

Introduction

Table 1.1 lists the 41 published species currently comprising the genus *Legionella*, and the 23 species which have been associated with human disease. The Mip protein was one of the first virulence determinants reported (Cianciotto, *et al.*, 1989b). Knockout mutations in the *mip* gene of *L. pneumophila*, *L. micdadei* and *L. longbeachae*, showed reduced ability to survive immediately after uptake into phagocytic cell lines or amoebae, and are attenuated in an animal model of virulence (Cianciotto, *et al.*, 1989b, Doyle, *et al.*, 1998, O'Connell, *et al.*, 1996). Mip exhibits peptidyl-prolyl *cis/trans* isomerase (PPIase) activity of the FKBP class (Fischer, *et al.*, 1992), but the exact role of this protein in the intracellular life-cycle is not understood. *Legionella*-derived protein maturation or trafficking, host receptor recognition, or inhibition of host defence mechanisms have been postulated (Hacker and Fischer, 1993). Alternatively, it is possible that PPIase function has no role in intracellular infection and the function of the protein is due to an as yet undetermined activity (Wintermeyer, *et al.*, 1995). Since Mip-like analogues have been detected widely in both procaryotes and eucaryotes (Fischer, *et al.*, 1992, Horne, *et al.*, 1997, Horne and Young, 1995), they are likely to have a significant cellular role (Hacker and Fischer, 1993).

The *mip* sequences determined for *L. pneumophila* (Engleberg, *et al.*, 1989), *L. micdadei* (Bangsborg, *et al.*, 1991) and *L. longbeachae* (Doyle, *et al.*, 1998) show marked

similarity to each other, although *L. micdadei* has a 30 bp insert in the 5' region of the Open Reading Frame (ORF). These sequences also show complete conservation of 7 amino acids in the C-terminal portion of the protein, which has been determined by site-directed mutagenesis to be important for PPIase activity, and is postulated to form the active site of the protein (Hacker and Fischer, 1993). Using Southern hybridisation, *mip*-like genes have been detected in many other *Legionella* species (Bangsborg, *et al.*, 1991, Riffard, *et al.*, 1996). However, stringency conditions (30-50% base pair mismatch) suggest that differences between species exist in this gene, which may indicate functional variability, and could account for the ability of certain species to cause human disease.

This chapter reports the sequence of the *mip* gene, and the inferred amino acid sequence, from an additional 35 *Legionella* species, allowing a comparison of the gene between species, to determine if structural differences within the Mip protein may account for the variable association of different species with disease.

Method and sequencing strategy

The following type strains were examined for the presence of a Mip-like protein using the whole cell method to prepare the protein extracts, followed by PAGE and western blot analysis using *L. longbeachae* sg 1 Mip antisera: *L. anisa*, *L. bozemanii* sg 1, *L. cherrii*, *L. cincinnatiensis*, *L. dumoffii*, *L. erythra*; *L. fairfieldensis*, *L. jamestowniensis*, *L. jordanis*, *L. longbeachae* sg 1, *L. micdadei*, *L. pneumophila* sg 1 Philadelphia-1, *L. quinlivanii*, *L. rubrilucens*, *L. sainthelensi* sg 1; *L. santicrucis*, and *L. tucsonensis*. DH5α[pIMVS26], known to express *L. longbeachae* sg 1 Mip, was included as a control. Mip-like proteins were detected in *L. anisa*, *L. cincinnatiensis*, *L. sainthelensi* sg 1, and *L. santicrucis*, as well as the control strain.

Chromosomal DNA from these four *Legionella* species were digested O/N at 37°C with the following restriction endonucleases: *Sph*I, *Xba*I, *Eco*RI, *Kpn*I, *Sma*I, *Cla*I, *Hind*III, *Bam*HI and *Sac*I. The restricted fragments were separated by size using electrophoresis, transferred to nylon membrane using Southern transfer, and probed by hybridisation with a Digoxigenin-labelled *L. longbeachae* sg 1 *mip* probe, generously provided by R. Doyle, IMVS, Adelaide, South Australia. All restriction endonucleases produced restricted fragments which hybridised, although many were above 4 kb in size. However, the following fragments were considered suitable for cloning into a plasmid vector: *L. cincinnatiensis* (*Sph*I fragment of 1.7 kb), *L. dumoffii* (*Sph*I fragment of 1.5 kb) and *L. santicrucis* (*Sph*I fragment of 1.7 kb). *L. anisa* DNA produced a 2.5 kb and a 4.9 kb fragment following digestion with *Cla*I and *Sac*I endonucleases respectively, which hybridised to the *mip* probe. A double digestion with both produced a fragment of 2.1 kb which hybridised (*L. anisa*, *Cla*I - *Sac*I fragment of 2.1 kb), and was also chosen for cloning into the plasmid vector.

The four *mip*-containing DNA fragments were ligated into pGEM®-7Zf(-) plasmid and transformed into DH5α. Transformants were selected using colony transfer and immuno-blotting. The presence of a single insert was confirmed with and without restriction endonuclease digestion of the recovered plasmid, and expression of Mip was confirmed by PAGE and western immuno-blot.

Four transformants were chosen, one for each species (see Table 2.1), and the plasmids purified using the large scale method (Method 1). A set of nested deletions in 200 bp steps was prepared for each of the four plasmids, and PEG-precipitated, small scale plasmid purifications of each were prepared for sequencing. Sequencing of the nested deletions from each of the four plasmid was performed using Dye Primer Sequencing and M13 forward and reverse primers. Unambiguous consensus sequences of each plasmid insert were assembled from overlapping sequence in both forward and reverse direction, and the resulting sequence

examined for the presence of Open Reading Frames (ORF). The following *mip*-like ORFs were detected. Those published for other species, namely *L. pneumophila*, *L. micdadei* and *L. micdadei*, the ORF size and pair-wise percentage homology are included:

	Legionella species	ORF size	Pair-wise percentage homology						
1	*L. anisa*	696 bp	100						
2	*L. dumoffii*	702 bp	89.2	100					
3	*L. santicrucis*	699 bp	87.8	86.3	100				
4	*L. cincinnatiensis*	699 bp	87.3	86.4	96.4	100			
5	*L. longbeachae* sg 1	699 bp	86.6	85.7	94.3	94.7	100		
6	*L. pneumophila* (Phil-1)	702 bp	78.7	78.9	79.5	78.5	79.0	100	
7	*L. micdadei*	729 bp	73.4	73.2	72.8	73.5	73.8	73.7	100
			1	2	3	4	5	6	7

Sequencing strategy for other species

An alignment of all seven *mip* gene-containing sequences, including the up- and down- stream regions, was examined for regions of homology. A series of forward and reverse consensus primers were designed and constructed targeting these regions, in both up- and down-stream regions, and within the *mip* gene sequence. These primers were used in various combinations of forward and reverse primer pairs in amplification reactions containing chromosomal DNA from the remaining 33 *Legionella* species. Any well-amplified single product close to the predicted size when compared to that expected from known sequences was sequenced with dye terminator sequencing and the homology with known *mip* gene sequences examined. If necessary, additional internal primers were designed to complete the sequencing of these amplified fragments. Partial fragments of the *mip* gene from new species were compared with already completed sequences, and new primers designed from the most homologous sequence which may amplify additional regions of incomplete sequence. Again, any well-amplified single product close to the predicted size was sequenced, and

compared to the sequence already obtained. Progressively, the *mip* gene sequence from an additional 32 *Legionella* species were sequenced with this strategy. In total, 97 primers, ranging from 15 to 29 bases long, were constructed to produce and sequence enough amplicons for unambiguous sequence to be determined in either or both directions to recover the whole gene including, if possible, the flanking regions. Gene amplification (Saiki, *et al.*, 1988) consisted of 35 cycles of denaturation at 94°C for 1 minute, annealing for 1 minute at various temperatures from 40°C to 60°C depending on primers and species, and extension at 72°C for 2 min.

Results

The open reading frame

Table 3.1 lists the type strains from the 35 *Legionella* species from which the *mip* gene sequence was recovered, together with the three sequences already published. The ATCC strain number, and the GeneBank storage number of the recovered or published sequences are included, as well as the relevant *16S rRNA* sequence GeneBank accession numbers used in Chapter Four.

An alignment of the inferred amino acid sequences of all 38 species (35 determined in this study, and the three sequences already published) are given in Fig. 3.1. None of the primers used produced an amplicon for *L. geestiana*, so the *mip* sequence for *L. geestiana* could not be determined by the techniques used. Two sequences are partial: the first four codons for *L. londiniensis*, and the C-terminal codons for *L. jamestowniensis* could not be recovered by the techniques used because specific primers for these regions could not be ascertained to enable amplification for sequencing. The complete sequences ranged from 232 to 251 amino acids. Sequence conservation ranges from 82 to 99 percent (69-97% at nucleotide level). Inferred amino acid sequences demonstrate a typical signal sequence, with minor interspecies differences. Fourteen species have a hyper-variable insert of up to 17

amino acids immediately adjacent to the signal sequence. One species, *L. israelensis*, has an additional 8 amino acids at the C-terminal region. Excluding these two regions, 116 of the remaining 225 sites are totally conserved for all species, usually grouped in regions of near total or total conservation, especially in the C-terminal portion of the protein, and interspersed with either single sites or short regions exhibiting variation. Such variation is typically conservative, such as serine (S), asparagine (N) or threonine (T) at site 111, and valine (V) and isoleucine (I) at sites 70, 138, 149, 214 and 216. The alignment of the inferred amino acid sequences was analysed using OMIGA 1.1 (Oxford Molecular Ltd., Oxford, UK) to determine if functional differences in the Mip protein could account for the variable association of species with disease. Figure 3.2 presents the alignment, with residues classified by chemical type, and the involvement in secondary structure predicted by the GOR II Method (Garnier, *et al.*, 1978, Garnier, *et al.*, 1996) displayed under each sequence. Figure 3.3 presents three hydropathy predictions for the inferred amino acid sequence for two representative species for each of the two disease association groups. Kyte-Doolittle hydropathy (Kyte and Doolittle, 1982), Goldman/Engelman/Steitz (GES) hydrophilicity (Engelman, *et al.*, 1986) and von Heijne hydrophilicity (von Heijne, 1981) are displayed. Figure 3.4 presents structural predictions on the same set of inferred amino acid sequences, including Argos (Argos, *et al.*, 1982) and von Heiijne (von Heijne, 1992) transmembrane helices, Karplus protein flexibility (Karplus and Shultz, 1985), Parker antigenicity (Parker, *et al.*, 1986) and Thornton protrusion index antigenicity (Thornton, *et al.*, 1986). Each of the figures confirm significant conservation of the secondary structure, and do not reveal any consistent difference between the two disease association groups. The 7 amino acids (highlighted in Fig. 3.1) determined to be crucial for PPIase activity in *L. pneumophila* (Hacker and Fischer, 1993), and conserved for *L micdadei* and *L. longbeachae*, are also totally conserved in the additional 35 species. The regions of highest homology encompass these sites. No single amino acid or region was observed that identified sequences from those species for which there is an association with disease. Among species with an association with human disease, 138 of 251 sites were

invariant, compared with 121 invariant sites for species never associated with disease. An analysis of the variant residues showed these differences were the result of one or several species in each group varying from the most common amino acid at each position. No one site contained an amino acid common to Mip for all species associated with disease, but totally substituted for those species never associated with disease, or vice versa.

The sequences for *L. fairfieldensis* and *L. worsleiensis* reported here are different from those published by Riffard, *et al.* (Riffard, *et al.*, 1996). Since their reported sequences are identical to various *L. pneumophila* serogroups (Ratcliff, *et al.*, 1998), laboratory contamination of either cultures or the amplification process with *L. pneumophila* seems the most likely explanation. Even the independently obtained *16S rRNA* data do not support such a high level of similarity of these species with *L. pneumophila* (Hookey, *et al.*, 1996).

A comparison of codon usage with that determined for *E. coli* (Hénaut and Danchin, 1996) revealed little similarity with any of the three gene classes, namely those involved with metabolic processes, those highly and continuously expressed during exponential growth, and those involved in horizontal transfer of DNA. In *Legionella*, codons terminating in T were usually predominant as for *E. coli*, but with higher frequency, and A was often the next most prevalent terminating base. In *E. coli*, A is often the least common terminating base.

Flanking sequences

While the putative ribosomal binding site (RBS) and spacing to the initiation codon was relatively conserved in the upstream region of the *mip* ORF, the remaining upstream sequence generally exhibited a marked reduction in homology (as low as 55% similarity between some species) compared with homology within the ORF. For this reason, it was not possible to produce amplicons spanning substantial portions of this region for many species. However, for those species where the ascertained sequence extended into the upstream region

to include the expression regulation signal sites, it was possible to determine putative -10 and -35 boxes. Confirmation of such sites has only been reported for *L. pneumophila* (Engleberg, *et al.*, 1989), *L. micdadei* (Bangsborg, *et al.*, 1991) and *L. longbeachae* (Doyle, *et al.*, 1998). In the downstream regions, a similar reduction in homology was observed. Moreover, stem loop structures exhibiting features of a rho-independent transcriptional terminator could be observed with all species where sequence data were available for this region.

Discussion

The conservation of the amino acid sequence of the Mip protein is substantial. All of the amino acids defined as essential for PPIase activity are totally conserved. Similarly, large regions are also highly conserved, especially surrounding the residues known to be involved in PPIase activity. Synonymous substitutions are very common, but non-synonymous substitutions appear to be tolerated by natural selection only if the change was conservative, or in non-critical regions of the protein. Since the most conserved regions are associated with PPIase activity, the results suggest PPIase activity is an important function of this protein, a point which may be in conflict with the results of Wintermeyer, *et al.* (1995). Using site-directed mutants defective in PPIase function in Mip, they reported that PPIase activity does not contribute to intracellular survival of *L. pneumophila* in U937 cells. In a recent review, Abu Kwaik, *et al.* (1998a) interpreted these results to infer that the PPIase function of Mip is not involved in its function in intracellular infection. An hypothesis to accommodate these apparently discrepant results would require that Mip exhibits at least two functions, one as a PPIase which is not critical during intracellular survival, and at least one other non-PPIase function important during intracellular infection. It would be interesting to construct residue specific Mip mutants that expressed the same phenotype as the knockout mutants, to investigate which residues if any are important for this second hypothetical function, and how they relate to those involved with PPIase activity. The conserved regions adjacent to the signal sequence could encompass such a hypothetical second functional domain, as it is

possible this region is extended out from the cell surface, and could be involved in attachment or signalling processes with the host cell. However, Wintermeyer, *et al.* (1995) did observe a 50-100 fold reduction in invasion rate, similar to that observed previously with a deletion mutant (Cianciotto, *et al.*, 1989b, Cianciotto, *et al.*, 1990b, Cianciotto and Fields, 1992). Thus it is entirely possible that Mip has no enzymic function other than as a PPIase, and that its role is not associated with intracellular multiplication, but rather during the survival events immediately following uptake into the host cells, as suggested by Cianciotto, *et al.* (1989b). Susa, *et al.* (1996) were able to demonstrate that Mip is maximally expressed 4-8 hours following uptake of the bacteria by a macrophage cell line. Mip was undetectable at one and 32 hours post infection, and only in low amounts at two and 16 hours. The significant upregulation at 4-8 hours confirms the hypothesis of Cianciotto, *et al.* (1989b) that Mip is not involved in the uptake phase of infection. Furthermore, Susa, *et al.* (1996) postulate that Mip is involved in intracellular multiplication, but in the light of the findings of Wintermeyer, *et al.*, (1995) perhaps only during the early growth phase, and suggest a signalling role between bacteria and host as a possible role. In this model, the totally conserved regions within the protein adjacent to the signal sequence could be involved in non-enzymic functions such as forming the dimeric form of the active protein. The analysis of the dimerization of Mip with small-angle X-ray solution scattering, confirms that the contact regions between the monomers forming the dimer involves primarily the N-terminal domains, especially the long α-helices (Schmidt, *et al.*, 1995). Either way, these regions must be essential for functional integrity, and site specific mutations within these regions could reveal important information relating to the role of Mip.

The apparent functional conservation across species strongly suggests Mip is a protein critical to the intracellular life of the organism, and given the wide-spread presence and conservation in both prokaryotes and eukaryotes of similar FKBP immunophilins, it would appear to belong to an important class of proteins. Further, since putatively functional Mip

proteins are found in *Legionella* species never associated with disease, the results support the notion that the ultimate virulence potential of each species is multifactorial, and is dependent upon other proteins as well as host and environmental factors. This concept is supported by the results of Ludwig, *et al.,* (1994) who demonstrated that *L. pneumophila* strains with identical or near identical *mip* sequence and Mip PPIase activity have substantially different virulence characteristics. In addition, O'Connell, *et al.* (1996) demonstrated that *L. jamestowniensis* and *L. parisiensis* demonstrate low ID_{50} values in a macrophage invasion model, even though they have never been associated with human disease.

The functional significance of the hyper-variable region is not known, and there is no relationship between the presence or size of the hypervariable region and association with disease. The sequence within the hypervariable regions is not homologous to any other region within the *mip* gene or to any sequence currently submitted to GeneBank. There is however good agreement with the protein size estimations made by Cianciotto, *et al.* (1990a) based on western immunoblot, and the presence of the hypervariable region.

Susa, *et al.* (1996) made a brief mention of unpublished results which suggested that the effect of a non-functional Mip within *L. pneumophila* could be reversed with cytosolic Mip. Their results demonstrated that mastocytoma cells transfected with the *mip* gene were able to support intracellular growth of a Mip-deficient *L. pneumophila* mutant. If this observation is reproducible in a variety of hosts, it would have substantial implications for the understanding of the function of Mip, implying a principal function as a extracellular protein, and not just a surface-exposed outer membrane protein.

Table 3.1 Type strains and sequences used in this study.

[1] sequences determined during this study, with the exception of those from *L. longbeachae*, *L. micdadei*, and *L. pneumophila*.

[2] Accession numbers have been interchanged to be consistent with ATCC isolate numbers and species (see text).

Organism	ATCC number	GeneBank sequence accession number rRNA gene	*mip* gene
L. adelaidensis	ATCC 49625	Z49715	U91606
L. anisa	ATCC 35292	Z32635	U91607
L. birminghamensis	ATCC 43702	Z49717	U91608
L. bozemanii sg 1	ATCC 33217	Z49718	U91609
L. brunensis	ATCC 43878	Z32636	U92227
L. cherrii	ATCC 35252	Z49720	U91635
L. cincinnatiensis	ATCC 43753	Z49721	U91636
L. dumoffii	ATCC 33279	Z32637	U91637
L. erythra	ATCC 35303	M36027	U92203
L. fairfieldensis	ATCC 49588	Z49722	U92204
L. feeleii sg 1	ATCC 35072	Z49740	U92205
L. gormanii	ATCC 33297	Z32639	U91638
L. gratiana	ATCC 49413	Z49725	U92206
L. hackeliae sg 1	ATCC 35250	M36028	U92207
L. israelensis	ATCC 43119	Z32640	U92208
L. jamestowniensis	ATCC 35298	Z49726	U92228
L. jordanis	ATCC 33623	Z32667	U92209
L. lansingensis	ATCC 49751	Z49727	U92210
L. londiniensis	ATCC 49505	Z49730[2]	U92229
L. longbeachae sg 1	ATCC 33462	M36029	X83036
L. maceachernii	ATCC 35300	Z32641	U92211
L. micdadei	ATCC 33218	M36032	S62141
L. moravica	ATCC 43877	Z49729	U92212
L. nautarum	ATCC 49506	Z49728[2]	U92213
L. oakridgensis	ATCC 33761	Z32642	U92214
L. parisiensis	ATCC 35299	Z49731	U92215
L. pneumophila Philadelphia-1	ATCC 33152	M36023	S42595
L. quateirensis	ATCC 49507	Z49732	U92216
L. quinlivanii	ATCC 43830	Z49733	U92217
L. rubrilucens	ATCC 35304	Z32643	U92218
L. sainthelensi sg 1	ATCC 35248	Z49734	U92219
L. santicrucis	ATCC 35301	Z49735	U92220
L. shakespearei	ATCC 49655	Z49736	U92221
L. spiritensis	ATCC 35249	M36030	U92222
L. steigerwaltii	ATCC 35302	Z49737	U92223
L. tucsonensis	ATCC 49180	Z32644	U92224
L. wadsworthii	ATCC 33877	Z49738	U92225
L. worsleiensis	ATCC 49508	Z49739	U92226
Coxiella burnetii		M21291	U14170

Figure 3.1 Aligned inferred amino acid sequences of the Mip protein from 38 species of *Legionella*. Symbols are as follows: ✳ = identical amino acid to *L. pneumophila*; X = undetermined amino acid; blank space corresponds to an alignment gap; ▼ = amino acid associated with enzyme activity; ◆ = signal sequence cleavage site, determined by N-terminal analysis of *L. longbeachae mip* (Doyle, *et al.*, 1998).

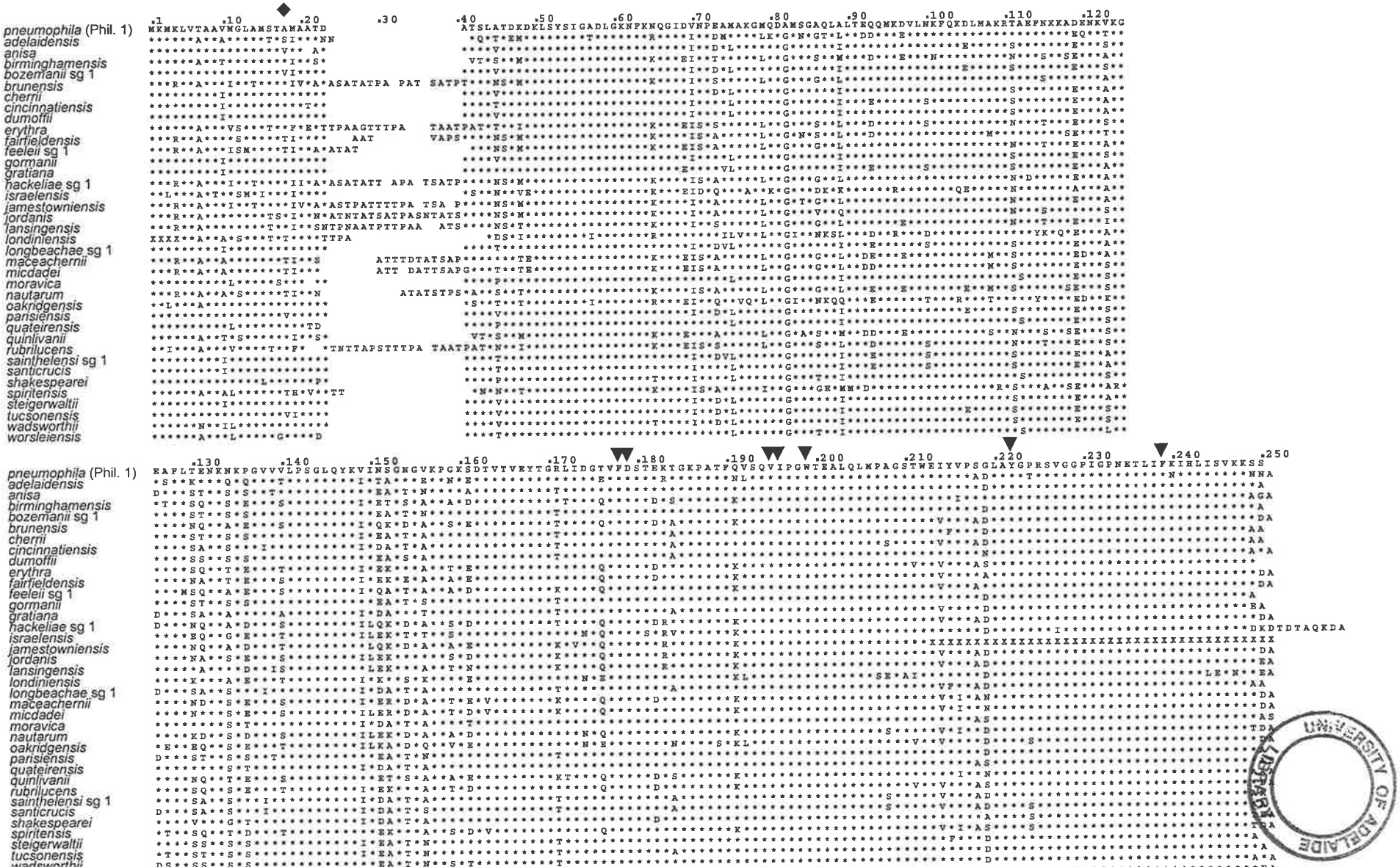
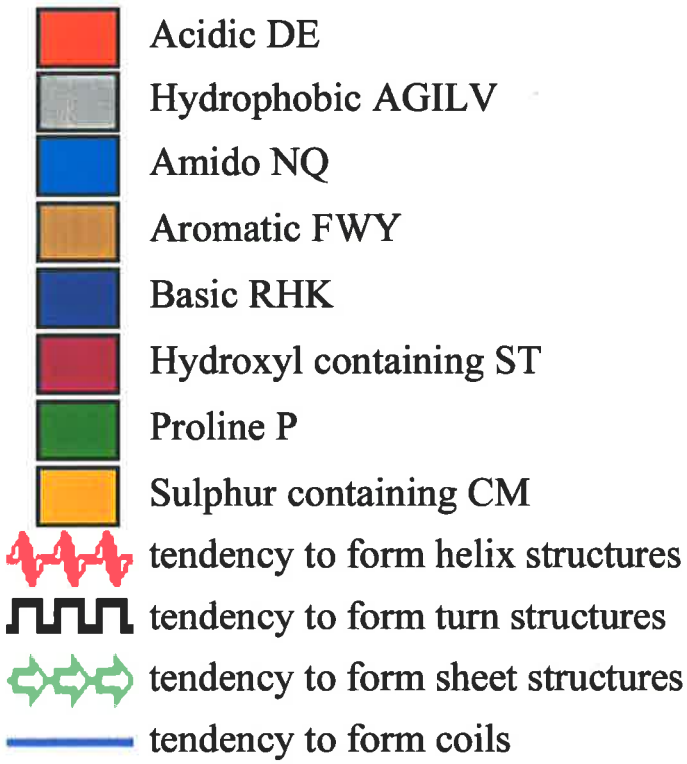
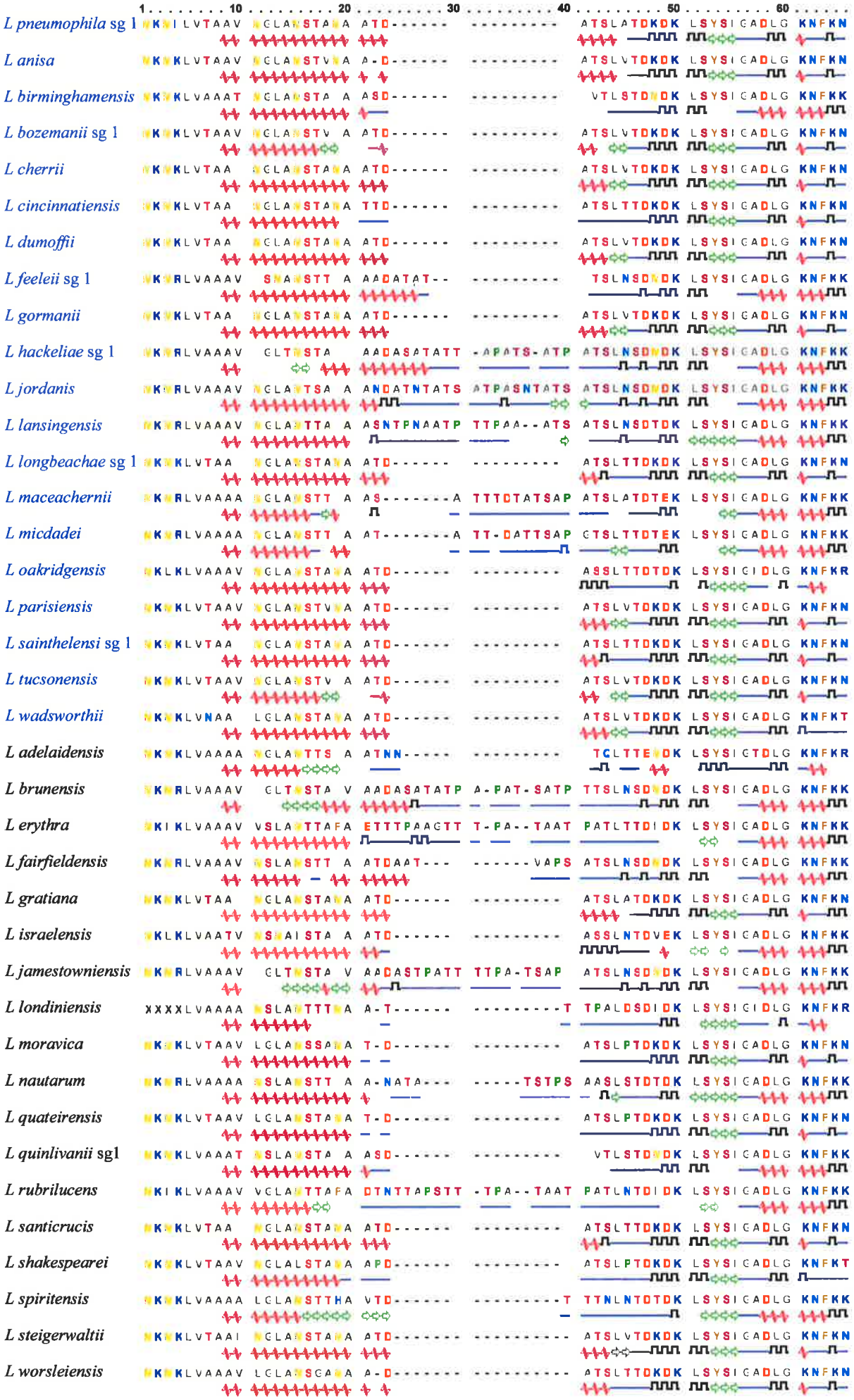
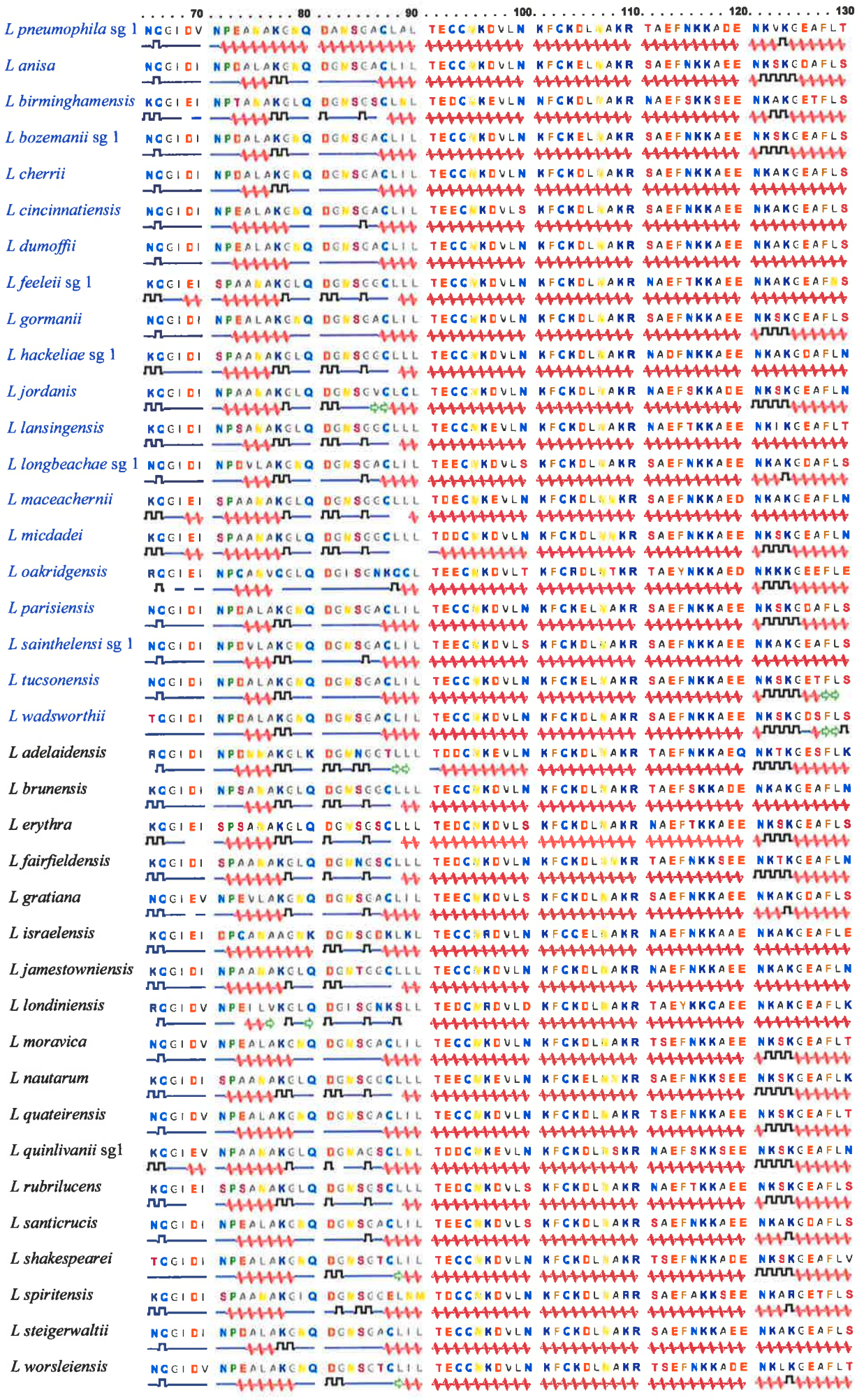

pneumophila (Phil. 1)
adelaidensis
anisa
birminghamensis
bozemanii sg 1
brunensis
cherrii
cincinnatiensis
dumoffii
erythra
fairfieldensis
feeleii sg 1
gormanii
gratiana
hackeliae sg 1
israelensis
jamestowniensis
jordanis
lansingensis
londiniensis
longbeachae sg 1
maceachernii
micdadei
moravica
nautarum
oakridgensis
parisiensis
quateirensis
quinlivanii
rubrilucens
sainthelensi sg 1
santicrucis
shakespearei
spiritensis
steigerwaltii
tucsonensis
wadsworthii
worsleiensis
UNIVERSITY OF ADELAIDE LIBRARY

Figure 3.2 Diagram (four pages) of the alignment of inferred amino acid sequences of the Mip protein from 38 species of *Legionella*. Species named in blue have a documented association with human disease; the remainder have been isolated only from environmental sources. Residues are coloured by chemical type. Symbols indicate secondary structures predicted by the GOR II Method (Garnier, *et al.*, 1978, Garnier, *et al.*, 1996).




1 10 20 30 40 50 60
L pneumophila sg 1
L anisa
L birminghamensis
L bozemanii sg 1
L cherrii
L cincinnatiensis
L dumoffii
L feeleii sg 1
L gormanii
L hackeliae sg 1
L jordanis
L lansingensis
L longbeachae sg 1
L maceachernii
L micdadei
L oakridgensis
L parisiensis
L sainthelensi sg 1
L tucsonensis
L wadsworthii
L adelaidensis
L brunensis
L erythra
L fairfieldensis
L gratiana
L israelensis
L jamestowniensis
L londiniensis
L moravica
L nautarum
L quateirensis
L quinlivanii sg1
L rubrilucens
L santicrucis
L shakespearei
L spiritensis
L steigerwaltii
L worsleiensis


70
80
90
100
110
120
130
L pneumophila sg 1
L anisa
L birminghamensis
L bozemanii sg 1
L cherrii
L cincinnatiensis
L dumoffii
L feeleii sg 1
L gormanii
L hackeliae sg 1
L jordanis
L lansingensis
L longbeachae sg 1
L maceachernii
L micdadei
L oakridgensis
L parisiensis
L sainthelensi sg 1
L tucsonensis
L wadsworthii
L adelaidensis
L brunensis
L erythra
L fairfieldensis
L gratiana
L israelensis
L jamestowniensis
L londiniensis
L moravica
L nautarum
L quateirensis
L quinlivanii sg1
L rubrilucens
L santicrucis
L shakespearei
L spiritensis
L steigerwaltii
L worsleiensis


130. . . . . 140. . . . . 150. . . . . 160. . . . . 170. . . . . 180. . . . . 190. . . . .
L pneumophila sg 1
L anisa
L birminghamensis
L bozemanii sg 1
L cherrii
L cincinnatiensis
L dumoffii
L feeleii sg 1
L gormanii
L hackeliae sg 1
L jordanis
L lansingensis
L longbeachae sg 1
L maceachernii
L micdadei
L oakridgensis
L parisiensis
L sainthelensi sg 1
L tucsonensis
L wadsworthii
L adelaidensis
L brunensis
L erythra
L fairfieldensis
L gratiana
L israelensis
L jamestowniensis
L londiniensis
L moravica
L nautarum
L quateirensis
L quinlivanii sg1
L rubrilucens
L santicrucis
L shakespearei
L spiritensis
L steigerwaltii
L worsleiensis

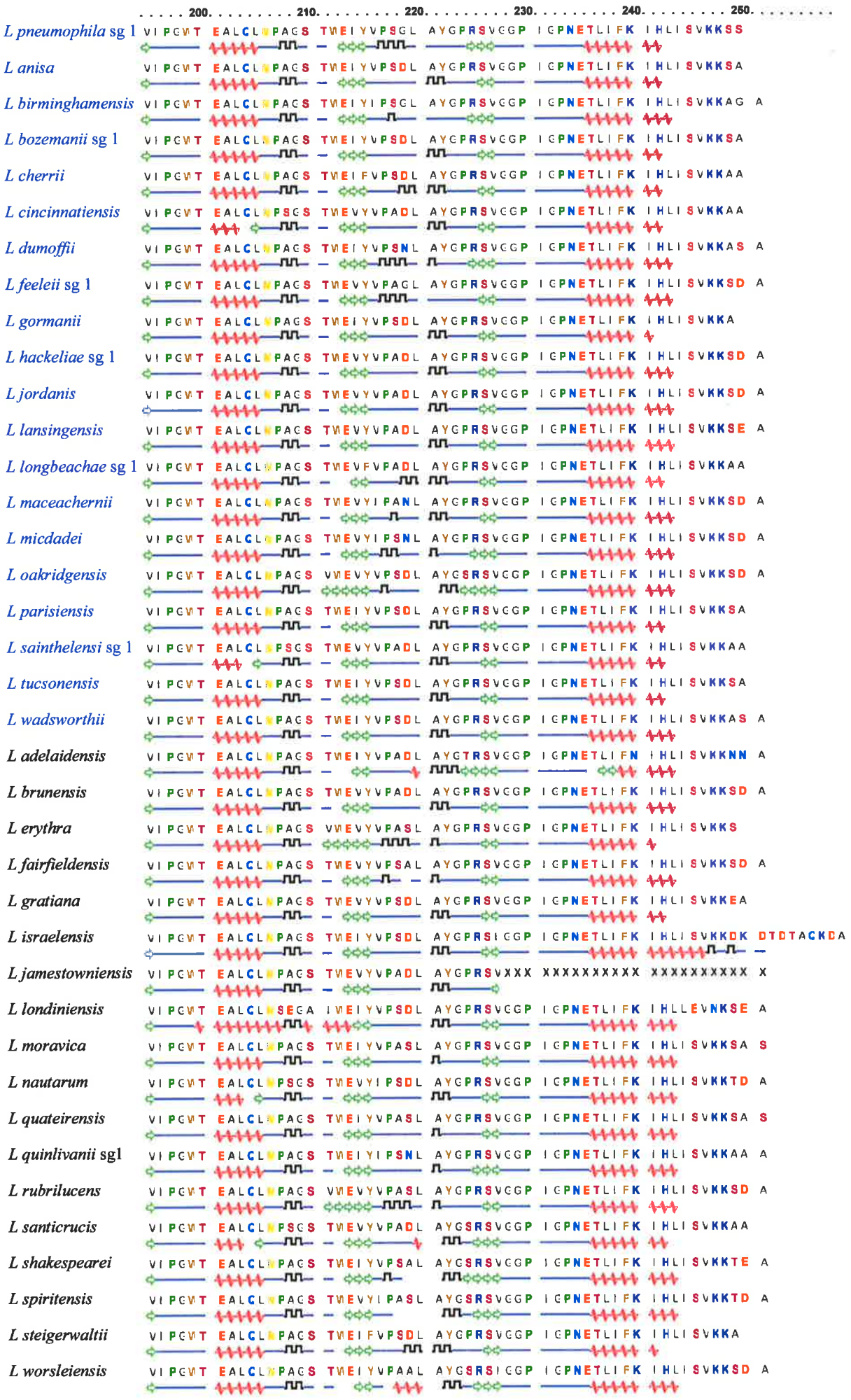
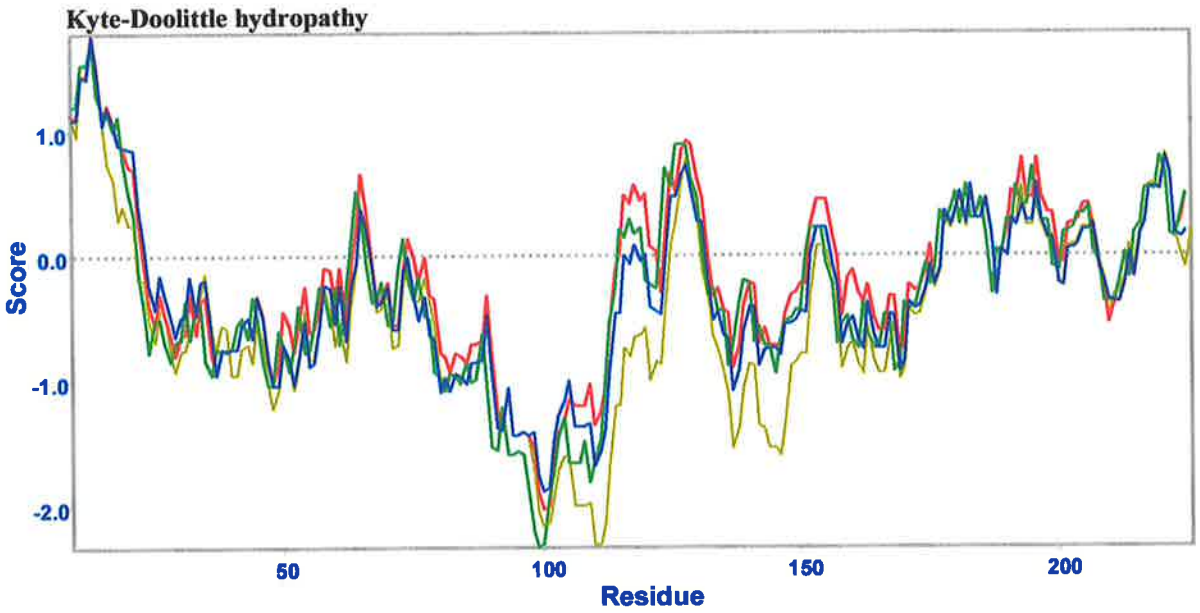
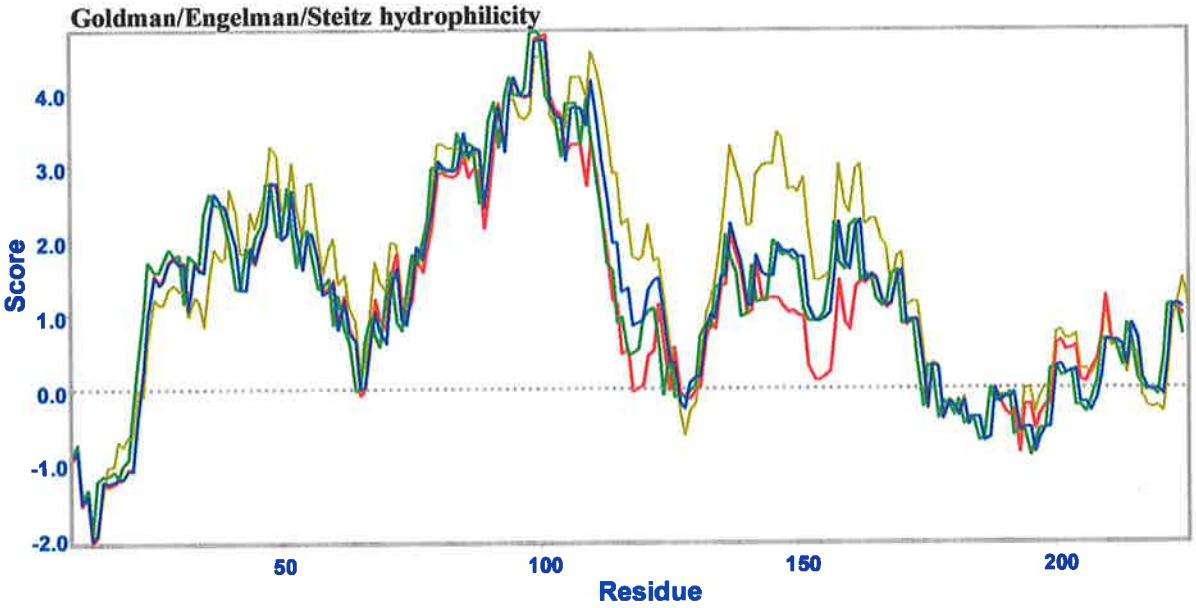

200
210
220
230
240
250
L pneumophila sg 1
L anisa
L birminghamensis
L bozemanii sg 1
L cherrii
L cincinnatiensis
L dumoffii
L feeleii sg 1
L gormanii
L hackeliae sg 1
L jordanis
L lansingensis
L longbeachae sg 1
L maceachernii
L micdadei
L oakridgensis
L parisiensis
L sainthelensi sg 1
L tucsonensis
L wadsworthii
L adelaidensis
L brunensis
L erythra
L fairfieldensis
L gratiana
L israelensis
L jamestowniensis
L londiniensis
L moravica
L nautarum
L quateirensis
L quinlivanii sg1
L rubrilucens
L santicrucis
L shakespearei
L spiritensis
L steigerwaltii
L worsleiensis

Figure 3.3 Hydropathy predictions of the Mip protein for four species of *Legionella*. *L. pneumophila* sg 1 (Philadelphia-1) and *L. longbeachae* sg 1 are associated with disease. *L. adelaidensis* and *L. moravica* have never been associated with disease. Produced by OMIGA 1.1. References are as follows: Kyte-Doolittle hydropathy (Kyte and Doolittle, 1982); Goldman/Engelman/Steitz (GES) hydrophilicity (Engelman, *et al.*, 1986); von Heijne hydrophilicity (von Heijne, 1981).


Kyte-Doolittle hydropathy
Score
1.0
0.0
-1.0
-2.0
50
100
150
200
Residue


Goldman/Engelman/Steitz hydrophilicity
Score
4.0
3.0
2.0
1.0
0.0
-1.0
-2.0
50
100
150
200
Residue

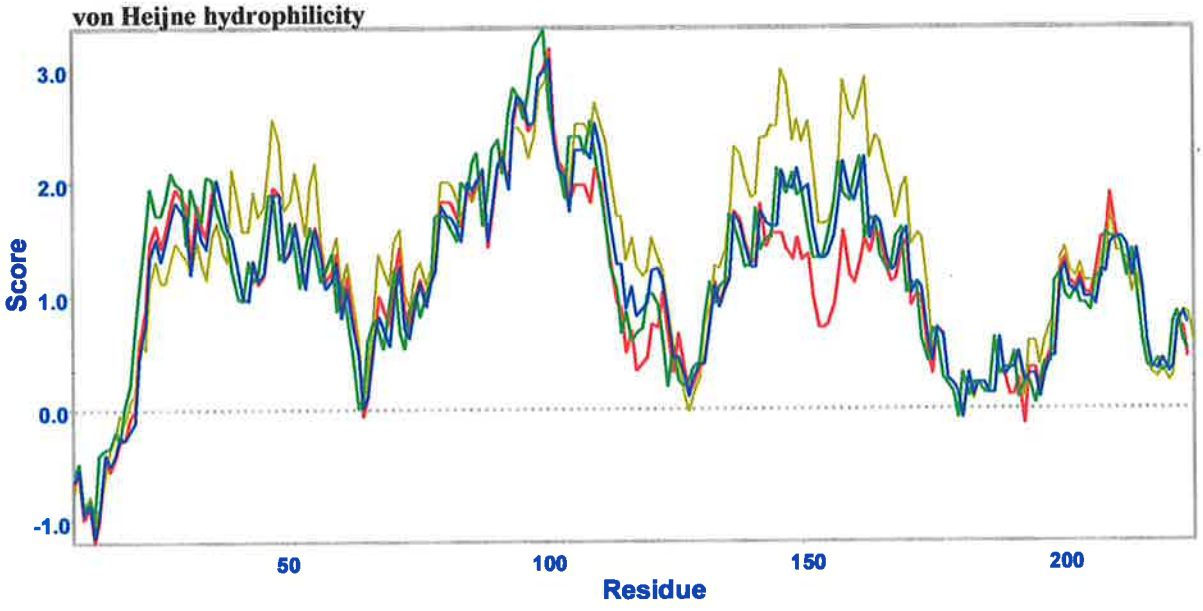
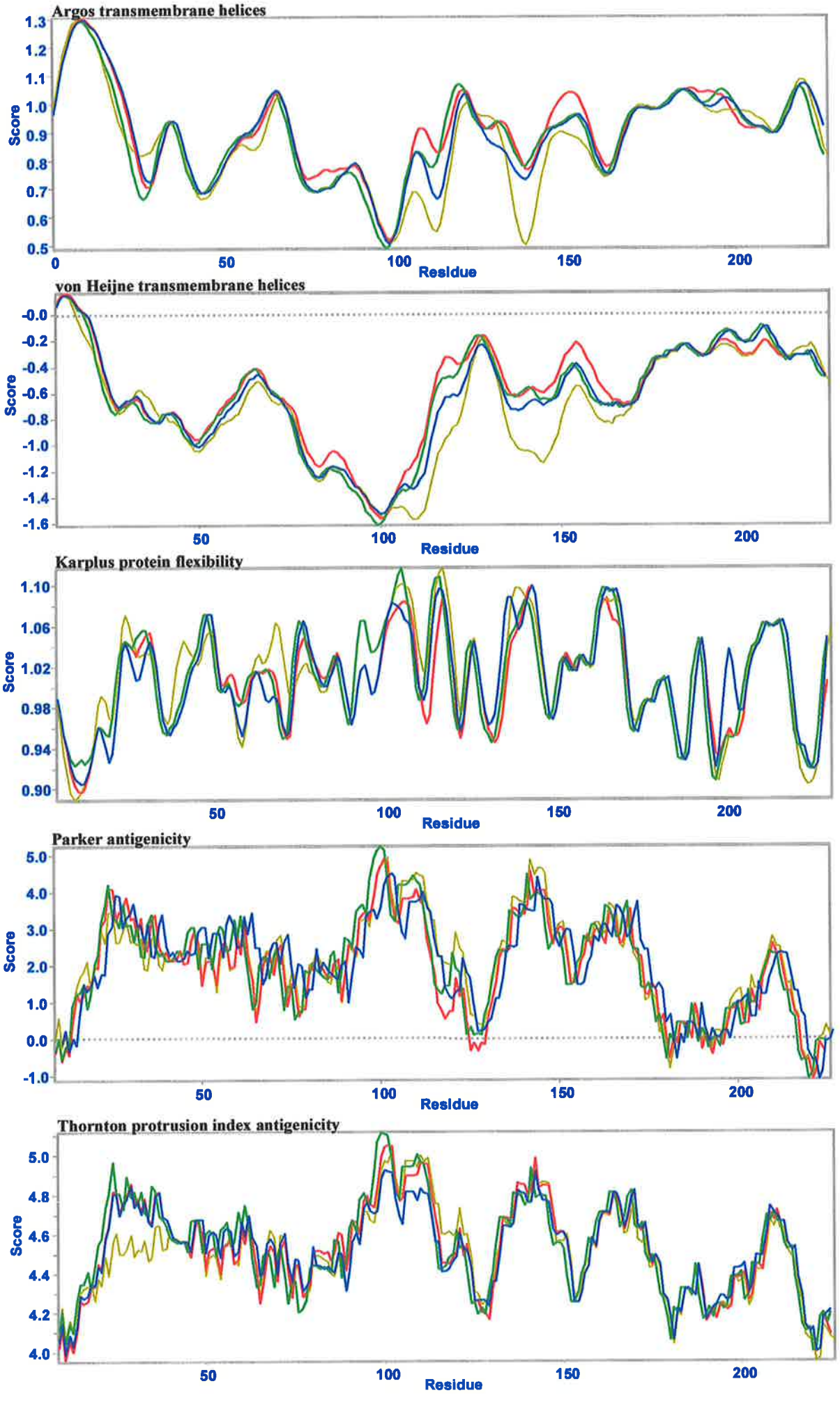

von Heijne hydrophilicity
Score
3.0
2.0
1.0
0.0
-1.0
50
100
150
200
Residue

Figure 3.4 Structural predictions of the Mip protein for four species of *Legionella*. *L. pneumophila* sg 1 (Philadelphia-1) and *L. longbeachae* sg 1 are associated with disease. *L. adelaidensis* and *L. moravica* have never been associated with disease. Produced by OMIGA 1.1. References are as follows: Argos (Argos, *et al.*, 1982) and von Heiijne (von Heijne, 1992) transmembrane helices; Karplus protein flexibility (Karplus and Shultz, 1985); Parker antigenicity (Parker, *et al.*, 1986); Thornton protrusion index antigenicity (Thornton, *et al.*, 1986).


Argos transmembrane helices
Score
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0
50
100
150
200
Residue
von Heijne transmembrane helices
Score
-0.0
-0.2
-0.4
-0.6
-0.8
-1.0
-1.2
-1.4
-1.6
50
100
150
200
Residue
Karplus protein flexibility
Score
1.10
1.06
1.02
0.98
0.94
0.90
50
100
150
200
Residue
Parker antigenicity
Score
5.0
4.0
3.0
2.0
1.0
0.0
-1.0
50
100
150
200
Residue
Thornton protrusion index antigenicity
Score
5.0
4.8
4.6
4.4
4.2
4.0
50
100
150
200
Residue

Chapter Four

Phylogeny of the genus *Legionella*

Introduction

Most nucleotide sequence-derived phylogenetic analyses of micro-organisms including *Legionella* (Hookey, *et al.*, 1996) are based on *rRNA* gene sequences (Woese, 1991). However, analyses which are based on a single gene may inadvertently become corrupted by the effects of selection and/or horizontal transmission on the gene (Woese, 1991). This is especially true for *rRNA* genes which may be predisposed to such effects because they are present in multiple copy number, and possess regions of high homology (Shen and Huang, 1986, Strätz, *et al.*, 1996). Recent comparisons of nucleotide sequences of protein-encoding genes with *rRNA* genes reinforce this concern (Amann, *et al.*, 1988, Kamla, *et al.*, 1996). This chapter describes a comparison of *16S rRNA* based phylogeny of *Legionella* with data from another gene, *mip*. Furthermore, the utility of generally used phenotypic and biochemical characteristics, such as auto-fluorescence (Wilkinson, 1988), cell wall ubiquinone and fatty acid constituents (Jantzen, *et al.*, 1993, Lambert and Moss, 1989, Wilkinson, *et al.*, 1990) are compared for their ability to discern evolutionary relationships within the genus.

Phylogenetic analysis

Nucleotide sequences of the putative reading frame of the *mip* gene from 38 species of *Legionella* and the outgroup *Coxiella burnetii* were aligned. Both the hyper-variable region

(nucleotide 55-132; codons 19 - 44) and the additional 8 nucleotides present at the carboxy terminal end of *L. israelensis* were excluded giving a total of 667 aligned sites. The *16S rRNA* sequences for the same taxa that were available for the *mip* data set were also aligned, giving a total of 1405 aligned sites. Transition/transversion (ts/tv) ratios, nucleotide frequencies, codon usage and evolutionary distances between sequences were calculated, before performing a phylogenetic analysis of the two data sets using both maximum parsimony (MP) and neighbour joining (NJ) methods. Bootstrap analysis was used to estimate the reliability of nodes. The phylogenetic analyses on the two nucleotide data sets were done independently to assess the degree of conflict between the resultant trees. The number of transitions and transversions were compared with increasing evolutionary distance to determine if the substitutions were accumulating in a linear manner, or were saturated. Additional details of the analysis method are given in the Materials and Methods (Chapter Two).

Results

The transition/transversion (ts/tv) ratio for each of the data sets estimated by maximum likelihood was approximately 1.0. For the *mip* sequences, excluding the outgroup and the hyper-variable region, 56% (378 of 675 sites) of sites are variable. Of these, 57% (217/378), 27% (100/378) and 16% (61/378) occur in the third, first and second codon positions respectively which equates to 96% (217/225), 44% (100/225) and 27% (61/225) of third, first and second codon positions exhibiting variability, respectively. Such variability, however, usually occurs with relatively few species at any specific site, especially for first and second codon positions.

The codon nucleotides coding for alignment positions containing non-synonymous substitutions were examined to determine the types and numbers of substitutions which might achieve the change sometime in the evolutionary history of the genus. Some frequent examples are presented in Fig. 4.1. Many require only one substitution to achieve the non-

synonymous change, e.g. a substitution in the first codon position alters the amino acid residue from serine (S) to alanine (A), or valine (V) to isoleucine (I). Other alignment positions contain amino acids which would require two nucleotide substitutions to have occurred for one amino acid to be replaced by the other eg arginine (R) to asparagine (N). However, the amino acid residue encoded by the codon which is only one substitution variant from both is also present eg Lysine (K). Theoretically, K evolutionally links R and N, although the order in which the substitutions might have occurred would be difficult to predict. Some of these alignment positions contain five different amino acid residues, but all the "intermediate" residues for which the codons are only one substitution apart are present. However, for the cluster which contains I, lysine (K), glutamine (Q), leucine (L) and A, A is two nucleotide substitutions variant from all other residues present in the same position, but the "intermediate" residue/s encoded for a codon only one substitution variant from one of the codons encoding the other four amino acids is no longer present. The "linking" codon/s must have been present at this position some time in the evolutionary history of the genus, and has either been lost from the *mip* gene sequences now existing, or may still be present in the wild strain population.

The frequency of substitutions at each codon position is graphically presented in Figs. 4.2 and 4.3. Figure 4.2 includes only the non-synonymous substitutions, and indicates the limited number of substitutions at the amino acid level. Figure 4.3 includes both synonymous and non synonymous substitutions, and demonstrates the much greater number of synonymous substitutions compared with non-synonymous ones.

To determine the degree of substitution saturation, the number of both ts and tv substitutions at each codon position was determined for each evolutionary distance present. The results, graphically displayed in Fig. 4.4 reveal that the increase in frequency of both ts and tv third codon substitutions slows for the larger genetic distances, indicating that the third

position is saturated for both ts and tv substitutions. In contrast, substitutions at the first and second positions appear to accumulate in a linear fashion, indicating that they are not saturated for change. As a consequence, alternative MP analyses were conducted with third position changes either included or excluded. The amino acid divergence, expressed as p-distance (defined as the number of variant residues/total number of residues compared), excluding codons 19-44 inclusive from the hypervariable region, ranged from 0.0090 to 0.2634 within the ingroup, and from 0.5024 to 0.5645 between the ingroup and outgroup.

Maximum parsimony analysis

For the *mip* data, including the outgroup, 51% (340/675) of all sites were parsimony informative, with 38% (130/340) of these being first and second position sites and 62% (210/340) being third position sites. The heuristic search option was used to find the most parsimonious tree(s) as there were too many taxa in the data set to use exhaustive or branch and bound searches. Three equally most parsimonious trees of length 2620 steps were obtained and are summarised by a strict consensus tree presented in Fig. 4.5. Bootstrap (100 pseudo-replications) support for these data are included. A second analysis, with the third codon position excluded, recovered 420 equally most parsimonious trees of length 762 steps and are summarised by a strict consensus tree presented in Fig. 4.6. For the *16S rRNA* data, 14.4% (202/1406) of all sites were parsimony informative. A heuristic search found 12 equally most parsimonious trees of length 1250 steps. These are summarised by a strict consensus tree presented in Fig. 4.7.

Distance analysis

For the *mip* data set, Jukes-Cantor distances (Jukes and Cantor, 1969) within the ingroup ranged from 0.0470 to 0.3922 and between the ingroup and outgroup from 0.6919 to 0.7871. Average nucleotide frequencies (%) are A = 33.4, T = 26.7, G = 22.7, and C = 17.2. For the *16S rRNA* data set, Jukes-Cantor distances within the ingroup ranged from 0.0023 to

0.0934, and between the ingroup and outgroup from 0.1367 to 0.1700. Average nucleotide frequencies (%) are A = 25.8, T = 20.5, G = 32.2 and C = 21.5. Following the recommendations of Kumar, *et al.* (1993) for these statistics, the Tajima-Nei distance (Tajima and Nei, 1984) and the Jukes-Cantor distance was used to estimate evolutionary distances for the *mip* and *16S rRNA* data sets respectively. The NJ trees constructed from these distances are presented in Figs. 4.8 (*mip*) and 4.9 (*rRNA*) to compare with the MP analyses.

Congruence among distance and MP trees.

Both distance and MP of the *16S rRNA* data recover a clade, hereafter called (*anisa - worsleiensis*), that includes *L. anisa*, *L. bozemanii*, *L. cherrii*, *L. cincinnatiensis*, *L. dumoffii*, *L. gormanii*, *L. gratiana*, *L. jordanis*, *L. longbeachae*, *L. moravica*, *L. parisiensis*, *L. pneumophila*, *L. quateirensis*, *L. sainthelensi*, *L. santicrucis*, *L. shakespearei*, *L. steigerwaltii*, *L. tucsonensis*, *L. wadsworthii*, and *L. worsleiensis*. Within this clade, both analyses recover a ((*cincinnatiensis*, *longbeachae*, *sainthelensi*, *santicrucis*) *gratiana*) clade with identical topology. Similarly, both analyses recover a (*moravica*, *quateirensis*, *shakespearei*, *worsleiensis*) clade but topology within it differs totally between the analyses. Bootstrap values greater than 90 in the MP analyses support all of the above groups. Additionally the following clades, with bootstrap support from the MP analysis are recovered by both analyses; (*erythra*, *rubrilucens*) 99, (*feelei* (*maceachernii*, *micdadei*) 100) <50, (*fairfieldensis*, *nautarum*) <50, (*birminghamensis*, *quinlivanii*) 72, (*oakridgensis* (*israelensis*, *londiniensis*) 90) <50 and (*hackeliae*, *jamestowniensis*) 71.

Both distance and MP (separate analyses including and excluding third codon position) of the *mip* data also recover the (*anisa - worsleiensis*) clade. Within this clade, all three analyses recover the following clades, with bootstrap support from the MP analysis that includes all three codon positions; ((*anisa*, *bozemanii*, *parisiensis*, *tucsonensis*) 83, *cherrii*, *dumoffii*, *gormanii*, *steigerwaltii*, *wadsworthii*) 89; ((*cincinnatiensis*, *longbeachae*,

sainthelensi, *santicrucis*) 96, *gratiana*) 97; (((*moravica*, *quateirensis*) 93, *worsleiensis*) 60, *shakespearei*) <50. Additionally, the following clades, with bootstrap support from the MP analysis, are recovered by both analyses; ((*erythra*, *rubrilucens*) 100, *spiritensis*) 51; (((*maceachernii*, *micdadei*) 100, *nautarum*) 99, *fairfieldensis*) <50; (*birminghamensis*, *quinlivanii*) 100; and (*jordanis*, *lansingensis*) <50, (*brunensis*, *hackeliae*, *jamestowniensis*) <50) <50. Sets of relationships recovered by all analyses from both *mip* and *rRNA* data sets are summarised in Fig. 4.10. The distribution of some distinguishing physiological and biochemical characteristics are also listed.

Non-congruence among MP and distance trees.

There are many instances of non-congruent placement of taxa between the trees generated from the two data sets, but the alternative placements do not result in strong bootstrap support. Sampling error, i.e. too few informative nucleotides sampled in one or both data sets, is the most likely explanation. Alternatively, the results may reflect fundamentally different evolutionary processes for each gene.

Discussion

Although few well supported sets of phylogenetic relationships were apparent from the analyses of the nucleotide data sets (Fig. 4.10), there are sufficient sets of resolved relationships to allow comparison with relationships recovered from phenotypic and biochemical characters. The unreliability of using single phenotypic or biochemical characters to infer relationships (Selander, *et al.*, 1991) can be illustrated with reference to the ubiquinone groupings. Ubiquinone groups A, C and D are present both within and external to the (*anisa-worsleiensis*) clade, and consequently cannot be indicators of evolutionary relationship. This may be due to the way in which the phenotype is determined (arbitrary classification) or it may be that the character is genuinely homoplastic, i.e. subject to convergence. The arbitrary nature of some classifications for *Legionella* can be understood

when the criteria on which they are based are examined. *L. israelensis* has been grouped alternately with ubiquinone group A species (Wilkinson, *et al.*, 1990) or with group D species (Lambert and Moss, 1989). The discrepancy is determined solely on a small difference in the amount of ubiquinone 13 (Q13) detected. Additionally, although the two ubiquinone group C species *L. oakridgensis* and *L. wadsworthii* both contain major amounts of Q10, they have distinctly different amounts of Q9, Q11 and Q12. In fact, the ubiquinone pattern for *L. wadsworthii* is much closer to that found for group B species, to which it is related.

However, using all the phenotypic and biochemical characters in combination may provide indications of evolutionary relationships. For example, a grouping of *Legionella* with the character-states blue-white autofluorescence, DNA group 2, ubiquinone group B, hydroxy-fatty acid group 3 and fatty acid group II is well supported by the *mip* data, but the *16S rRNA* data neither support nor reject this relationship. Furthermore the above species share DNA, ubiquinone and hydroxy-fatty acid character states with the ((*cincinnatiensis*, *longbeachae*, *sainthelensi*, *santicrucis*) *gratiana*) clade. Again, this relationship is well supported by the *mip* data but the 16S rRNA data neither support nor reject this relationship. The known association with disease for each species is independent of any phylogenetic relationship determined in the present study.

Ultimately, few of the 36 nodes were strongly resolved by both data sets, despite the availability of over two kilobases of nucleotide sequence. The paucity of topological resolution is probably due to the low proportion of informative sites in the *16S rRNA* gene for a data set with 38 terminal taxa and 36 internal nodes. However, reliance on the apparent better resolution of the *mip* gene sequences to infer relationship risks mistaking the evolutionary history of the genus for what may in reality be the evolutionary history of a single gene (Cilia, *et al.*, 1996). Additionally, the poor support for many of the ancestral nodes

based on the *mip* data set presumably reflects the dominant influence of the third position of the codon, where many of the ancestral substitutions have themselves been substituted.

The *16S rRNA* gene is a prime target of bacterial phylogenetic studies, but the very characteristics that have caused its selection also increase the likelihood of recombination (Shen and Huang, 1986, Woese, 1991), and limit its usefulness at recovering evolutionary relationships at lower taxonomic levels, e.g. *Bacillus* (Ash, *et al.*, 1991, Rossler, *et al.*, 1991) and *Nitrosospira* (Utaker, *et al.*, 1995). There is clearly a need to sample additional protein encoding genes that have higher rates of replacement substitutions but the choice of gene must be influenced by the likelihood with which individual loci are subject to recombination (Nelson and Selander, 1994). The *16S rRNA* locus still has an effective role in establishing higher level evolutionary relationships, and in establishing individual genetic identifications.

Figure 4.1. Examples of nucleotide substitutions which would produce non-synonymous amino acid residue substitutions commonly found in the same position in the inferred amino acid alignment.

Transitional substitution

Transversional substitution

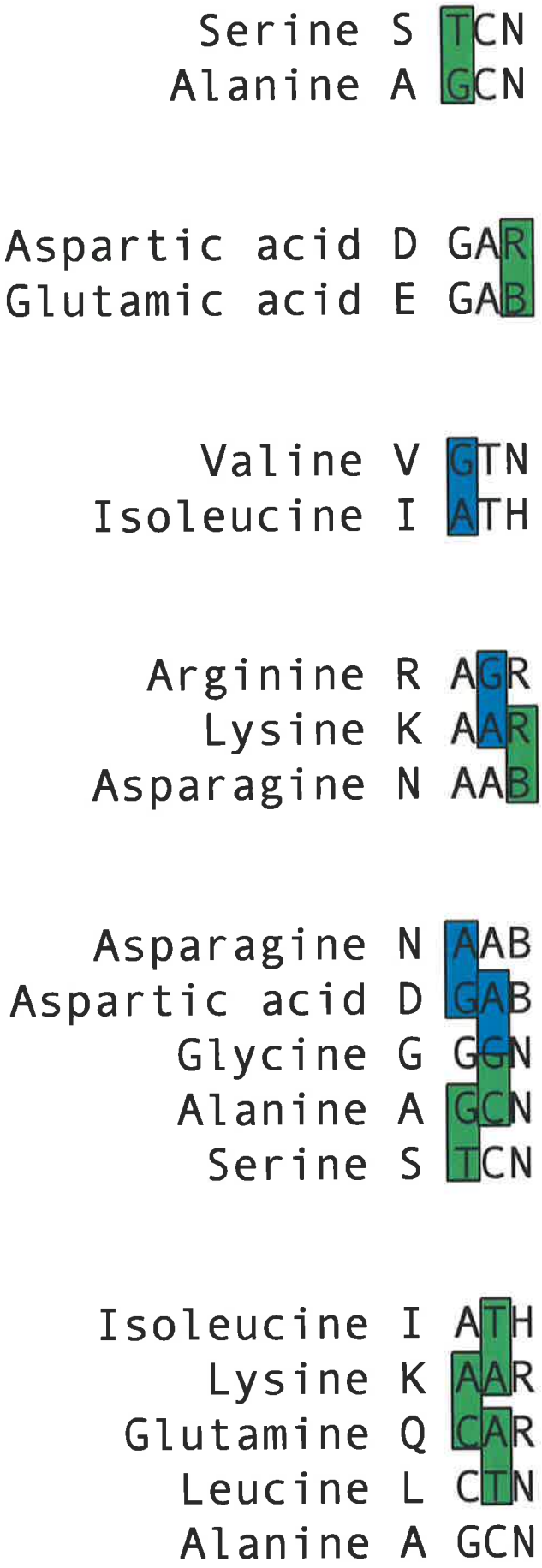
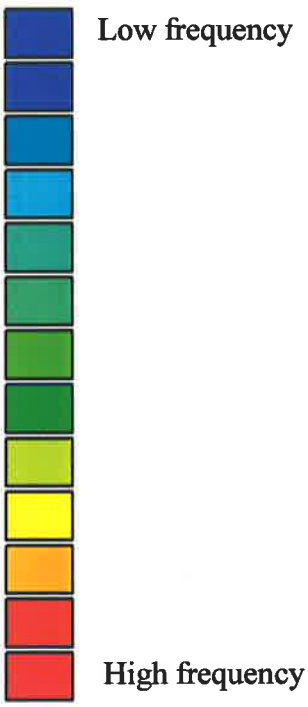

Serine S TCN
Alanine A GCN
Aspartic acid D GAR
Glutamic acid E GAB
Valine V GTN
Isoleucine I ATH
Arginine R AGR
Lysine K AAR
Asparagine N AAB
Asparagine N AAB
Aspartic acid D GAB
Glycine G GGN
Alanine A GCN
Serine S TCN
Isoleucine I ATH
Lysine K AAR
Glutamine Q CAR
Leucine L CTN
Alanine A GCN

Figure 4.2. Rate of non-synonymous substitutions at each amino acid position. Species named in blue have a documented association with human disease; the remainder have been isolated only from environmental sources. Produced by OMIGA 1.1. Each residue is represented by a block coloured to indicate the frequency of non-synonymous substitutions at that position, as follows.



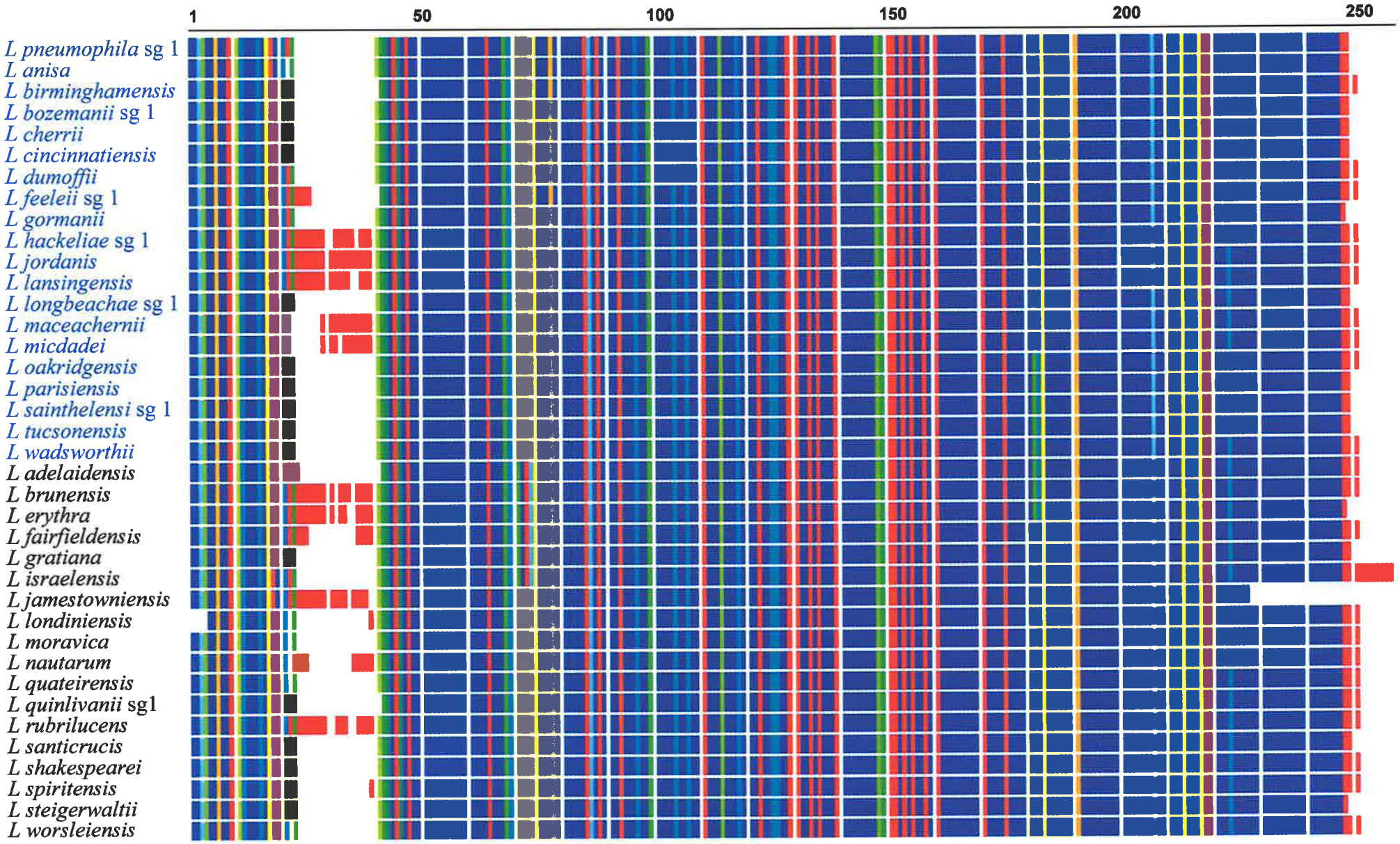
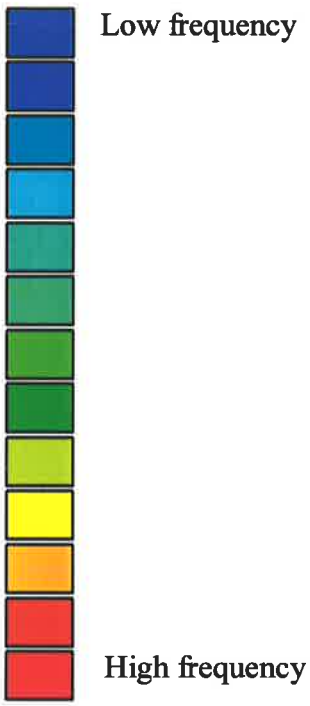

1
50
100
150
200
250
L pneumophila sg 1
L anisa
L birminghamensis
L bozemanii sg 1
L cherrii
L cincinnatiensis
L dumoffii
L feeleii sg 1
L gormanii
L hackeliae sg 1
L jordanis
L lansingensis
L longbeachae sg 1
L maceachernii
L micdadei
L oakridgensis
L parisiensis
L sainthelensi sg 1
L tucsonensis
L wadsworthii
L adelaidensis
L brunensis
L erythra
L fairfieldensis
L gratiana
L israelensis
L jamestowniensis
L londiniensis
L moravica
L nautarum
L quateirensis
L quinlivanii sg1
L rubrilucens
L santicrucis
L shakespearei
L spiritensis
L steigerwaltii
L worsleiensis

Figure 4.3. Rate of both synonymous and non-synonymous substitutions at each amino acid position. Species named in blue have a documented association with human disease; the remainder have been isolated only from environmental sources. Produced by OMIGA 1.1. Each residue is represented by a block coloured to indicate the frequency of all substitutions at that position, as follows:



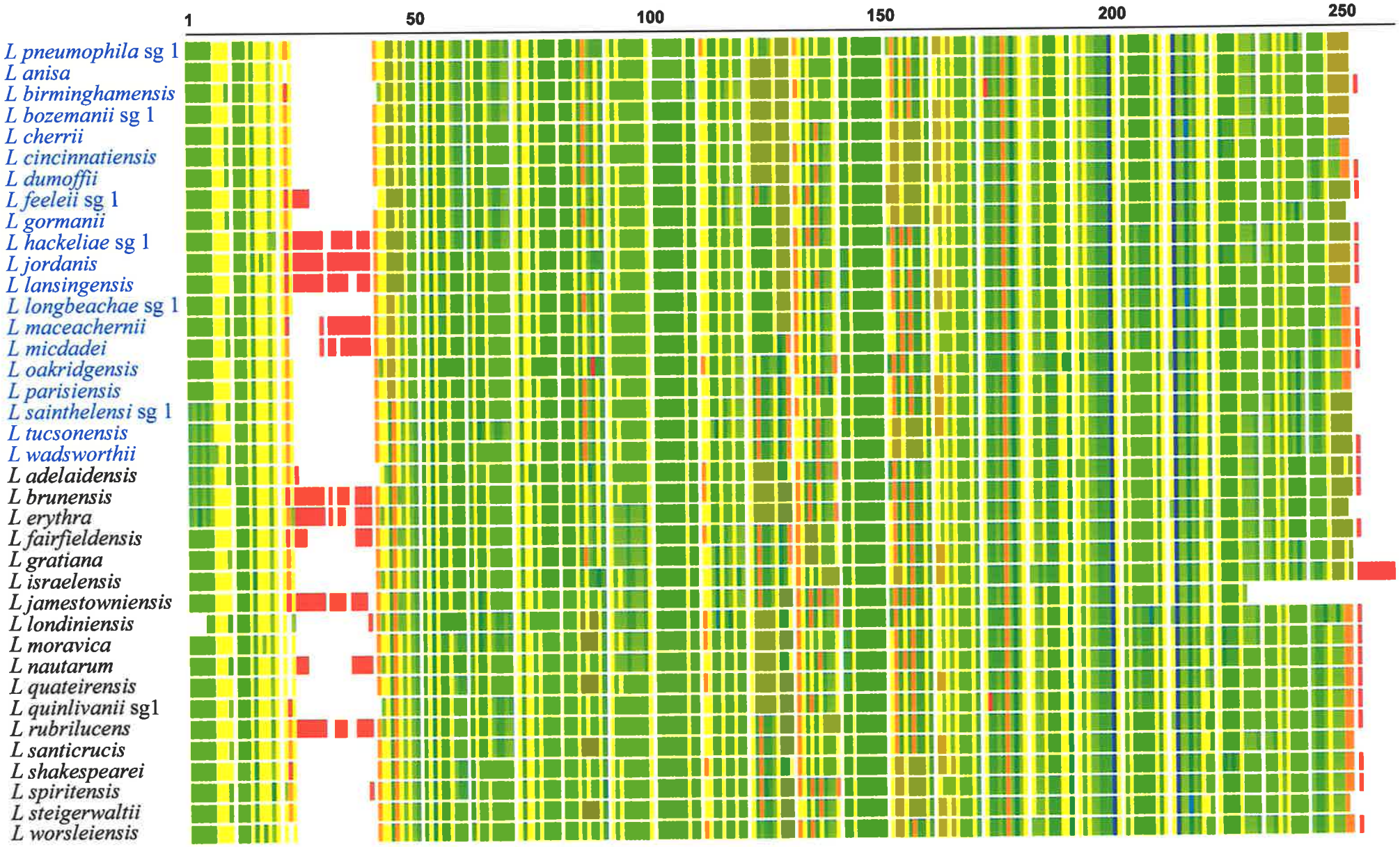

1
50
100
150
200
250
L pneumophila sg 1
L anisa
L birminghamensis
L bozemanii sg 1
L cherrii
L cincinnatiensis
L dumoffii
L feeleii sg 1
L gormanii
L hackeliae sg 1
L jordanis
L lansingensis
L longbeachae sg 1
L maceachernii
L micdadei
L oakridgensis
L parisiensis
L sainthelensi sg 1
L tucsonensis
L wadsworthii
L adelaidensis
L brunensis
L erythra
L fairfieldensis
L gratiana
L israelensis
L jamestowniensis
L londiniensis
L moravica
L nautarum
L quateirensis
L quinlivanii sg1
L rubrilucens
L santicrucis
L shakespearei
L spiritensis
L steigerwaltii
L worsleiensis

Figure 4.4. The number of transitional and transversional substitutions at each codon position with increasing evolutionary distance. Each data set is represented by a second order polynomial regression line of best fit, as there are too many data points to present the raw data.

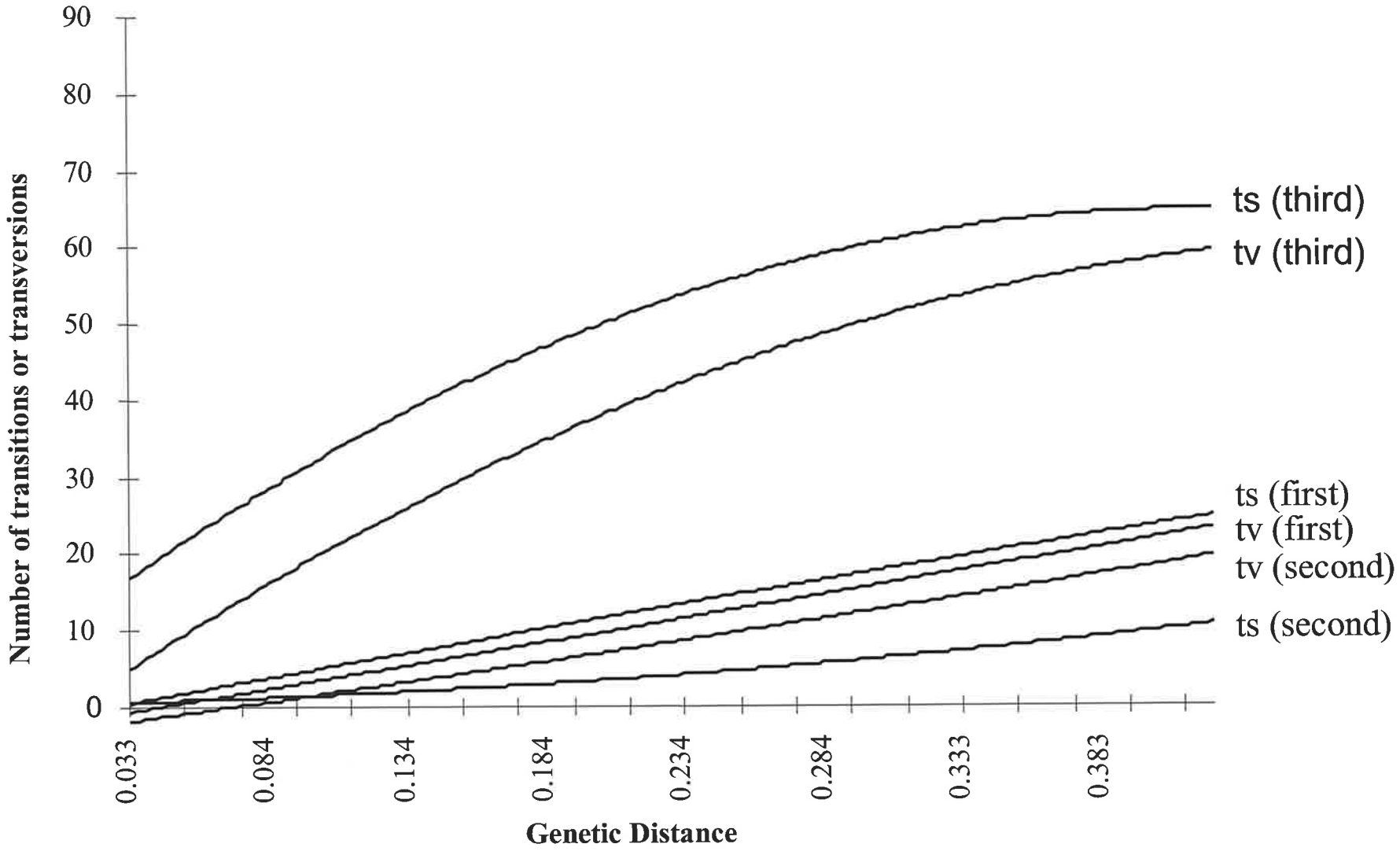

90
80
70
60
50
40
30
20
10
0
Number of transitions or transversions
0.033
0.084
0.134
0.184
0.234
0.284
0.333
0.383
Genetic Distance
ts (third)
tv (third)
ts (first)
tv (first)
tv (second)
ts (second)

Figure 4.5. Strict consensus tree summarising the three equally most parsimonious trees of length 2620 steps found using all codon positions in the aligned *mip* data. Bootstrap values greater than 50% are indicated.

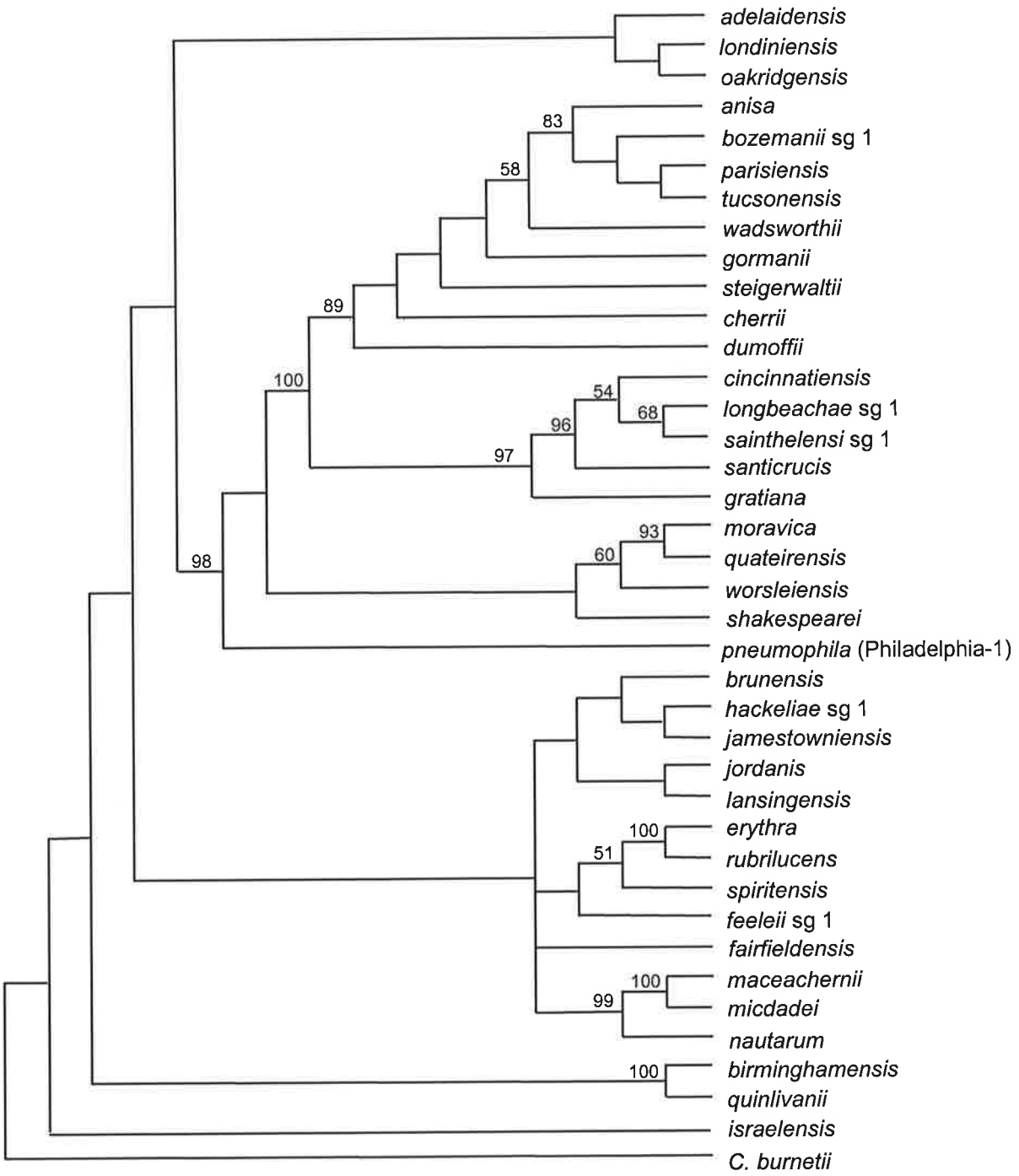

adelaidensis
londiniensis
oakridgensis
anisa
bozemanii sg 1
parisiensis
tucsonensis
wadsworthii
gormanii
steigerwaltii
cherrii
dumoffii
cincinnatiensis
longbeachae sg 1
sainthelensi sg 1
santicrucis
gratiana
moravica
quateirensis
worsleiensis
shakespearei
pneumophila (Philadelphia-1)
brunensis
hackeliae sg 1
jamestowniensis
jordanis
lansingensis
erythra
rubrilucens
spiritensis
feeleii sg 1
fairfieldensis
maceachernii
micdadei
nautarum
birminghamensis
quinlivanii
israelensis
C. burnetii
83
58
89
100
54
68
96
97
93
60
98
100
51
100
99
100

Figure 4.6. Strict consensus tree summarising the 420 equally most parsimonious trees of length 762 steps found using only the first and second codon positions for the *mip* data.

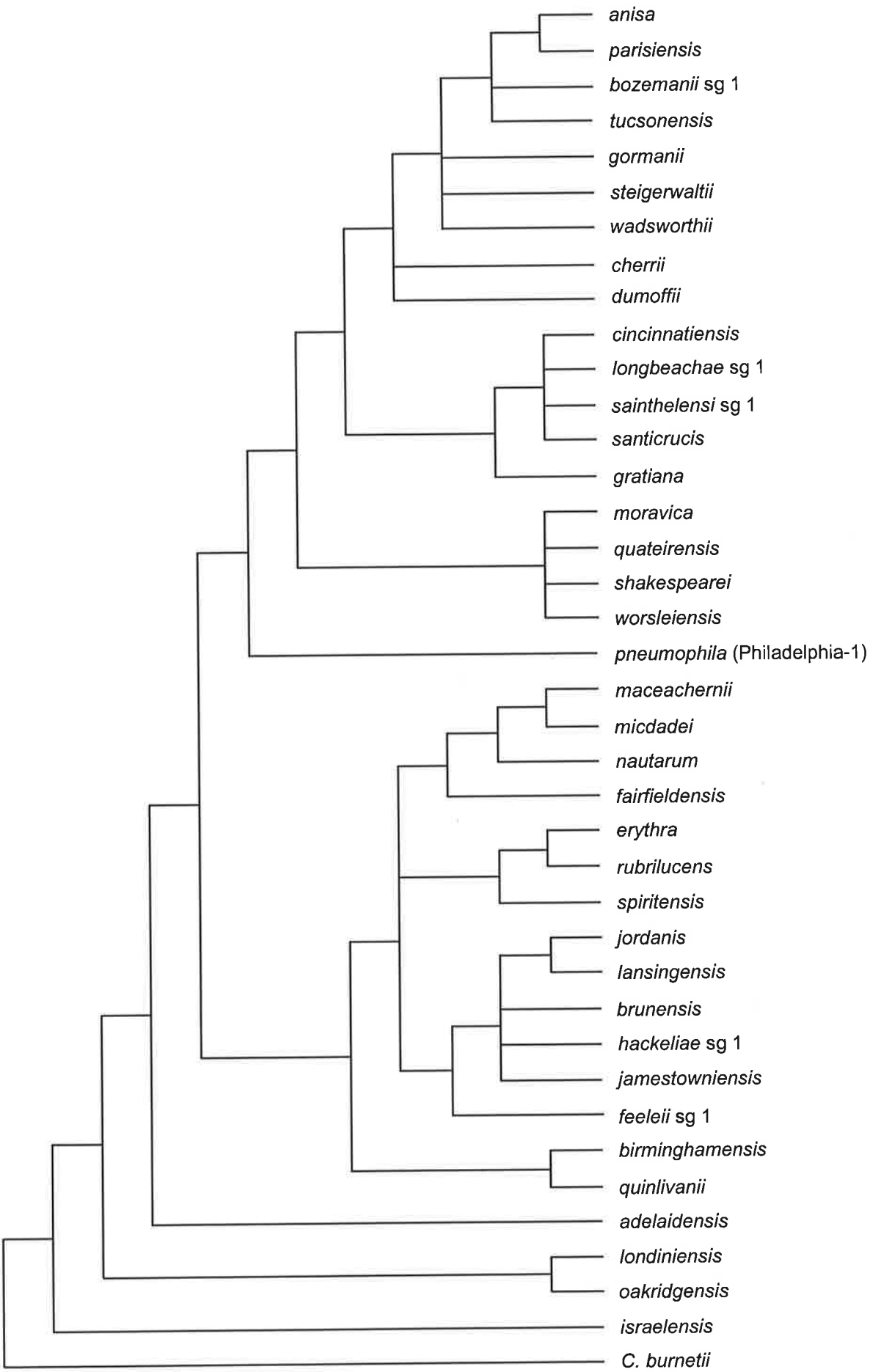

anisa
parisiensis
bozemanii sg 1
tucsonensis
gormanii
steigerwaltii
wadsworthii
cherrii
dumoffii
cincinnatiensis
longbeachae sg 1
sainthelensi sg 1
santicrucis
gratiana
moravica
quateirensis
shakespearei
worsleiensis
pneumophila (Philadelphia-1)
maceachernii
micdadei
nautarum
fairfieldensis
erythra
rubrilucens
spiritensis
jordanis
lansingensis
brunensis
hackeliae sg 1
jamestowniensis
feeleii sg 1
birminghamensis
quinlivanii
adelaidensis
londiniensis
oakridgensis
israelensis
C. burnetii

Figure 4.7. Strict consensus tree summarising the 12 equally most parsimonious trees of length 1250 steps found using all sites in the aligned *16S rRNA* sequence data. Bootstrap values greater than 50% are indicated.

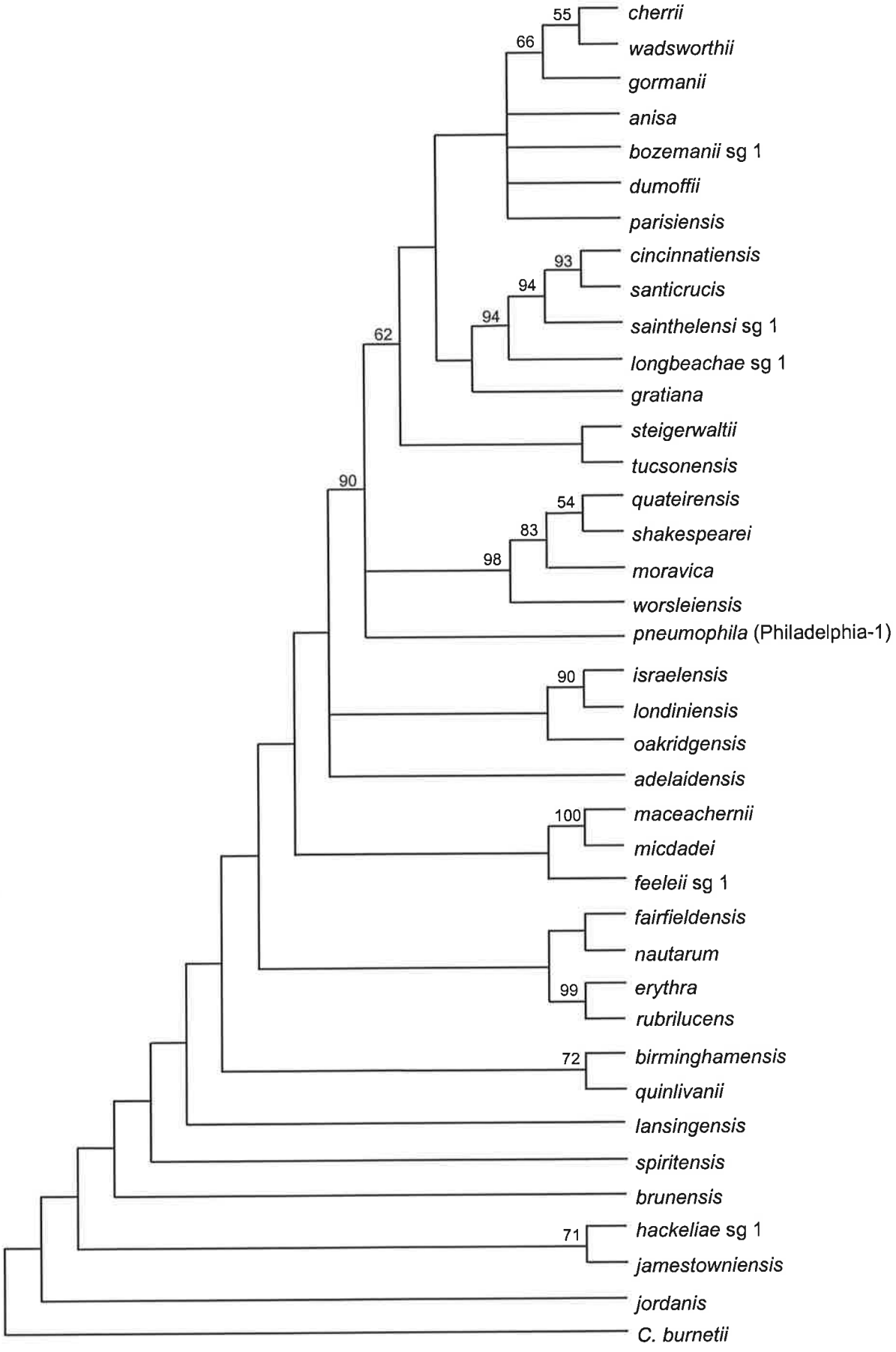

55
66
cherrii
wadsworthii
gormanii
anisa
bozemanii sg 1
dumoffii
parisiensis
93
94
94
62
cincinnatiensis
santicrucis
sainthelensi sg 1
longbeachae sg 1
gratiana
steigerwaltii
tucsonensis
90
54
83
98
quateirensis
shakespearei
moravica
worsleiensis
pneumophila (Philadelphia-1)
90
israelensis
londiniensis
oakridgensis
adelaidensis
100
maceachernii
micdadei
feeleii sg 1
fairfieldensis
nautarum
99
erythra
rubrilucens
72
birminghamensis
quinlivanii
lansingensis
spiritensis
brunensis
71
hackeliae sg 1
jamestowniensis
jordanis
C. burnetii

Figure 4.8. Neighbour-joining tree constructed from Tajima-Nei distances (Tajima and Nei, 1984) among the *mip* sequences and distances of 38 species of *Legionella*. Bootstrap values greater than 50% are indicated.

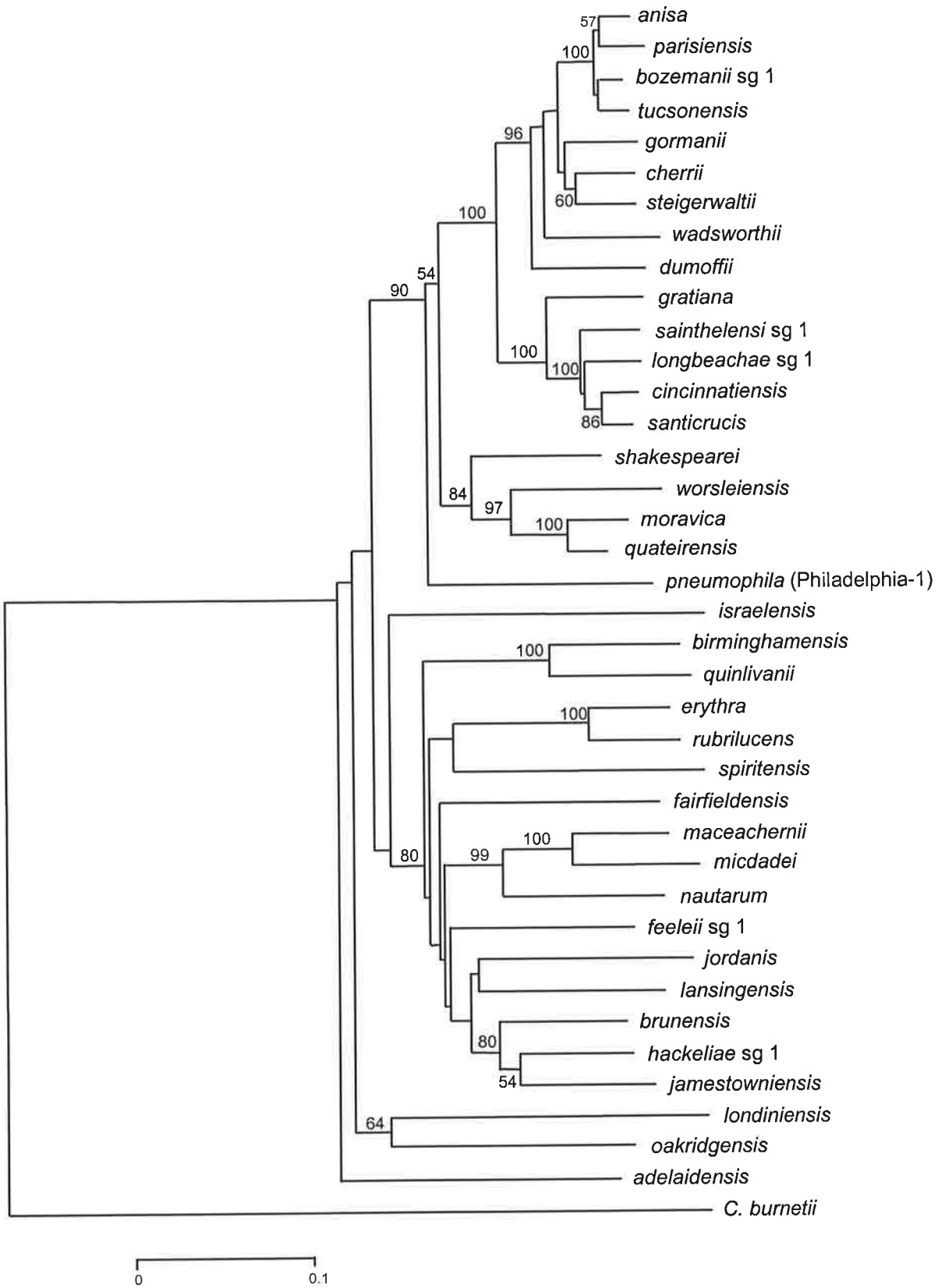

anisa
parisiensis
bozemanii sg 1
tucsonensis
gormanii
cherrii
steigerwaltii
wadsworthii
dumoffii
gratiana
sainthelensi sg 1
longbeachae sg 1
cincinnatiensis
santicrucis
shakespearei
worsleiensis
moravica
quateirensis
pneumophila (Philadelphia-1)
israelensis
birminghamensis
quinlivanii
erythra
rubrilucens
spiritensis
fairfieldensis
maceachernii
micdadei
nautarum
feeleii sg 1
jordanis
lansingensis
brunensis
hackeliae sg 1
jamestowniensis
londiniensis
oakridgensis
adelaidensis
C. burnetii
57
100
96
60
100
54
90
100
100
86
84
97
100
100
80
99
100
80
54
64
0
0.1

Figure 4.9. Neighbour-joining tree constructed from Jukes-Cantor distances (Jukes and Cantor, 1969) among the *16S rRNA* sequences of 38 species of *Legionella*. Bootstrap values greater than 50% are indicated.

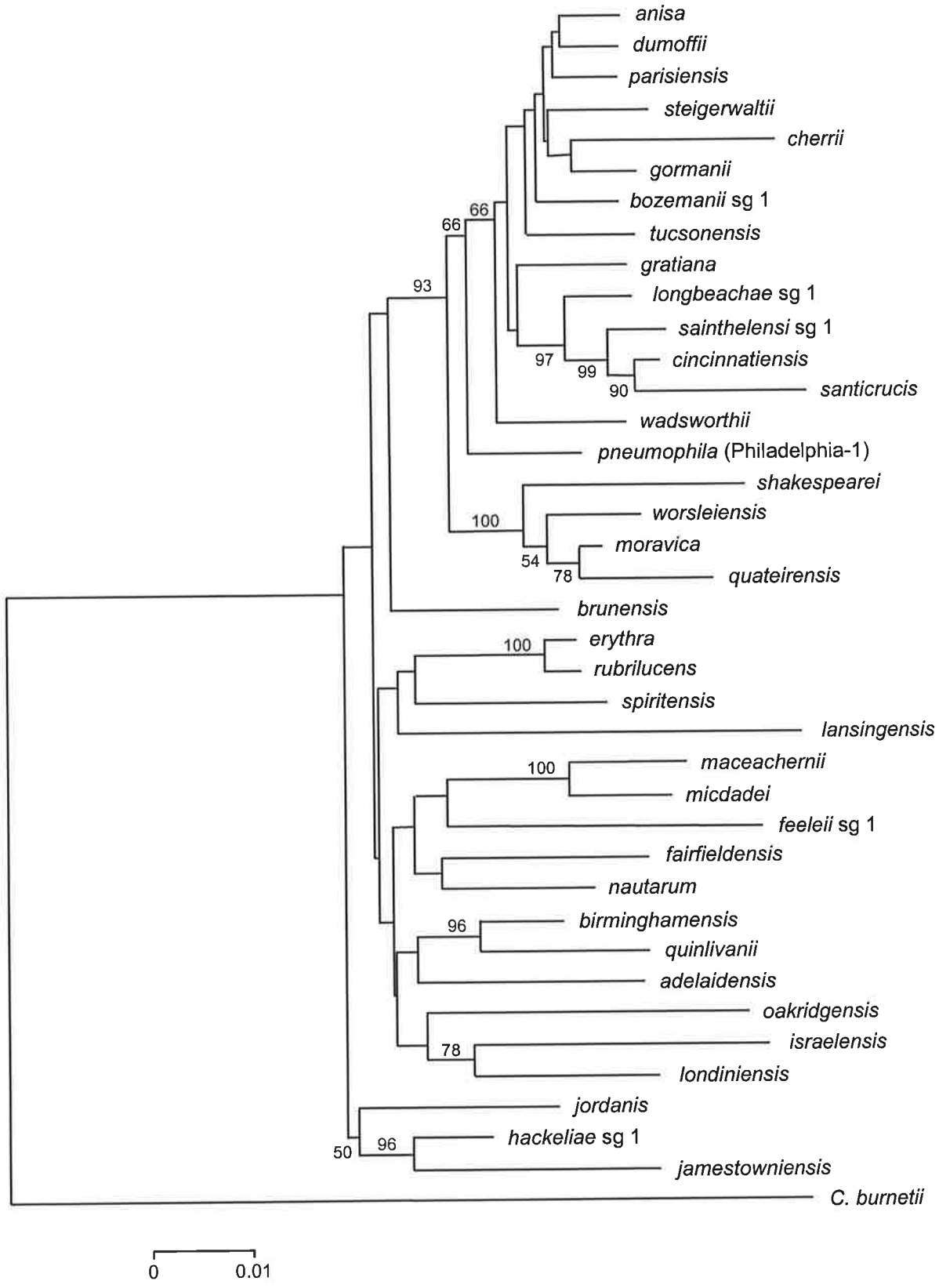
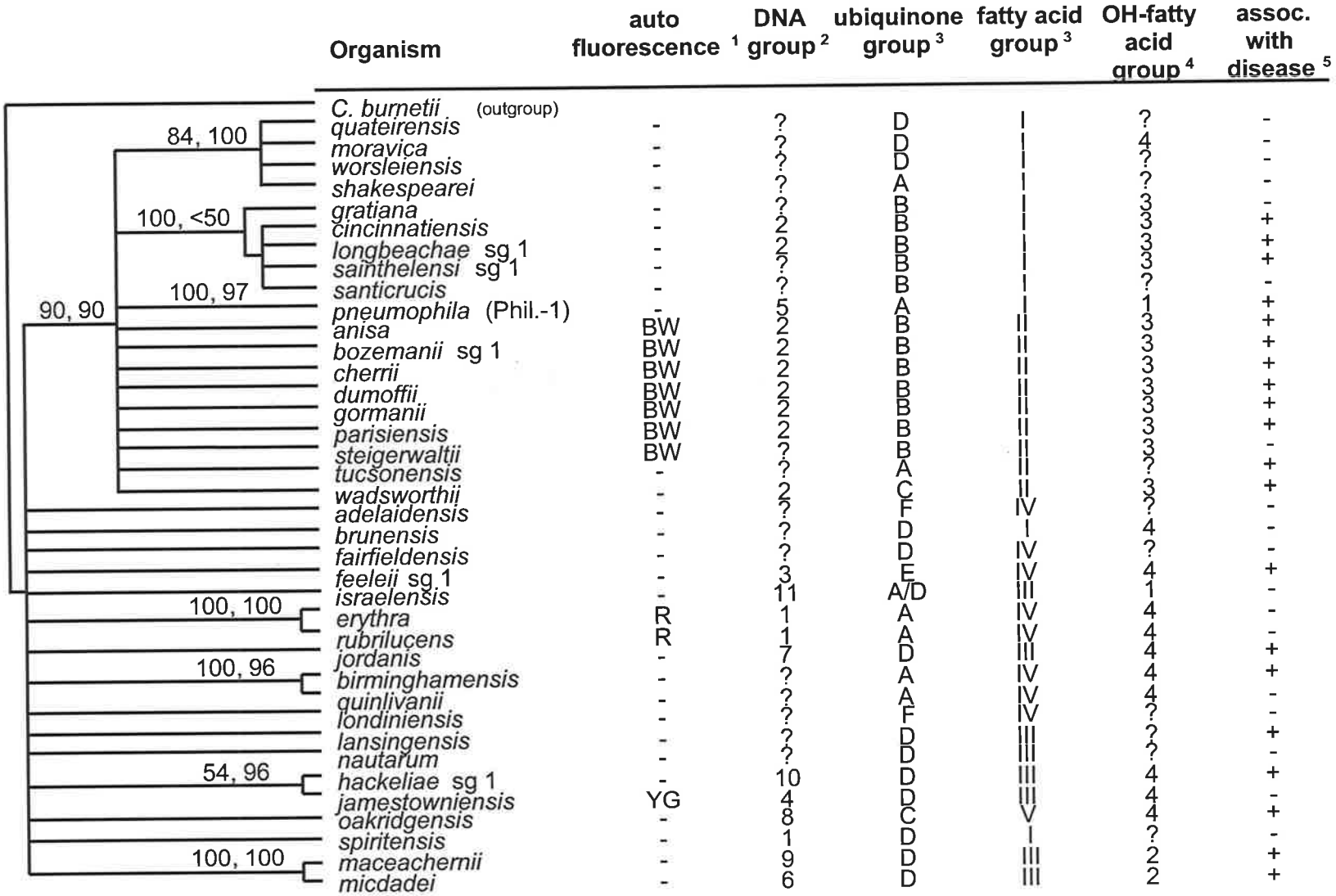

anisa
dumoffii
parisiensis
steigerwaltii
cherrii
gormanii
bozemanii sg 1
tucsonensis
gratiana
longbeachae sg 1
sainthelensi sg 1
cincinnatiensis
santicrucis
wadsworthii
pneumophila (Philadelphia-1)
shakespearei
worsleiensis
moravica
quateirensis
brunensis
erythra
rubrilucens
spiritensis
lansingensis
maceachernii
micdadei
feeleii sg 1
fairfieldensis
nautarum
birminghamensis
quinlivanii
adelaidensis
oakridgensis
israelensis
londiniensis
jordanis
hackeliae sg 1
jamestowniensis
C. burnetii
66
66
93
97
99
90
100
54
78
100
100
96
78
50
96
0
0.01

Fig. 4.10. A tree summarising relationships receiving support from both the *mip* and *16S rRNA* data sets, compared with phenotypic and biochemical characteristics. Values above branches represent bootstrap values for the *mip* (left-hand value) and *16S rRNA* (right-hand value) data. For explanation of phenotypic and biochemical markers, and group definitions, see the following; [1] (Wilkinson, 1988); [2] (Brenner, 1987); [3] (Lambert and Moss, 1989, Wilkinson, *et al.*, 1990); [4] (Jantzen, *et al.*, 1993); [5] (O'Connell, *et al.*, 1996).



Organism	auto fluorescence [1]	DNA group [2]	ubiquinone group [3]	fatty acid group [3]	OH-fatty acid group [4]	assoc. with disease [5]
C. burnetii (outgroup)						
quateirensis	-	?	D	I	?	-
moravica	-	?	D	I	4	-
worsleiensis	-	?	D	I	?	-
shakespearei	-	?	A	I	?	-
gratiana	-	?	B	I	3	-
cincinnatiensis	-	2	B	I	3	+
longbeachae sg 1	-	2	B	I	3	+
sainthelensi sg 1	-	?	B	I	3	+
santicrucis	-	?	B	I	?	-
pneumophila (Phil.-1)	-	5	A	I	1	+
anisa	BW	2	B	II	3	+
bozemanii sg 1	BW	2	B	II	3	+
cherrii	BW	2	B	II	3	+
dumoffii	BW	2	B	II	3	+
gormanii	BW	2	B	II	3	+
parisiensis	BW	2	B	II	3	+
steigerwaltii	BW	?	B	II	3	-
tucsonensis	-	?	A	II	?	+
wadsworthii	-	2	C	II	3	+
adelaidensis	-	?	F	IV	?	-
brunensis	-	?	D	I	4	-
fairfieldensis	-	?	D	IV	?	-
feeleii sg 1	-	3	E	IV	4	+
israelensis	-	11	A/D	III	1	-
erythra	R	1	A	IV	4	-
rubrilucens	R	1	A	IV	4	-
jordanis	-	7	D	III	4	+
birminghamensis	-	?	A	IV	4	+
quinlivanii	-	?	A	IV	4	-
londiniensis	-	?	F	IV	?	-
lansingensis	-	?	D	III	?	+
nautarum	-	?	D	III	?	-
hackeliae sg 1	-	10	D	III	4	+
jamestowniensis	YG	4	D	III	4	-
oakridgensis	-	8	C	V	4	+
spiritensis	-	1	D	I	?	-
maceachernii	-	9	D	III	2	+
micdadei	-	6	D	III	2	+

Chapter Five

Further phylogenetic analysis of *Legionella* species within the (*anisa-worsleiensis*) clade

Introduction

Chapter Four examined the phylogenetic relationships between species within the *Legionella* genus. Comparing the sequence data from the *mip* gene with that published for the *16S rRNA* gene, over two kilobases of sequence data were analysed with both distance and maximum parsimony (MP) phylogenetic methods. However, few of the 36 nodes were strongly resolved. The *16S rRNA* gene sequence data set contained relatively few informative sites and the *mip* gene sequence data set, while demonstrating better resolution, was potentially influenced by the substitution-saturated third-codon position.

Of the clades which were at least partially resolved, the largest, designated the *anisa – worsleiensis* clade, contained 19 taxa arranged in three sub-clades, although the internal arrangement of each varied within the two data sets. Resolution of this smaller, more recently evolved clade might be possible if additional phylogenetically informative data could be determined. Sequencing additional genetically stable "house-keeping" gene/s would overcome the risk of assuming that the evolutionary history of a single gene (*mip*) is identical to that for the organism (Cilia, *et al.*, 1996).

An analysis of the sequence and publication data-bases suggested two genes which might be phylogenetically suitable, and where the sequences had been determined for more than one

Legionella species, an important consideration for determining consensus primers for gene amplification. The RNA polymerase B-subunit (*rpoB*) gene has been determined for *L. pneumophila* and *L. longbeachae*, and consensus primers which amplified a portion of the gene from additional species have been designed (Bangsborg, J. T., 1999, personal communication). Additionally, the Zinc metalo-protease (*mspA*) gene (also termed the *proA* gene, and hereafter termed the *mspA* gene) has been determined for the same two species (Black, *et al.*, 1990, Gene Bank X8305). Although not strictly a house-keeping gene, *mspA* should also not be subject to the high rates of mutation or horizontal gene transfer found in some genes, e.g. those involved with outer membrane antigenicity. An analysis of the two *mspA* sequences determined a similarity of 87.5% and 73% at the protein and nucleotide levels respectively, with homologous regions of sufficient length to be suitable targets for consensus primers.

This chapter reports the use of additional sequence from the *mspA* gene to further resolve the topology of the *anisa – worsleiensis* clade.

Sequencing strategy

Consensus primers were designed to several regions of high similarity in an alignment of the *mspA* gene from *L. pneumophila* and *L. longbeachae*, as follows, the location with respect to *L. pneumophila mspA* ORF being given in parenthesis:

Msp-F1: 5' – GG(AG) TTT CCA GTT T(AT)T GGT GG – 3'	(247 – 266)
Msp-F2: 5' – CCA ACA GCA ATT ATT GAT GC – 3'	(559 – 578)
Msp-F3: 5' - TC(GT) CCA ATG CAA TTA GT(CT) ATG C – 3'	(997 – 1018)
Msp-R1: 5' – GGT GTC CAA TAA TCC ATA TT(CG) GC – 3'	(1502 – 1480)
Msp-R2: 5' – TT(AC) ACC ATA ACA TCA AAA GCC – 3'	(1478 – 1458)

The consensus primers for the *rpoB* gene were as follows:

RIFU1: 5′ - CG(CT) CG(CT) GTT CG(CT) TC(ATGC) GT(AT) GG(ACT) GA - 3′

RIFD1 5′ - AA(AGT) CCA ATA TT(AT) GG(AGT) CCT TC(ATGC) GG - 3′

Results

mspA *sequence analysis*

The two *rpoB* gene primers and the six possible combinations of the *mspA* gene primers were used to prime PCR reactions containing total cellular DNA extracted from the type strains of the majority of *Legionella* species, as well as a number of wild strains. The *rpoB* primers did not amplify all strains in the *anisa – worsleiensis* clade, and many that did amplify produced secondary amplification products. Amplification at higher annealing temperatures did not remove the secondary products without significantly reducing the target product. Of the *mspA* primers, Msp-R1 was unable to prime amplification reactions for all but a few species, irrespective of the forward primer used. However, Msp-F3 in combination with Msp-R2 proved to be the most useful primer set, priming amplification in all but a few rare, currently uncharacterised species, to produce a single product of approximately 460 bp. Msp-F1 and Msp-F2 amplified larger products, which could give more sequence information, but some species failed to amplify, perhaps because they target sites in the 621 bp signal peptide region, which showed lower similarity than the enzyme-coding region. The amplicons produced from the 19 species in the *anisa – worsleiensis* clade, and *L. micdadei* were sequenced by dye-terminator chemistry primed with Msp-F3 and aligned to produce an alignment of 438 sites. To reduce costs, amplicons were only sequenced in the forward direction unless sequence ambiguity or unusual sequence required confirmation with sequence in the reverse direction. This decision was based on the lack of observed conflicting bases between forward and reverse sequences for the *mip* sequences, where both directions had been sequenced. *L. micdadei*, which was included as an outgroup, contained and additional 3 bp forming an in-frame codon not present in the sequence from the species in the *anisa* –

worsleiensis clade, located at sites 46-48. Excluding the outgroup, 42% (183/435) of sites were variable, 71% (130/183), 15% (27/183) and 14% (26/183) occurring at the third, first and second codon positions respectively.

To determine the relative degree of substitution saturation, the number of both transition (ts) and transversion (tv) substitutions at each codon position was determined for each evolutionary distance present. The corrected distance measure GTR+I+G was used. The results, displayed in Fig. 5.1 reveal that the increase in frequency of ts and tv third codon substitutions slows for the larger genetic distances, indicating that the third position is saturated for both types of substitutions. The first codon position is also demonstrating a degree of saturation for transitions. The remainder appear to accumulate in a linear fashion, indicating that they are not saturated.

An Incongruence Length Difference (IDL) test (Partition Homogeneity Test; PAUP 4.062) was performed on the data sets from *msp*, *mip* and *16S rRNA* genes, to determine if any of the data sets could be combined for a more informative analysis (Farris, *et al.*, 1995). The result indicated that the *msp* and *mip* data sets could be combined (P = 0.47 for the null hypothesis that they can be combined), but not with the *16S rRNA* data set (P = 0.01 – 0.02). After combining the *msp* and *mip* data sets, and excluding the outgroup and the *mip* hypervariable region, 40% (456/1149) of sites were variable, 72% (326/456), 17% (79/456) and 11% (51/456) occurring at the third, first and second codon positions respectively. The third codon position for the combined data set is similarly saturated for ts and tv substitutions as for the individual data sets. The *mip* hyper-variable region and the 3 additional bases present in the *L. micdadei mspA* gene were excluded from all subsequent analyses.

Maximum parsimony (MP) analysis

Two analyses were conducted on the combined *mip-msp* data set, one which included all codon positions, and the second with the third codon position excluded, to determine the influence of the more variable third codon position on the tree topology. Including the outgroup, 34% (388/1149) of all sites were parsimony informative, with 27% (103/388) of these being first and second position sites, and 73% (285/388) being third codon sites. The heuristic search option was used to find the most parsimonious tree/s, because of the large number of taxa in the data set. For the complete data set, one most parsimonious tree of length 5144 steps was obtained (Fig. 5.2a). The second analysis, with the third codon position excluded, recovered 9 equally most parsimonious trees of length 316 steps and are summarised by a strict consensus tree presented in Fig. 5.2b. Bootstrap support (2000 pseudo-replications, heuristic search) are included in trees from both analyses.

Maximum likelihood (ML) and distance analysis

To determine if a similar topology could be recovered from a different phylogenetic method, the completed data set (minus the variable bases excluded as above) was also analysed by both maximum likelihood (ML) and distance methods. Firstly, the analysis model which best fitted the data was estimated to be GTR+I+G (General Time Reversible where I is the proportion of invariant sites and G is the γ shape parameter, a measure of the spread of change rates at sites along the alignment; PAUP v2 Model Test) (Rodriguez, *et al.*, 1990). Using ML and quartet puzzling (QP), the best-fit parameters were determined to be as follows:

Estimated base frequencies:

A = 0.314402 C = 0.164464 G = 0.223269 T = 0.297865

Estimated R matrix:

1.8264948	4.820743	1.617673
	1.0267696	8.7781691
		1

Estimated proportion of invariant sites (I) = 0.453304

Estimated value of the γ shape parameter (G) = 1.093153

Figure 5.3a presents the ML phylogenetic dendrogram based on these parameters. Bootstrap (100 pseudo-replicates; heuristic search) support for these data are included. Lastly, a neighbour joining (NJ) analysis was performed using GTR+I+G as the distance measure and the best-fit parameters determined above. Figure 5.3b presents the NJ dendrogram, including bootstrap support (2000 pseudo-replicates, heuristic search).

Congruence and non-congruence among MP, ML and NJ trees

In the previous analyses, utilising *mip* and *16S rRNA* gene sequence data, two clades were recovered, namely the ((*cincinnatiensis*, *longbeachae*, *sainthelensi*, *santicrucis*) *gratiana*) clade and the (*moravica*, *quateirensis*, *shakespearei*, *worsleiensis*) clade, but the topology within each differed between the analyses.

The *mip-msp* data set also recovers the ((*cincinnatiensis*, *longbeachae*, *sainthelensi*, *santicrucis*) *gratiana*) clade. All four analyses also recover a (*longbeachae*, *sainthelensi*) clade. However, *sainthelensi* and *santicrucis* remain unresolved. Although all analyses but NJ structure the clade as ((((*longbeachae*, *sainthelensi*) *santicrucis*) *cincinnatiensis*) *gratiana*), the NJ analysis with the best-fit GTR+I+G distance measure, groups *santicrucis* and *cincinnatiensis* together with a bootstrap of 76%. However, the former topology for MP and ML is not well supported by bootstrap analysis (>70%). The very short branch lengths between these nodes indicates that few substitutions separate these taxa, and the different topology probably reflects bias inherent within the various analysis methods. As a consequence the relationship of *santicrucis* and *cincinnatiensis* within this clade remains unresolved.

For the (*moravica*, *quateirensis*, *shakespearei*, *worsleiensis*) clade, all analyses recover a ((*moravica*, *worsleiensis*) *quateirensis*) clade well supported by bootstrap analysis except for the MP analysis which excludes the third codon position. This analysis recovers the same topology, but not with high bootstrap support. The rate of accumulation of substitutions at first and second codon positions is much slower than for the third codon position because most substitutions are non-synonymous, and are subject to functional constraints. As a consequence, more recent evolutionary events commonly receive only limited support from substitutions in the first two codon positions. Bootstrap analysis primarily measures the degree of conflict for a node within the data, rather than the absolute level of support. If nodes are supported by only a few substitutions, as is likely in this situation, bootstrap support will be low even if there is no conflicting support for alternative topologies. The remaining *shakespearei* node is consistently located external to the ((*moravica*, *worsleiensis*) *quateirensis*) clade but does not receive high bootstrap support from any of the analyses.

Among the remaining taxa, for which relationships were unresolved in the previous analyses (see Chapter Four), some structure is evident. An ((*anisa*, *parisiensis*) *bozemanii*, *tucsonensis*) clade is recovered by all four analyses, although only well supported by ML. The relationship of *bozemanii* and *tucsonensis* within the clade differs between analyses. Support from substitutions within the first two codon positions would be expected to be limited, for the reasons stated above. When all substitutions are included, both MP and ML support a sister relationship for these two species, but this is not supported by the NJ analysis. It is probable that there are too few substitutions determining these relationships for them to be confidently resolved at this time. In addition a (*cherrii*, *steigerwaltii*) clade is well supported when all codon positions are included. These two clades form part of a larger, well supported (((*anisa*, *parisiensis*) *bozemanii*, *tucsonensis*) (*cherrii*, *steigerwaltii*) *dumoffii*, *gormanii*,

wadsworthii)) clade, but the internal topology differs with each analysis. These sets of relationships are summarised in Fig 5.4.

Discussion

The additional analysis of the (*anisa-worsleiensis*) clade has confirmed the two clades determined by the *mip/16S rRNA* sequence data analysis, namely the ((*cincinnatiensis*, *longbeachae*, *sainthelensi*, *santicrucis*) *gratiana*) clade and the (*moravica*, *quateirensis*, *shakespearei*, *worsleiensis*) clade, as well as revealing some of the topology within each. Further, the additional protein-encoding gene sequence data has determined a third clade, comprising ((*anisa*, *parisiensis*) *bozemanii*, *tucsonensis*). A fourth clade, comprising (*cherrii*, *steigerwaltii*) is well supported when all codon positions are included. As this would be a recent divergence with respect to other *Legionella* species divergences, it is not surprising that it is not well supported by first and second codon data, where enzymic functions constrain permitted mutations.

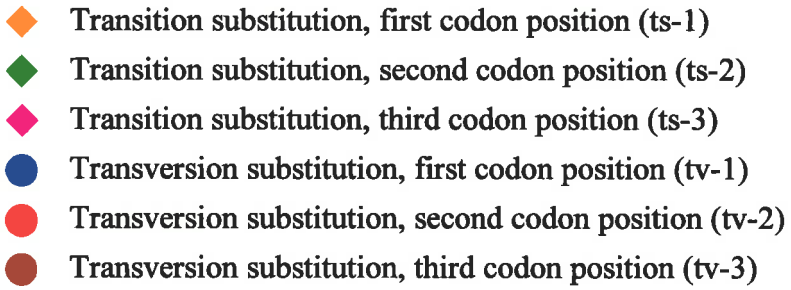
It has already been noted that while *rRNA* gene sequence is the predominant target for bacterial phylogenetic studies, recombination events limit its usefulness in recovering evolutionary relationships at lower taxonomic levels (Ash, *et al.*, 1991, Rossler, *et al.*, 1991, Utaker, *et al.*, 1995). The major clades mentioned above are now supported by three sequence data sets. Thus, these clades are highly likely to reflect the true evolutionary history of the organisms, and not just the evolutionary history of one or a small group of genes. However, much of the internal topology of the clades remains unresolved among the data sets. Still more sequence from appropriate genes is required if the complete evolutionary history of the *Legionella* genus is to be inferred with any degree of confidence. However, two theoretically conserved protein-encoding genes have been examined for phylogenetic information, and both show saturation of base substitutions at the third codon position. Cohan's Coalescence Model of gene diversity purging predicts that substitution saturation in the third codon position of

other proteins would be widespread and perhaps universal. As a consequence significant weight would need to be given to the relationships inferred from first and second codon substitutions. Because of the number of *Legionella* species, it is likely that data from a large number of proteins will be required for the topology to be resolved with confidence.

The availability of sequence from multiple genes is increasing as sequencing costs substantially reduce. However, multi-locus sequence comparisons are still relatively rare, and most are concerned with the comparisons of relationships inferred by *rRNA* genes and a single protein-encoding gene, at a genus level (Amann, *et al.*, 1988, Eisen, 1995, Kamla, *et al.*, 1996), or the evolutionary history of procaryotes in general (Ludwig, *et al.*, 1993, Ludwig, *et al.*, 1998), and the issues of horizontal gene transfer (Shen and Huang, 1986, Strätz, *et al.*, 1996, Woese, 1991). Recent comparisons of multiple protein-encoding genes have been reported for a number of bacteria, such as *Neisseria meningitidis* (Maiden, *et al.*, 1998), *Streptococcus pneumoniae* (Enright and Spratt, 1998), *Streptococcus pyogenes* and *Staphylococcus aureus* (Enright and Spratt, 1999), but the focus is more on typing strains involved with infectious outbreaks, using a technique termed multilocus sequence typing MLST (Enright and Spratt, 1999). The current paucity of intra-genus phylogenies based on the sequence of multiple protein-encoding genes is indicative of the difficulties which still exist with determining the sequence of multiple genes from every species within a bacterial genus. The difficulty and expense experienced in this study while determining the sequence of the *mip* gene for all *Legionella* species, is one example. Similarly, the *mspA* gene primers determined in this study do not universally amplify sequence from all *Legionella* species. Additional investigation would be required, perhaps involving the cloning of the *mspA* gene from several species, before a completely universal set of primers could be determined. However, such multi-locus studies must be performed if the evolutionary history of bacteria genera is to be determined.

Whether it is possible to recover the complete evolutionary history of any organism is currently being discussed by some phylogeneticists. Some argue that issues of substitution saturation and horizontal gene transfer have resulted in genomes being an incomprehensible amalgam of genes with an evolutionary history so complicated it can not be represented as a tree (Doolittle, 1999). Other phylogeneticists point to recent phylogenetic analyses based on whole genome sequences. Although only a few genomes have been completely sequenced, a recent comparison of the complete genomes of 11 microorganisms produced an evolutionary tree which was in broad agreement with that determined from small subunit *rRNA* genes (Fitz-Gibbon and House, 1999). They concluded that despite apparent rampant lateral gene transfer among microorganisms, the results indicate the existence of a single robust underlying evolutionary history for these organisms. Irrespective of the consequence of future studies which include additional complete genomes as they are determined, the understanding of the evolutionary process will ultimately be enriched and not impoverished by sequence-based phylogenetics (Doolittle, 1999).

Figure 5.1 The number of transitional (ts) and transversional (tv) substitutions at each of the three codon positions within the *mspA* gene, with relation to increasing evolutionary distance (GTR+I+G).



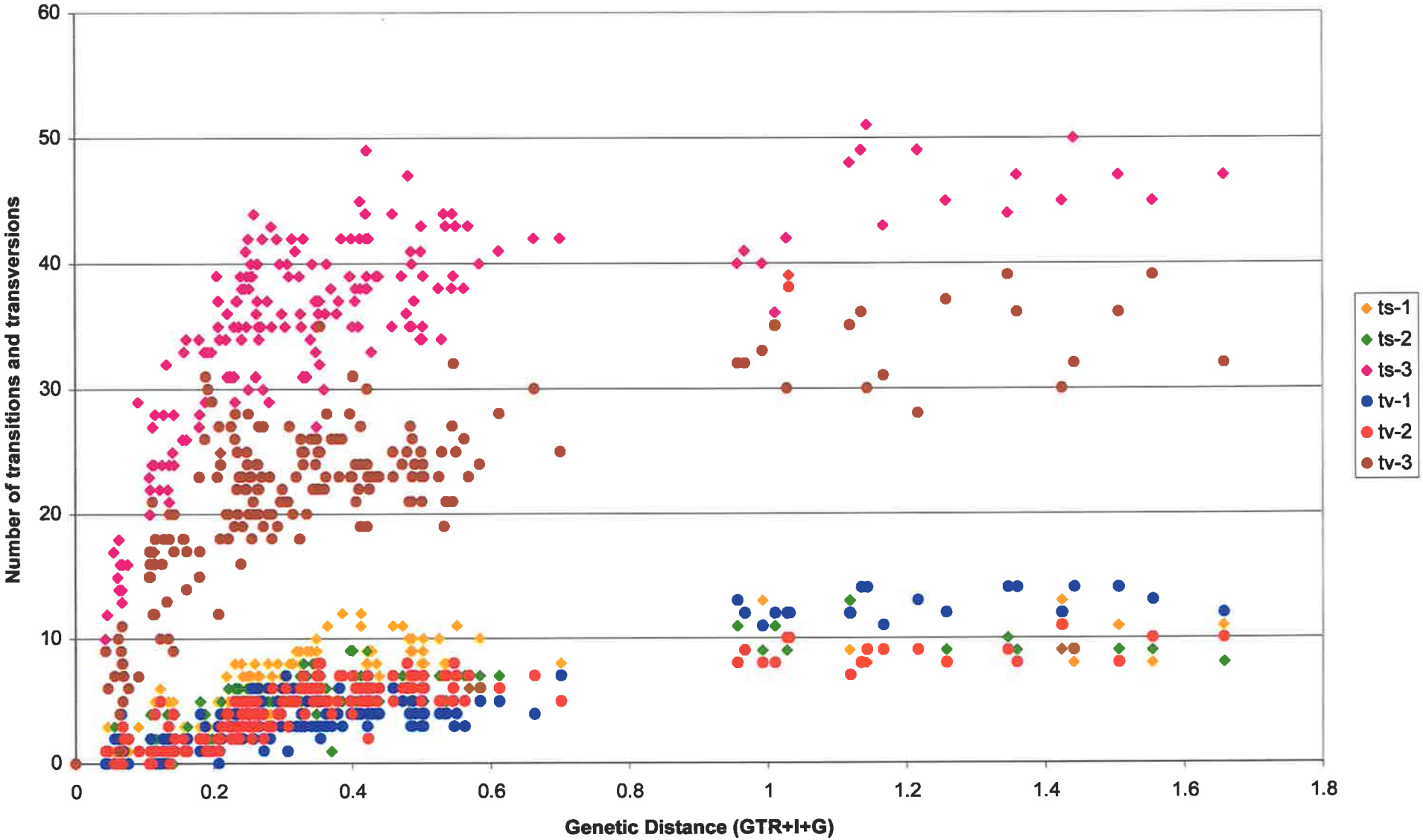
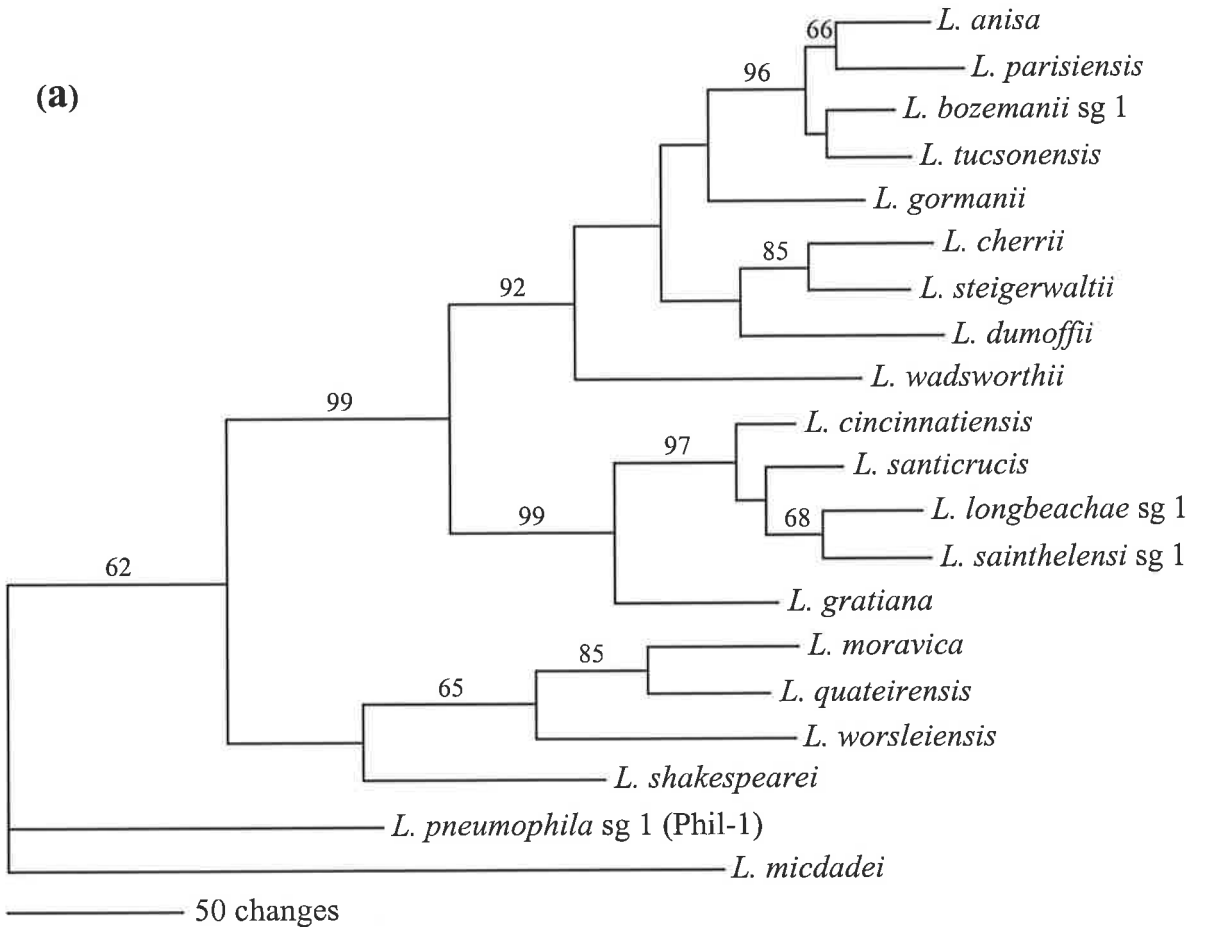

Number of transitions and transversions
60
50
40
30
20
10
0
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
Genetic Distance (GTR+I+G)
ts-1
ts-2
ts-3
tv-1
tv-2
tv-3

Figure 5.2. The most parsimonious tree found for the (*anisa-worsleiensis*) clade *mip-msp* data set (a) all codon positions, one tree of length 5144 steps, and (b) first and second codon positions only, strict consensus tree summarising the nine equally most parsimonious trees of length 316 steps. Bootstrap values greater than 50% are indicated (2000 pseudo-replicates, heuristic search)


(a)
66
96
92
85
99
97
99
68
62
65
85
L. anisa
L. parisiensis
L. bozemanii sg 1
L. tucsonensis
L. gormanii
L. cherrii
L. steigerwaltii
L. dumoffii
L. wadsworthii
L. cincinnatiensis
L. santicrucis
L. longbeachae sg 1
L. sainthelensi sg 1
L. gratiana
L. moravica
L. quateirensis
L. worsleiensis
L. shakespearei
L. pneumophila sg 1 (Phil-1)
L. micdadei
50 changes

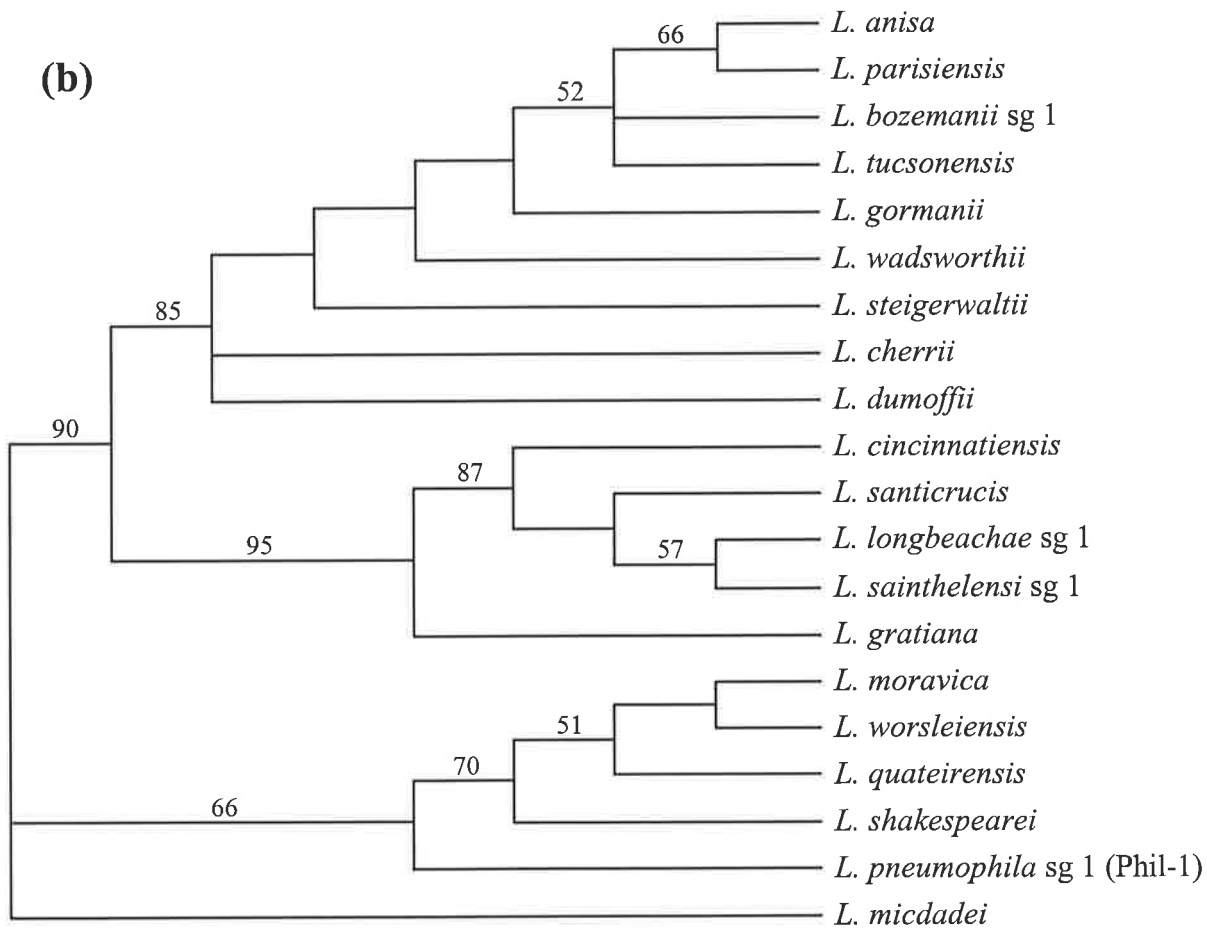
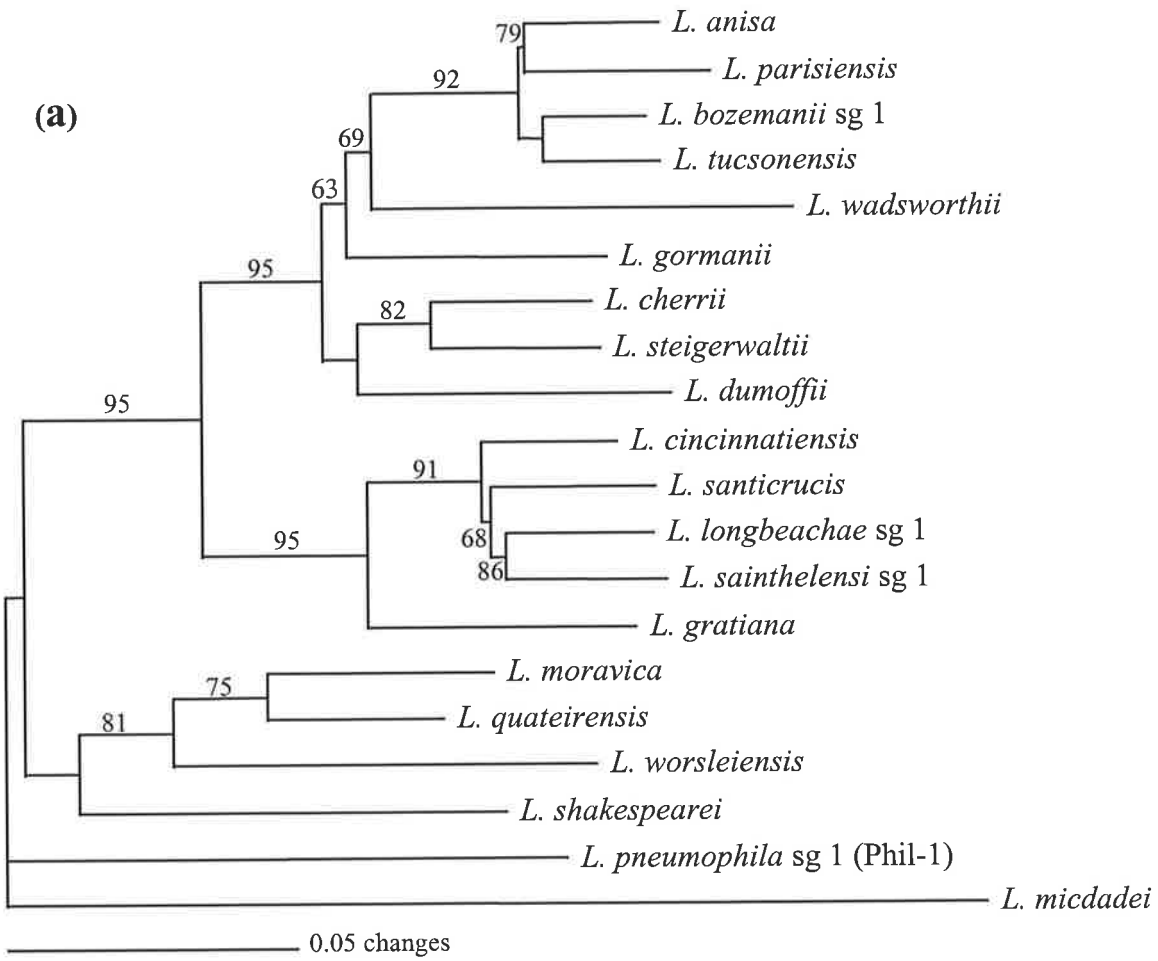

(b)
66
52
85
90
87
95
57
51
70
66
L. anisa
L. parisiensis
L. bozemanii sg 1
L. tucsonensis
L. gormanii
L. wadsworthii
L. steigerwaltii
L. cherrii
L. dumoffii
L. cincinnatiensis
L. santicrucis
L. longbeachae sg 1
L. sainthelensi sg 1
L. gratiana
L. moravica
L. worsleiensis
L. quateirensis
L. shakespearei
L. pneumophila sg 1 (Phil-1)
L. micdadei

Figure 5.3. (a) Maximum likelihood tree for the (*anisa-worsleiensis*) clade *mip-msp* data set, all codon positions, General Time Reversible (GTR+I+G) distances (Rodriguez, *et al.*, 1990) among the sequences, and (b) Neighbour-joining tree for the same data set and distance measure. Bootstrap values greater than 50% are indicated (ML, 100 pseudo-replicates, heuristic search; NJ, 2000 pseudo-replicates, heuristic search)


(a)
L. anisa
L. parisiensis
L. bozemanii sg 1
L. tucsonensis
L. wadsworthii
L. gormanii
L. cherrii
L. steigerwaltii
L. dumoffii
L. cincinnatiensis
L. santicrucis
L. longbeachae sg 1
L. sainthelensi sg 1
L. gratiana
L. moravica
L. quateirensis
L. worsleiensis
L. shakespearei
L. pneumophila sg 1 (Phil-1)
L. micdadei
79
92
69
63
95
82
95
91
95
68
86
75
81
0.05 changes

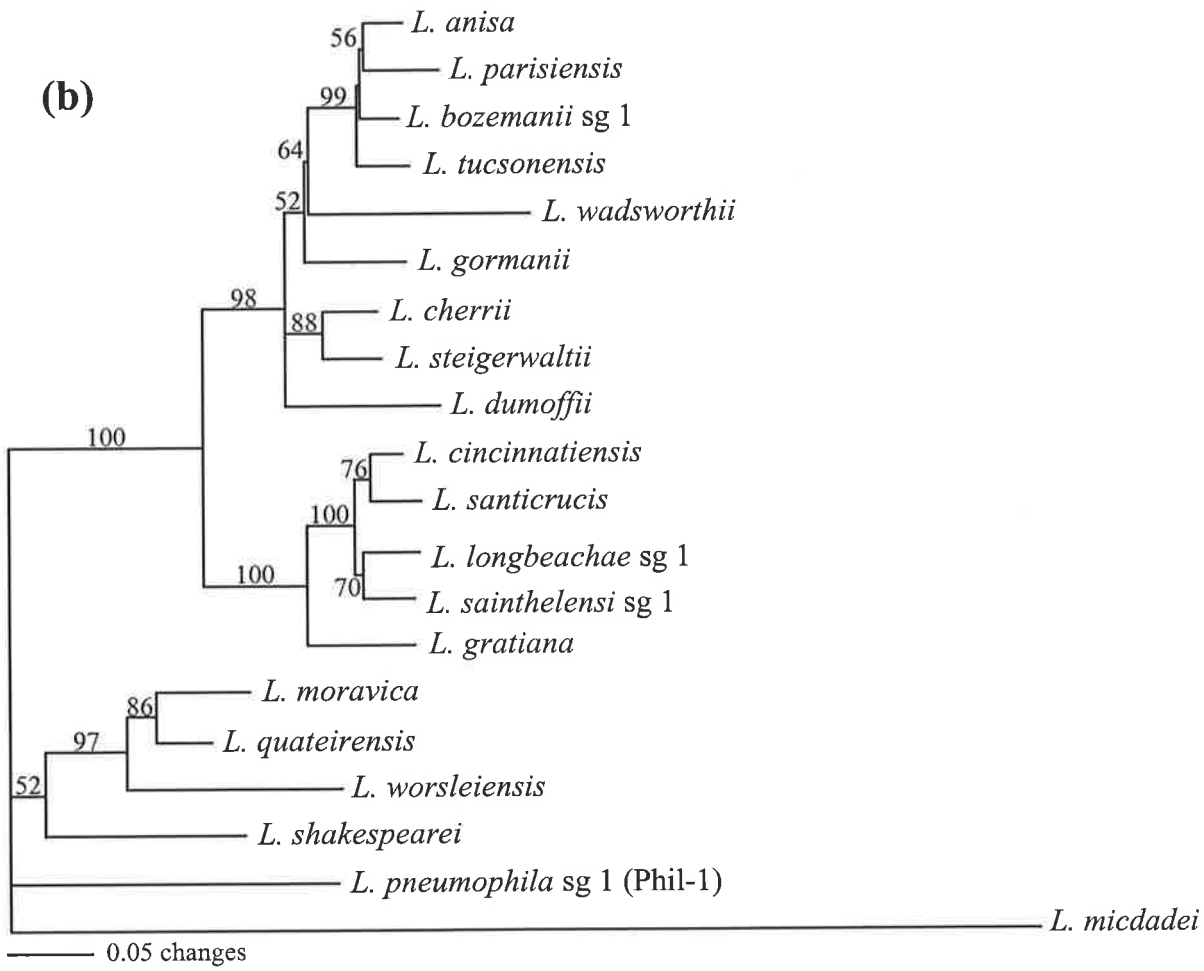
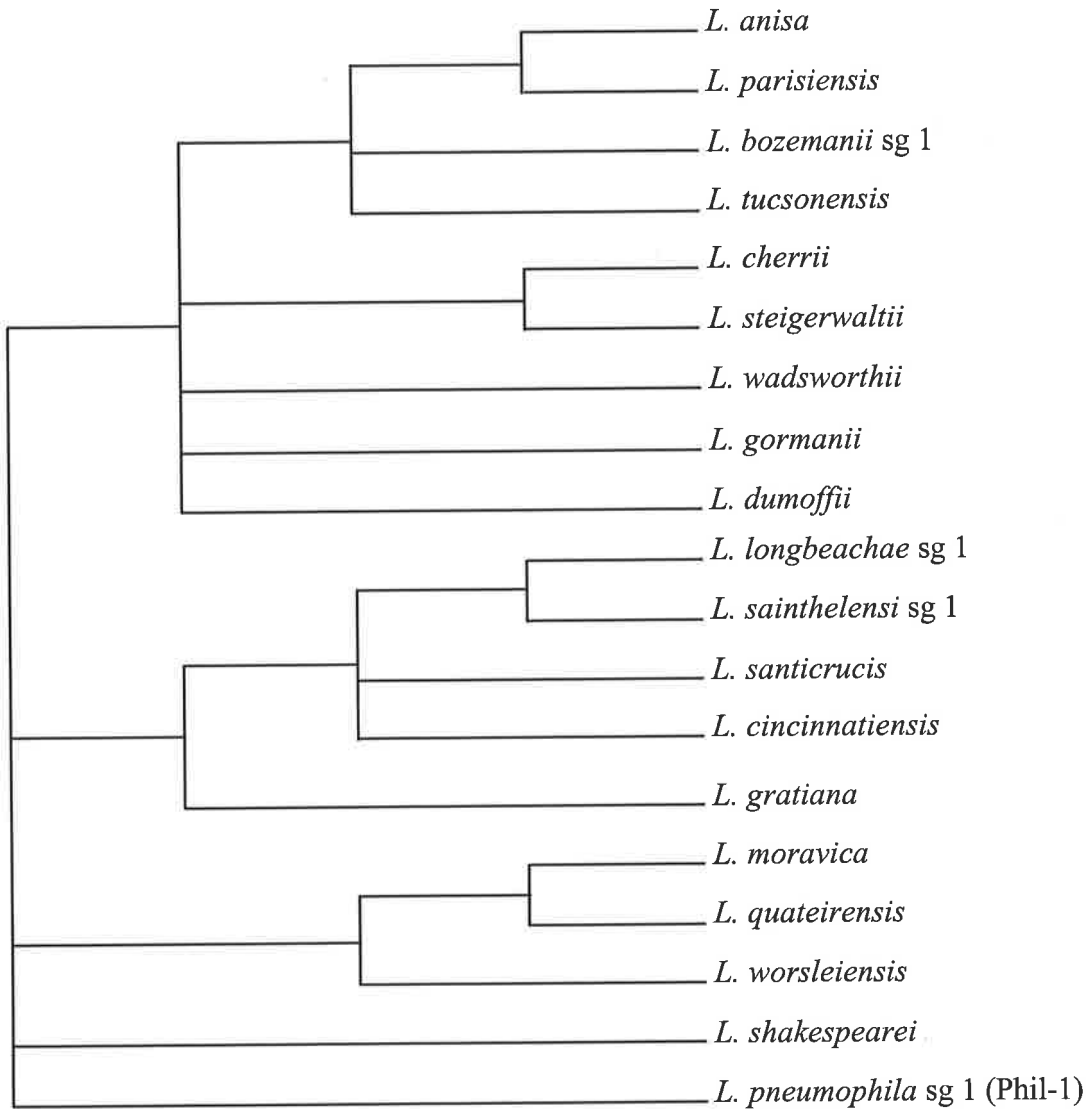

(b)
L. anisa
L. parisiensis
L. bozemanii sg 1
L. tucsonensis
L. wadsworthii
L. gormanii
L. cherrii
L. steigerwaltii
L. dumoffii
L. cincinnatiensis
L. santicrucis
L. longbeachae sg 1
L. sainthelensi sg 1
L. gratiana
L. moravica
L. quateirensis
L. worsleiensis
L. shakespearei
L. pneumophila sg 1 (Phil-1)
L. micdadei
56
99
64
52
98
88
100
76
100
100
70
86
97
52
0.05 changes

Figure 5.4. A tree summarising relationships within the (*anisa-worsleiensis*) clade receiving consistent support from the four analyses of the *mip-msp* data set.


L. anisa
L. parisiensis
L. bozemanii sg 1
L. tucsonensis
L. cherrii
L. steigerwaltii
L. wadsworthii
L. gormanii
L. dumoffii
L. longbeachae sg 1
L. sainthelensi sg 1
L. santicrucis
L. cincinnatiensis
L. gratiana
L. moravica
L. quateirensis
L. worsleiensis
L. shakespearei
L. pneumophila sg 1 (Phil-1)

Chapter Six

A *mip* sequence-based classification scheme

Introduction

The genus *Legionella* comprises approximately forty species, at least seven of which have more than one serotype (Table 1.1). Approximately half of the species have been associated with human disease. *Legionella*-like organisms isolated from clinical specimens, or from the environment during the course of an outbreak, need to be identified to elucidate the disease process and to identify the source. Legionellae have proved to be relatively unreactive when utilising traditional biochemical tests, necessitating more complex identification methods (Brenner, 1987, Brenner, *et al.*, 1985, Mauchline and Keevil, 1991, Wilkinson, 1988). Serology-based methods are widely used in clinical laboratories, but antigen cross-reactivity limits specificity, and restricts their confident use to a few frequently isolated species (Maiwald, *et al.*, 1998, Veríssimo, *et al.*, 1999). This is especially true for countries where legionellosis caused by species other than *L. pneumophila* is common (Doyle, *et al.*, 1998). More complex classification schemes have been proposed (Mauchline and Keevil, 1991, Veríssimo, *et al.*, 1999), the most successful being one based on the range and proportion of cellular fatty acids and ubiquinones (Diogo, *et al.*, 1999, Jantzen, *et al.*, 1993, Lambert and Moss, 1989, Wait, 1988, Wilkinson, *et al.*, 1990). As additional species have been characterised, this method has become less discriminating, as apparently unique patterns were proved to be shared by several species (Wilkinson, *et al.*, 1990). The inclusion of hydroxylated fatty acids has improved discrimination, but requires the analysis of both mono-

and di-hydroxylated fatty acids, and individual patterns are complex making analysis difficult (Jantzen, *et al.*, 1993).

Gene sequence-based phylogenic (genotypic) schemes have become widely used for organisms that are difficult to classify (Dewhirst, *et al.*, 1992, Lawson, *et al.*, 1993, Ludwig and Schleifer, 1994, Poyart, *et al.*, 1998, Ruimy, *et al.*, 1994, Sallen, *et al.*, 1996). Compared to phenotypic methods, genotypic schemes have the advantage of containing more discriminatory data that is easily reproducible between laboratories, and with less ambiguity because of the "digital" nature of the data. Many genotypic schemes utilise sequence variation in the *16S rRNA* (Collins, *et al.*, 1994, Dewhirst, *et al.*, 1992, Lawson, *et al.*, 1993, Ludwig and Schleifer, 1994, Ruimy, *et al.*, 1994, Sallen, *et al.*, 1996), because of the ease with which regions can be amplified and sequenced using universal primers. The *16S rRNA* sequences of *Legionella* species have been reported (Hookey, *et al.*, 1996), as have the sequences of the *mip* gene (See Chapter Three). Other gene sequences have been determined for *Legionella* (Black, *et al.*, 1990, Hoffman, *et al.*, 1992a, Segal and Shuman, 1997), but only the *rRNA* genes and the *mip* gene have been comprehensively determined for most species, an essential prerequisite for any gene to be the basis of a genotyping scheme. In Chapter Three, the species variation among both the *16S rRNA* and *mip* genes was reported. The *mip* gene showed over twice the variation at the DNA level (56% of base sites), compared with that of *16S rRNA* (23% of base sites), with a pair-wise comparison of species producing a variation of 3% - 31% (mean 20%) between species pairs for the *mip* gene, compared with 1% - 10% (mean 6%) for the *16S rRNA* gene. These statistics infer that the *mip* gene contains more discriminatory data than the *16S rRNA*.

Additionally, the *mip* gene appeared to be genetically relatively stable, with no evidence of homologous recombination, in that identical or near identical sequence was not found for the gene from phenotypically divergent species. With respect to genetic stability, the

mip gene may therefore behave like a housekeeping gene, which are known to be more stable than other gene classes (Achtman, 2000). Homologous recombination would severely compromise a sequence-based classification scheme (Achtman, 2000), and is a theoretical possibility at least for *rRNA* gene targets (Strätz, *et al.*, 1996). Thus the genetic stability but greater mutational variation of the *mip* gene (compared with *16S rRNA* genes) suggests that it is an ideal target for a classification scheme, with results likely to be more discriminating in identifying species and more resilient to clonal variation within each species. It may even be possible to discriminate between serogroups where present, or demonstrate distinct intra-species clonal groups.

This chapter reports the use of the *mip* gene to develop a sequence based classification scheme for *Legionella*, the first proposed for this genus. Further, it reports the comparison of sequences from species with additional serogroups, to determine if serogroups can be discriminated. Similarly, it reports the comparison of sequences from wild strains isolated on several continents, for which there is confirmatory phenotypic, or DNA hybridisation identification data, to test the robustness of the scheme for variation within strains of the same species. Lastly, isolates which appear phenotypically or from DNA hybridisation studies to be different from currently characterised species were tested to determine if a sequence-based classification scheme can clarify their identity. Some of these unusual isolates have been previously reported (Wilkinson, *et al.*, 1990).

Summary of methodology and sequencing strategy

The type strains and the *mip* gene sequences used to construct the classification scheme are given in Table 3.1. Table 6.1 presents additional reference type strains and wild strains, used to test the robustness of the classification scheme, and the GenBank/EMBL/DDBJ accession numbers of sequences determined from those strains during the study, and already published (Ratcliff, *et al.*, 1998). A total of 350 wild-strain isolates (or

DNA) obtained from clinical and environmental sources from Australia, Europe, Israel, Kenya, Japan, Singapore and the USA, were used in the study (includes LLAP DNA samples). Both extracted chromosomal DNA, and heated cell lysates from these strains were used in the amplification reaction.

From an alignment of the *mip* sequence of 40 type strains, forward and reverse primers were chosen to target the RBS/ORF region and the PPIase site of the *mip* gene respectively, to amplify a fragment of 661 to 715 bases, depending on the presence and size of the hypervariable region, which is species dependent (see Fig. 3.1). This equates to nearly 90% of the gene from the 5'end. While the inferred amino acid sequence is highly conserved in the targeted regions, the primers required redundancies to account for variation at the third codon site.

The primer sequences were as follows:

Forward primer (Legmip_f) (27mer):

5' - GGG(AG)ATT(ACG)TTTATGAAGATGA(AG)A(CT)TGG - 3'

Reverse primer(Legmip_r) (23mer):

5' - TC(AG)TT(ATCG)GG(ATG)CC(ATG)AT(ATCG)GG(ATCG)CC(ATG)CC - 3'

Gene amplification consisted of 35 cycles of denaturation at 94°C for 1 minute, annealing for 2 minutes at 58°C, and extension for 2 minutes at 72°C. Products were visualised by gel electrophoresis, purified using the QIAquick PCR Purification Kit, and sequences were determined using Dye Terminator chemistry. Legmip_f proved to be unreliable in priming the sequencing reaction, so a new forward sequencing primer, designated Legmip_fs, was designed to overlap Legmip_f but stepped into the amplicon by 7 bases at the 3' end. Its sequence is as follows (26mer):

5' - TTTATGAAGATGA(AG)A(CT)TGGTC(AG)CTGC - 3'

The sequences determined were compared with the type strains using the Homology Search option in GeneCompar 2.0 to find related sequences, and the Cluster Analysis module from the same program was used to produce the UPGMA phylogenetic dendrogram.

Results

Utilising the PCR primers Legmip_f and Legmip_r, all type strains gave a single amplification product of the predicted size (661-715 bp depending on the species and presence of the hypervariable insert), with the exception of *L. geestiana* which failed to produce any product. The sequence of each amplicon was determined, using both Legmip_r and Legmip_fs which reliably primed the sequencing reactions to produce unambiguous sequence after approximately fifteen bases, until the end of the fragment. The combination of Legmip_f as the forward amplifying primer with Legmip_r to produce the amplicon, and Legmip_fs as the forward sequencing primer consistently produced clean, unambiguous sequence. During the course of this study, all amplicons from type and reference strains were sequenced in both forward and reverse directions. To reduce costs, amplicons from many wild strains were only sequenced in the forward direction unless sequence ambiguity or unusual sequence required confirmation with sequence in the reverse direction. This decision was based on the lack of observed conflicting bases between forward and reverse sequences for the *mip* sequences, where both directions had been sequenced. Amplicon sequences from the type strains were identical to the published sequence, and contained sufficient variation (3.6% - 30.5%) to uniquely identify each species.

A small 194 bp fragment of the *mip* gene from *L. geestiana* was however recently amplified and sequenced, utilising a "universal" forward primer targeting the homologous region at amino-acid residues 164-170 in Fig. 3.1, (Geesmip-f: 5'-GTNACNGTNGANTANA CNGG-3', where N represents the inclusion of all four nucleotides) and Legmip-r as the

reverse primer. A similarly "universal" forward primer targeting the homologous region at residues 50-65 failed to produce valid sequence when used in conjunction with Legmip-r.

To test the specificity of the primers, DNA extracted from the following organisms was tested; *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Escherichia coli*, *Haemophilus influenzae*, *Salmonella enterica* (multiple serovars), *Shigella flexneri*, *Proteus mirabilis*, *Aeromonas hydrophila*, *Pseudomonas aeruginosa*, *Yersinia enterocolitica*, *Campylobacter* (multiple species), *Helicobacter pylori*, *Corynebacterium diphtheriae*, *Mycobacterium abscessus*, *Vibrio parahaemolyticus* and *Nocardia asteroides*. No amplification was detected, despite the presence of genes coding for Mip-like analogues in some species (Horne, *et al.*, 1997, Horne and Young, 1995, Rahfeld, *et al.*, 1996, Wong, *et al.*, 1997).

Heat lysis proved adequate in extracting DNA for amplification, although boiling or steaming gave slightly more consistent results than microwave heating. Prolonged boiling or steaming for 15 minutes did not appear to degrade the DNA, as measured by the amount of amplicon produced by subsequent amplification.

Table 6.1 presents the results of additional reference strains, as well as the comparison of results from the wild strains. It includes strains that conform to a type strain, and those for which a speciation can not be determined, which may represent novel species. For a few of these latter strains, amplification produced only limited product, which sequenced poorly. Lowering the annealing temperature to 50°C produced more product, which in turn produced longer unambiguous sequence. Sequencing was also attempted after amplification at a below optimal annealing temperature to produce multiple products of different size, and excellent sequence was still obtained with Legmip_fs, but not with Legmip_r, as would be predicted. Isolates LC3644 and Int-6, thought by the isolating laboratory to be strains of *L. geestiana*,

based on cultural characteristics and serology, were not able to be amplified, even with the annealing temperature reduced to 40°C. However, sequence was obtained from an amplicon produced by Geesmip-f and Legmip-r. These results have been included in Table 6.1, even though the sequence length of 194 bp is much shorter than that for other strains.

A UPGMA phylogenetic dendogram of the inter- and intra-species *mip* sequence similarity is presented in Fig. 6.1. The similarity determined for the shorter sequence from *L. geestiana* and the two *L. geestiana* wild strains have been included. The type strains are easily discriminated, and the wild strain isolates group within the same species as determined by serological, chromatographical or molecular identification techniques, except for a few where the phenotypic identification was based solely on serology, and cross-reactivity prevented unambiguous speciation. In addition, the strains thought to be new species from their phenotypic identification, were discriminated from all currently characterised species.

DNA from a number of LLAP strains was also able to be tested to determine if such strains contained a *mip* gene which could be amplified with the Legmip primers, and if so, to determine how similar the sequence was to culturable legionellae. The *mip* gene was detected in the *L. lytica* type strain and all of the 9 uncharacterised LLAP strains tested. Most strains loosely clustered with *L. lytica* in several distinct groups, but LLAP-1 and LLAP-10 were more closely related to other legionellae. These results are included in Table 6.1 and Fig 6.1.

Discussion

For a sequence based classification scheme to be successful, the gene amplification needs to be specific for the genus but universal for the individual species, and the subsequent inter-species sequence variation needs to be sufficient to discriminate clearly between species, even after allowing for the intra-species sequence variation found among wild strains.

The Legmip primers reported here produced a single amplicon of the estimated size, and in sufficient quantity to enable accurate sequence determination, for all isolates tested, whether type or wild strains, with the exception of the *L. geestiana* strains. However, the fact that 194 bp sequences of the *mip* gene from these three strains were obtained confirms that *L. geestiana* does possess a *mip* gene similar but not identical to that from other *Legionella* species. The failure of the Legmip primers to produce an amplicon is therefore related to primer specificity and not to the absence of the *mip* gene. Several attempts to clone and sequence the entire gene from *L. geestiana* have failed (unpublished data), so the modifications required to the sequence of the Legmip primers to enable amplification of sequence from *L. geestiana* strains are known at this time. This has not been an impediment as no locally derived wild strain failed to amplify with the Legmip primers. Similarly, other laboratories should experience little problem, as such isolates are likely to be very rare.

Figure 6.1 shows that all of the species could be easily discriminated, and that the wild strain isolates grouped within the same species as determined by serological, chromatographical or molecular identification techniques, except for a few where the phenotypic identification was based solely on serology, and cross-reactivity prevented unambiguous speciation. Many of these strains had been retained and stored at -70°C because of these non-typical reactions. For instance LC2763, believed to be a *L. bozemanii*, also showed strong cross reactivity with *L. parisiensis*-specific antiserum. LC3936 reacted equally well with *L. cincinnatiensis*, *L. santicrucis*, and both serogroups of *L. sainthelensi* and *L. longbeachae* specific antiserum. LC4042 reacted weakly with only *L. spiritensis* specific antiserum. However, this isolate was red autofluorescent, suggesting the isolate is likely to be related to either *L. erythra* or *L. rubrilucens*, the only red autofluorescent species characterised at the time of isolation. The genotypic identification of other strains from the

same laboratory, where the serological identification had been confirmed by ribotyping and/or DNA hybridisation, was identical in every case.

These results demonstrate how ambiguous serological identification can be for some species, and reveal a significant potential flaw with serological identification. LC4042 is now known to be a wild strain of a newly characterised species, *L. taurinensis* (Lo Presti, *et al.*, 1999). However, at the time of isolation, which was prior to the recognition of *L. taurinensis*, serological identification of such strains could only be as the species which is most similar antigenically. Strains of uncharacterised species would be recognised as novel only if they were significantly antigenically different from all known species. Otherwise, they would be mis-identified based on the strongest cross reactions. In contrast, genotypic schemes inherently recognise such strains as not only unique but also quantify the difference to all other species.

Legionella genomospecies clustered with *L. quinlivanii* with which it shares 69% DNA homology (Benson, *et al.*, 1996). With a difference of 3.6%, *L. bozemanii* and *L. tucsonensis* showed the least inter-species sequence variation among currently characterised species. However, the four *L. bozemanii* wild strains and *L. bozemanii* sg 2, with a intra-species variation of 1.3% or less, were easily discriminated from *L. tucsonensis*. Similarly, *L. cincinnatiensis* wild strains (maximum intra-species variation of 1.6%) were easily discriminated from *L. santicrucis* strains (inter-species variation of 4.1%). Some species, such as *L. anisa*, *L. longbeachae*, *L. micdadei* and *L. londiniensis*, for which there were several wild strains tested, showed very little sequence difference, suggesting a clonal population, compared to *L. feelei*, *L. jamestowniensis*, *L. rubrilucens* and *L. quinlivanii*, which appeared to demonstrate multiple divergent clonal populations. This finding is especially interesting given that the *L. longbeachae* strains were isolated on several continents, whereas all but one of the *L. quinlivanii* isolates, including the type strain came from one region within Australia.

The mono-phyletic nature of *L. longbeachae* has been demonstrated by other workers (Bull, *et al.*, 1997). *Legionella jamestowniensis*, *L. rubrilucens* and *L. quinlivanii* were generally quite divergent from their nearest neighbour, so the greater sequence variation among wild strains did not compromise the resolution of their speciation. This contrasts with chromatographic identification schemes, where ubiquinone and fatty acid profiles do not discriminate easily between some species, for example *L. erythra* and *L. rubrilucens* (Wilkinson, *et al.*, 1990). In fact, it was this lack of discrimination when using such methods which was the motivation for this study.

LC0777C is a very interesting European strain. It reacted poorly serologically, and while there was moderate homology to *L. gormanii* to which it genotyped, there was also moderate homology to *L. tucsonensis* and *L. feeleii* sg 2. With 4.2% sequence variation from *L. gormanii*, LC0777C demonstrates more genotypic variation from *L. gormanii* than wild strain variation in sister species. This is consistent with the reduced serological reactivity, and may represent an additional serogroup, a new genotype clone or even a novel species. Three additional strains have recently been recognised as belonging to the same cluster.

However the multiple serogroups for species where they have been determined, were not able to be confidently discriminated by this scheme in every circumstance. For example, the sequences for *L. hackeliae* sg 1 and 2 are identical. Similarly, many of the serogroups of *L. pneumophila* are identical to each other, and as a species seem to mainly cluster into two closely related clonal populations, with serogroup 1 strains represented in both. *L. pneumophila* serogroup 5 and three non-serogrouped wild strains fall outside these two clonal populations. These later two strains may represent a new serogroup or novel species, being both serologically and genotypically different to other *L. pneumophila* strains. IMVS-D1/77, a serogroup 13 strain was located with the serogroup 13 type strain. For *L. bozemanii*, IMVS-A8E7, a serogroup 2 isolate and IMVS-A5F7, a serogroup 1 isolate both cluster with the

serogroup 2 type strain. Two other serogroup 2 isolates, IMVS-D2/7 and IMVS-K7B4 produced a similar sequence, but were different to the serogroup 2 type strain. Similarly, some serogroup 1 strains clustered uniquely together. Consequently, while these latter genotype groups may be indicative of a particular serogroup, as may some of the unique *L. pneumophila* serogroup sequences, the differences are small, and it is possible that such differences may be within the normal clonal variation found in the wild strain population independent of a specific serogroup. The testing of more strains of these species would be necessary to determine if these few differences are characteristic of specific serogroups.

In contrast, the differences in sequence from the two serogroups of *L. sainthelensi* do appear discriminatory, with two wild strains producing a sequence identical to the serogroup 2 type strain. Strain LC4261 has been serologically confirmed as *L. sainthelensi* sg 2, although it also demonstrated serological cross-reactivity with *L. santicrucis*, and to a lesser extent *L. cincinnatiensis* and *L. longbeachae*. Similarly, although the *L. longbeachae* serogroup 2 type strain differs from the serogroup 1 type strain by only two bases, the 45 serogroup 1 strains and three serogroup 2 strains all produced sequence identical to their type strains. Therefore, further testing of additional serogroup 2 wild strains may prove this small difference between the two serogroups to be definitive.

Two strains with identical *mip* gene sequence, IMVS-896 and IMVS-915, were 96.2% similar to *L. rubrilucens*, but had atypical ubiquinone patterns. Additional strains from Europe and the USA also possessed identical *mip* gene sequence, but were thought to be *L. spiritensis* strains, based on serological reactivity, even though they were red pigmented. Recently, one of these D3356 (Turin), isolated in France, was characterised as a novel species, *L. taurinensis* (Lo Presti, *et al.*, 1999), recognising the uniqueness of these strains. This again emphasises the ability of sequence-based classification schemes to recognise the uniqueness of strains which belong to uncharacterised novel species.

Wilkinson, *et al.* (1990) reported the existence of strains representing six uncharacterised species, designated Species A to F. These strains were amplified and the product sequenced, to determine how well the genotyping scheme would identify them. Species F is the type strain of *L. adelaidensis* (Benson, *et al.*, 1991). Species A and B give unique sequences, supporting the ubiquinone and fatty acid profile data that they are indeed novel species. Species C, D and E cluster closely with *L. londiniensis*, *L. waltersii* and *L. brunensis* respectively. However these latter three species were either not characterised or isolates not available at the time of testing (Benson, *et al.*, 1996, Dennis, *et al.*, 1993, Wilkinson, *et al.*, 1988). A comparison of the ubiquinone and fatty acid profiles for Species C, D and E, and *L. londiniensis*, *L. waltersii* and *L. brunensis* confirm that the scheme has classified them correctly. The speciation of Species C and E has been independently confirmed with DNA/DNA hybridisation (R. Benson, personal communication, 1997). Additional to these uncharacterised species, other isolates, stored because they possessed unusual or unique ubiquinone and fatty acid profiles, were also tested. Isolates IMVS-911 and IMVS-960, designated Species H, produced sequences identical to each other, but unique from all characterised species. A third isolate D2897 has recently been added to this species. Similarly IMVS-959, designated Species I, produced a unique sequence. These three results are in complete agreement with the ubiquinone and fatty acid profiles, and these strains are likely to represent two new species not currently described. Species G (isolates IMVS-823 and IMVS-895) produced identical sequences, but grouped moderately close to *L. shakespearei* (5.4% variation). These two isolates produced ubiquinone and fatty acid profiles similar to *L. shakespearei*, a species which had not been described at the time of their storage (Verma, *et al.*, 1992). LC3043 and LC3044, serologically similar to each other, were found to also have sequence identical to each other, and identical to the two Species G strains. Interestingly, they demonstrated weak serological homology only with *L. shakespearei*. A fifth uncharacterised isolate from Japan was recently added to this novel group which is likely

to represent an additional novel species. Species J (isolate IMVS-933) has grouped with *L. feeleii*. While the isolate produces a similar ubiquinone profile to *L. feeleii* (Moss, *et al.*, 1983) the fatty acid profile shows some differences. The significance of this result is unclear, and identity of this isolate also needs to be further elucidated. Three additional isolates, designated Species L (LC1863), Species M (LC4046) and Species N (LC4048) when published (Ratcliff, *et al.*, 1998) were all found to have unique sequence, consistent with the phenotypic classification. Two additional isolates have subsequently been determined to also be Species N strains.

Additional unusual strains more recently submitted from laboratories in Europe and the USA have also been tested, as the early results looked promising. Some of these strains are known to be novel, based on DNA/DNA hybridisation comparisons. Two related, novel strains, D1844 and D3362 (based on DNA/DNA hybridisation) have *mip* sequence which is identical to IMVS-C1/946. Two additional German strains, Dr-36 and Dr-47, also have *mip* sequence identical to these three strains. The *mip* sequence of the five strains is 96.9% similar to that from the *L. anisa* type strain. These results suggest that all five strains represent a novel species. Similarly BA019T/3, which is 98.3% related, based on *mip* sequence similarity, is either related to this novel species, or is itself novel. Most of these strains were thought to be either *L. anisa* or *L. bozemanii* based on phenotypic characteristics. Prior to the examination of strains D1844 and D3362, even IMVS-C1/946 was thought to be a variant strain of *L. anisa* (Ratcliff, *et al.,* 1998) based on the *mip* sequence results. However, its difference was quantitatively recognised. The availability of strains D1844 and D3362 enabled the *mip* sequence difference to be recognised as indicating uniqueness.

Other strains which are either known (based on DNA/DNA hybridisation) or thought to represent novel species based on significant diversity in their *mip* sequence are as follows: IMVS-C7C3, -C7D1 & -97/L11, FM-3-661 & D2440, LC2720, IMVS-3376, MT-53, D276c,

D1541, D1620 & D2862, D4131 & 99-119, D4522e, D4728, D4750a & D4751, 81-029, 89-2081, 91-010, 91-004, 91-010, 91-028, 99-113 and 99-121. It is interesting that a number of these novel strains, along with Species M and N, form a diverse clade which is very different from currently characterised species, based on *mip* sequence similarity. Although none have been associated with disease, they have been isolated from very different geographical regions. Their importance and ecology would make an interesting study. Strain D4728, known by DNA/DNA hybridisation to be unique, is currently the strain with a *mip* sequence closest (1.3%) to a currently characterised species, namely *L. longbeachae*. This result indicates that while many of the above strains are easily recognised as unique based on the *mip* sequence result, some species are genetically much closer to their sister species, with a correspondingly smaller *mip* sequence variation. Whether there is an absolute *mip* sequence variation which is indicative of species' uniqueness is not yet known. However, as more strains are characterised, and the topology of the genus better understood, the concept of a definitive cut-off will be unnecessary.

The *mip* sequence similarity for a number of the LLAP strains has recently been included in the formal classification of three new *Legionella* species (Adeleke, *et al.*, 2000). The *mip* sequence results were in complete agreement with serological, DNA/DNA hybridisation and fatty acid and ubiquinone results. LLAP-2 and D4313, slow-growing but culturable strains were very similar to LLAP-6, the type strain for *L. rowbothamensis*. A second slow growing Czech strain TE-1 also belongs to this species, based on the *mip* sequence. These findings suggest that at least some LLAP strains may be culturable, if similar conditions are used as for the culturable strains such as incubation at 30°C in 5% CO_2 or in a candle jar, for 7 – 10 days (V. Drasar, personal communication, 1998). LLAP-7 and LLAP-9 are strains of *L. lytica*, and LLAP-1 and LLAP-10 are the type strains of the two new species *L. drosanskii* and *L. fallonii*, respectively. LLAP-4 and LLAP-12 are strains of an additional novel species yet to be characterised.

Although it is not possible to be certain until the strains are properly classified, this study has revealed a possible 30 novel species, in addition to the 44 species currently recognised. This makes *Legionella* a very diverse and complex genus containing many more species than most bacterial genera.

In conclusion, the scheme was able to unambiguously discriminate between 44 *Legionella* species, and correctly group 26 additional serogroups or reference strains within those species. Additionally, 350 wild strains isolated within Australia, Europe, the USA, Japan, Singapore, Israel and Kenya, were grouped consistently with their phenotypic identification, including strains which are thought to represent 32 currently uncharacterised novel species. Two isolates grouped genotypically different from their phenotypic classification, but it is probable that DNA hybridisation would support the genotypic classification. There were no regional clonal variations detected which would indicate laboratories in other countries would be troubled with ambiguous classification. The scheme is technically simple for a laboratory with even basic molecular capabilities and equipment, if access to a sequencing laboratory is available, especially given that heat lysis is quite adequate in "extracting" DNA suitable for amplification. Although many wild strain amplification products were sequenced in both directions during this study, routinely only one direction would be necessary. While both sequencing primers performed extremely well, Legmip_fs would be the preferred sequencing primer, as it proved to be unaffected by non-specific amplification.

There was no evidence of genetic recombination horizontally across species in the sequences from the approximately 400 strains used in this study, and while still theoretically possible, such events would be unlikely to affect the classification of wild strains in practice.

However, unusual results or critical isolates could be confirmed with other phenotypic methods or *rRNA* gene sequences.

Table 6.1. Isolates used to test the robustness of the classification scheme.

[a] Wild strains were isolated within Australia (uncoloured, prefixed IMVS), USA (coloured maroon, supplied by R. Benson) UK (coloured red, prefixed LC, supplied by T. Harrison), Germany (coloured blue, prefixed Dr, supplied by J. Helbig), the Czech Republic (coloured yellow, supplied by V. Drasar), Rome (coloured green, prefixed Rm, supplied by M. Castellani-Pastoris), and Japan (coloured aqua, supplied by A Saito) unless indicated in parenthesis in the table.

[b] sequences determined and submitted to the GeneBank database during this study.

NA: not applicable, either isolate is the ATCC reference strain, or a wild strain with identical sequence to a type or reference strain.

[c] Phenotypic identification included colonial morphology, autofluorescence, and slide agglutination or immunofluorescence using polyvalent antisera for all isolates. Australian strains were additionally identified with chromatographic analysis of "whole cell" fatty acids and ubiquinones (Wilkinson, *et al.,* 1990). Strains coloured red, prefixed LC and marked * were additionally identified with ribotyping and, with the exception of LC0455, DNA hybridisation. Strains coloured maroon and marked # were identified withDNA hybridisation. Identification enclosed in parenthesis relates solely to similarly marked isolate.

NC: not characterised, identification does not conform to currently characterised species, and are potentially novel species.

[d] number of nucleotide (DNA) or amino acid (AA) differences from the type strain. Sequence differences for *L. pneumophila* isolates are with respect to Philadelphia-1. Sequences from *L. geestiana* are only 194 bp long.

[e] sequence identical to *L. pneumophila* serogroup 1 (Knoxville-1) and other strains.

[f] sequence identical to *L. pneumophila* serogroup 1 (Wadsworth) and other strains.

[g] number of differences from closest characterised species.

Legionella isolate [a]	GenBank Acc. No. [b]	Phenotypic identification [c]	Sequence identification	No. of differences [d] DNA	AA
L. anisa					
IMVS-A5F5, -C4B9, -C4H2, -0759, 89-2071, LC3934, Dr-1, Dr-4, Dr-6 {Italy}, Dr-8, Rm-11	NA	*anisa*	*anisa*	0	0
IMVS-946, [D1844#, D3362#], Dr-36 {Italy}, Dr-47 {Czech}	AF022312	*anisa/bozemanii* [NC]	?NC	18	1
BA-10T/3		*anisa/bozemanii*	?NC	16	2
L. birminghamensis					
IMVS-C4D5, Ji-58-682	NA	*birminghamensis*	*birminghamensis*	0	0
IMVS-K7H8	AF047743	*birminghamensis*	*birminghamensis*	1	0
L. bozemanii					
LC4348, Dr-50 {UK}	NA	*bozemanii* sg 1	*bozemanii*	0	0
Rm-5		*bozemanii*	*bozemanii*	1	1
IMVS-D2/50, -D2/62		*bozemanii*	*bozemanii*	3	1
sg 2 ATCC35545, IMVS-A5F7, [IMVS-A5A1, -A8E7], Dr-5 {Sweden}, D4691, D4781d {Japan}, UL-7T/2	AF022308, AF022309	NA, *bozemanii* sg 1, [sg 2]	*bozemanii*	2	1
IMVS-A5I7, -K7B3, -D2/64, D4781a {Japan}, D4787, D4634, Dr-16 {Spain}, Dr-25 {Belgium}, Dr-27 {Belgium}, Dr-28 {France}, Rm-22, Rm-24, [LC2763]	AF022310	*bozemanii* sg 1 [or *parisiensis*]	*bozemanii*	5	1
IMVS-K7B4, -D2/7, -D2/8, -94/2, -32935	AF022311	*bozemanii* sg 2	*bozemanii*	5	1
L. brunensis					
PLE-3/2	NA	*brunensis*	*brunensis*	0	0
Species E (IMVS-594)	AF022350	NC	*brunensis*	6	0
KV-11, Ji-646, FS-60, D2683 {Czech}, D3386		*brunensis*	*brunensis*	20	3
L. cherrii					
IMVS-C8H1		?	*cherrii*	0	0
D1486		*cherrii*	*cherrii*	8	1
LC3664, D4744	NA	*cherrii*	*cherrii*	9	1
L. cincinnatiensis					
IMVS-C4B3	NA	*cincinnatiensis*	*cincinnatiensis*	0	0
IMVS-K5B7	AF022358	*cincinnatiensis*	*cincinnatiensis*	3	0
IMVS-K8D7	AF022359	*cincinnatiensis*	*cincinnatiensis*	1	0
IMVS-K8I2	AF047745	*cincinnatiensis*	*cincinnatiensis*	10	3
Dr-68 {UK}, Dr-72 {UK}, 98-085, [LC3936]	AF047746	*cincinnatiensis*, [cross-reactivity]	*cincinnatiensis*	1	0
L. drosanskii (LLAP-1)		NC	new species		
L. dumoffii					
Dr-23, Dr-26, Dr-46	NA	*dumoffii*	*dumoffii*	0	0
IMVS-K7D6, -C7A3, PM-10-702, Dr-34 {Italy}, Dr-57 {NZ}, Sa-3/4, LC0455* {Singapore}	AF022313	*dumoffii*	*dumoffii*	2	0
L. erythra					
IMVS-926, LC0709, LC3719,	NA	*erythra*	*erythra*	0	0
LC1317*, Dr-11 {Italy}	AF047747	*erythra*	*erythra*	3	1
L. fallonii (LLAP-10)		NC	new species		
L. feeleii					
sg 2 ATCC35849, LC4210 {Singapore}, D3971	AF022341	NA	*feeleii*	2	0

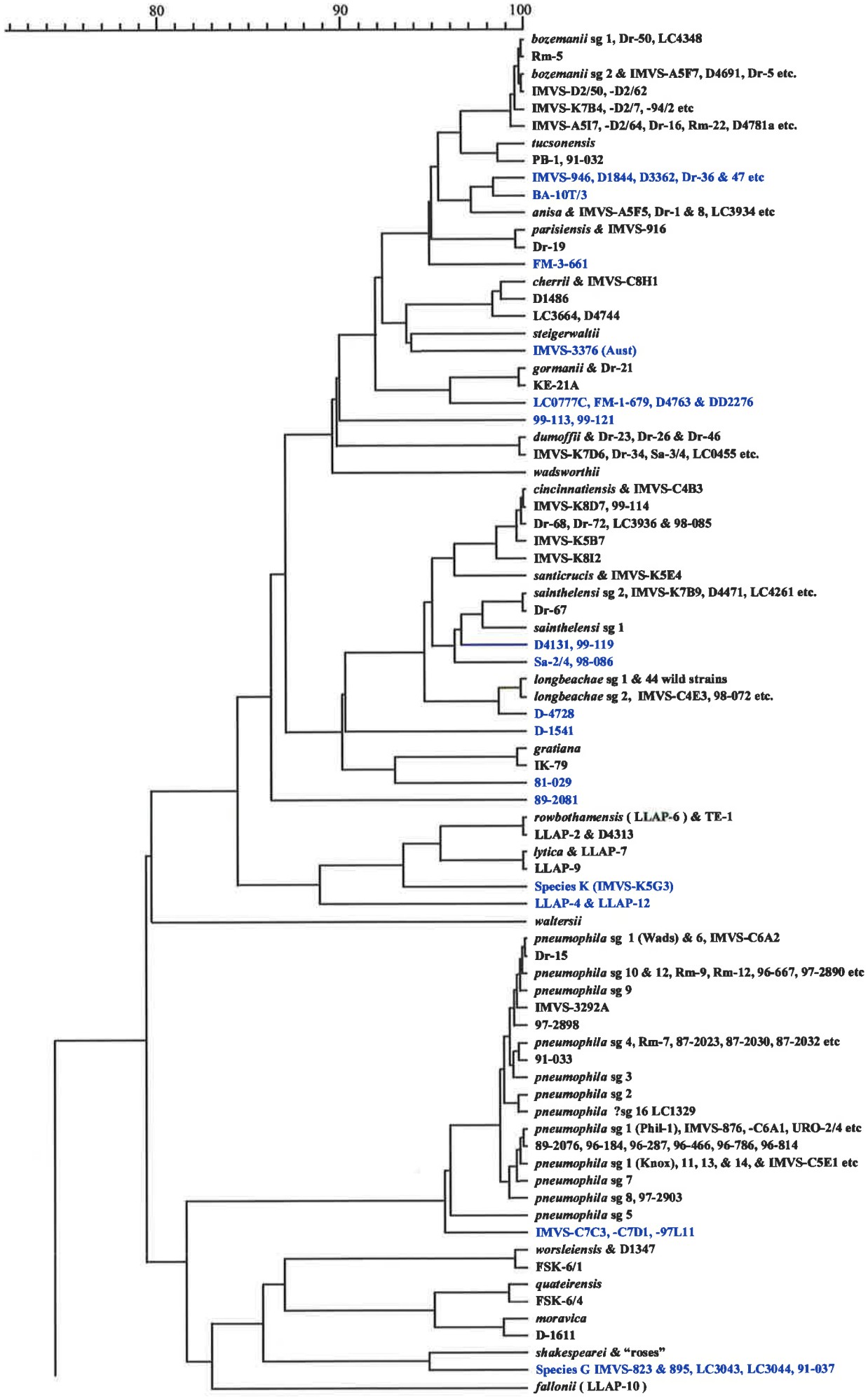
Legionella isolate [a]	GenBank Acc. No. [b]	Phenotypic identification [c]	Sequence identification	No. of differences [d]	
				DNA	AA
D3131		?	*feeleii*	3	0
C9815		?	*feeleii*	10	1
KE-16/1, 96-696		*feeleii* sg 1	*feeleii*	9	1
D4781b {Japan}, [KE-17/2]		*feeleii* [sg 2]	*feeleii*	10	1
IMVS 853, -865, -913, -97/L20, -C7D3, Sa-1/3	AF022340	*feeleii*	*feeleii*	11	1
Species J (IMVS-933)	AF022354	NC	*feeleii*	12	1
L. geestiana					
Int-6		*geestiana*	*geestiana*	2	0
LC3644 {Czech}	NA	*geestiana*	*geestiana*	3	0
L. gormanii					
Dr-21		*gormanii*	*gormanii*	0	0
KE-21A		*gormanii*	*gormanii*	2	0
L. gratiana	NA				
IK-79		?NC	*gratiana*	3	1
L. hackeliae					
sg 2 ATCC35999	NA	NA	*hackeliae*	0	0
L. jamestowniensis					
IMVS-724, -935	AF022339	*jamestowniensis*	*jamestowniensis*	2	0
IMVS-707, -708, -871, -945	AF022337	*jamestowniensis*	*jamestowniensis*	24	2
IMVS-857	AF022338	*jamestowniensis*	*jamestowniensis*	27	4
L. jordanis					
LC3940 {Singapore}, FM-165, Dr-59 {NZ}, Dr-70 {UK}	NA	*jordanis*	*jordanis*	0	0
L. londiniensis					
Species C (IMVS-449), -755, -914, -967, LC4049 {France}, Dr-38 {France}	NA	*londiniensis*	*londiniensis*	0	0
IMVS 912	AF022346	*londiniensis*	*londiniensis*	1	0
L. longbeachae					
IMVS-D1/3, -D1/5, -D1/55, -D1/65, -D2/35, -D2/43, D2/58, -A5F3, -A5G3, -A5G7, -A8H1, -A7E3, -K4E8, -K4F1, -K6G5, -L5F7, -S323, -4399827, -A/24, D63, D493, D880, D1028, D1056, D1620, D1624 (Israel), D1738, D1750, D1751, D1820, D1959, D1992, D4782, Dr-3 {Denmark}, Dr-22 {Italy}, Ly-9981, KE-20[e], 96-003,98-055, 98-066, 98-075, 98-082, 98-083, 98-084	NA	*longbeachae* sg 1	*longbeachae*	0	0
sg 2 ATCC33484, [IMVS-C4E7, 98-072, 98-073]		NA, [*longbeachae* sg 2]	*longbeachae*	2	0
L. lytica					
LLAP-7		NC	*lytica*	0	0
LLAP-9		NC	*lytica*	1	0
L. maceachernii					
IMVS 984 & IMVS-A4H7, LC4349, Mo-7, Dr-12 {Italy}, Dr-65 {Italy}	AF022315	*maceachernii*	*maceachernii*	7	0
IMVS-910, -943, -962, -C8F6, -K8F5, D4340	AF022314	*maceachernii*	*maceachernii*	6	0
L. micdadei					
IMVS-A5E3, -C4F7, -K5A1, -K5D3, -K5E1, -K7B3, Dr-45, D4307 {Israel}, D4310, D4363, D4534, M097-017C1, LC0858 {Italy}	AF023175	*micdadei*	*micdadei*	0	0
IMVS-541, D4309, Dr-14 {Czech}, Rm-2	AF047749	*micdadei*	*micdadei*	1	0
IMVS-K5A5		*micdadei*	*micdadei*	1	0

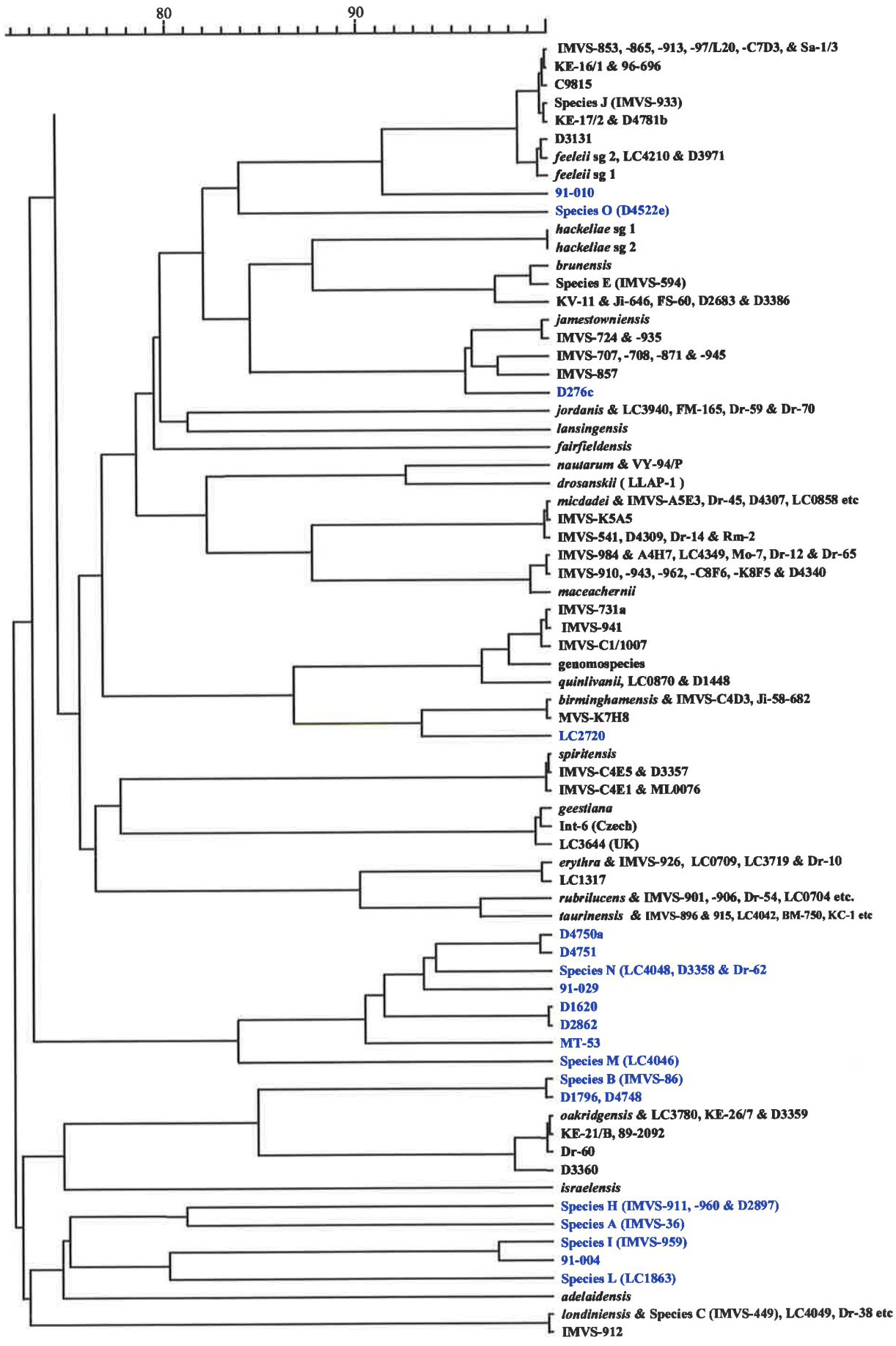
Legionella isolate[a]	GenBank Acc. No.[b]	Phenotypic identification[c]	Sequence identification	No. of differences[d] DNA	AA
L. moravica					
D1611		*moravica*	*moravica*	8	0
L. nautarum					
VY-94/P		*nautarum*	*nautarum*	0	0
L. oakridgensis					
LC3780 {Singapore}, KE-26/7, D3359 {France}	NA	*oakridgensis*	*oakridgensis*	0	0
KE-21B, 89-2092		?	*oakridgensis*	1	1
Dr-60 {NZ}		? *oakr/sainthel/hackeliae*	*oakridgensis*	1	0
D3360 {France}		*oakridgensis*	*oakridgensis*	12	3
L. parisiensis					
IMVS-916, Dr-10 {Spain}	NA	*parisiensis*	*parisiensis*	0	0
Dr-19 {UK}		*parisiensis*	*parisiensis*	3	2
L. pneumophila					
sg 1 Bellingham-1 ATCC 43111	AF022329	NA	*pneumophila*	9	1
sg 1 Knoxville-1 ATCC 33153	AF022332	NA	*pneumophila*	2	0
sg 1 OLDA ATCC43109	AF022335	NA	*pneumophila*	0	0
sg 1 Allentown ATCC43106	AF022330	NA	*pneumophila*	2	0
sg 1 Benidom 030 E ATCC43108	AF022336	NA	*pneumophila*	9	1
sg 1 Camperdown-1 ATCC43113	AF022331	NA	*pneumophila*	0	0
sg 1 Heysham 1 ATCC43107	AF022334	NA	*pneumophila*	0	0
sg 1 Oxford 4032 E ATCC43110	AF022333	NA	*pneumophila*	0	0
sg 1 Wadsworth 130b	Engleberg, *et al.*, 1989	NA	*pneumophila*	9	1
sg 2 ATCC33154	AF022316	NA	*pneumophila*	11	1
sg 3 ATCC33155	AF022317	NA	*pneumophila*	12	1
sg 4 ATCC33156	AF022318	NA	*pneumophila*	11	1
sg 5 ATCC33216	AF022319	NA	*pneumophila*	30	5
sg 6 ATCC33215	AF022320	NA	*pneumophila*	9	1
sg 7 ATCC33823	AF022321	NA	*pneumophila*	3	0
sg 8 ATCC35096	AF022322	NA	*pneumophila*	6	0
sg 9 ATCC35289	AF022323	NA	*pneumophila*	8	1
sg 10 ATCC43283	AF022324	NA	*pneumophila*	8	1
sg 11 ATCC43130	AF022325	NA	*pneumophila*	2	0
sg 12 ATCC43290	AF022326	NA	*pneumophila*	8	1
sg 13 ATCC43736	AF022327	NA	*pneumophila*	2	0
sg 14 ATCC43703	AF022328	NA	*pneumophila*	2	0
IMVS-C7C3, -C7D1 –97L11	AF023173	*pneumophila* sg ?	*pneumophila*/?NC	31	3
IMVS-C6A1, -876, URO-2/4, MO-70/1	NA	*pneumophila* sg 1	*pneumophila*	0	0
89-2076, 96-184, 96-287, 96-466, 96-786, 96-814		*pneumophila*	*pneumophila*	1	0
IMVS-C5E1, -C5E3, -A4C7, -D2/21, [- D1/77], D1909	NA[e]	*pneumophila* sg 1 [sg13]	*pneumophila*	2	0
LC1329	AF047750	*pneumophila* sg 16	*pneumophila*	6	1
IMVS-3292A		*pneumophila*	*pneumophila*	7	3
97-2903		*pneumophila*	*pneumophila*	6	0

Legionella isolate[a]	GenBank Acc. No.[b]	Phenotypic identification[c]	Sequence identification	No. of differences[d]	
				DNA	AA
Rm-9, Rm-12, 96-667, 97-2890		*pneumophila*	*pneumophila*	8	1
Dr-15 {Czech}		*pneumophila*	*pneumophila*	8	1
IMVS-C6A2	NA[f]	*pneumophila* sg 1	*pneumophila*	9	1
97-2898		*pneumophila*	*pneumophila*	9	1
91-033		*pneumophila*	*pneumophila*	10	1
87-2023, 87-2030, 87-2032, 87-2036, 96-163, 96-326, 96-438, 96-817		*pneumophila*	*pneumophila*	11	1
L. quateirensis					
FSK-6/4		*worsleiensis* (atypical)	*quateirensis*	6	2
L. quinlivanii					
LC0870, D1448 (Aus)	NA	*quinlivanii*	*quinlivanii*	0	0
IMVS-C1/1007	AF022347	*quinlivanii*	*quinlivanii*	19	3
IMVS-731a	AF022348	*quinlivanii*	*quinlivanii*	17	3
IMVS-941	AF022349	*quinlivanii*	*quinlivanii*	23	3
genomospecies ATCC51913	AF022356	NA	*quinlivanii*	20	3
L. rowbothamensis (LLAP-6)			new species		
TE-1		NC	*rowbothamensis*	0	0
LLAP-2, D4313		NC	*rowbothamensis*	1	0
L. rubrilucens					
IMVS-901, -906, -939, Dr-54, Dr-63, LC0704, LC0805H*, LC1092*, LC4557	NA	*rubrilucens*	*rubrilucens*	0	0
L. sainthelensi					
sg 2 ATCC49322, IMVS-K7B9, D4471, D1489, [LC4261]	AF022357	*sainthelensi* [sg 2 or *santicrucis*]	*sainthelensi*	14	0
Dr-67 {UK}		*sainthelensi*	*sainthelensi*	1	0
D4131, 99-119		*sainthelensi*	?*sainthelensi*/ ?NC	20	1
Sa-2/4, 98-086		*sainthelensi* sg 1	?*sainthelensi*/ ?NC	24	2
L. santicrucis					
IMVS-K5E4	NA	*santicrucis*	*santicrucis*	0	0
L. shakespearei					
"roses"	NA	?NC	*shakespearei*	0	0
L. spiritensis					
IMVS-C4E5, D3357	AF047751	*spiritensis*	*spiritensis*	1	0
IMVS-C4E1, ML0076	AF047752	*spiritensis*	*spiritensis*	2	0
L. taurinensis					
IMVS 896, -915, Dr-30 {Italy}, Dr-32 {France}, Dr-39 {Italy}, Dr-42 {Czech}, Rm-10, Rm-19, [LC4042 {France}, LC4045 {France}, BM-750, KC-1, KE-9]	AF022342	*rubrilucens*, [?*spiritensis*]	*taurinensis*	0	0
L. tucsonensis					
PB-1, 91-032		*tucsonensis*	*tucsonensis*	9	1
L. waltersii ATCC51914	AF022355				
Species D (IMVS-500, -532)	AF022343	NC	*waltersii*	0	0
L. worsleiensis					
D1347	NA	*worsleiensis*	*worsleiensis*	0	0
FSK-6/1		*worsleiensis*	*worsleiensis*	4	0

Legionella isolate [a]	GenBank Acc. No. [b]	Phenotypic identification [c]	Sequence identification	No. of differences [d] DNA	AA
Species not characterised					
Species A (IMVS-36)	AF022344	NC	NC	160[g]	33[g]
Species B (IMVS-86 ,-934)	AF022345	NC	NC	95[g]	21[g]
D1796[#], D4748 {Sweden}		NC	Species B	2	0
Species G (IMVS-823, 895)	AF022351	NC/?*shakespearei*	NC	33[g]	3[g]
LC3043, LC3044, 91-037					
Species H (IMVS-911, -960, D2897)	AF022352	NC	NC	144[g]	46[g]
Species I (IMVS-959)	AF022353	NC	NC	146[g]	57[g]
Species K (IMVS-K5G3)	AF047753	NC	NC	81[g]	17[g]
Species L (LC1863, {Kenya})	AF047754	NC	NC	123[g]	32[g]
Species M (LC4046* {France})	AF047756	NC	NC	151[g]	54[g]
Species N (LC4048*{France}, D3358, Dr-62 {France})	AF047755	NC	NC	143[g]	43[g]
LC0777C, FM-1-679, D4763 {Sweden}, D2276 {Czech}	AF047748	? *tucsonensis*/*feeleii* sg 2	?*gormanii*/?NC	26[g]	4[g]
LC2720 {Singapore}	AF047744	*birminghamensis*	?NC	41[g]	1[g]
D4522e		NC	NC	100[g]	25[g]
FM-3-661, D2440 {Czech}		NC	NC	28[g]	4[g]
D4728		NC	NC	9[g]	1[g]
D276c[#]		NC	NC	28[g]	3[g]
D1541[#]		NC	NC	57[g]	11[g]
D1620[#], D2862[#]		NC	NC	155[g]	49[g]
D4750a (Canada)		NC	NC	165[g]	51[g]
D4751 (Canada)		NC	NC	154[g]	51[g]
IMVS-3376		NC	NC	37[g]	7[g]
91-028		NC	NC	156[g]	55[g]
81-029		NC	NC	41[g]	7[g]
89-2081		NC	NC	75[g]	18[g]
91-010		NC	NC	57[g]	11[g]
91-004		NC	NC	146[g]	54[g]
99-113, 99-121		NC	NC	56[g]	14[g]
LLAP-4, LLAP12		NC	NC	66[g]	11[g]

Figure 6.1. UPGMA phylogenetic dendrogram of sequence similarities found amongst type and wild strains of *Legionella*. The vertical bar joining 2 isolates or clusters indicates level of similarity. Isolates coloured blue represent potentially novel species.


80
90
100
bozemanii sg 1, Dr-50, LC4348
Rm-5
bozemanii sg 2 & IMVS-A5F7, D4691, Dr-5 etc.
IMVS-D2/50, -D2/62
IMVS-K7B4, -D2/7, -94/2 etc
IMVS-A5I7, -D2/64, Dr-16, Rm-22, D4781a etc.
tucsonensis
PB-1, 91-032
IMVS-946, D1844, D3362, Dr-36 & 47 etc
BA-10T/3
anisa & IMVS-A5F5, Dr-1 & 8, LC3934 etc
parisiensis & IMVS-916
Dr-19
FM-3-661
cherrii & IMVS-C8H1
D1486
LC3664, D4744
steigerwaltii
IMVS-3376 (Aust)
gormanii & Dr-21
KE-21A
LC0777C, FM-1-679, D4763 & DD2276
99-113, 99-121
dumoffii & Dr-23, Dr-26 & Dr-46
IMVS-K7D6, Dr-34, Sa-3/4, LC0455 etc.
wadsworthii
cincinnatiensis & IMVS-C4B3
IMVS-K8D7, 99-114
Dr-68, Dr-72, LC3936 & 98-085
IMVS-K5B7
IMVS-K8I2
santicrucis & IMVS-K5E4
sainthelensi sg 2, IMVS-K7B9, D4471, LC4261 etc.
Dr-67
sainthelensi sg 1
D4131, 99-119
Sa-2/4, 98-086
longbeachae sg 1 & 44 wild strains
longbeachae sg 2, IMVS-C4E3, 98-072 etc.
D-4728
D-1541
gratiana
IK-79
81-029
89-2081
rowbothamensis (LLAP-6) & TE-1
LLAP-2 & D4313
lytica & LLAP-7
LLAP-9
Species K (IMVS-K5G3)
LLAP-4 & LLAP-12
waltersii
pneumophila sg 1 (Wads) & 6, IMVS-C6A2
Dr-15
pneumophila sg 10 & 12, Rm-9, Rm-12, 96-667, 97-2890 etc
pneumophila sg 9
IMVS-3292A
97-2898
pneumophila sg 4, Rm-7, 87-2023, 87-2030, 87-2032 etc
91-033
pneumophila sg 3
pneumophila sg 2
pneumophila ?sg 16 LC1329
pneumophila sg 1 (Phil-1), IMVS-876, -C6A1, URO-2/4 etc
89-2076, 96-184, 96-287, 96-466, 96-786, 96-814
pneumophila sg 1 (Knox), 11, 13, & 14, & IMVS-C5E1 etc
pneumophila sg 7
pneumophila sg 8, 97-2903
pneumophila sg 5
IMVS-C7C3, -C7D1, -97L11
worsleiensis & D1347
FSK-6/1
quateirensis
FSK-6/4
moravica
D-1611
shakespearei & "roses"
Species G IMVS-823 & 895, LC3043, LC3044, 91-037
fallonii (LLAP-10)


80
90
IMVS-853, -865, -913, -97/L20, -C7D3, & Sa-1/3
KE-16/1 & 96-696
C9815
Species J (IMVS-933)
KE-17/2 & D4781b
D3131
feeleii sg 2, LC4210 & D3971
feeleii sg 1
91-010
Species O (D4522e)
hackeliae sg 1
hackeliae sg 2
brunensis
Species E (IMVS-594)
KV-11 & Ji-646, FS-60, D2683 & D3386
jamestowniensis
IMVS-724 & -935
IMVS-707, -708, -871 & -945
IMVS-857
D276c
jordanis & LC3940, FM-165, Dr-59 & Dr-70
lansingensis
fairfieldensis
nautarum & VY-94/P
drosanskii (LLAP-1)
micdadei & IMVS-A5E3, Dr-45, D4307, LC0858 etc
IMVS-K5A5
IMVS-541, D4309, Dr-14 & Rm-2
IMVS-984 & A4H7, LC4349, Mo-7, Dr-12 & Dr-65
IMVS-910, -943, -962, -C8F6, -K8F5 & D4340
maceachernii
IMVS-731a
IMVS-941
IMVS-C1/1007
genomospecies
quinlivanii, LC0870 & D1448
birminghamensis & IMVS-C4D3, Ji-58-682
MVS-K7H8
LC2720
spiritensis
IMVS-C4E5 & D3357
IMVS-C4E1 & ML0076
geestiana
Int-6 (Czech)
LC3644 (UK)
erythra & IMVS-926, LC0709, LC3719 & Dr-10
LC1317
rubrilucens & IMVS-901, -906, Dr-54, LC0704 etc.
taurinensis & IMVS-896 & 915, LC4042, BM-750, KC-1 etc
D4750a
D4751
Species N (LC4048, D3358 & Dr-62
91-029
D1620
D2862
MT-53
Species M (LC4046)
Species B (IMVS-86)
D1796, D4748
oakridgensis & LC3780, KE-26/7 & D3359
KE-21/B, 89-2092
Dr-60
D3360
israelensis
Species H (IMVS-911, -960 & D2897)
Species A (IMVS-36)
Species I (IMVS-959)
91-004
Species L (LC1863)
adelaidensis
londiniensis & Species C (IMVS-449), LC4049, Dr-38 etc
IMVS-912

Chapter Seven

Ecological diversity of *Legionella* strains

Introduction

The availability of gene sequence from multiple strains of a bacterial species has demonstrated that the sequence varies from strain to strain as a result of random mutational errors during DNA replication. However, the permissibility of some substitutions is very much constrained by the functionality of the gene product. Non-synonymous substitutions can only survive in the gene pool if the effect on the function of the resultant protein is neutral or positive. In contrast, synonymous substitutions, most frequently in the third-codon position where the genetic code is more redundant, are not similarly constrained, and so can accumulate at a much higher rate.

The existence of mutational substitutions has lead to the concept of "sequence space", a multi-dimensional visualization of the sequence variation as a cloud or cluster encompassing all of the strains within a species (Ambler, 1996). The sequence variation from strains of related species within the genus also cluster as a sequence "cloud" nearby but not overlapping, with the distance between them approximating the degree of evolutionary relatedness. Based on experimental data, the space between the species clusters is thought to be void of intermediate forms, with the boundaries between clusters demarcated, although some species clusters are "tighter" than others; that is the intra-cluster sequence variation is less (Maynard Smith, 1970, Spratt, *et al.*, 1995). Cohan has proposed a Coalescence Model to explain this

phenomena, as an outcome of natural selection enabling dominant strains within a cluster to periodically purge competing strains from within an ecological niche (Cohan, 1994a, Cohan, 1995, Palys, *et al.*, 1997). The genetic diversity (the sequence cluster) is purged from within the population, at all loci, as only the genome sequence of the dominant strain remains within that population. Other populations inhabiting different niches are unaffected by the diversity purging, but undergo their own periodic natural selection-based genetic purging. Repetitive rounds of mutation and natural selection promotes the distinctness of ecological populations as separate sequence clusters, with the tightness of the sequence cluster reflecting whether mutation or strain purging has been the most recent (See Chapter One and Fig. 1.4). The recent awareness of the possible frequency of genetic exchange has lead to a recent reanalysis of the model to determine if genetic exchange seriously effects the outcome (Majewski and Cohan, 1999). The authors concluded that in certain situations, such as frequent genetic exchange or globally advantageous mutations, it is possible that diversity between two populations might be restricted, but it is not possible to homogenize two sequence clusters into one.

Such a concept leads to a rigorous definition of a species as only those strains within an ecological niche which are competing with each other and can dominate or be dominated by sister strains, i.e. strains from a single sequence cluster. However, for over 20 years, a bacterial species has been somewhat arbitrarily defined as a phenotypically distinct group of strains with at least 70% whole chromosomal similarity, determined by DNA/DNA hybridization (Johnson, 1986, Wayne, *et al.*, 1987). The 70% similarity cut-off value has no scientific basis, other than that its selection empirically gave similar results to bacterial classifications current at the time. As a consequence, it could be predicted that currently defined species could contain strains from more than one sequence cluster, if they are closely related, perhaps reflecting recent divergence from each other relative to the divergence of more distinct strains.

In Chapters Three and Six the *mip* sequence of approximately 400 type and wild strains of *Legionella* were reported, representing 45 recognized species, and several novel species. As would be predicted from Cohan's Coalescent Model, the species type strain sequences were distinct. Wild strains were not available for some species, but for those species where multiple wild strains were available, the sequences determined for those strains formed distinct sequence clusters, although several species contain multiple closely related sequence clusters. However, a surprising phenomenon was observed for *L. longbeachae*. For *L. longbeachae* sg 1, 46 strains isolated from Australia, USA, Europe, Israel and Japan were examined, and found to contain identical sequence for the 614 bp of the *mip* sequence examined. It would be expected that some variation would be present, at least as a result of synonymous change, and especially between strains from diverse geographical regions. That no variation was detected among these strains raises very interesting questions relating to the ecology of this and perhaps other *Legionella* species.

In this chapter, the sequence similarity demonstrated by *L. longbeachae* strains is examined further. Cohan's Coalescence Model predicts that significant similarity would be similarly present in other genes. As described in Chapter Five, the sequence of the *mspA* (*proA*) gene for *L. pneumophila* and *L. longbeachae* has been reported (GeneBank M31884 and X83035 respectively), and suitable primers targeting approximately 450 bp of the gene have been designed for determining the sequence from strains from other species. Sufficient wild strains are available for a number of *Legionella* species, namely *L. longbeachae, L. anisa, L. bozemanii, L. dumoffii, L. feelei, L. micdadei*, and the newly described species *L. taurinensis*, to allow the comparative analysis of sequence similarity within both the *mip* and *mspA* genes for strains isolated from several continents.

Sequencing strategy

The *mspA* primers Msp-F3 and Msp-R2 (See Chapter Five) were assessed for their ability to prime an amplification reaction. When successful, amplicons of approximately the correct size and containing sequence with significant similarity to known *mspA* gene sequences from a much larger selection of wild strains were similarly examined. Representative strains from each of the seven species were selected for *mspA* gene sequencing, based on the *mip* sequence clustering, to include strains from the different countries of origin within each of the *mip* sequence-determined intraspecies clusters. To reduce costs, amplicons were only sequenced in the forward direction with dye-terminator chemistry using Msp-F3 to prime the reactions, unless sequence ambiguity or unusual sequence required confirmation with sequence in the reverse direction. This decision was based on the lack of observed conflicting bases between forward and reverse sequences for the *mip* sequences from the type strains, where both directions had been sequenced. The sequences were analyzed and UPGMA dendrograms produced with GeneCompar 2.0 Advanced Cluster Module. Similarity was determined as the number of identical bases between each sequence pair.

Results

The *mip* sequences for the seven species were compiled from the sequences used to validate the genotyping scheme presented in Chapter Six.

The *mspA*-specific primers produced amplicons of the expected size by PCR for each strain tested. The amplicons were sequenced, and the *mspA* sequence compared for each strain within the seven species being examined. Figs. 7.1 - 7.4 present a comparison of the UPGMA similarity dendrograms for every strain tested for the seven representative species, based on *mip* (left dendrogram, reported in Chapter Four) and *mspA* (right dendrogram) sequence. All

strains included in the *mip* data set are included in each figure. Strains highlighted in blue are those also belonging to the *mspA* data set.

mip *data*

The remarkable *mip* sequence similarity among *L. longbeachae* strains is clearly demonstrated in Fig. 7.1a. Serogroup 1 strains from Australia, Germany, the Czech Republic, Israel, Japan and the USA did not contain a single substitution within the sequence segment examined, among 46 strains. That even synonymous changes are absent implies that this is not a random event. Even though only four *L. longbeachae* sg 2 strains were tested, all four *mip* sequences were still identical to each other and differed by only two substitutions from the sg 1 strains. Strain D4728 is included to complete the clade, but is thought to represent a novel species (see Chapter Six).

An examination of the other species reveals that such absolute similarity is not unique to *L. longbeachae*, although some clades are more complex. The newly described *L. taurinensis* species, represented in this study by 14 strains from Australia, France, the Czech Republic, and Italy, also shows similar absolute *mip* sequence similarity (Fig 7.1b). The *L. anisa* clade (Fig. 7.2a) comprises three internal clades, although BA-10T/3 (clade 3) and probably clade 2 are not *L. anisa* strains but represent novel species (see Chapter Six). However, among the five strains in clade 2, as well as among the 12 strains in the largest clade which includes the type strain (clade 1), the *mip* sequences are identical. The *L. micdadei* clade also comprises three internal clades of 1, 4 and 14 strains, each clade separated by only one base substitution. The *mip* sequences for clades containing multiple strains are identical. Similarly, *L. bozemanii* has four internal clades of 13, 10, 4 and 5 strains respectively, although clades 2 and 3 are further subdivided (Fig 7.3). However, the *mip* sequence within each internal clade or sub-clade containing more than one strain are also identical to each other, even though clades 1, 2b and 3a contain strains isolated in different

countries. *L. dumoffii* contains two internal clades of 4 and 7 strains, and again the *mip* sequence within each is identical, even though strains come from different countries (Fig. 7.4a). Only the *L. feeleii* clade (Fig. 7.4b) shows a degree of sequence diversity, but even then, some *mip* sequences from strains from different countries are identical. In summary, the level of *mip* sequence similarity first evident among *L. longbeachae* sg 1 strains, is also evident within other species, but because the clades contain fewer strains, or are more complex, the absolute intra-clade sequence identity is less obvious. In fact, a re-examination of the classification scheme dendrogram (Table 6.1 and Fig. 6.1) reveals multiple identical sequence from either type or wild strains for *birminghamensis*, *brunensis* (2 clusters), *cherrii*, *cincinnatiensis* (3 clusters), *erythra*, *fairfieldensis*, *gormanii* (3 clusters), *jamestowniensis* (2 clusters), *jordanis*, *londiniensis*, *nautarum*, *oakridgensis* (2 clusters), *parisiensis*, *pneumophila* (7 clusters), *quinlivanii*, *rowbothamensis*, *rubrilucens*, *sainthelensi*, *santicrucis*, *spiritensis* (2 clusters), *tucsonensis*, *worsleiensis* (2 clusters) and Species G, H and N. In virtually every circumstance, the strains sharing identical sequence originated from different countries. Absolute *mip* sequence identity within one or several internal clades for each *Legionell*a species is the norm, not the exception.

mspA *data*

Cohan's Coalescent Model predicts that diversity purging demonstrated within one gene should also be evident in other genes from strains occupying the same ecological niche. This study demonstrates this is so, with a similar level of sequence similarity being evident within the sequence of the *mspA* gene segment examined, although there are some intriguing relationships demonstrated as well. For *L. longbeachae*, *L. taurinensis* and *L. anisa*, the *mspA* sequence dendrograms exactly match that based on the *mip* sequence data (Figs 7.1 and 7.2a), for the strains examined by both analyses. All *L. longbeachae* sg 1 strain *mspA* sequences are identical, as do the two sg 2 strains, although only separated by one base substitution. Similarly, D4728 is relatively distant, separated by 6 and 5 base substitutions from sg 1 and 2

strains respectively. All *L. taurinensis* strains tested contain identical *mspA* sequences, for the segment of the gene examined, as are the sequences from strains in the two larger *L. anisa* clades. Strain BA-10T/3 is also separated from other strains, although the *mspA* sequence is much more similar to the next closest clade than suggested by the *mip* sequence analysis. However, while the absolute sequence similarity determined for the *mip* gene segment is repeated for the internal clades within the *L. micdadei* (Fig. 7.2b), *L. bozemanii* (Fig. 7.3) and *L. dumoffii* (Fig. 7.4a) clades for the *mspA* sequence, several strains (highlighted in red) belong to different clades. To ensure this was not the result of organism, DNA or amplicon contamination, these strains were re-cultured from stock suspensions, and twice re-cultured from single discrete colonies, before the DNA was re-extracted and amplifications and sequencing reactions performed with fresh reagents. Identical sequence was determined for each and compared with that determined previously. Two *L. micdadei* strains, D4363 and D4534, from *mip* clade 2 group in *mspA* clade 1 (Fig 7.2b). Similarly, two *L. bozemanii mip* clade 2 strains, IMVS-D2/50 (*mip* clade 2a) and IMVS-A8E7 (*mip* clade 2b) group in *mspA* clade 1 (Fig 7.3), as does the type strain for *L. dumoffii* (Fig 7.4a).

Once laboratory error has been excluded, there are two possible explanations for such a result, namely convergent evolution or homologous recombination producing lateral gene transfer. While neither can be excluded, convergent evolution is highly unlikely. While the two *L. micdadei* strains would need only one substitution in either gene segment to move clades, the *L. bozemanii* and *L. dumoffii* strains would require at least two, and as many as five base substitutions. For the correct substitutions to occur at the correct sites, by chance alone, is highly unlikely, especially given the absolute level of sequence similarity at other sites, and among strains. The role and frequency of lateral gene transfer within procaryotes has only recently become appreciated (Cohan, 1996, Lan and Reeves, 1996, Strätz, *et al.*, 1996), and is a much more likely explanation for these results. Mechanisms for genetic exchange have been reported for *L. pneumophila* (Dreyfus and Iglewski, 1985, Mintz, *et al.*,

1992), including the *icm/dot* locus (Segal, *et al.*, 1998, Vogel, *et al.*, 1998), and the presence of plasmids in *L. longbeachae* encoding *tra* genes (Doyle, R., 2000, personal communication) but specific examples of horizontal gene transfer have not been reported. These four strains demonstrate the first report of lateral gene transfer for *Legionella*. The detection of evidence for lateral gene transfer does not contradict the evidence for diversity purging, for the donors of the transferred gene elements are most likely to be the dominant strains from closely related niches.

Lastly, the *L. feeleii* clade strains demonstrate a similar level of *mspA* sequence diversity as revealed by the *mip* sequence data, although some of the internal tree topology varies. This is in fact not surprising, as the number of substitutions is often too few for UPGMA or phylogenetic methods to determine the true topology. This also applies to the relationships between some clades implied for the other species. However, the exact inter-clade topology is not important for these analyses.

Discussion

These results strongly support the thesis that most and perhaps all species of *Legionella* obey Cohan's Coalescence Model, and that diversity purging is occurring within the clades within each species clade. That *L. feeleii* strains (and probably *L. jamestowniensis* strains, based on the *mip* sequence data presented in Chapter Four) demonstrate more sequence variability could be evidence that these strains are in the more diverse phase of the repeated cycles of diversity purging (natural selection) and accumulation (mutation). An alternative explanation is that a similar level of sequence similarity to that of other species does in fact exist for each sub-clade, but more strains would need to be examined to reveal the identical sequence clusters within the more complex topology of the *L. feeleii* clade. The observations for other *Legionella* species suggest this explanation is less likely than the former.

Reports of intracluster nucleotide sequence similarity are relatively rare. Roberts and Cohan (1995) investigated sequence variation in natural populations of two *Bacillus* species, but nucleotide substitutions were inferred from restriction site variation. However, they did detect inferred nucleotide sequence identity for strains of both *B. subtilis* and *B. mojavensis* from two geographically isolated regions, namely the Mojave Desert in the USA and the Sahara Desert in Tunisia in Africa. A number of human pathogens have been examined using Multilocus sequence typing (MLST), which uses the composite nucleotide sequence from 450-500 bp internal fragments from as many as seven loci, to identify the different alleles within the population with a discrimination approaching that of MLEE (Enright and Spratt, 1999). Hypervirulent strains (dominant strains in Cohan's Coalescence Model), defined by identical MLST sequence, have been detected for *N. meningitidis, S. pneumoniae, S. pyogenes* and *S. aureus* (Enright and Spratt, 1999, Spratt and Maiden, 1999, Spratt, *et al.*, 1995). These strains are presumed to be disseminated globally by human carriers, to cause epidemics in susceptible human populations (ecological niche). A number of studies have examined the distance between bacterial species, but use inferred amino acid sequences, as the nucleotide sequences are too divergent (Ambler, 1996, Ochi, 1995). These studies target species that are defined as "tight" by Ambler, such as *P. aeruginosa* and *Rhodobacter spaeroides* based on an examination cytochromes, azurin and RNA proteins. No nucleotide sequences similarity analysis is performed, but a search of GeneBank by this researcher reveals that identity at the nucleotide level similar to that for *mip* and *mspA* is not present.

The results presented in this study for *L. longbeachae* are in conflict with the conclusions of a recent report that Australian *L. longbeachae* strains are not clonal (Montanaro-Punzengruber, *et al.*, 1999). However, the authors were using PFGE to discriminate between strains, and in many cases differences between strains were the result of only one or two fragment differences. This technique is a much more sensitive measure of the

variation within the total cellular DNA, and includes mutations, recombination events, and probably the presence of plasmids and perhaps even phage DNA. Such events cause small modifications to the cellular DNA much more frequently than the purging effects of strain domination. As a consequence, it is not surprising that variation can be detected in *L. longbeachae* strain PFGE patterns, even though the underlying clonality as a result of diversity purging is very marked. It is also imperative that where dramatic differences occur, the species designations of strains are confirmed acurately to ensure an error has not been made.

In addition, differences have been observed in the virulence of *L. longbeachae* strains in guinea pigs, using aerosol inoculation (R. Doyle, 2000, personal communication). At least three virulence types were detected; avirulent, moderately virulent and highly virulent. However, extra-chromosomal elements were also detected in many strains by the same worker. The contribution of these elements to the virulence potential of *L. longbeachae* strains is not yet known. It would be consistent with the findings reported here, and the variation in PFGE patterns discussed above, if such extra-chromosomal elements were significantly contributing to the virulence of *L. longbeachae* strains in mammalian hosts, overlaying the mono-phyletic nature of the species' strains.

If *Legionella* conform to Cohan's Coalescence Model, two very interesting questions arise. Firstly, given that diversity purging necessarily requires a strain to dominate an ecological niche, then how is it possible that the same dominant strain is present not only in different regions and countries, but also in different continents? This infers that the strains are migrating globally, so that the same strain dominates similar ecological niches across the world. It also infers a migration vehicle to transport the strains. Although water-borne, no evidence of long-term legionellae survival in sea water exists, and even if valid, it still does not explain how legionellae can get from the sea to inhabit freshwater environments far

inland. It is also unlikely for such migration to be based on human activity or travel, as the ecological diversity within the *Legionella* genus is too substantial to have developed solely during the last few hundred years, the period of frequent global movement of people and goods. It is also highly unlikely that free living legionellae would survive the desiccation which would occur by being carried between continents in wind currents.

The most likely vehicle is one originally proposed as the infectious "invasion unit" in human disease by Shuman, *et. al.* (1998), namely protozoa infected with legionellae being carried in the wind currents. Although protozoal trophozoites would similarly be susceptible to desiccation, protozoa can encyst to produce an environmentally protective "migration vehicle" in which legionellae could survive long enough to be transported such long distances. Viable legionellae have been observed within protozoan cysts (Adeleke, *et al.*, 1996), and cysts of protozoa known to host legionellae have been detected in air currents. Rogerson and Detwiler (1999) detected protozoan cysts in near-surface air of South Dakota, USA, in rates as high as 1.08 cysts m^{-3}, and noted that the numbers were highest on days of high wind and lower relative humidity. In an earlier study on the role of airborne transmission of pathogenic amoebae causing amoebic meningoencephalitis in Nigeria during the dusty Harmattan period (a hot, dusty wind blowing from the Sahara in January and February), Lawande (1983) was able to culture a total of 38 strains of amoebae on settle plates exposed 10 m above the ground for various periods of up to four hours. Among the strains of amoebae cultured there were many known to support the intracellular growth of legionellae. In fact the possibility of amoebic cysts being lifted and carried in air currents, and then inhaled to cause amoebic infection was first suggested over 60 years ago (Hewit, 1937).

Although no evidence exists for inter-continental movement of amoebic cysts, circumstantial evidence exists that suggests that it is possible. Firstly, Robinson has noted that among the *Acanthamoeba* found in Antarctica, most of which have never been classified, nor

phylogenetic relationships determined, there are two distinct populations, varying in their temperature adaptation. Some strains are cold adapted, and may represent a resident *Acanthamoeba* population. The other population is not cold adapted, and respond to temperature similar to *Acanthamoeba* from temperate regions. These strains may represent a migrant population, disbursed from more temperate regions (Dr. Brett Robinson, Australian Water Quality Centre, South Australia, 2000, personal communication).

In addition, evidence exists that other biotics, such as plant pollen and even miniscule spiders can traverse oceans in high-altitude wind currents. Airborne plant pollen and fungal spores from South America have both been detected in Antarctica during air sampling (Marshall, 1996, Marshall, 1997). Further, the peak concentrations of such biotics coincided with the seasonal arrival of the air mass from South America. In addition, pollen grains have been collected from the atmosphere above the mid-Atlantic Ocean (Erdtman, 1937 in Rogerson and Detwiler, 1999), and fungal spores from the Arctic atmosphere above 1000 metres (Meier, 1935 in Rogerson and Detwiler, 1999).

Secondly, and perhaps even more convincing, is the current debate on the ecological diversity of protozoa. One group of workers extrapolate from studies of ciliated protozoa to propose that the number of species worldwide is in fact quite modest, and cite the presumed rapid transport of protozoal cysts in air as one reason to propose that protozoa (and smaller, more numerous species, including bacteria) are globally ubiquitous, global diversity relatively low, and at least in the case of ciliated protozoa, most species have already been described (Finlay, 1998, Finlay and Esteban, 1998). They propose that because of the efficient dispersal, microbes should occupy similar habitats in geographically isolated locations throughout the world, and that most strains are cosmopolitan in that they occupy appropriate similar habitats on different continents. The results presented here are completely consistent with this view. However, it would be difficult to absolutely prove global migration of legionellae. Culture or

PCR detection of legionellae from air samples from the upper atmosphere with sufficient consistency to be confident the results are not the consequence of laboratory contamination would be very expensive and time consuming. However, the sequence similarity among *Legionella* strains is very strong circumstantial evidence that global migration of legionellae, probably within protozoan cysts, occurs continuously, enabling each newly-emergent dominant strain to inhabit similar ecological niches across the globe.

A similar finding was made by Zwart, *et al.* (1998) when studying the sequence variation found among *16S rRNA* sequences from bacteria in fresh water lakes. High sequence similarity (97-100%) was found in *16S rRNA* sequences of strains forming several clusters from the α and β subdivisions of the Proteobacteria found in three geographically isolated lakes in the Netherlands, and in Alaska and New York State in the USA. The classification and life cycle of the bacteria detected is unknown as they have never been cultured. The authors cite the level of sequence identity among these organisms as evidence of the global disbursal, and that they are able to successfully compete in a wide variety of fresh water environments. However, endosymbiosis is common within both subdivisions of Proteobacteria. Endosymbiotic examples in the α-Proteobacteria include *Caediacter caryophilia* and the ciliate *Paramecium* (Springer, *et al.*, 1993) and many α-Proteobacteria with arthropods (Wenseleers, 1999; http://www.kuleuven.ac.be/bio/ento/wenselrf.htm), and for β-Proteobacteria, *Crithidia* Spp. and the trypanosomatid protozoa *Blastocrithidia culicis* (Du, *et al.*, 1994). The possibility remains that disbursal of these fresh water Protobacteria may be facilitated by a symbiotic host.

The second very interesting question to arise out of the implications of *Legionella* strains conforming to Cohan's Coalescence Model is what is a true *Legionella* species? The model defines a species as a cluster of related strains which form an ecologically distinct unit

because of the effects of the repetitive rounds of diversity purging and accumulation (Palys, *et al.*, 1997). Such units conform to sequence similarity clusters where the average divergence between strains of different clusters is far greater than the average divergence between strains of the same cluster. This approach at classifying bacterial diversity is supported by current evolutionary theory, unlike the arbitrariness of a particular level of total chromosomal hybridization. In fact, some evidence exists that the 70% cut-off does not necessarily yield groups of strains which conform to real ecological units (Vauterin, *et al.*, 1995). A review of the basis of speciation of strains within *Legionella*, based on this model, would subdivide many current species classifications in to two and often more "true", ecologically distinct, species. The clinical relevance of such an approach is uncertain, and may simply add to an already confusing number of species. However, it may enable the detection of particular highly infectious "clades" from the background of environmental strains unlikely to cause disease.

Figure 7.1. UPGMA dendrograms of (a) *L. longbeachae* and (b) *L. taurinensis* strains, based on sequence from the *mip* (left-hand dendrogram) and *mspA* (right-hand dendrogram) genes. Strains highlighted in blue were examined in both sequence sets. D4728 has been included with the *L. longbeachae* strains although it is novel to complete the clade and to demonstrate that the divergence from *L. longbeachae* determined by the *mip* gene sequence is mirrored by the *mspA* gene sequence.

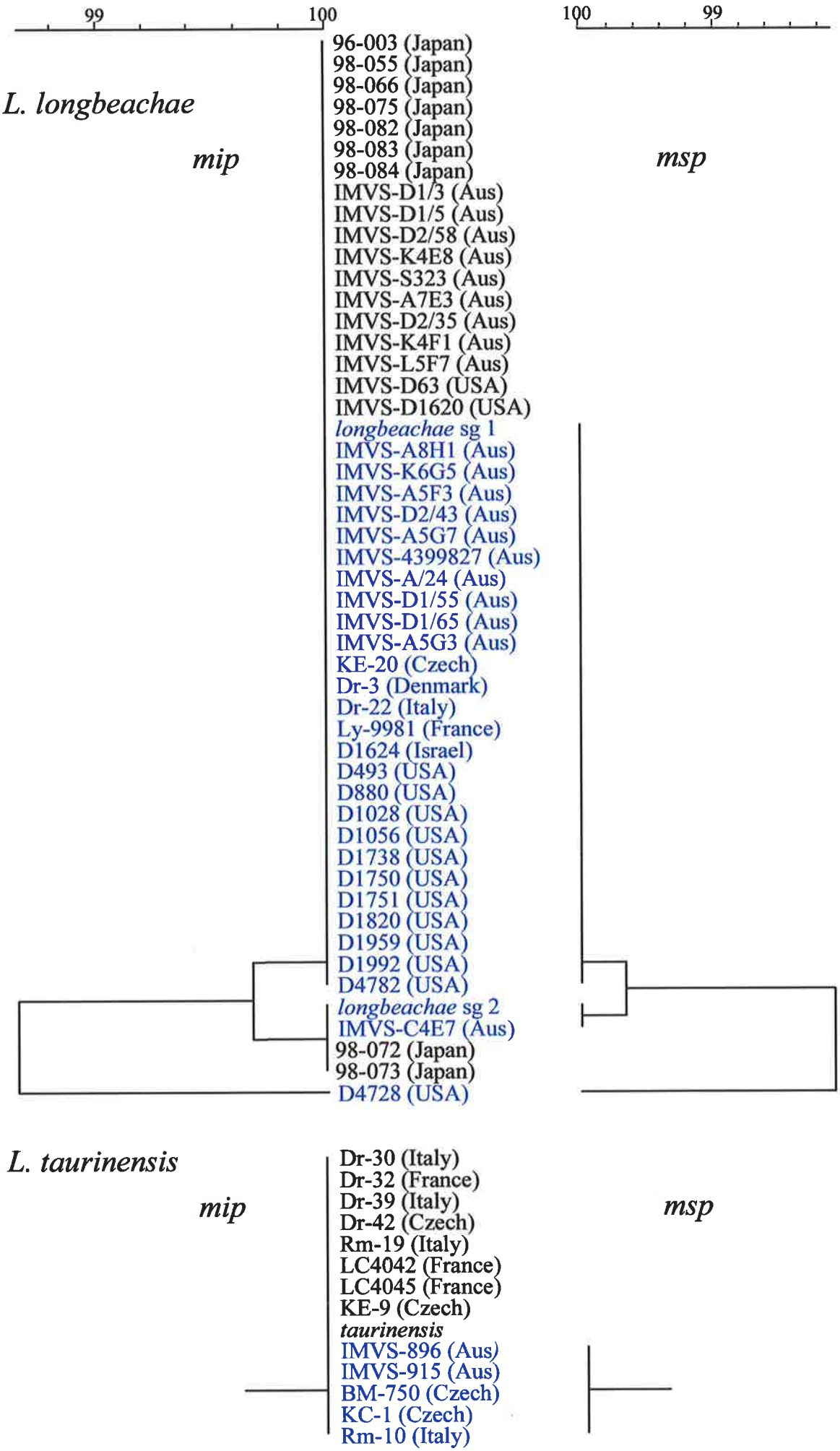
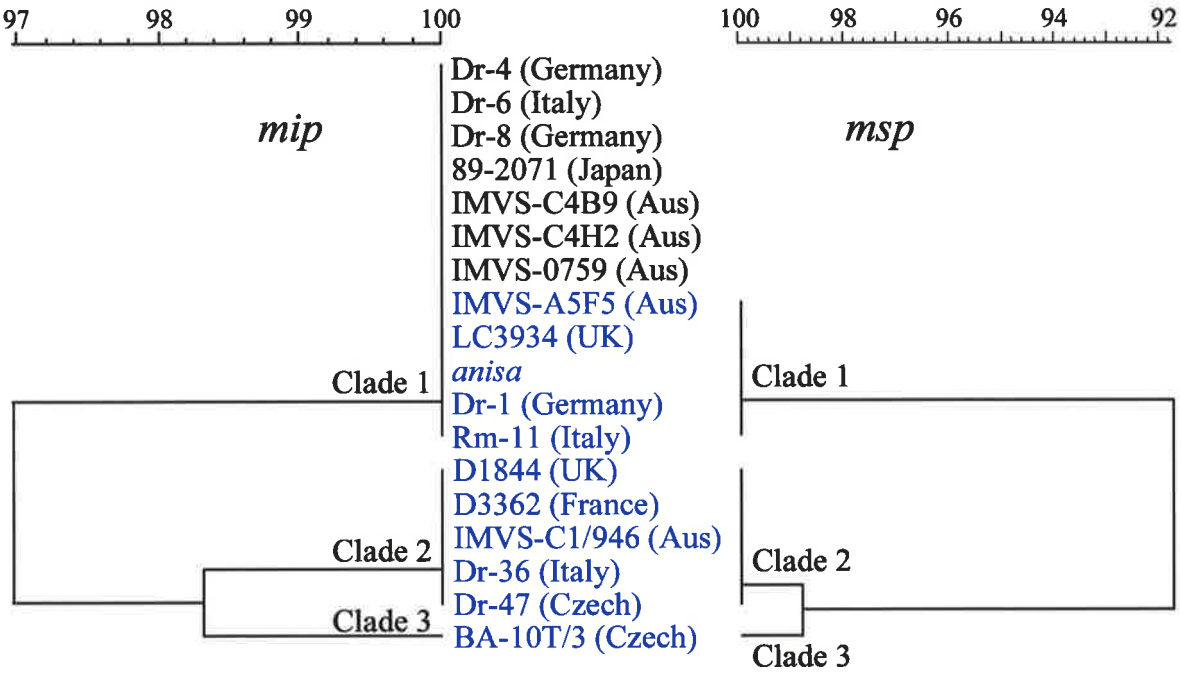

99
100
100
99
L. longbeachae
mip
msp
96-003 (Japan)
98-055 (Japan)
98-066 (Japan)
98-075 (Japan)
98-082 (Japan)
98-083 (Japan)
98-084 (Japan)
IMVS-D1/3 (Aus)
IMVS-D1/5 (Aus)
IMVS-D2/58 (Aus)
IMVS-K4E8 (Aus)
IMVS-S323 (Aus)
IMVS-A7E3 (Aus)
IMVS-D2/35 (Aus)
IMVS-K4F1 (Aus)
IMVS-L5F7 (Aus)
IMVS-D63 (USA)
IMVS-D1620 (USA)
longbeachae sg 1
IMVS-A8H1 (Aus)
IMVS-K6G5 (Aus)
IMVS-A5F3 (Aus)
IMVS-D2/43 (Aus)
IMVS-A5G7 (Aus)
IMVS-4399827 (Aus)
IMVS-A/24 (Aus)
IMVS-D1/55 (Aus)
IMVS-D1/65 (Aus)
IMVS-A5G3 (Aus)
KE-20 (Czech)
Dr-3 (Denmark)
Dr-22 (Italy)
Ly-9981 (France)
D1624 (Israel)
D493 (USA)
D880 (USA)
D1028 (USA)
D1056 (USA)
D1738 (USA)
D1750 (USA)
D1751 (USA)
D1820 (USA)
D1959 (USA)
D1992 (USA)
D4782 (USA)
longbeachae sg 2
IMVS-C4E7 (Aus)
98-072 (Japan)
98-073 (Japan)
D4728 (USA)
L. taurinensis
mip
msp
Dr-30 (Italy)
Dr-32 (France)
Dr-39 (Italy)
Dr-42 (Czech)
Rm-19 (Italy)
LC4042 (France)
LC4045 (France)
KE-9 (Czech)
taurinensis
IMVS-896 (Aus)
IMVS-915 (Aus)
BM-750 (Czech)
KC-1 (Czech)
Rm-10 (Italy)

Figure 7.2. UPGMA dendrograms of (a) *L. anisa* and (b) *L. micdadei* strains, based on sequence from the *mip* (left-hand dendrogram) and *mspA* (right-hand dendrogram) genes. Strains highlighted in blue were examined in both sequence sets. Strains highlighted in red are located in different clades by the two sequence sets.

L. anisa



L. micdadei

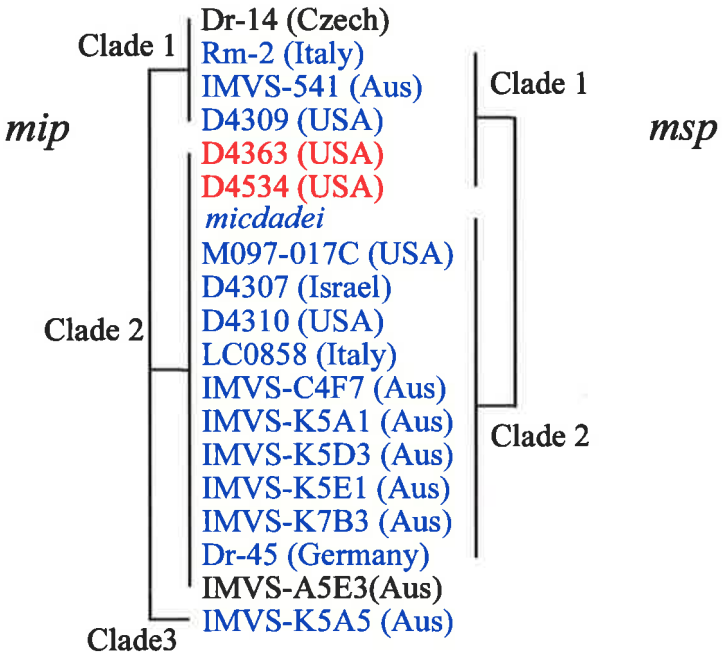
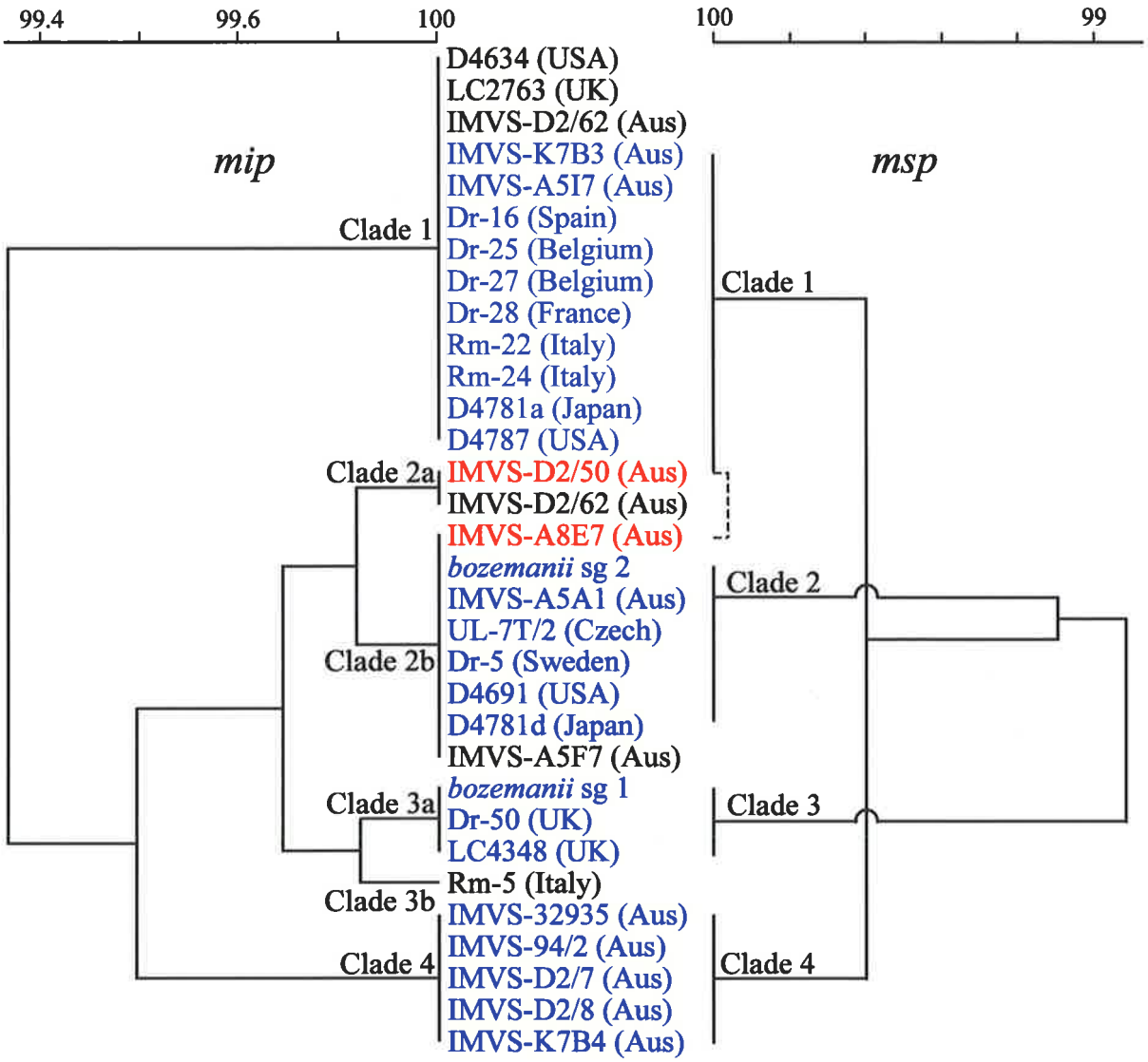


Figure 7.3. UPGMA dendrograms of *L. bozemanii* strains, based on sequence from the *mip* (left-hand dendrogram) and *mspA* (right-hand dendrogram) genes. Strains highlighted in blue were examined in both sequence sets. Strains highlighted in red are located in different clades by the two sequence sets. The table of nucleotide bases for each sequence set represents all the sequence positions exhibiting base substitutions for each of the clades, on which the dendrogram topology is based.

L. bozemanii

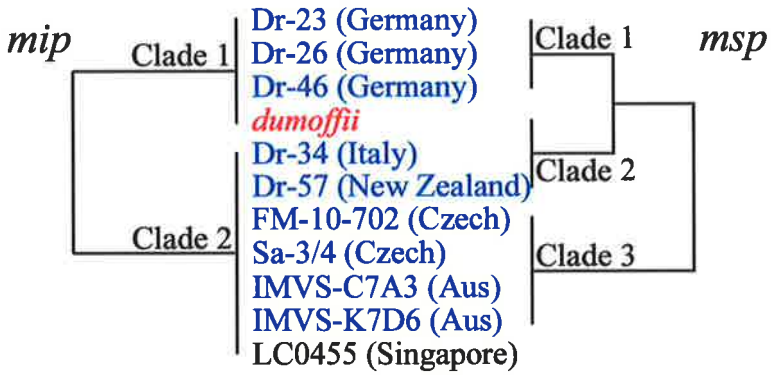


Alignment position	66	99	324	330	394	453	489	556	597
Clade 1	C	A	T	C	A	C	T	C	A
Clade 2a	T	A	T	C	A	T	C	T	A
Clade 2b	T	A	C	C	A	T	C	T	A
Clade 3a	T	A	C	C	G	C	C	T	A
Clade 3b	T	A	C	C	A	C	C	T	A
Clade 4	T	G	C	T	A	T	C	T	C

Alignment position	29	71	95	110	242	260	278	308
Clade 1	T	G	T	C	C	C	A	T
Clade 2	C	A	G	C	C	T	A	C
Clade 3	C	A	T	T	C	C	G	C
Clade 4	C	A	T	T	T	C	A	C

Figure 7.4. UPGMA dendrograms of (a) *L. dumoffii* and (b) *L. feeleiii* strains, based on sequence from the *mip* (left-hand dendrogram) and *mspA* (right-hand dendrogram) genes. Strains highlighted in blue were examined in both sequence sets. The table of nucleotide bases for each sequence set represents all the sequence positions exhibiting base substitutions for each of the clades, on which the dendrogram topology is based.

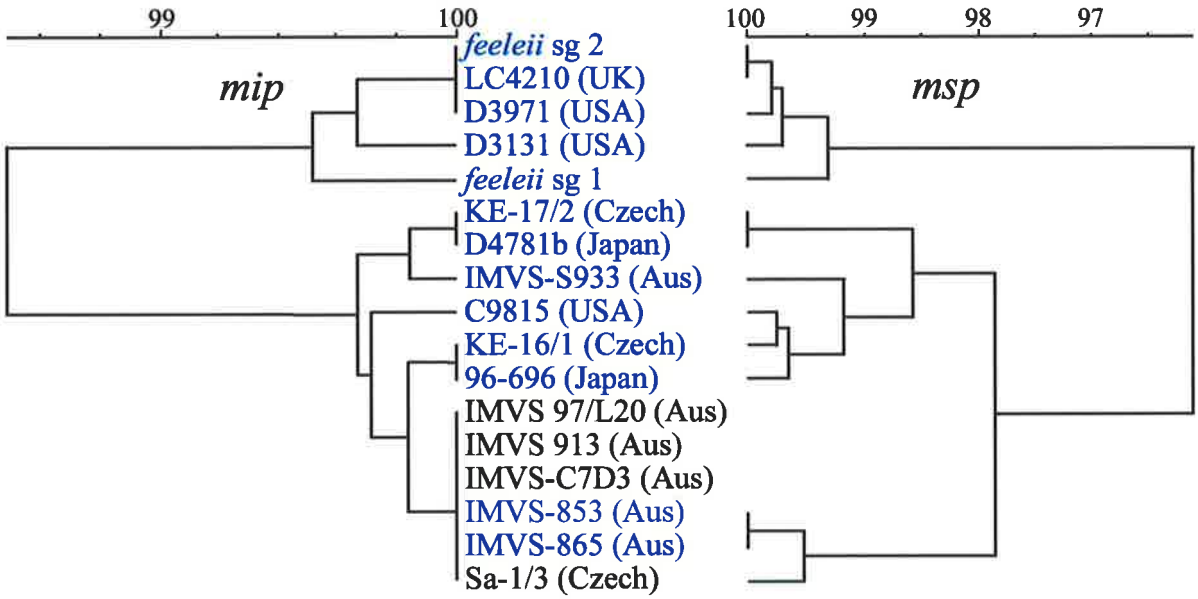
L. dumoffii



Alignment position	447	504
Clade 1	A	T
Clade 2	G	C

Alignment position	212	382	409
Clade 1	A	T	T
Clade 2	C	T	T
Clade 3	C	C	C

L. feeleii



Chapter Eight

Summary Discussion

This thesis presents information which was either determined or inferred from *mip* and *mspA* (*proA*) genes from the genus *Legionella*, to further understanding of this important intracellular pathogen. In total, sequence from approximately 400 type and wild strains of *Legionella* was determined and analyzed. The principal target was the *mip* gene, coding for a ~24-kDa outer-membrane protein. At the commencement of this study, the Mip protein from *L. pneumophila* was the only protein from *Legionella* with a documented association with virulence. For comparison, a segment of the *mspA* gene from a select number of strains and species was also sequenced, to further evaluate the information obtained from the *mip* gene sequences. In addition, the published *16SrRNA* gene sequences from 38 type strains was also included in some analyses.

Functional analysis

Firstly, the inferred amino acid sequence of Mip from 38 type strains was aligned and compared, to determine if there was a difference in the sequence from those species associated with disease, and those which had only been isolated from the environment. While differences were present, the overall level of amino acid homology in the inferred sequence alignment was very high, ranging from 82 – 99%. Furthermore, all residues known to be involved in PPIase activity (the only enzyme function determined for the protein) were totally conserved, as well as large regions in the C terminal region where the PPIase active site is located.

Because of the level of conservation found in these regions, it can be inferred that they also are functionally important for PPIase activity. A hyper-variable insert of up to 17 amino acids was determined immediately adjacent to the signal sequence. The function of this region is unknown, but as it was found in species with and without an association with disease, it can not be specifically related to virulence. However, there were regions of near or absolute homology in the N terminal region which must be functionally important. As X-ray crystallography has determined that the contact points between the monomers occurs in this region, it is reasonable to infer that these homologous regions are highly likely to be associated with the maintaining the dimeric form of the active protein.

Some workers have postulated that the PPIase function is not the main function of this protein in intracellular life (Abu Kwaik, *et al.*, 1998a, Wintermeyer, *et al.*, 1995). However, similar levels of reduction in invasion rates for the *mip* mutant used, which was solely deficient in PPIase activity, were in fact observed, compared to that determined by Cianciotto, *et al.* (1989b and 1990b) who first proposed a role in virulence for the Mip protein. To date there is no evidence for an additional enzymic function in Mip, and it is likely PPIase activity is its main and perhaps only function. However, the exact role for Mip within the intracellular life-cycle of *Legionella*, in both amoebae and human phagocytic cells, remains to be elucidated.

Phlyogeny

Secondly, the same set of sequence data has been used to infer the evolutionary history of the species within the genus. A very conservative approach was chosen, as it is very easy to generate apparently reasonable, but phylogenetically unsupported tree topology simply by analyzing the data with one of the many computer phylogenic software programs with no regard for the nature of the data being examined and whether it conforms to the assumptions being made by the algorythm chosen to infer the evolutionary relationships. After examining

the aligned sequence data for the relevant parameters, both NJ and MP methods were chosen, and the topology from each was compared to that based on published *16S rRNA* sequences, by the same two methods. Because the third codon position in the *mip* sequence was so saturated with substitutions that their rate of accumulation was no longer linear, the analyses were also performed with substitutions in this position excluded. This is similar to performing the analysis on the inferred amino acid sequences. Because the survival of non-synonymous substitutions in the gene pool is significantly constrained by the protein's function, thereby reducing the rate of accumulation of substitutions to often below levels which would become saturated, phylogenic topologies inferred by substitutions in only the first and second codon positions are thought to be more robust for ancestral linkages. For the same reasons of function, *rRNA* gene sequence is similarly thought to be robust in predicting ancestral links. For the data presented here, the robustness of the nodes determined by the analyses was assessed with bootstrap values. The values obtained for the ancestral nodes confirmed that ancestral topology was being inferred by too few substitutions, for all the data sets. That there was also little consensus among the trees from the various analyses for the ancestral topologies confirms that the data was inadequate in predicting ancestral linkages.

However, some clades were consistently predicted by all the analyses. The main clade, called the *anisa – worsleiensis* clade, contained 19 taxa, and while some internal topology was consistently predicted and supported with significant bootstrap values, not all of the topology was clear, even for the two analyses (MP and NJ) on the same data set. Ultimately very few of the 36 nodes were strongly resolved by both data sets. Four other clades were consistently predicted by all analyses, being pairs of very closely related species.

A comparison of the resolved nodes with phenotypic and biochemical characters demonstrates the unreliability of using a single phenotypic or biochemical character to infer relationships. Fatty acid and ubiquinone profiles, still in current use to classify legionellae,

while sharing a crude similarity with the resolved nodes, never the less locate some strains in the wrong groupings e.g. *L. wadsworthii* grouped with *L. oakridgensis*. Many of the groupings for these characters are arbitrary, such as a major proportion of one of the fatty acids or ubiquinones, chosen empirically, with no basis in any evolutionary theory. In the above example, even though *L. wadsworthii* is grouped with *L. oakridgensis* because both have major amounts of Q10, *L. wadsworthii*, unlike *L. oakridgensis* also has significant amounts of Q9, Q11, and Q12, as do those strains with which it is genetically related.

Further analysis of the anisa – worsleiensis *clade*

To further resolve the *anisa – worsleiensis* clade, a 438 bp segment of the *mspA* gene was determined for the 19 species, and included in the phylogenetic analysis. The *mspA* data set was determined to be sufficiently homogeneous with the *mip* data set, that the two could be combined for a more informative analysis. For this analysis, because the number of taxa was much less, and also the availability of a more powerful computer, a ML analysis could also be performed, in addition to NJ and MP methods. Further, resolution of the *anisa – worsleiensis* clade was possible, but not all nodes were able to be resolved. Fig 5.4 represents the summary of relationships receiving consistent support from all of the analyses.

Classification scheme

Because of the resolution of the *mip* sequence data at the nucleotide level, and absence of any apparent evidence of lateral gene transfer, the *mip* sequence was able to be used to assist in the classification of *Legionella* strains, and to genotype unknown legionellae isolates. A universal primer set was determined which would amplify approximately 90% of the *mip* gene (661 – 715 bp, depending on the presence and size of the hypervariable region), with the exception *of L. geestiana*. Only a small fragment of 194 bases has been able to determined for *L. geestiana*, in spite of repeated attempts with gene amplification and cloning. However, this

species is rarely isolated, and suspected *Legionella* isolates not amplifying with the universal primer set can be confirmed as *L. geestiana* with a second primer set.

A forward sequencing primer Legmip_fs was designed to overlap the forward amplifying primer Legmip_f, but off-set by 6 bases into the amplicon at the 3' end. This primer proved so good at priming the dye terminator sequencing reaction to produce clean, unambiguous sequence for the length of the amplicon, even if contaminating amplification products were also present, that routine classification of unknown legionellae strains only need be sequenced with this primer.

Using the universal Legmip primer set, all type strains produced an amplicon of the correct size, from which unambiguous sequence could be determined using the sequencing forward primer. The resolution of all of the species type strains was excellent, with the two closest characterised species, *L. bozemanii* and *L. tucsonensis*, being easily resolved (3.6% sequence variation). The scheme was validated with strains from as many countries as possible, and wild strains were obtained for every species except *L. gratiana*, *L. hackeliae*, *L. lansingensis L. steigerwalti* and *L. wadsworthii*. All wild strains grouped with the relevant type strain, with the exception of some strains identified by serotyping alone. As serological cross-reactions is well documented, this was not considered a problem. In fact, many times the originating laboratory re-assessed their identification, and either agreed with the genotype classification, or could not unambiguously classify it as another species due to serological cross-reactions. The most compelling evidence for the accuracy of the *mip* genotyping scheme is that every wild strain identified by a genetic method such as *rRNA* sequence, ribotyping or DNA/DNA hybridisation was similarly identified by the *mip* genotyping scheme.

In addition to wild strains of known species, strains though to represent novel species were also assessed by the scheme, and their uniqueness confirmed. The exact number novel

species represented by these strains is not yet clear, as many type close to known strains. Their true identity requires clarification by a complete assessment by at least most of the methods currently used to describe a new species. However, to date 32 potentially novel species have been detected, including LLAPs, some represented by multiple strains isolated in more than one continent, and many already have corroborating evidence of their novel classification from other methods, including DNA-DNA hybridization.

LLAPs are also able to be classified by this scheme, and the inclusion of *mip* genotyping results in the recent formal publication of three new novel species, namely *L. drosanskii*, *L. rowbothamensis*, and *L. fallonii* serves to highlight the validity with which the method is recognized by other specialists in the field. At least one other LLAP novel species has been confirmed by this scheme, comprising LLAP-4 and LLAP-12, and other strains thought to represent novel species exist. The inclusion of the *mip* genotype in the formal publication of these species has been discussed.

Ecological diversity

The last component of the work covered in this thesis really only became evident when the genotype of each of the many wild strains was determined. Most evident for *L. longbeachae* sg 1 strains, but shown in this thesis to be the norm, and not the exception, the *mip* sequence for most strains belonging to an ecological cluster have been determined to be identical. That sequence diversity is periodically purged from an ecological cluster when a dominant strain arises which out competes other strains, has been predicted by Cohan's Coalescence Model. However, what is very surprising is that many of these strains possessing unique sequence have been isolated in different regions, countries and continents, implying that progeny of the same selectively advantaged strains dominate similar ecological niches globally. The *mip* sequence identity for a selected number of these strains was confirmed by sequence from a segment of the *mspA* gene, also supporting the prediction of Cohan's

Coalescence Model that diversity is purged from every gene. However, five strains are located in different sub-clades by the two sequence data sets. This is interpreted as evidence for horizontal gene transfer, increasingly being recognized as a common phenomena in procaryotes. Although conjugation mechanisms are present in *Legionella*, specific examples of chromosomal gene transfer have not yet been reported for this genus, although it must certainly occur.

To account for the global domination of certain selectively advantaged strains, it is proposed that legionellae migrate within, and are thus protected from desiccation by, amoebic cysts. Legionellae within amoebic trophozoites, cysts or expelled vesicles has already been proposed as the "infectious unit" in human disease to account for the infectious dose paradox. Amoebic cysts have also been in air currents as much as 10 metres above the ground, and airborne carriage of amoebae has been suggested to account for the rise of amoebic meningitis during the harmattan season in Nigeria. No evidence yet exists for inter-continental carriage of amoebic cysts. However, other biotics such as plant pollen and fungal spores have been detected in the air mid-ocean, and being carried to the Antarctic from South America. In addition, evidence exists that the global diversity of amoebae may be low, consistent with this proposal of global amoebic migration. Thus there exists significant corroborating evidence that legionellae can migrate globally. No alternative hypothesis satisfactorily accounts for the level of sequence identity detected here within ecological units.

These results also raise significant implications as to the true nature of a *Legionella* species. Based on the evolutionary genetic theory underlying Cohan's Coalescence Model, each ecological unit is the true species. This implies that the number of true *Legionella* species is significantly more than currently recognised. A reassessment of the speciation of strains in the various sub-clades determined here would clarify this issue further.

This study has highlighted several issues which should be investigated further. Firstly, the phylogenetic analysis of the genus has left a number of nodes unresolved, such that the evolutionary history of some species remains indefinite. Sequencing additional genes may resolve the evolutionary relationships of these species further, although the current sequence data suggests that quite a number of genes will need to be analysed. This is however consistent with the trend proposed by Enright and Spratt (1999) to use MLST techniques targeting multiple genes (seven in strains analysed by them, but potentially many more in the future) to type bacterial strains. As the length of sequence able to be recovered by a sequencing reaction increase, and the costs reduce, the use of MLST will increase. The availability of sequence from many more genes and many more strains will make the further phylogenetic analysis of *Legionella* possible.

The sequence-based classification scheme proposed by this study will also benefit from the trend towards using MLST. As resolving as the classification scheme appears, it still relies on the data from one gene, and is thus potentially susceptible to the effects of horizontal gene transfer. The evidence of horizontal gene transfer detected in the study now makes this a real rather than a theoretical issue, although the actual examples would have not caused misclassification. The increase in confidence which will result from genotyping strains with sequence from many genes will make MLST obligatory for good laboratory practice, once the technique is practical and cost effective. It is possible that MLST will bring to light significant examples of horizontal gene transfer, further emphasising the validity of this approach. During the study, nearly 30 potentially novel species have been identified, and while attempts were made to obtain the novel strains from many laboratories across the world, including the large culture collections, it is still likely that more exist. The speciation of all of these strains by the criteria traditionally used for *Legionella* is a daunting task. However, if it is not possible for all to be properly classified, then al least those novel species for which there are multiple strains from different geographical regions should be formally classified. Discussions

have already taken place with the laboratories where the novel strains were isolated, and with several laboratories in Europe and the USA possessing the relevant expertise, to initiate the task of formal classification of these novel species.

The examination of intraspecies strains reported in this study has revealed a surprising level of sequence identity within many species. Extending this analysis to include strains from species not examined in detail, and additional strains of *L. feeleii* and *L. jamestowniensis*, will clarify the issue of diversity purging within these clusters further. Attempts to recover legionellae from air-borne amoebic cysts, while potentially difficult, would also clarify the remarkable possibility of global "hitch-hiking" as the dispersal mechanism. Given that increasing numbers of bacterial species are being implicated in endosymbiotic associations with protozoa, the issues of dispersal and diversity purging within these species also requires elucidation.

Bibliography

Abu Kwaik, Y. 1996. The phagosome containing *Legionella pneumophila* within the protozoan *Hartmannella vermiformis* is surrounded by the rough endoplasmic reticulum. Appl Environ Microbiol. **62:**2022-2028.

Abu Kwaik, Y., L.-Y. Gao, B. J. Stone, and O. S. Harb. 1998a. Invasion of mammalian and protozoan cells by *Legionella pneumophila*. Bull Inst Pasteur. **96:**237-247.

Abu Kwaik, Y., L.-Y. Gao, B. J. Stone, C. Venkataraman, and O. S. Harb. 1998b. Invasion of protozoa by *Legionella pneumophila* and its role in bacterial ecology and pathogenesis. Appl Environ Microbiol. **66:**3127-3133.

Abu Kwaik, Y., C. Venkataraman, O. S. Harb, and L.-Y. Gao. 1998c. Signal transduction in the protozoan host *Hartmannella vermiformis* upon attachment and invasion by *Legionella micdadei*. Appl Environ Microbiol. **64:**3134-3139.

Achtman, M. 2000. A phylogenetic perspective on molecular epidemiology. *In* M. Sussman (ed.), Molecular medical microbiology. Academic Press, London, England. In press.

Adeleke, A., J. M. Pruckler, R. F. Benson, T. J. Rowbotham, M. A. Halablab, and B. S. Fields. 1996. *Legionella*-like amebal pathogens - phylogenetic status and possible role in respiratory disease. Emerg Infect Dis. **2:**225-230.

Adeleke, A. A., B. S. Fields, R. F. Benson, M. I. Daneshvar, J. M. Pruckler, R. M. Ratcliff, T. G. Harrison, R. S. Weyant, R. J. Birtles, D. Raoult, and M. A. Halablab. 2000. *Legionella drosanskii* sp. nov., *Legionella rowbothamensis* sp. nov. and *Legionella fallonii* sp. nov: three unusual new *Legionella* species. Int J Syst Evol Microbiol. Manuscript submitted for publication.

Aderem, A., and D. M. Underhill. 1999. Mechanisms of phagocytosis in macrophages. Annu Rev Immunol. **17:**593-623.

Amann, R., W. Ludwig, and K. H. Schleifer. 1988. Beta-subunit of ATP-synthase: a useful marker for studying the phylogenetic relationship of eubacteria. J Gen Microbiol. **134:**2815-2821.

Ambler, R. P. 1996. The distance between bacterial species in sequence space. J Mol Evol. **42:**617-630.

Andrews, H. L., J. P. Vogel, and R. R. Isberg. 1998. Identification of linked *Legionella pneumophila* genes essential for intracellular growth and evasion of the endocytic pathway. Infect Immun. **66:**950-958.

Argos, P., J. K. Rao, and P. A. Hargrave. 1982. Structural prediction of membrane-bound proteins. Eur J Biochem. **128:**565-575.

Ash, C. J., A. E. Farrow, S. Wallbanks, and M. D. Collins. 1991. Phylogenetic heterogeneity of the genus *Bacillus* revealed by comparative analysis of small-subunit-ribosomal RNA sequences. Lett Appl Microbiol. **13:**202-206.

Atlas, R. M. 1993. Handbook of microbiological media, p. 938, CRC Press, Boca Ratson, FL, USA.

Attridge, S. R., and D. Rowley. 1983. The role of flagellum in the adherence of *Vibrio cholerae*. J Infect Dis. **147:**864-872.

Avise, J. C. 1994a. Interpretive tools, pp. 92-138, Molecular markers, natural history and evolution. Chapman and Hall, New York, NY, USA.

Avise, J. C. 1994b. Introduction, pp. 3-15, Molecular markers, natural history and evolution. Chapman and Hall, New York, NY, USA.

Bangsborg, J. M., N. P. Cianciotto, and P. Hindersson. 1991. Nucleotide sequence analysis of the *Legionella micdadei mip* gene, encoding a 30-kilodalton analog of the *Legionella pneumophila* Mip protein. Infect Immun. **59:**3836-3840.

Bangsborg, J. M., P. Hindersson, G. Shand, and N. Hoiby. 1995. The *Legionella micdadei* flagellin: expression in *Escherichia coli* K 12 and DNA sequence of the gene. Acta Pathol Microbiol Immunol Scand. **103:**869-877.

Bansal, N. S., and F. McDonell. 1997. Identification and DNA fingerprinting of *Legionella* strains by random amplified polymorphic DNA analysis. J Clin Microbiol. **35:**2310-2314.

Barker, J., and M. R. Brown. 1994. Trojan horses of the microbial world: protozoa and the survival of bacterial pathogens in the environment. Microbiology. **140:**1253-1259.

Barker, J., I. D. Farrell, and J. G. Hutchison. 1986. Factors affecting growth of *Legionella pneumophila* in liquid media. J Med Microbiol. **22:**97-100.

Barker, J., P. A. Lambert, and M. R. Brown. 1993. Influence of intra-amoebic and other growth conditions on the surface properties of *Legionella pneumophila*. Infect Immun. **61:**3503-3510.

Baron, P. A., and K. Willeke. 1986. Respirable droplets from whirlpools: measurements of size distribution and estimation of disease potential. Environ Res. **39:**8-18.

Bäumler, A. J. 1997. The record of horizontal gene transfer in *Salmonella*. Trends Microbiol. **5:**318-322.

Bellinger-Kawahara, C., and M. A. Horwitz. 1990. Complement component C3 fixes selectively to the major outer membrane protein (MOMP) of *Legionella pneumophila* and mediates phagocytosis of liposome-MOMP complexes by human monocytes. J Exp Med. **172:**1201-1210.

Bender, L., M. Ott, A. Debes, U. Rdest, J. Heesemann, and J. Hacker. 1991. Distribution, expression, and long-range mapping of legiolysin gene (*lly*)-specific DNA sequences in legionellae. Infect Immun. **59:**3333-3336.

Benson, R. F., W. L. Thacker, B. B. Plikaytis, and H. W. Wilkinson. 1987. Cross-reactions in *Legionella* antisera with *Bordetella pertussis* strains. J Clin Microbiol. **25:**594-596.

Benson, R. F., W. L. Thacker, R. P. Waters, P. A. Quinlivan, W. R. Mayberry, D. J. Brenner, and H. W. Wilkinson. 1989. *Legionella quinlivanii* sp. nov. isolated from water. Curr Microbiol. **18:**195-197.

Benson, R. F., W. L. Thacker, F. C. Fang, B. Kanter, W. R. Mayberry, and D. J. Brenner. 1990. *Legionella sainthelensi* serogroup 2 isolated from patients with pneumonia. Res Microbiol. **141:**453-463.

Benson, R. F., W. L. Thacker, J. A. Lanser, N. Sangster, W. R. Mayberry, and D. J. Brenner. 1991. *Legionella adelaidensis*, a new species isolated from cooling tower water. J Clin Microbiol. **29:**1004-1006.

Benson, R. F., W. L. Thacker, M. I. Daneshvar, and D. J. Brenner. 1996. *Legionella waltersii* sp. nov. and an unnamed *Legionella* genomospecies isolated from water in Australia. Int J Syst Bacteriol. **46:**631-634.

Bercovier, H., B. Fattal, and H. Shuval. 1986. Seasonal distribution of legionellae isolated from various types of water in Israel. Isr J Med Sci. **22:**644-646.

Berdal, B. P., and O. Olsvik. 1983. *Legionella* extracellular protease activity on chromogenic peptides: a simplified procedure for biochemical enzyme identification. Acta Pathol Microbiol Immunol Scand. **91:**89-91.

Berger, K. H., and R. R. Isberg. 1993. Two distinct defects in intracellular growth complemented by a single genetic locus in *Legionella pneumophila*. Mol Microbiol. **7:**7-19.

Berger, K. H., J. J. Merriam, and R. R. Isberg. 1994. Altered intracellular targeting properties associated with mutations in the *Legionella pneumophila dotA* gene. Mol Microbiol. **14:**809-822.

Berk, S. G., R. S. Ting, G. W. Turner, and R. J. Ashborn. 1998. Production of respirable vesicles containing live *Legionella pneumophila* cells by two *Acanthamoeba* spp. Appl. Environ. Microbiol. **64:**279-286.

Birtles, R. J., N. Doshi, N. A. Saunders, and T. G. Harrison. 1991. Second serogroup of *Legionella quinlivanii* isolated from two unrelated sources in the United Kingdom. J Appl Bacteriol. **71:**402-406.

Black, W. J., F. D. Quinn, and L. S. Tompkins. 1990. *Legionella pneumophila* zinc metalloprotease is structurally and functionally homologous to *Pseudomonas aeruginosa* elastase. J Bacteriol. **172:**2608-2613.

Blander, S. J., and M. A. Horwitz. 1989. Vaccination with the major secretory protein of *Legionella pneumophila* induces cell-mediated and protective immunity in a guinea pig model of Legionnaires' disease. J Exp Med. **169:**691-705.

Blander, S. J., L. Szeto, H. A. Shuman, and M. A. Horwitz. 1990. An immunoprotective molecule, the major secretory protein of *Legionella pneumophila*, is not a virulence factor in a guinea pig model of Legionnaires' disease. J Clin Invest. **86:**817-824.

Bollin, G. E., J. F. Plouffe, M. F. Para, and B. Hackman. 1985. Aerosols containing *Legionella pneumophila* generated by shower heads and hot-water faucets. Appl Environ Microbiol. **50:**1128-1131.

Bornstein, N., D. Marmet, M. Surgot, M. Nowicki, H. Meugnier, J. Fleurette, E. Ageron, F. Grimont, P. A. Grimont, W. L. Thacker, R. F. Benson, and D. J. Brenner. 1989. *Legionella gratiana* sp. nov. isolated from French spa water. Res Microbiol. **140:**541-552.

Bornstein, N., D. Marmet, M. H. Dumaine, M. Surgot, and J. Fleurette. 1991. Detection of flagella in 278 *Legionella* strains by latex reagent sensitized with antiflagellum immunoglobulins. J Clin Microbiol. **29:**953-956.

Bozue, J. A., and W. Johnson. 1996. Interaction of *Legionella pneumophila* with *Acanthamoeba castellanii*: uptake by coiling phagocytosis and inhibition of phagosome-lysosome fusion. Infect Immun. **64:**668-673.

Brand, B. C., A. B. Sadosky, and H. A. Shuman. 1994. The *Legionella pneumophila icm* locus: a set of genes required for intracellular multiplication in human macrophages. Mol Microbiol. **14:**797-808.

Breiman, R. F., B. S. Fields, G. N. Sanden, L. Volmer, A. Meier, and J. S. Spika. 1990. Association of shower use with Legionnaires' disease. Possible role of amoebae. JAMA. **263:**2924-2926.

Brenner, D. J. 1987. Classification of the *Legionellae*. Semin Respir Infect. **2:**190-205.

Brenner, D. J., A. G. Steigerwalt, and J. E. McDade. 1979. Classification of the Legionnaires' disease bacterium: *Legionella pneumophila*, genus novum, species nova, of the family *Legionellaceae*, family nova. Ann Intern Med. **90:**656-658.

Brenner, D. J., A. G. Steigerwalt, G. W. Gorman, R. E. Weaver, J. C. Feeley, L. G. Cordes, H. W. Wilkinson, C. Patton, B. M. Thomason, and K. R. Lewallen Saseville. 1980. *Legionella bozemanii* sp. nov. and *Legionella dumoffii* sp.nov.: classification of two additional species of *Legionella* associated with human pneumonia. Curr Microbiol. **4:**111-116.

Brenner, D. J., A. G. Steigerwalt, G. W. Gorman, H. W. Wilkinson, W. F. Bibb, M. Hackel, R. L. Tyndall, J. Campbell, J. C. Feeley, W. L. Thacker, P. Skaliy, W. T. Martin, B. J. Brake, B. S. Fields, H. V. McEachern, and L. Corcoran. 1985. Ten new species of *Legionella*. Int J Syst Bacteriol. **35:**50-59.

Brieland, J., M. McClain, L. Heath, C. Chrisp, G. Huffnagle, M. LeGendre, M. Hurley, J. Fantone, and C. Engleberg. 1996. Coinoculation with *Hartmannella vermiformis* enhances replicative *Legionella pneumophila* lung infection in a murine model of Legionnaires' disease. Infect Immun. **64:**2449-2456.

Brown, M. C. M., A. Weston, J. R. Saunders, and G. O. Humphreys. 1979. Transformation of *Escherichia coli* C600 by plasmid DNA at different phases of growth. FEMS Microbiol Lett. **5:**219-222.

Brown, M. R. W., and J. Barker. 1999. Unexplored reservoirs of pathogenic bacteria: protozoa and biofilms. Trends Microbiol. **7:**46-50.

Bull, J. Z., G. R. Nimmo, and P. Giffard. 1997. Genetic diversity and clonal population structure of *Legionella longbeachae* serogroup 1 in Australia. Microbiol Aust. **18:**A120.

Burns, D. L. 1999. Biochemistry of type IV secretion. Curr Opin Microbiol. **2:**25-29.

Butler, C. A., E. D. Street, T. P. Hatch, and P. S. Hoffman. 1985. Disulphide-bonded outer membrane proteins in the genus *Legionella*. Infect Immun. **48:**14-18.

Byrd, T. F., and M. A. Horwitz. 1989. Interferon gamma-activated human monocytes downregulate transferrin receptors and inhibit the intracellular multiplication of *Legionella pneumophila* by limiting the availability of iron. J Clin Invest. **83:**1457-1465.

Byrne, B., and M. S. Swanson. 1998. Expression of *Legionella pneumophila* virulence traits in response to growth conditions. Infect Immun. **66:**3029-3034.

Cameron, S., D. Roder, C. Walker, and J. Feldheim. 1991. Epidemiological characteristics of *Legionella* infection in South Australia: implications for disease control. Aust NZ J Med. **21:**65-70.

Campbell, J., W. F. Bibb, M. A. Lambert, S. Eng, A. G. Steigerwalt, J. Allard, C. W. Moss, and D. J. Brenner. 1984. *Legionella sainthelensi*: a new species of *Legionella* isolated from water near Mt. St. Helens. Appl Environ Microbiol. **47:**369-373.

Cao, Y., J. Adachi, and M. Hasegawa. 1998. Comment on the Quartet Puzzling Method for finding maximum-likelihood tree topologies. Mol Biol Evol. **15:**87-89.

Cavalli-Sforza, L. L., and A. W. F. Edwards. 1967. Phylogenetic analysis: models and estimation procedures. Am J Hum Genet. **19:**230-257.

Chandler, F. W., I. L. Roth, C. S. Callaway, J. L. Bump, B. M. Thomason, and R. E. Weaver. 1980. Flagella on Legionnaires' disease bacteria: ultrastructural observations. Ann Intern Med. **93:**711-714.

Cherry, W. B., G. W. Gorman, L. H. Orrison, C. W. Moss, A. G. Steigerwalt, H. W. Wilkinson, S. E. Johnson, R. M. McKinney, and D. J. Brenner. 1982. *Legionella jordanis*: a new species of *Legionella* isolated from water and sewage. J Clin Microbiol. **15:**290-297.

Cianciotto, N., B. I. Eisenstein, N. C. Engleberg, and H. Shuman. 1989a. Genetics and molecular pathogenesis of *Legionella pneumophila*, an intracellular parasite of macrophages. Mol Biol Med. **6:**409-424.

Cianciotto, N. P., B. I. Eisenstein, C. H. Mody, G. B. Toews, and N. C. Engleberg. 1989b. A *Legionella pneumophila* gene encoding a species-specific surface protein potentiates initiation of intracellular infection. Infect Immun. **57:**1255-1262.

Cianciotto, N. P., J. M. Bangsborg, B. I. Eisenstein, and N. C. Engleberg. 1990a. Identification of *mip*-like genes in the genus *Legionella*. Infect Immun. **58:**2912-2918.

Cianciotto, N. P., B. I. Eisenstein, C. H. Mody, and N. C. Engleberg. 1990b. A mutation in the *mip* gene results in an attenuation of *Legionella pneumophila* virulence. J Infect Dis. **162:**121-126.

Cianciotto, N. P., and B. S. Fields. 1992. *Legionella pneumophila mip* gene potentiates intracellular infection of protozoa and human macrophages. Proc Natl Acad Sci USA. **89:**5188-5191.

Cianciotto, N. P., J. K. Stamos, and D. W. Kamp. 1995a. Infectivity of *Legionella pneumophila mip* mutant for alveolar epithelial cells. Curr Microbiol. **30:**247-250.

Cianciotto, N. P., W. O'Connell, G. A. Dasch, and L. P. Mallavia. 1995b. Detection of *mip*-like sequences and Mip-related proteins within the family *Rickettsiaceae*. Curr Microbiol. **30:**149-153.

Ciesielski, C. A., M. J. Blaser, and W. L. Wang. 1986. Serogroup specificity of *Legionella pneumophila* is related to lipopolysaccharide characteristics. Infect Immun. **51:**397-404.

Cilia, V., B. Lafay, and R. Christen. 1996. Sequence heterogeneities among 16S rRNA sequences, and their effect on phylogenetic analyses at the species level. Mol Biol Evol. **13:**451-461.

Cirillo, J. D., S. Falkow, and L. S. Tompkins. 1994. Growth of *Legionella pneumophila* in *Acanthamoeba castellanii* enhances invasion. Infect Immun. **62:**3254-3261.

Clemens, D. L., and M. A. Horwitz. 1992. Membrane sorting during phagocytosis: selective exclusion of major histocompatibility complex molecules but not complement receptor CR3 during conventional and coiling phagocytosis. J Exp Med. **175:**1317-1326.

Coers, J., C. Monahan, and C. R. Roy. 1999. Modulation of phagosome biogenesis by *Legionella pneumophila* creates an organelle permissive for bacterial growth. Nat Cell Biol. **1:**451-453.

Cohan, F. M. 1994a. The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. Am Nat. **143:**965-986.

Cohan, F. M. 1994b. Genetic exchange and evolutionary divergence in prokaryotes. Trends Ecol Evol. **9:**175-180.

Cohan, F. M. 1995. Does recombination constrain neutral divergence among bacterial taxa? Evolution. **49:**164-175.

Cohan, F. M. 1996. The role of genetic exchange in bacterial evolution. ASM News. **62:**631-636.

Cohan, F. M. 1998. Genetic structure of bacterial populations. *In* R. Singh and C. Krimbas (eds.), Evolutionary genetics from molecules to morphology. Cambridge University Press, Cambridge, England.

Collins, M. T. 1986. *Legionella* infections in animals. Isr J Med Sci. **22:**662-673.

Collins, M. D., P. A. Lawson, A. Willems, J. J. Cordoba, J. Fernandez-Garayzabal, P. Garcia, J. Cai, H. Hippe, and J. A. Farrow. 1994. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int J Syst Bacteriol. **44:**812-826.

Conlan, J. W., and L. A. Ashworth. 1986b. The relationship between the serogroup antigen and lipopolysaccharide of *Legionella pneumophila*. J Hyg (Lond). **96:**39-48.

Coxon, P. Y., J. T. Summersgill, J. A. Ramirez, and R. D. Miller. 1998. Signal transduction during *Legionella pneumophila* entry into human monocytes. Infect Immun. **66:**2905-2913.

De Weger, L. A., C. I. M. van der Vlugt, A. H. M. Wijjes, P. A. M. Bakker, B. Schippers, and B. Lugtenberg. 1987. Flagella of a plant-growth-stimulating *Pseudomonas fluorescens* strain are required for colonization of potato roots. J Bacteriol. **169:**2769-2773.

Dennis, P. J., D. J. Brenner, W. L. Thacker, R. Wait, G. Vesey, A. G. Steigerwalt, and R. F. Benson. 1993. Five new *Legionella* species isolated from water. Int J Syst Bacteriol. **43:**329-37.

Dennis, P. J., A. E. Wright, D. A. Rutter, J. E. Death, and B. P. Jones. 1984. *Legionella pneumophila* in aerosols from shower baths. J Hyg (Lond). **93:**349-353.

Descheemaeker, P., C. Lammens, B. Pot, P. Vandamme, and H. Goosens. 1997. Evaluation of arbitrarily primed PCR analysis and pulsed-field gel electrophoresis of large genomic DNA fragments for identification of enterococci important in human disease. J Syst Bacteriol **47:**555-561.

Dewhirst, F. E., B. J. Paster, I. Olsen, and G. J. Fraser. 1992. Phylogeny of 54 representative strains of species in the family *Pasteurellaceae* as determined by comparison of 16S rRNA sequences. J Bacteriol. **174:**2002-2013.

Diogo, A., A. Veríssimo, M. F. Nobre, and M. S. da Costa. 1999. Usefulness of fatty acid composition for differentiation of *Legionella* species. J Clin Microbiol. **37:**2248-2254.

Doolittle, W. F. 1999. Phylogenetic classification and the universal tree. Science. **284:** 2124-2129.

Dower, W. J., J. F. Miller, and C. W. Ragsdale. 1988. High efficiency transformation of *Escherichia coli* by high voltage electroporation. Nuc Acids Res. **16:**6127-6145.

Dowling, J. N., A. K. Saha, and R. H. Glew. 1992. Virulence factors of the family *Legionellaceae*. Microbiol Rev. **56:**32-60.

Doyle, R. M., T. W. Steele, A. M. McLennan, I. H. Parkinson, P. A. Manning, and M. W. Heuzenroeder. 1998. Sequence analysis of the *mip* gene of the soilborne pathogen *Legionella longbeachae*. Infect Immun. **66:**1492-1499.

Dreyfus, L. A., and B. H. Iglewski. 1985. Conjugation-mediated genetic exchange in *Legionella pneumophila*. J Bacteriol. **161:**80-84.

Du, Y., D. A. Maslov, and K. P. Chang. 1994. Monophyletic origin of beta-division proteobacterial endosymbionts and their coevolution with insect trypanosomatid protozoa *Blastocrithidia culicis* and *Crithidia* spp. Proc Natl Acad Sci USA. **91:**8437-8441.

Echeita, M. A., and M. A. Usera. 1998. Chromosomal rearrangements in *Salmonella enterica* serotype Typhi affecting molecular typing in outbreak investigations. J Clin Microbiol. **36:**2123-2126.

Edelstein, P. H. 1993. Legionnaires' disease. Clin Infect Dis. **16:**741-749.

Edelstein, P. H., D. J. Brenner, C. W. Moss, A. G. Steigerwalt, E. M. Francis, and W. L. George. 1982. *Legionella wadsworthii* species nova: a cause of human pneumonia. Ann Intern Med. **97:**809-813.

Edelstein, P. H., and M. A. Edelstein. 1989. Evaluation of the Merifluor-*Legionella* immunofluorescent reagent for identifying and detecting 21 *Legionella* species. J Clin Microbiol. **27:**2455-2458.

Efron, B., and G. Gong. 1983. A leisurely look at the bootstrap, the jackknife, and cross-validation. Am Statist. **37:**36-48.

Efron, B., and R. J. Tibshirani. 1993. An Introduction to the Bootstrap. Chapman and Hall, New York, USA.

Ehret, W., and G. Ruckdeschel. 1985a. Membrane proteins of *Legionellaceae*. I. Membrane proteins of different strains and serogroups of *Legionella pneumophila*. Zentralbl Bakteriol Mikrobiol Hyg [A.] **259:**433-445.

Ehret, W., and G. Ruckdeschel. 1985b. Membrane proteins of *Legionellaceae*. II. Serogroup and species specific antigens in the outer membrane of *Legionella pneumophila*. Zentralbl Bakteriol Mikrobiol Hyg [A]. **260:**184-196.

Ehret, W., and G. Ruckdeschel. 1985c. Molecular weight of the major outer membrane protein of *Legionella pneumophila*. Eur J Clin Microbiol. **4:**592-593.

Eisen, J. A. 1995. The RecA protein as a model molecule for molecular systematic studies of bacteria: comparison of trees of RecAs and 16S rRNAs from the same species. J Mol Evol. **41:**1105-1123.

Elliott, J. A., and W. Johnson. 1981. Immunological and biochemical relationships among flagella isolated from *Legionella pneumophila* serogroups 1, 2, and 3. Infect Immun. **33:**602-610.

Elliott, J. A., and W. C. Winn, Jr. 1986. Treatment of alveolar macrophages with cytochalasin D inhibits subsequent growth of *Legionella pneumophila*. Infect Immun. **51:**31-36.

Engelman, D. M., T. A. Steitz, and A. Goldman. 1986. Identifying nonpolar transbilayer helices in amino acid sequences of membrane proteins. Annu Rev Biophys Biophys Chem. **15:**321-353.

Engleberg, N. C., D. J. Drutz, and B. I. Eisenstein. 1984a. Cloning and expression of *Legionella pneumophila* antigens in *Escherichia coli*. Infect Immun. **44:**222-227.

Engleberg, N. C., E. Pearlman, and B. I. Eisenstein. 1984b. *Legionella pneumophila* surface antigens cloned and expressed in *Escherichia coli* are translocated to the host cell surface and interact with specific anti-*Legionella* antibodies. J Bacteriol. **160:**199-203.

Engleberg, N. C., C. Carter, D. R. Weber, N. P. Cianciotto, and B. I. Eisenstein. 1989. DNA sequence of *mip*, a *Legionella pneumophila* gene associated with macrophage infectivity. Infect Immun. **57:**1263-1270.

Enright, M. C., and B. G. Spratt. 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology. **144:**3049-3060.

Enright, M. C., and B. G. Spratt. 1999. Multilocus sequence typing. Trends Microbiol. **7:**482-487.

Erdtman, G. 1937. Pollen grains recovered from the atmosphere over the Atlantic. Acta Hort Gothoburg. **12:**185.

Fabbi, M., M. C. Pastoris, E. Scanziani, S. Magnino, and L. DiMatteo. 1998. Epidemiological and environmental investigations of *Legionella pneumophila* infection in cattle and case report of fatal pneumonia in a calf. J Clin Microbiol. **36:**1942-1947.

Fallon, R. J. 1990. *Legionella*, pp. 276-287. *In* M. T. Parker and B. I. Duerden (eds.), Topley and Wilson's Principles of Bacteriology, Virology and Immunity, Eighth ed, vol. 2. Edward Arnold, London, England.

Fang, G. D., V. L. Yu, and R. M. Vickers. 1989. Disease due to the *Legionellaceae* (other than *Legionella pneumophila*). Historical, microbiological, clinical, and epidemiological review. Medicine Baltimore. **68:**116-132.

Farris, J. S., M. Kallersjo, A. G. Kluge, and C. Bult. 1995. Testing significance of incongruence. Cladistics. **10:**315-319.

Feeley, J. C., R. J. Gibson, G. W. Gorman, N. C. Langford, J. K. Rasheed, D. C. Mackel, and W. B. Baine. 1979. Charcoal-yeast extract agar: primary isolation medium for *Legionella pneumophila*. J Clin Microbiol. **10:**437-441.

Felsenstein, J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol. **17:**368-376.

Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. **39:**783-791.

Felsenstein, J. 1993. PHYLIP (Phylogeny Inference Package), Version 3.5c ed. Distributed by author. Department of Genetics, University of Washington, Seattle, USA.

Fernandez, R. C., S. M. Logan, S. H. Lee, and P. S. Hoffman. 1996. Elevated levels of *Legionella pneumophila* stress protein Hsp60 early in infection of human monocytes and L929 cells correlate with virulence. Infect Immun. **64:**1968-1976.

Fields, B. S. 1993. *Legionella* and protozoa: interaction of a pathogen and its natural host, pp. 129-136. *In* J. M. Barbaree, R. F. Beiman, and A. P. Dufour (eds.), *Legionella*: current status and emerging perspectives. American Society for Microbiology, Washington, DC, USA.

Fields, B. S. 1996. The molecular ecology of legionellae. Trends Microbiol. **4:**286-290.

Fields, B. S., S. R. Fields, J. N. Loy, E. H. White, W. L. Steffens, and E. B. Shotts. 1993. Attachment and entry of *Legionella pneumophila* in *Hartmannella vermiformis*. J Infect Dis. **167:**1146-1150.

Finlay, B. J. 1998. The global diversity of protozoa and other small species. Int J Parasitol. **28:**29-48.

Finlay, B. J., and G. F. Esteban. 1998. Freshwater protozoa: biodiversity and ecological function. Biodiversity and Conservation. **7:**1163-1186.

Fischer, G., H. Bang, B. Ludwig, K. Mann, and J. Hacker. 1992. Mip protein of *Legionella pneumophila* exhibits peptidyl-prolyl *cis/trans* isomerase (PPlase) activity. Mol Microbiol. **6:**1375-1383.

Fitz-Gibbon, S. T., and C. H. House. 1999. Whole genome-based phylogenetic analysis of free-living microorganisms. Nucleic Acids Res. **27:**4218-4222.

Fraser, D. W., and J. E. McDade. 1979. Legionellosis. Sci Am. **241:**82-89.

Fraser, D. W., T. R. Tsai, W. Orenstein, W. E. Parkin, J. Beecham, R. G. Sharrar, J. Harris, G. F. Mallison, S. M. Martin, J. E. McDade, C. C. Shepard, and P. S. Brachman. 1977. Legionnaires' disease. Description of an epidemic of pneumonia. N Engl J Med. **297:**1189-1197.

Gabay, J. E., and M. A. Horwitz. 1985. Isolation and characterization of the cytoplasmic and outer membranes of the Legionnaires' disease bacterium (*Legionella pneumophila*). J Exp Med. **161:**409-422.

Gabay, J. E., E. M. Blake, W. D. Niles, and M. A. Horwitz. 1985. Purification of *Legionella pneumophila* major outer membrane protein and demonstration that it is a porin. J Bacteriol. **162:**85-91.

Gao, L.-Y., and Y. Abu Kwaik. 1999a. Activation of caspase 3 during *Legionella pneumophila*-induced apoptosis. Infect Immun. **67:**4886-4894.

Gao, L.-Y., and Y. Abu Kwaik. 1999b. Apoptosis in macrophages and alveolar epithelial cells during early stages of infection by *Legionella pneumophila* and its role in cytopathogenicity. Infect Immun. **67:**862-870.

Gao, L.-Y., O. S. Harb, and Y. Abu Kwaik. 1997. Utilization of similar mechanisms by *Legionella pneumophila* to parasitize two evolutionary distant hosts, mammalian and protozoan cells. Infect Immun. **65:**4738-4746.

Gao, L.-Y., O. S. Harb, and Y. Abu Kwaik. 1998. The identification of macrophage-specific infectivity loci (*mil*) of *Legionella pneumophila* that are not required for infectivity of protozoa. Infect Immun. **66:**883-892.

Garduño, R. A., A. G. Faulkner, M. A. Trevors, N. Vats, and P. S. Hoffmen. 1998a. Immunolocalization of Hsp60 in *Legionella pneumophila*. J Bacteriol. **180:**505-513.

Garduño, R. A., E. Garduño, and P. S. Hoffman. 1998b. Surface-associated Hsp60 chaperonin of *Legionella pneumophila* mediates invasion in a HeLa cell model. Infect Immun. **66:**4602-4610.

Garger, S. J., O. M. Griffith, and L. K. Grill. 1983. Rapid purification of plasmid DNA by a single centrifugation in a two step caesium chloride-ethidium bromide gradient. Biochem Biophys Res Comm. **117:**835-842.

Garnier, J., D. J. Osguthorpe, and B. Robson. 1978. Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins. J Mol Biol. **120:**97-120.

Garnier, J., J. F. Gilbrat and B. Robson. 1996. GOR method for predicting protein secondary structure from amino acid sequence. Methods Enzymol. **266:**540-553

Glick, T. H., M. B. Gregg, B. Berman, G. Mallison, W. W. Rhodes, Jr., and I. Kassanoff. 1978. Pontiac fever: an epidemic of unknown etiology in a health department: I. Clinical and epidemiological aspects. Am J Epidemiol. **107:**149-160.

Gorman, G. W., J. C. Feeley, A. Steigerwalt, P. H. Edelstein, C. W. Moss, and D. J. Brenner. 1985. *Legionella anisa*: a new species of *Legionella* isolated from potable waters and a cooling tower. Appl Environ Microbiol. **49:**305-309.

Gosting, L. H., K. Cabrian, J. C. Sturge, and L. C. Goldstein. 1984. Identification of a species-specific antigen in *Legionella pneumophila* by a monoclonal antibody. J Clin Microbiol. **20:**1031-1035.

Griffin, F. M., J. A. Griffin, J. E. Leider, and S. C. Silverstein. 1975. Studies on the mechanism of phagocytosis. I. Requirements for circumferential attachment of particle-bound ligands to specific receptors on the macrophage plasma membrane. J Exp Med. **142:**1263-1280.

Groothuis, D. G., A. H. Havelaar, and H. R. Veenendaal. 1985. A note on legionellas in whirlpools. J Appl Bacteriol. **58:**479-481.

Hacker, J., and G. Fischer. 1993. Immunophilins: structure-function relationship and possible role in microbial pathogenicity. Mol Microbiol. **10:**445-456.

Haertl, R., and G. Bandlow. 1991. Subtyping of *Legionella pneumophila* serogroup 1 isolates by small-fragment restriction endonuclease analysis. Eur J Clin Microbiol Infect Dis. **10:**630-635.

Hägele, S., J. Hacker, and B. C. Brand. 1998. *Legionella pneumophila* kills human phagocytes but not protozoan host cells by inducing apoptotic cell death. FEMS Microbiol Lett. **169:**51-58.

Hales, L. M., and H. A. Shuman. 1999. *Legionella pneumophila* contains a type II general secretion pathway required for growth in amoebae as well as for secretion of the Msp protease. Infect Immun. **67:**3662-3666.

Hanahan, D. 1985. Bacterial transformation, pp. 109-135. *In* D. M. Glover (ed.), DNA Cloning. IRL Press, Oxford, England.

Harb, O. S., C. Venkataraman, B. J. Haack, L.-Y. Goa, and Y. Abu Kwaik. 1998. Heterogeneity in the attachment and uptake mechanisms of the Legionnaires' disease bacterium, *Legionella pneumophila*, by protozoan hosts. Appl Environ Microbiol. **64:**126-132.

Harrison, T. G., and N. A. Saunders. 1994. Taxonomy and typing of *Legionellae*. Med. Microbiol. **5:**79-90.

Harrison, T. G., and A. G. Taylor. 1988. Identification of legionellae by serological methods, pp. 113-135. *In* T. G. Harrison and A. G. Taylor (eds.), A Laboratory Manual for *Legionella*. John Wiley & Sons Ltd., London, England.

Harrison, T. G., N. A. Saunders, A. Haththotuwa, N. Doshi, and A. G. Taylor. 1990. Typing of *Legionella pneumophila* serogroups 2-14 strains by analysis of restriction fragment length polymorphisms. Lett Appl Microbiol. **11:**189-192.

Hawkes, R., E. Niday, and J. Gordon. 1982. A dot-immunobinding assay for monoclonal and other antibodies. Anal Biochem. **119:**142-147.

Hay, J. C., and R. H. Scheller. 1997. SNAREs and NSF in targeted membrane fusion. Curr Opin Cell Biol. **9:**505-512.

Hebert, G. A., A. G. Steigerwalt, and D. J. Brenner. 1980. *Legionella micdadei* species nova: classification of a third species of *Legionella* associated with human pneumonia. Curr Microbiol. **3:**255-257.

Helbig, J. H., B. Ludwig, P. C. Lück, A. Groh, W. Witzleb, and J. Hacker. 1995. Monoclonal antibodies to *Legionella Mip* proteins recognize genus- and species-specific epitopes. Clin Diagn Lab Immunol. **2:**160-165.

Hell, W., A. Essig, S. Bohnet, S. Gatermann, and R. Marre. 1993. Cleavage of tumor necrosis factor-alpha by *Legionella* exoprotease. Acta Pathol Microbiol Immunol Scand. **101:**120-126.

Hénaut, A., and A. Danchin. 1996. Analysis and predictions from *Escherichia coli* sequences, or *E. coli in silico*, p 2052. *In* F. C. Neidhardt, R. Curtis III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umnarger (eds.), In *Escherichia coli* and *Salmonella*: cellular and molecular biology, vol. II. American Society for Microbiology, Washington, DC, USA.

Henning, U., H. Schwarz, and R. Chen. 1979. A radioimmunological screening screening method for specific membrane proteins. Anal Biochem. **97:**153-157.

Herwaldt, L. A., G. W. Gorman, T. McGrath, S. Toma, B. Brake, A. W. Hightower, J. Jones, A. L. Reingold, P. A. Boxer, P. W. Tang, C. W. Moss, H. Wilkinson, D. J. Brenner, A. G. Steigerwalt and C. V. Broome. 1984. A new *Legionella* species, *Legionella feeleii* species nova, causes Pontiac fever in an automobile plant. Ann Intern Med. **100:**333-338.

Heuner, K., L. Bender-Beck, B. C. Brand, P. C. Lück, K. H. Mann, R. Marre, M. Ott, and J. Hacker. 1995. Cloning and genetic characterization of the flagellum subunit gene (*flaA*) of *Legionella pneumophila* serogroup 1. Infect Immun. **63:**2499-2507.

Heuner, K., B. C. Brand, and J. Hacker. 1999. The expression of the flagellum of *Legionella pneumophila* is modulated by different environmental factors. FEMS Microbiol Lett. **175:**69-77.

Hewit, R. 1937. Natural habitat and distribution of *Hartmannella castellanii* (Douglas), reported contaminant of bacterial cultures. J Parasitol. **23:**491-495.

Hickey, E. K., and N. P. Cianciotto. 1997. An iron- and fur-repressed *Legionella pneumophila* gene that promotes intracellular infection and encodes a protein with similarity to the *Escherichia coli* aerobactin synthetases. Infect Immun. **65:**133-143.

High, A. S., S. D. Torosian, and F. G. Rodgers. 1993. Cloning, nucleotide sequence and expression in *Escherichia coli* of a gene (*ompM*) encoding a 25-kDa major outer-membrane protein (MOMP) of *Legionella pneumophila*. J Gen Microbiol. **139:**1715-1721.

Hindahl, M. S., and B. H. Iglewski. 1986. Outer membrane proteins from *Legionella pneumophila* serogroups and other *Legionella* species. Infect Immun. **51:**94-101.

Hoffman, P. S., L. Houston, and C. A. Butler. 1990. *Legionella pneumophila htpAB* heat shock operon: nucleotide sequence and expression of the 60-kilodalton antigen in *L. pneumophila*-infected HeLa cells. Infect Immun. **58:**3380-3387.

Hoffman, P. S., M. Ripley, and R. Weeratna. 1992a. Cloning and nucleotide sequence of a gene (*ompS*) encoding the major outer membrane protein of *Legionella pneumophila*. J Bacteriol. **174:**914-920.

Hoffman, P. S., J. H. Seyer, and C. A. Butler. 1992b. Molecular characterisation pf the 28 and 31-kilodalton subunits of the *Legionella pneumophila* major outer membrane protein. J Bacteriol. **174:**908-913.

Hookey, J. V., N. A. Saunders, N. K. Fry, R. J. Birtles, and T. G. Harrison. 1996. Phylogeny of *Legionellaceae* base on small-subunit ribosomal DNA sequences and proposal of *Legionella lytica* comb. nov. for *Legionella*-like amoebal pathogens. Int J Syst Bacteriol. **46:**526-531.

Horne, S. M., and K. D. Young. 1995. *Escherichia coli* and other species of the *Enterobacteriaceae* encode a protein similar to the family of Mip-like FK506-binding proteins. Arch Microbiol. **163:**357-365.

Horne, S. M., T. J. Kottom, L. K. Nolan, and K. D. Young. 1997. Decreased intracellular survival of an *fkpA* mutant of *Salmonella typhimurium* Copenhagen. Infect Immun. **65:**806-810.

Horwitz, M. A. 1984. Phagocytosis of Legionnaires' disease bacterium (*Legionella pneumophila*) occurs by a novel mechanism: engulfment with a pseudopod coil. Cell. **36:**27-33.

Horwitz, M. A. 1992. Interactions between macrophages and *Legionella pneumophila*. Curr Top Microbiol Immunol. **181:**265-282.

Horwitz, M. A., and S. C. Silverstein. 1983. Intracellular multiplication of Legionnaires' disease bacteria (*Legionella pneumophila*) in human monocytes is reversibly inhibited by erythromycin and rifampin. J Clin Invest. **71:**15-26.

Hughes, M. S., and T. W. Steele. 1994. Occurrence and distribution of *Legionella* species in composted plant materials. Appl Environ Microbiol. **60:**2003-2005.

Hultgren, S. J., S. Abraham, M. Caparon, P. Falk, J. W. St. Geme III, and S. Normark. 1993. Pilis and nonpilus bacterial adhesions: assembly and function in cell recognition. Cell. **73:**887-901.

Husmann, L. K., and W. Johnson. 1994. Cytotoxicity of extracellular *Legionella pneumophila*. Infect Immun. **62:**2111-2114.

Jacob, T., J. C. Escallier, M. V. Sanguedolce, C. Chicheportiche, P. Bongrand, C. Capo, and J. L. Mege. 1994. *Legionella pneumophila* inhibits superoxide generation in human monocytes via the down-modulation of alpha and beta protein kinase C isotypes. J Leukoc Biol. **55:**310-312.

James, B. W., W. S. Mauchline, P. J. Dennis, and C. W. Keevil. 1997. A study of iron acquisition mechanisms of *Legionella pneumophila* grown in chemostat culture. Curr Microbiol. **34:**238-243.

Jantzen, E., A. Sonesson, T. Tangen, and J. Eng. 1993. Hydroxy-fatty acid profiles of *Legionella* species: diagnostic usefulness assessed by principal component analysis. J Clin Microbiol. **31:**1413-1419.

Jernigan, D. B., L. I. Sanders, K. B. Waites, E. S. Brookings, R. F. Benson, and P. G. Pappas. 1994. Pulmonary infection due to *Legionella cincinnatiensis* in renal transplant recipients: two cases and implications for laboratory diagnosis. Clin Infect Dis. **18:**385-389.

Johnson, J. L. 1986. Nucleic acids in bacterial classification, pp. 972-975. *In* P. H. A. Sneath, N. S. Mair, N. E. Sharpe, and J. G. Holt (eds.), Bergey's manual of systematic bacteriology, Vol. 2. Williams and Wilkins, Balimore, MD., USA.

Jones, C. J., and S. Aizawa. 1991. The bacterial flagellum and flagellar motor: structure, assembly and function. Adv microb Physiol. **32:**109-172.

Jones, C. J., K. J. Edwards, S. Castaglione, M. O. Winfield, F. Sala, C. van de Wiel, G. Bredemeijer, B. Vosman, M. Matthes, A. Daly, R. Brettschneider, P. Bettini, M. Buiatti, E. Maestri, A. Malceschii, N. Marmiroli, R. Aert, G. Volckaert, T. Rueda, R. Linacero,

A. Vazques, and A. Karp. 1997. Reproducibility testing of RAPD, AFLP and SSR markers in plants by a network of European laboratories. Mol Breed. **3:**381-390.

Joshi, A. D., and M. S. Swanson. 1999. Comparative analysis of *Legionella pneumophila* and *Legionella micdadei* virulence traits. Infect Immun. **67:**4134-4142.

Jukes, T. H. and C. R. Cantor. 1969. Evolution of protein molecules. pp. 21-132. *In* H. Munro (ed.), Mammalian protein molecules. Academic Press, New York, USA.

Kamla, V., B. Henrich, and U. Hadding. 1996. Phylogeny based on elongation factor Tu reflects the phenotypic features of mycoplasmas better than that based on *16S rRNA*. Gene. **171:**83-87.

Karplus, P. A., and G. E. Shultz. 1985. Prediction of chain flexibility in proteins. A tool for the selection of peptide antigens. Naturwissenschaften. **72:**212-213.

Kaufmann, A. F., J. E. McDade, C. M. Patton, J. V. Bennett, P. Skaliy, J. C. Feeley, D. C. Anderson, M. E. Potter, V. F. Newhouse, M. B. Gregg, and P. S. Brachman. 1981. Pontiac fever: isolation of the etiologic agent (*Legionella pneumophilia*) and demonstration of its mode of transmission. Am J Epidemiol. **114:**337-347.

Kirby, J. E., and R. R. Isberg. 1998. Legionnaires' disease: the pore macrophage and the legion of terror within. Trends Microbiol. **6:**256-258.

Kirby, J. E., J. P. Vogel, H. L. Andrews, and R. R. Isberg. 1998. Evidence for pore forming ability by *Legionella pneumophila*. Mol Microbiol. **27:**323-336.

Koeleman, J. G. M., J. Stoof, D. J. Biesmans, P. H. M. Savelkoul, and C. M. J. E. Vandenbroucke-Grauls. 1997. Comparison of ARDRA, RAPD and AFLP fingerprinting for the identification of *Acinetobacter* genomic species and typing of *Acinetobacter baumannii*. J Clin Microbiol. **36:**2522-2529.

Krinos, C., A. S. High and F. G. Rogers. 1999. Role of the 25-kDa major outer membrane protein of *L. pneumophila* in attachment to U937 cells and its potential as a virulence factor for chick embryos. J Appl Microbiol **86:**237-244.

Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetic Analysis. Version 1.0. The Pennsylvania State University, University Park, PA. USA.

Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydropathic character of a protein. J Mol Biol. **157:**105-132.

Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. **227:**680-683.

Lambert, M. A., and C. W. Moss. 1989. Cellular fatty acid compositions and isoprenoid quinone contents of 23 *Legionella* species. J Clin Microbiol. **27:**465-473.

Lan, R., and P. R. Reeves. 1996. Gene transfer is a major factor in bacterial evolution. Mol Biol Evol. **13:**47-55.

Lanser, J., M. Adams, R. Doyle, P. Hewitt, and N. Sangster. 1992. Genetic characterization of *Legionella pneumophila* serogroup 1 associated with respiratory disease in Australia. Appl Environ Microbiol. **58:**706-708.

Lawande, R. V. 1983. Recovery of soil amoebae from the air during the harmattan in Zaria, Nigeria. Annals Trop Med Parasitol. **77:**45-49.

Lawson, P. A., P. Llop-Perez, R. A. Hutson, H. Hippe, and M. D. Collins. 1993. Towards a phylogeny of the clostridia based on *16S rRNA* sequences. FEMS Microbiol Lett. **113:**87-92.

Lee, J. V., and A. A. West. 1991. Survival and growth of *Legionella* species in the environment. J Appl Bacteriol. **70:**121S-129S.

Lehman, P. F., D. Lin, and B. A. Lasker. 1992. Genotypic identification and characterisation of species within the genus *Candida* by using random amplified polymorphic DNA. J Clin Microbiol. **30:**3249-3254.

Liles, M. R., P. H. Edelstein, and N. P. Cianciotto. 1999. The prepilin peptidase is required for protein secretion by and the virulence of the intracellular pathogen *Legionella pneumophila*. Mol Microbiol. **31:**959-970.

Liles, M. R., V. K. Viswanathan, and N. P. Cianciotto. 1998. Identification and temperature regulation of *Legionella pneumophila* genes involved in type IV pilus biogenesis and type II protein secretion. Infect Immun. **66:**1776-1782.

Lim, I., N. Sangster, D. P. Reid, and J. A. Lanser. 1989. *Legionella longbeachae* pneumonia: a report of two cases. Med J Aust. **150:**599-601.

Lin, Y. E., R. D. Vidic, J. E. Stout, and V. L. Yu. 1998. *Legionella* in water distribution systems. J American Water Works Association. **90:**112-121.

Lo Presti, F., S. Riffard, F. Vandenesch, M. Reyrolle, E. Ronco, P. Ichai, and J. Etienne. 1997. The first clinical isolate of *Legionella parisiensis*, from a liver transplant patient with pneumonia. J Clin Microbiol. **35:**1706-1709.

Lo Presti, F., S. Riffard, S. Jarraud, F. Le Gallou, H. Richet, F. Vandenesch, and J. Etienne. 1998a. Two first human isolates of *Legionella oakridgensis*: identification and epidemiological investigation. European Working Group on *Legionella* Infections 13th Meeting, Helsinki, Finland, May 31 - June 2, proffered paper.

Lo Presti, F., S. Riffard, F. Vandenesch, and J. Etienne. 1998b. Identification of *Legionella* species by random amplified polymorphic DNA profiles. J Clin Microbiol. **36:**3193-3197.

Lo Presti, F., S. Riffard, H. Meugnier, M. Reyrolle, Y. Lasne, P. A. D. Grimont, F. Grimont, F. Vandenesch, J. Etienne, F. Fleurette, and J. Freney. 1999. *Legionella taurinensis* sp. nov., a new species antigenically similar to *Legionella spiritensis*. Int J Syst Bacteriol. **49:**397-403.

Ludwig, W., and K. H. Schleifer. 1994. Bacterial phylogeny based on *16S* and *23S rRNA* sequence analysis. FEMS Microbiol Rev. **15:**155-173.

Ludwig, B., J. Rahfeld, B. Schmidt, K. Mann, E. Wintermeyer, G. Fischer, and J. Hacker. 1994. Characterization of Mip proteins of *Legionella pneumophila*. FEMS Microbiol Lett. **118:**23-30.

Ludwig, W., J. Neumaier, N. Klugbauer, E. Brockmann, C. Roller, S. Jilg, K. Reetz, I. Schachtner, A. Ludvigsen, and M. Bachleitner. 1993. Phylogenetic relationships of bacteria based on comparative sequence analysis of elongation factor Tu and ATP-synthase beta-subunit genes. Antonie Van Leeuwenhoek. **64:**285-305.

Ludwig, W., O. Strunk, S. Klugbauer, N. Klugbauer, M. Weizenegger, J. Neumaier, M. Bachleitner, and K. H. Schleifer. 1998. Bacterial phylogeny based on comparative sequence analysis. Electrophoresis. **19:**554-568.

Lugtenberg, B., J. Meijers, R. Peters, P. van der Hoek, and L. van Alphen. 1975. Electrophoretic resolution of the "major outer membrane protein" of *Escherichia coli* K-122 into four bands. FEBS Lett. **58:**254-258.

Lundemose, A. G., D. A. Rouch, S. Birkelund, G. Christiansen, and J. H. Pearce. 1992. *Chlamydia trachomatis* Mip-like protein. Mol Microbiol. **6:**2539-2548.

Mahoney, F. J., C. W. Hoge, T. A. Farley, J. M. Barbaree, R. F. Breiman, R. F. Benson, and L. M. McFarland. 1992. Communitywide outbreak of Legionnaires' disease associated with a grocery store mist machine. J Infect Dis. **165:**736-739.

Maiden, M. C., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci USA. **95:**3140-3145.

Maiwald, M., J. H. Helbig, and P. C. Lück. 1998. Laboratory methods for the diagnosis of *Legionella* infections. J Microbiol Method. **33:**59-79.

Majewski, J., and F. M. Cohan. 1999. Adapt globally, act locally: the effect of selective sweeps on bacterial sequence diversity. Genetics. **152:**1459-1474.

Mangione, E. J., R. S. Remis, K. A. Tait, H. B. McGee, G. W. Gorman, B. B. Wentworth, P. A. Baron, A. W. Hightower, J. M. Barbaree, and C. V. Broome. 1985. An outbreak of Pontiac fever related to whirlpool use, Michigan 1982. JAMA. **253:**535-539.

Manning, P. A., M. W. Heuzenroeder, J. Yeadon, D. I. Leavesley, P. R. Reeves, and D. Rowley. 1986. Molecular cloning and expression in *Escherichia coli* K-12 of the O antigens of the Inaba and Ogawa serotypes of the *Vibrio cholerae* O1 lipopolysaccharides and their potential for vaccine development. Infect Immun. **53:**272-277.

Marra, A., S. J. Blander, M. A. Horwitz, and H. A. Shuman. 1992. Identification of a *Legionella pneumophila* locus required for intracellular multiplication in human macrophages. Proc Natl Acad Sci USA. **89:**9607-9611.

Marshall, W. A. 1996. Biological particles over Antarctica. Nature. **383:**680.

Marshall, W. A. 1997. Seasonality in the Antarctic airborn fungal spores. Appl Environ Microbiol. **63:**2240-2245.

Marston, B. J., H. B. Lipman, and R. F. Breiman. 1994. Surveillance for Legionnaires' disease. Arch Intern Med. **154:**2417-2422.

Maruta, K., M. Ogawa, H. Miyamoto, K. Izu, and S. Yoshida. 1998. Entry and intracellular localization of *Legionella dumoffii* in Vero cells. Microbial Pathogen. **24:**65-73.

Mauchline, W. S., and C. W. Keevil. 1991. Development of the BIOLOG substrate utilization system for identification of *Legionella* spp. Appl Environ Microbiol. **57:**3345-3349.

Maynard-Smith, J. 1970. Do bacteria have population genetics? Symp Soc Gen Microbiol. **52:**1-12.

McDade, J. E., and C. C. Shepard. 1979. Virulent to avirulent conversion of Legionnaires' disease bacterium (*Legionella pneumophila*) - its effect on isolation techniques. J Infect Dis. **139:**707-711.

McDade, J. E., C. C. Shepard, D. W. Fraser, T. R. Tsai, M. A. Redus, W. R. Dowdle, and the *Legionella* Ivestigation Team. 1977. Legionnaires' Disease: isolation of a bacterium and demonstration of its role in other respiratory disease. N Engl J Med. **297:**1197-1203.

McDade, J. E., D. L. Brenner, and F. M. Bozeman. 1979. Legionnaires' disease bacterium isolated in 1947. Ann Intern Med **90:**659-661

McKinney, R. M., R. K. Porschen, P. H. Edelstein, M. L. Bissett, P. P. Harris, S. P. Bondell, A. G. Steigerwalt, R. E. Weaver, M. E. Ein, D. S. Lindquist, R. S. Kops, and D. J. Brenner. 1981. *Legionella longbeachae* species nova, another etiologic agent of human pneumonia. Ann Intern Med. **94:**739-743.

Meier, C. 1935. Collecting micro-organisms from the arctic atmosphere. With field notes and material by Charles A. Lindberg. Sci Mon, NY. **40:**5.

Mintz, C. S., B. S. Fields, and C. H. Zou. 1992. Isolation and characterization of a conjugative plasmid from *Legionella pneumophila*. J Gen Microbiol. **138:**1379-1386.

Mo, Y.-Y., N. P. Cianciotto, and L. P. Mallavia. 1995. Molecular cloning of a *Coxiella burnetii* gene encoding a macrophage infectivity potentiator (Mip) analogue. Microbiology. **141:**2861-2871.

Moffat, J. F., W. J. Black, and L. S. Tompkins. 1994a. Further molecular characterization of the cloned *Legionella pneumophila* zinc metalloprotease. Infect Immun. **62:**751-753.

Moffat, J. F., P. H. Edelstein, D. P. Regula, Jr., J. D. Cirillo, and L. S. Tompkins. 1994b. Effects of an isogenic Zn-metalloprotease-deficient mutant of *Legionella pneumophila* in a guinea-pig pneumonia model. Mol Microbiol. **12:**693-705.

Montanaro-Punzengruber, J. C., L. Hicks, W. Meyer, and G. L. Gilbert. 1999. Australian isolates of *Legionella longbeachae* are not a clonal population. J Clin Microbiol. **37:**3249-3254.

Moro, A., F. Ruiz-Cabello, A. Fernandez-Cano, R. P. Stock, and A. Gonzalez. 1995. Secretion by *Trypanosoma cruzi* of a peptidyl-prolyl *cis/trans* isomerase involved in cell infection. EMBO J. **14:**2483-2490.

Morris, G. K., A. Steigerwalt, J. C. Feeley, E. S. Wong, W. T. Martin, C. M. Patton, and D. J. Brenner. 1980. *Legionella gormanii* sp. nov. J Clin Microbiol. **12:**718-721.

Moss, C. W., W. F. Bibb, D. E. Karr, G. O. Guerrant, and M. A. Lambert. 1983. Cellular fatty acid composition and ubiquinone content of *Legionella feeleii* sp. nov. J Clin Microbiol. **18:**917-919.

Muder, R. R., V. L. Yu, and A. H. Woo. 1986. Mode of transmission of *Legionella pneumophila*. A critical review. Arch Intern Med. **146:**1607-1612.

Muller, A., J. Hacker, and B. C. Brand. 1996. Evidence for apoptosis of human macrophage-like HL-60 cells by *Legionella pneumophila* infection. Infect Immun. **64:**4900-4906.

Nei, M. 1996. Phylogenetic analysis in molecular evolutionary genetics. Annu Rev Genet. **30:**371-403.

Nelson, K., and R. K. Selander. 1994. Intergeneric transfer and recombination of the 6-phosphogluconate dehydrogenase gene (*gnd*) in enteric bacteria. Proc Natl Acad Sci USA. **91:**10227-10231.

Newhall, W. J., and R. B. Jones. 1983. Disulphide-linked oligomers of the major outer membrane protein of chlamydiae. J Bacteriol. **154:**998-1001.

O'Brien, S. J., and R. S. Bhopal. 1993. Legionnaires' disease: the infective dose paradox. Lancet. **342:**5-6.

O'Brien, T. F., K. H. Mayer, J. D. Hopkins, and L. Chao. 1983. New concepts in medical bacteriology. J Med. **14:**271-280.

Ochi, K. 1995. Comparative ribosomal protein sequence analyses of a phylogenetically defined genus, *Pseudomonas*, and its relatives. Int J Syst Bacteriol. **45:**268-273.

O'Connell, W. A., J. M. Bangsborg, and N. P. Cianciotto. 1995. Characterization of a *Legionella micdadei mip* mutant. Infect Immun. **63:**2840-2845.

O'Connell, W. A., L. Dhand, and N. P. Cianciotto. 1996. Infection of macrophage-like cells by *Legionella* species that have not been associated with disease. Infect Immun. **64:**4381-4384.

Olsen, G. L., H. Natsuda, R. Hagstrom, and R. Overbeek. 1994. FastDNAML: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. CABIOS. **10:**41-48.

Orrison, L. H., W. B. Cherry, R. L. Tyndall, C. B. Fliermans, S. B. Gough, M. A. Lambert, L. K. McDougal, W. F. Bibb, and D. J. Brenner. 1983. *Legionella oakridgensis*: unusual new species isolated from cooling tower water. Appl Environ Microbiol. **45:**536-545.

Page, R. D. M., and E. C. Holmes. 1998. Molecular evolution. A phylogenetic approach. Blackwell Science Ltd, Oxford, England.

Pahl, A., and U. Keller. 1992. FK-506-binding proteins from streptomycetes producing immunosuppressive macrolactones of the FK-506 type. J Bacteriol. **174:**5888-5894.

Palys, T., L. K. Nakamura, and F. M. Cohan. 1997. Discovery and classification of ecological diversity in the bacterial world: the role of DNA sequence data. Int J Syst Bacteriol. **47:**1145-1156.

Parker, J. M., D. Guo, and R. S. Hodges. 1986. New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessible sites. Biochemistry. **25:**5425-5432.

Pascual, C., P. A. Lawson, J. A. Farrow, M. N. Gimenez, and M. D. Collins. 1995. Phylogenetic analysis of the genus *Corynebacterium* based on *16S rRNA* gene sequences. Int J Syst Bacteriol. **45:**724-728.

Paton, A. W., R. M. Ratcliff, R. M. Doyle, J. Seymour Murray, D. Davos, J. A. Lanser, and J. C. Paton. 1996. Molecular microbiological investigation of an outbreak of hemolytic-uremic syndrome caused by dry fermented sausage contaminated with Shiga-like toxin-producing *Escherichia coli*. J Clin Microbiol. **34:**1622-1627.

Pearlman, E., N. C. Engleberg, and B. I. Eisenstein. 1985. Identification of protein antigens of *Legionella pneumophila* serogroup 1. Infect Immun. **47:**74-79.

Petty, H. R., and R. F. Todd. 1993. Receptor-receptor interactions of complement receptor type 3 in neutrophil membranes. J Leukocyte Biol. **54:**492-494.

Pissavin, C., and N. Hugouvieux-Cotte-Pattat. 1997. Characterization of a periplasmic peptidyl-prolyl *cis/trans* isomerase in *Erwinia chrysanthemi*. FEMS Microbiol Let. **157:**59-65.

Poyart, C., G. Quesne, S. Coulon, P. Berche, and P. Trieu-Cuot. 1998. Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. J Clin Microbiol. **36:**41-47.

Pruckler, J. M., R. F. Benson, M. Moyenuddin, W. T. Martin, and B. S. Fields. 1995. Association of flagellum expression and intracellular growth of *Legionella pneumophila.* Infect Immun. **63:**4928-4932.

Purcell, M., and H. A. Shuman. 1998. The *Legionella pneumophila icmGCDJBF* genes are required for killing of human macrophages. Infect Immun. **66:**2245-2255.

Quinn, F. D., and L. S. Tompkins. 1989. Analysis of a cloned sequence of *Legionella pneumophila* encoding a 38-kDa metalloprotease possessing haemolytic and cytotoxic activities. Mol Microbiol. **3:**797-805.

Rahfeld, J. U., K. P. Rücknagel, B. Schelbert, B. Ludwig, J. Hacker, K. Mann, and G. Fischer. 1994a. Confirmation of a the existence of a third family among peptidyl-prolyl *cis/trans* isomerases. Amino acid sequence and recombinant production of Parvulins. FEBS Lett. **352:**180-184.

Rahfeld, J. U., A. Schierhorn, K. Mann, and G. Fischer. 1994b. A novel peptidyl-prolyl *cis/trans* isomerase from *Escherichia coli*. FEBS Lett. **343:**65-69.

Rahfeld, J. U., K. P. Rücknagel, G. Stoller, S. M. Horne, A. Schierhorn, K. D. Young, and G. Fischer. 1996. Isolation and amino acid sequence of a new 22-kDa FKBP-like

peptidyl-prolyl *cis/trans* isomerase of *Escherichia coli*. Similarity to Mip-like proteins of pathogenic bacteria. J Biol Chem. **271:**22130-22138.

Ralph, D., M. McClelland, J. Welsh, G. Baranton, and P. Perolat. 1993. Leptospira species characterised by arbitrarily primed polymerase chain reaction (PCR) and by mapped restriction polymorphisms in PCR-amplified *rRNA* genes. J Bacteriol. **175:**973-981.

Ratcliff, R. M., J. A. Lanser, P. A. Manning, and M. W. Heuzenroeder. 1998. Sequence-based classification scheme for the genus *Legionella* targeting the *mip* gene. J Clin Microbiol. **36:**1560-1567.

Rechnitzer, C., and M. Blom. 1989. Engulfment of the Philadelphia strain of *Legionella pneumophila* associated with pseudopod coils on human phagocytes. Comparison with other *Legionella* strains and species. Acta Pathol Microbiol Immunol Scand. **97:**105-114.

Rechnitzer, C., and A. Kharazmi. 1992. Effect of *Legionella pneumophila* cytotoxic protease on human neutrophil and monocyte function. Microb Pathog. **12:**115-125.

Rechnitzer, C., A. Williams, J. B. Wright, A. B. Dowsett, N. Milman, and R. B. Fitzgeorge. 1992. Demonstration of the intracellular production of tissue-destructive protease by *Legionella pneumophila* multiplying within guinea-pig and human alveolar macrophages. J Gen Microbiol. **138:**1671-1677.

Retzlaff, C., Y. Yamamoto, S. Okubo, P. S. Hoffman, H. Friedman, and T. W. Klein. 1996. *Legionella pneumophila* heat-shock protein-induced increase of interleukin-1 beta mRNA involves protein kinase C signalling in macrophages. Immunology. **89:**281-288.

Riffard, S., F. Vandenesch, M. Reyrolle, and J. Etienne. 1996. Distribution of *mip*-related sequences in 39 species (48 serogroups) of *Legionellaceae*. Epidemiol Infect. **117:**501-506.

Riffard, S., F. Lo Presti, F. Vandenesch, F. Forey, M. Reyrolle, and J. Etienne. 1998. Comparative analysis of infrequent-restriction-site PCR and pulsed-field gel electrophoresis for epidemiological typing of *Legionella pneumophila* serogroup 1 strains. J Clin Microbiol. **36:**161-167.

Rittig, M. G., A. Krause, T. Haupl, U. E. Schaible, M. Modolell, M. D. Kramer, E. Lutjen-Drecoll, M. M. Simon, and G. R. Burmester. 1992. Coiling phagocytosis is the preferential phagocytic mechanism for *Borrelia burgdorferi*. Infect Immun. **60:**4205-4212.

Rittig, M. G., G. Burmester, and A. Krause. 1998a. Coiling phagocytosis: when the zipper jams, the cup is deformed. Trends Microbiol. **6:**384-388.

Rittig, M. G., K. Schröppel, K.-H. Seack, U. Sander, E.-N. N'Diaye, I. Maridonneau-Parini, W. Solbach, and C. Bogdan. 1998b. Coiling phagocytosis of trypanosomatids and fungal cells. Infect Immun. **66:**4331-4339.

Rittig, M. G., B. Wilske, and A. Krause. 1999. Phagocytosis of microorganisms by means of overshooting pseudopods: where do we stand? Microbes Infect. **1:**727-735.

Roberts, M. S., and F. M. Cohan. 1995. Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. Evolution. **49:**1081-1094.

Rockey, D. D., B. B. Chesebro, R. A. Heinzen, and T. Hackstadt. 1996. A 28-kDa major immunogen of *Chlamydia psittaci* shares identity with Mip proteins of *Legionella* spp. and

Chlamydia trachomatis-cloning and characterization of the *C. psittaci mip*-like gene. Microbiology. **142:**945-953.

Rodgers, F. G., P. W. Greaves, A. D. Macrae, and M. J. Lewis. 1980. Electron microscopic evidence of flagella and pili on *Legionella pneumophila*. J Clin Pathol. **33:**1184-1188.

Rodriguez, F., J. F. Oliver, A. Marín, and J. R. Medina. 1990. The general stochastic model of nucleotide substitutions. JTheor Biol. **142:**485-501.

Rogerson, A., and A. Detwiler. 1999. Abundance of airborne heterophile protists in ground level air of South Dakota. Atmospheric Research. **51:**35-44.

Rossler, D., W. Ludwig, K. H. Schleifer, C. Lin, T. J. McGill, J. D. Wisotzkey, J. R. Jurtshuk, and G. E. Fox. 1991. Phylogenetic diversity in the genus *Bacillus* as seen by *16S RNA* sequencing studies. System Appl Microbiol. **14:**266-269.

Rowbotham, T. J. 1980. Preliminary report on the pathogenicity of *Legionella pneumophila* for freshwater and soil amoebae. J Clin Pathol. **33:**1179-1183.

Rowbotham, T. J. 1983. Isolation of *Legionella pneumophila* from clinical specimens via amoebae, and the interaction of those and other isolates with amoebae. J Clin Pathol. **36:**978-986.

Rowbotham, T. J. 1986. Current views on the relationships between amoebae, *Legionellae* and man. Isr J Med Sci. **22:**678-689.

Roy, T. M., D. Fleming, and W. H. Anderson. 1989. Tularemic pneumonia mimicking Legionnaires' disease with false-positive direct fluorescent antibody stains for *Legionella*. South Med J. **82:**1429-1431.

Rudd, K. E., H. J. Sofia, and E. V. Koonin. 1995. A new family of peptidyl-prolyl isomerases. Trends Biochem. **20:**12-14.

Ruimy, R., V. Breittmayer, P. Elbaze, B. Lafay, O. Boussemart, M. Gauthier, and R. Christen. 1994. Phylogenetic analysis and assessment of the genera *Vibrio*, *Photobacterium*, *Aeromonas*, and *Plesiomonas* deduced from small-subunit *rRNA* sequences. Int J Syst Bacteriol. **44:**416-426.

Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. **239:**487-491.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol. **4:**406-425.

Sallen, B., A. Rajoharison, S. Desvarenne, F. Quinn, and C. Mabilat. 1996. Comparative analysis of *16S* and *23S rRNA* sequences of *Listeria* species. Int J Syst Bacteriol. **46:**669-674.

Salvesen, G. S., and V. M. Dixit. 1998. Capsases: Intracellular signaling by proteolysis. Cell. **91:**443-446.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989a. Appendix A: Bacterial media, antibiotics and bacterial strains, pp. A1-A13, Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbour Laboratory Press, USA.

Sambrook, J., E. F. Fritsch, and T. Maniatis. 1989b. Appendix E: Commonly used techniques in molecular cloning, p E.5, Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbour Laboratory Press, USA.

Saunders, N. A., T. G. Harrison, A. Haththotuwa, N. Kachwalla, and A. G. Taylor. 1990. A method for typing strains of *Legionella pneumophila* serogroup 1 by analysis of restriction fragment length polymorphisms. J Med Microbiol. **31:**45-55.

Savelkoul, P. H. M., H. J. M. Aarts, J. de Haas, L. Dijkshoorn, B. Duim, M. Otsen, J. L. W. Rademaker, L. Schouls, and J. A. Lenstra. 1999. Amplified-fragment length ploymorphism analysis: the state of the art. J Clin Microbiol. **37:**3083-3091.

Sawadogo, M., and M. W. Van Dyke. 1991. A rapid method for the purification of deprotected oligodeoxynucleotides. Nucleic Acids Res. **19:**74.

Schmidt, B., J. Rahfeld, A. Schierhorn, B. Ludwig, J. Hacker, and G. Fischer. 1994. A homodimer represents an active species of the peptidyl-prolyl *cis/trans* isomerase FKBP25mem from *Legionella pneumophila*. FEBS Lett. **352:**185-190.

Schmidt, B., S. König, D. Svergun, V. Volkov, G. Fischer, and M. H. Koch. 1995. Small-angle X-ray solution scattering study on the dimerization of the FKBP25mem from *Legionella pneumophila*. FEBS Lett. **372:**169-172.

Schmidt, B., T. Tradler, J. U. Rahfeld, B. Ludwig, B. Jain, K. Mann, K. P. Rücknagel, B. Janowski, A. Schierhorn, G. Kullertz, J. Hacker, and G. Fischer. 1996. A cyclophilin-like peptidyl-prolyl *cis/trans* isomerase from *Legionella pneumophila* - characterization, molecular cloning and overexpression. Mol Microbiol. **21:**1147-1160.

Schoonmaker, D., T. Heimberger, and G. Birkhead. 1992. Comparison of ribotyping and restriction enzyme analysis using pulsed-field gel electrophoresis for distinguishing *Legionella pneumophila* isolates obtained during a nosocomial outbreak. J Clin Microbiol. **30:**1491-1498.

Schreiber, S. L. 1991. Chemistry and biology of the immunophilins and their immunosuppressive ligands. Science. **251:**283-287.

Segal, G., and H. A. Shuman. 1997. Characterization of a new region required for macrophage killing by *Legionella pneumophila*. Infect Immun. **65:**5057-5066.

Segal, G., and H. A. Shuman. 1998. How is the intracellular fate of the *Legionella pneumophila* phagosome determined. Trends Microbiol. **6:**253-255.

Segal, G., and H. A. Shuman. 1999. *Legionella pneumophila* utilizes the same genes to multiply within *Acanthamoeba castellanii* and human macrophages. Infect Immun. **67:**2117-2124.

Segal, G., M. Purcell, and H. A. Shuman. 1998. Host cell killing and bacterial conjugation require overlapping sets of genes within a 22-kb region of the *Legionella pneumophila* genome. Proc Natl Acad Sci USA. **95:**1669-1674.

Segal, G., J. J. Russo, and H. A. Shuman. 1999. Relationships between a new type IV secretion system and the *icm/dot* virulence system of *Legionella pneumophila*. Mol Microbiol. **34:**799-809.

Selander, R. K., P. Beltran, and N. H. Smith. 1991. Evolutionary genetics of *Salmonella*, pp. 25-57. *In* R. K. Selander, A. G. Clark, and T. S. Whittam (eds.), Evolution at the Molecular Level. Sinauer Associates, Sunderland, MA., USA.

Shen, P., and H. V. Huang. 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics. **112:**441-457.

Shuman, H. A., M. Purcell, G. S. Hales, and L. A. Wiater. 1998. Intracellular multiplication of *Legionella pneumophila*: human pathogen or accidental tourist? Curr Top Microbiol Immunol. **225:**99-112.

Sneath, P. H. A., and R. R. Sokal. 1973. Numerical taxonomy. W. H. Freeman, San Francisco, CA, USA.

Sober, E. 1983. Parsimony in systematics: philosophical issues. Annu Rev Ecol Syst. **14:**335-357.

Sober, E. 1989. Restructuring the past: parsimony, evolution and inference. MIT Press, Cambridge, MA, USA.

Sonesson, A., H. Moll, E. Jantzen, and U. Zähringer. 1993. Long-chain alpha-hydroxy-(omega-1)-oxo fatty acids and alpha-hydroxy-1, omega-dioic fatty acids are cell wall

constituents of *Legionella* (*L. jordanis*, *L. maceachernii* and *L. micdadei*). FEMS Microbiol Lett. **106:**315-320.

Sonesson, A., E. Jantzen, K. Bryn, T. Tangen, J. Eng, and U. Zähringer. 1994a. Composition of 2,3-dihydroxy fatty acid-containing lipopolysaccharides from *Legionella israelensis*, *Legionella maceachernii* and *Legionella micdadei*. Microbiology. **140:**1261-1271.

Sonesson, A., E. Jantzen, T. Tangen, and U. Zähringer. 1994b. Chemical characterization of lipopolysaccharides from *Legionella feeleii*, *Legionella hackeliae* and *Legionella jordanis*. Microbiology. **140:**2663-2671.

Sonesson, A., E. Jantzen, T. Tangen, and U. Zähringer. 1994c. Chemical composition of lipopolysaccharides from *Legionella bozemanii* and *Legionella longbeachae*. Arch Microbiol. **162:**215-221.

Sonesson, A., E. Jantzen, T. Tangen, and U. Zähringer. 1994d. Lipopolysaccharides of *Legionella erythra* and *Legionella oakridgensis*. Can J Microbiol. **40:**666-671.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol. **98:**503-517.

Spitalny, K. C., R. L. Vogt, L. A. Orciari, L. E. Witherell, P. Etkind, and L. F. Novick. 1984a. Pontiac fever associated with a whirlpool spa. Am J Epidemiol. **120:**809-817.

Spitalny, K. C., R. L. Vogt, and L. E. Witherell. 1984b. National survey on outbreaks associated with whirlpool spas. Am J Public Health. **74:**725-726.

Spratt, B. G., and M. C. Maiden. 1999. Bacterial population genetics, evolution and epidemiology. Philos Trans R Soc Lond B Biol Sci. **354:**701-710.

Spratt, B. G., N. H. Smith, J. Zhou, M. O'Rourke, and E. Feil. 1995. The population genetics of the pathogenic *Neisseria*. Symp Soc Gen Microbiol. **52:**143-160.

Springer, N., W. Ludwig, R. Amann, H. J. Schmidt, H. D. Gortz, and K. H. Schleifer. 1993. Occurrence of fragmented *16S rRNA* in an obligate bacterial endosymbiont of *Paramecium caudatum*. Proc Natl Acad Sci USA. **90:**9892-9895.

Steele, T. W., and A. M. McLennan. 1996. Infection of *Tetrahymena pyriformis* by *Legionella longbeachae* and other *Legionella* species found in potting mixes. Appl Environ Microbiol. **62:**1081-1083.

Steele, T. W., J. Lanser, and N. Sangster. 1990a. Isolation of *Legionella longbeachae* serogroup 1 from potting mixes. Appl Environ Microbiol. **56:**49-53.

Steele, T. W., C. V. Moore, and N. Sangster. 1990b. Distribution of *Legionella longbeachae* serogroup 1 and other *Legionellae i*n potting soils in Australia. Appl Environ Microbiol. **56:**2984-2988.

Steinert, M., L. Emody, R. Amann, and J. Hacker. 1997. Resuscitation of viable but nonculturable *Legionella pneumophila* Philadelphia JR32 by *Acanthamoeba castellanii*. Appl Environ Microbiol. **63:**2047-2053.

Steinmetz, I., C. Rheinheimer, I. Hübner, and D. Bitter-Suermann. 1991. Genus-specific epitope on the 60-kilodalton *Legionella* heat shock protein recognised by monoclonal antibody. J Clin Microbiol. **29:**346-354.

Stone, B. J., and Y. Abu Kwaik. 1998. Expression of multiple pili by *Legionella pneumophila*: identification and characterization of a type IV pilin gene and its role in adherence to mamalian and protozoan cells. Infect Immun. **66:**1768-1775.

Stout, J. E., and V. L. Yu. 1997. Legionellosis. N Engl J Med. **337:**682-687.

Strätz, M., M. Mau, and K. N. Timmis. 1996. System to study horizontal gene exchange among microorganisms without cultivation of recipients. Mol Microbiol. **22:**207-215.

Strimmer, K., and A. von Haesler. 1996. Quartet Puzzling: A quartet maximum-likelihood method for reconstructing tree topologies. Mol Biol Evol. **13:**964-969.

Strimmer, K., N. Goldman, and A. von Haeseler. 1997. Bayesian probabilities and Quartet Puzzling. Mol Biol Evol. **14:**210-211.

Strom, M. S., and S. Lory. 1993. Structure-function and biogenesis of the type IV pili. Annu Rev Microbiol. **47:**565-596.

Summersgill, J. T., and J. W. Snyder. 1990. Merifluor-*Legionella* cross-reactivity [letter]. J Clin Microbiol. **28:**1680.

Susa, M., J. Hacker, and R. Marre. 1996. De novo synthesis of *Legionella pneumophila* antigens during intracellular growth in phagocytic cells. Infect Immun. **64:**1679-1684.

Swanson, M. S., and R. R. Isberg. 1993. Formation of the *Legionella pneumophila* replicative phagosome. Infect Agents Dis. **2:**224-226.

Swofford, D. 1993. PAUP - Phylogenetic Analysis Using Parsimony, 3.1 ed. Illinois Natural History Survey, Champaign, IL, USA.

Swofford, D. L., G. J. Olsen, P. J. Waddell, and D. M. Hillis. 1996. Phylogenetic Inference, pp. 407-514. *In* D. M. Hillis, D. Moritz, and B. K. Mable (eds.), Molecular Systematics, 2nd ed. Sinauer Associates, Sunderland, MA, USA.

Szeto, L., and H. A. Shuman. 1990. The *Legionella pneumophila* major secretory protein, a protease, is not required for intracellular growth or cell killing. Infect Immun. **58:**2585-2592.

Tajima, F., and M. Nei. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol. **1:**269-285.

Terranova, W., M. L. Cohen, and D. W. Fraser. 1978. 1974 outbreak of Legionnaires' disease diagnosed in 1977. Clinical and epidemiological features. Lancet **15:**122-124.

Thacker, S. B., J. V. Bennett, T. F. Tsai, D. W. Fraser, J. E. McDade, C. C. Shepard, K. H. Williams Jr., W. H. Stuart, H. B. Dull, and T. C. Eickhoff. 1978. An outbreak in 1965 of severe respiratory illness caused by the Legionnaires' disease bacterium. J Infect Dis **138:**512-519.

Thacker, W. L., H. W. Wilkinson, B. B. Plikaytis, A. G. Steigerwalt, W. R. Mayberry, C. W. Moss, and D. J. Brenner. 1985. Second serogroup of *Legionella feeleii* strains isolated from humans. J Clin Microbiol. **22:**1-4.

Thacker, W. L., R. F. Benson, J. L. Staneck, S. R. Vincent, W. R. Mayberry, D. J. Brenner, and H. W. Wilkinson. 1988. *Legionella cincinnatiensis* sp. nov. isolated from a patient with pneumonia. J Clin Microbiol. **26:**418-420.

Thacker, W. L., R. F. Benson, R. B. Schifman, E. Pugh, A. G. Steigerwalt, W. R. Mayberry, D. J. Brenner, and H. W. Wilkinson. 1989. *Legionella tucsonensis* sp. nov. isolated from a renal transplant recipient. J Clin Microbiol. **27:**1831-1834.

Thacker, W. L., R. F. Benson, L. Hawes, W. R. Mayberry, and D. J. Brenner. 1990. Characterization of a *Legionella anisa* strain isolated from a patient with pneumonia. J Clin Microbiol. **28:**122-123.

Thacker, W. L., R. F. Benson, L. Hawes, H. Gidding, B. Dwyer, W. R. Mayberry, and D. J. Brenner. 1991. *Legionella fairfieldensis* sp. nov. isolated from cooling tower waters in Australia. J Clin Microbiol. **29:**475-478.

Thacker, W. L., J. W. Dyke, R. F. Benson, D. H. Havlichek, Jr., B. Robinson Dunn, H. Stiefel, W. Schneider, C. W. Moss, W. R. Mayberry, and D. J. Brenner. 1992. *Legionella lansingensis* sp. nov. isolated from a patient with pneumonia and underlying chronic lymphocytic leukemia. J Clin Microbiol. **30:**2398-2401.

Thornton, J. M., M. S. Edwards, W. R. Tayler, and D. J. Barlow. 1986. Location of "continuous" antigenic determinants in the protruding regions of proteins. EMBO J. **5:**409-413.

Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitroellulose sheets: procedure and some applications. Proc Natl Acad Sci USA. **74:**4340-4354.

Utaker, J. B., L. Bakken, Q. Q. Jiang, and I. F. Ness. 1995. Phylogenetic analysis of seven new isolates of ammonia-oxidising bacteria based on *16S RNA* gene sequences. System Appl Microbiol. **18:**549-559.

Valsangiacomo, C., F. Baggi, V. Gaia, T. Balmelli, R. Peduzzi, and J. C. Piffaretti. 1995. Use of amplified fragment length polymorphism in molecular typing of *Legionella pneumophila* and application to epidemiological studies. J Clin Microbiol. **33:**1716-1719.

Vandamme, P., B. Pot, M. Gillis, P. De Vos, K. Kersters, and J. Swings. 1996. Polyphasic taxonomy, a consensus approach to bacterial systematics. Microbiol Rev. **60:**407-438.

Vauterin, L., B. Hoste, K. Kersters, and J. Swings. 1995. Reclassification of *Xanthomonas*. Int J Syst Bacteriol. **45:**472-489.

Venkataraman, C., B. J. Haack, S. Bondada, and Y. Abu Kwaik. 1997. Identification of a Gal/Gal/NAc lectin in the protozoan *Hartmannella vermiformis* as a potential receptor for attachment and invasion by the Legionnaires' disease bacterium. J Exp Med. **186:**537-547.

Venkataraman, C., L.-Y. Gao, S. Bondada, and Y. A. Kwaik. 1998. Identification of putative cytoskeletal protein homologues in the protozoan host *Hartmannella vermiformis* as substrates for induced tyrosine phosphatase activity upon attachment to the Legionnaires' disease bacterium, *Legionella pneumophila*. J Exp Med. **188:**505-514.

Veríssimo, A., P. V. Morais, A. Diogo, C. Gomes, and M. S. da Costa. 1999. Characterization of *Legionella* species by numerical analysis of whole-cell protein electrophoresis. Int J Syst Bacteriol. **46:**41-49.

Verma, U. K., D. J. Brenner, W. L. Thacker, R. F. Benson, G. Vesey, J. B. Kurtz, P. J. Dennis, A. G. Steigerwalt, J. S. Robinson, and C. W. Moss. 1992. *Legionella shakespearei* sp. nov., isolated from cooling tower water. Int J Syst Bacteriol. **42:**404-407.

Vogel, J. P., C. Roy, and R. R. Isberg. 1996. Use of salt to isolate *Legionella pneumophila* mutants unable to replicate in macrophages. Ann N Y Acad Sci. **797:**271-272.

Vogel, J. P., H. L. Andrews, S. K. Wong, and R. R. Isberg. 1998. Conjugative transfer by the virulence system of *Legionella pneumophila*. Science. **279:**873-876.

Vogt, R. L., P. J. Hudson, L. Orciari, E. M. Heun, and T. C. Woods. 1987. Legionnaire's disease and a whirlpool-spa [letter]. Ann Intern Med. **107:**596.

von Heijne, G. 1981. On the hydrophobic nature of signal sequences. Eur J Biochem. **116:**419-422.

von Heijne, G. 1992. Membrane protein structure prediction. Hydrophobicity analysis and the positive-inside rule. J Mol Biol. **225:**487-494.

Wait, R. 1988. Confirmation of the identity of *Legionellae* by whole cell fatty-acid and isoprenoid quinone profiles, pp. 69-101. *In* T. G. Harrison and A. G. Taylor (eds.), A laboratory manual for *Legionella*. John Wiley & Sons Ltd., London, England.

Walker, C., and P. Weinstein. 1992. Review of legionellosis in South Australia 1990-91. Commun Dis Intell. **16:**70-71.

Wang, N., J. P. Butler, and D. E. Inberg. 1993. Mechanotransduction across the cell surface and through the cytoskeleton. Science. **260:**1124-1127.

Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int J Syst Bacteriol. **37:**463-464.

Weinbaum, D. L., R. R. Benner, J. N. Dowling, A. Alpern, A. W. Pasculle, and G. R. Donowitz. 1984. Interaction of *Legionella micdadei* with human monocytes. Infect Immun. **46:**68-73.

Wiater, L. A., K. Dunn, F. R. Maxfield, and H. A. Shuman. 1998. Early events in phagosome establishment are required for intracellular survival of *Legionella pneumophila*. Infect Immun. **66:**4450-4460.

Wilkinson, H. W. 1988. Reactions of *Legionella* species on biochemical tests, p 28. *In* H. W. Wilkinson (ed.), Hospital-Laboratory Diagnosis of *Legionella* infections. Centres for Disease Control amd Prevention, Atlanta, GA., USA.

Wilkinson, H. W., W. L. Thacker, D. J. Brenner, and K. J. Ryan. 1985a. Fatal *Legionella maceachernii* pneumonia. J Clin Microbiol. **22:**1055-1056.

Wilkinson, H. W., W. L. Thacker, A. G. Steigerwalt, D. J. Brenner, N. M. Ampel, and E. J. Wing. 1985b. Second serogroup of *Legionella hackeliae* isolated from a patient with pneumonia. J Clin Microbiol. **22:**488-489.

Wilkinson, H. W., W. L. Thacker, R. F. Benson, S. S. Polt, E. Brookings, W. R. Mayberry, D. J. Brenner, R. G. Gilley, and J. K. Kirklin. 1987. *Legionella birminghamensis* sp. nov. isolated from a cardiac transplant recipient. J Clin Microbiol. **25:**2120-2122.

Wilkinson, H. W., V. Drasar, W. L. Thacker, R. F. Benson, J. Schindler, B. Potuznikova, W. R. Mayberry, and D. J. Brenner. 1988. *Legionella moravica* sp. nov. and *Legionella brunensis* sp. nov. isolated from cooling-tower water. Ann Inst Pasteur Microbiol. **139:**393-402.

Wilkinson, I. J., N. Sangster, R. M. Ratcliff, P. A. Mugg, D. E. Davos, and J. A. Lanser. 1990. Problems associated with identification of *Legionella* species from the environment and isolation of six possible new species. Appl Environ Microbiol. **56:**796-802.

Winans, S. C., D. L. Burns, and P. J. Christie. 1996. Adaption of a conjugal transfer system for the export of pathogenic macromolecules. Trends Microbiol. **4:**64-68.

Winiecka-Krusnell, J., and E. Linder. 1999. Free-living amoebae protecting *Legionella* in Water: the tip of an iceberg? Scand J Infect Dis. **31:**383-385.

Winiecka-Krusnell, J., E. Linder, M. Lundholm, E. Hjelm, and H. Hallander. 1998. Association between the presence of free-living amoebae and *L. pneumophila* in the hospital water distribution system. Presented at the Annual Meeting of Swedish Physician Society, Gothenburg, Sweden, 24th-26th Nov.

Winn, W. C. 1995. *Legionella*, pp. 533-544. *In* P. R. Murray, E. J. Baron, M. A. Pfaller, F. C. Tenover, and R. Y. Yolken (eds.), Manual of Clinical Microbiology, 6th ed. American Society for Microbiology, Washington, DC, USA.

Winn, W. C., and R. L. Myerowitz. 1981. The pathology of the *Legionella* pneumonias. A review of 74 cases and the literature. Hum Pathol. **12:**401-422.

Wintermeyer, E., U. Rdest, B. Ludwig, A. Debes, and J. Hacker. 1991. Characterization of legiolysin (*lly*), responsible for haemolytic activity, colour production and fluorescence of *Legionella pneumophila*. Mol Microbiol. **5:**1135-1143.

Wintermeyer, E., M. Flügel, M. Ott, M. Steinert, U. Rdest, K. H. Mann, and J. Hacker. 1994. Sequence determination and mutational analysis of the *lly* locus of *Legionella pneumophila*. Infect Immun. **62:**1109-1117.

Wintermeyer, E., B. Ludwig, M. Steinert, B. Schmidt, G. Fischer, and J. Hacker. 1995. Influence of site specifically altered Mip proteins on intracellular survival of *Legionella pneumophila* in eukaryotic cells. Infect Immun. **63:**4576-4583.

Woese, C. R. (ed.). 1991. Reconstruction of bacterial evolution with *rRNA*. Sinauer Associates, Sunderland, MA., USA.

Wong, K. H., C. W. Moss, D. H. Hochstein, R. J. Arko, and W. O. Schalla. 1979. "Endotoxicity" of the Legionnaires' disease bacterium. Ann Intern Med. **90:**624-627.

Wong, C. Y., M. W. Heuzenroeder, D. M. Quinn, and R. L. Flower. 1997. Cloning and characterization of two immunophilin-like genes, *ilpA* and *fkpA*, on a single 3.9-kilobase fragment of *Aeromonas hydrophila* genomic DNA. J Bacteriol. **179:**3397-3403.

Woo, A. H., A. Goetz, and V. L. Yu. 1992. Transmission of *Legionella* by respiratory equipment and aerosol generating devices. Chest. **102:**1586-1590.

Yang, Z. 1993. Maximum liklihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol. **10:**1396-1401.

Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Tends Ecol Evol. **11:**367-372.

Yu, V. L. 1995. *Legionella pnuemophila* (Legionnaires' Disease), 4th ed. Churchill Livingstone Inc., New York, NY, USA.

Zhang, H., D. W. Niesel, J. W. Peterson, and G. R. Klimpel. 1998. Lipoprotein release by bacteria: potential factor in bacterial pathogenesis. Infect Immun. **66:**5196-5201.

Zhu, W. T., and T. T. Loh. 1996. Differential effects of phorbol ester on apoptosis in HL-60 promyelocytic leukemic cells. Biochem Pharmacol. **51:**1229-1236.

Zuckman, D. M., J. B. Hung, and C. R. Roy. 1999. Pore-forming activity is not sufficient for *Legionella pneumophila* phagosome trafficking and intracellular growth. Mol Microbiol. **32:**990-1001.

Zwart, G., W. D. Hiorns, B. A. Methé, M. P. van Agterveld, R. Huismans, S. C. Nold, J. P. Zehr, and H. J. Laanbroek. 1998. Nearly identical *16S rRNA* sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. Syst Appl Microbiol. **21:**546-556.

Zychlinsky, A., and P. J. Sansonetti. 1997. Apoptosis as a proinflammatory event: what can we learn from bacteria-induced cell death? Trends Microbiol. **5:**201-204.

Appendices

Appendix 1. Distance matrices used for the neighbour joining phylogenetic analysis in Chapter Four. Distances are calculated on the combined *16S rRNA* and *mip* sequence data. Upper matrix is the Jukes-Cantor distances (Jukes and Cantor, 1969) for each species pair. Lower matrix is the uncorrected "p" distance for each species pair.

Legionella 16S rRNA and *mip* sequence

Jukes-Cantor distance matrix

	adelaidensis	*anisa*	*birminghamensis*	*bozemanii*-1	*brunensis*	*cherrii*	*cincinnatiensis*	*dumoffii*	*erythra*	*fairfieldensis*	*feeleii*-1	*gormanii*	*gratiana*	*hackeliae*-1	*israelensis*	*jamestowniensis*	*jordanis*	*lansingensis*	*londiniensis*	*longbeachae*-1	*maceachernii*
adelaidensis	-	0.1276	0.1195	0.1218	0.1227	0.143	0.1244	0.1355	0.1335	0.1277	0.1383	0.1302	0.1186	0.1173	0.1512	0.1181	0.1285	0.1433	0.1303	0.1305	0.1422
anisa	0.1173	-	0.1225	0.0213	0.1019	0.0533	0.063	0.0464	0.1138	0.1212	0.1189	0.0364	0.0626	0.1065	0.1361	0.1198	0.1214	0.1325	0.1405	0.0615	0.1331
birminghamensis	0.1104	0.113	-	0.1167	0.0963	0.1329	0.1216	0.1221	0.1066	0.1074	0.1114	0.1239	0.1188	0.0995	0.138	0.1029	0.1103	0.129	0.1227	0.121	0.1074
bozemanii-1	0.1124	0.021	0.1081	-	0.1036	0.0506	0.0554	0.0443	0.1144	0.1241	0.1149	0.0386	0.0577	0.1066	0.1348	0.1192	0.1192	0.1349	0.1387	0.0615	0.1271
brunensis	0.1132	0.0952	0.0904	0.0967	-	0.1145	0.1069	0.105	0.0927	0.1026	0.1054	0.1008	0.1035	0.0649	0.125	0.0834	0.0878	0.1034	0.1118	0.1064	0.0996
cherrii	0.1302	0.0515	0.1218	0.049	0.1062	-	0.0754	0.0554	0.1317	0.1407	0.1285	0.0479	0.0709	0.1196	0.1456	0.1229	0.1398	0.1481	0.1546	0.0695	0.144
cincinnatiensis	0.1146	0.0605	0.1123	0.0534	0.0996	0.0718	-	0.0689	0.124	0.1214	0.1204	0.0675	0.0463	0.1147	0.1322	0.122	0.126	0.1382	0.1438	0.0303	0.1267
dumoffii	0.1239	0.045	0.1127	0.043	0.098	0.0534	0.0658	-	0.1217	0.122	0.1197	0.0432	0.063	0.1223	0.1416	0.1183	0.1199	0.1417	0.1425	0.0686	0.1273
erythra	0.1223	0.1056	0.0994	0.1061	0.0872	0.1208	0.1143	0.1123	-	0.1042	0.1083	0.1201	0.1206	0.1035	0.1389	0.1059	0.1143	0.1145	0.1272	0.1239	0.1158
fairfieldensis	0.1174	0.1119	0.1001	0.1144	0.0959	0.1283	0.1121	0.1126	0.0973	-	0.1131	0.1321	0.1181	0.1079	0.1367	0.1157	0.1238	0.1142	0.1232	0.1262	0.1107
feeleii-1	0.1263	0.1099	0.1035	0.1065	0.0983	0.1181	0.1112	0.1106	0.1008	0.105	-	0.1214	0.1181	0.1031	0.1457	0.1138	0.1161	0.1272	0.1363	0.1245	0.1114
gormanii	0.1195	0.0356	0.1142	0.0376	0.0943	0.0464	0.0645	0.042	0.111	0.1211	0.1121	-	0.0604	0.1133	0.1373	0.1192	0.1252	0.1367	0.1406	0.0627	0.1358
gratiana	0.1097	0.06	0.1098	0.0555	0.0966	0.0677	0.0449	0.0604	0.1114	0.1093	0.1093	0.0581	-	0.1154	0.1401	0.1146	0.1302	0.1424	0.1252	0.0493	0.1217
hackeliae-1	0.1086	0.0993	0.0932	0.0994	0.0621	0.1105	0.1063	0.1129	0.0967	0.1005	0.0963	0.1052	0.1069	-	0.1344	0.0637	0.088	0.103	0.1282	0.1061	0.113
israelensis	0.1369	0.1244	0.1261	0.1234	0.1152	0.1323	0.1212	0.1291	0.1268	0.1249	0.1324	0.1255	0.1278	0.123	-	0.1344	0.1392	0.1414	0.1513	0.1447	0.1403
jamestowniensis	0.1093	0.1107	0.0962	0.1102	0.0789	0.1134	0.1126	0.1095	0.0988	0.1072	0.1056	0.1102	0.1063	0.061	0.123	-	0.0921	0.1118	0.1201	0.1245	0.1156
jordanis	0.1181	0.1121	0.1026	0.1102	0.0828	0.1276	0.1159	0.1108	0.106	0.1142	0.1076	0.1153	0.1195	0.083	0.1271	0.0867	-	0.1191	0.1305	0.1254	0.1155
lansingensis	0.1305	0.1215	0.1185	0.1234	0.0966	0.1344	0.1262	0.1291	0.1062	0.1059	0.117	0.125	0.1297	0.0962	0.1289	0.1039	0.1101	-	0.1613	0.1398	0.1386
londiniensis	0.1196	0.1281	0.1132	0.1266	0.1039	0.1397	0.1309	0.1298	0.117	0.1136	0.1246	0.1282	0.1153	0.1179	0.137	0.111	0.1198	0.1452	-	0.1423	0.14
longbeachae-1	0.1198	0.059	0.1117	0.059	0.0992	0.0664	0.0297	0.0656	0.1142	0.1162	0.1148	0.0601	0.0477	0.0989	0.1316	0.1147	0.1154	0.1276	0.1296	-	0.1286
maceachernii	0.1295	0.1219	0.1001	0.1169	0.0932	0.131	0.1166	0.1171	0.1073	0.1029	0.1036	0.1242	0.1123	0.1049	0.1279	0.1072	0.1071	0.1265	0.1277	0.1182	-
micdadei	0.1295	0.1245	0.1055	0.1224	0.1052	0.1352	0.1283	0.1236	0.1069	0.1055	0.1034	0.1246	0.1182	0.1089	0.1357	0.1073	0.106	0.1209	0.1202	0.1246	0.0544
moravica	0.1171	0.0772	0.1042	0.0767	0.0992	0.0925	0.0791	0.079	0.1019	0.1122	0.1142	0.0813	0.0852	0.1016	0.1292	0.1078	0.1038	0.1196	0.1086	0.077	0.1161
nautarum	0.112	0.1181	0.1056	0.1146	0.0953	0.1233	0.1173	0.1162	0.107	0.1011	0.109	0.1222	0.1189	0.0959	0.1126	0.0994	0.1007	0.1195	0.1223	0.117	0.0857
oakridgensis	0.124	0.1339	0.1176	0.1301	0.1193	0.1421	0.1268	0.1367	0.1163	0.1249	0.1334	0.1343	0.133	0.1199	0.1295	0.1237	0.1331	0.1341	0.1216	0.1307	0.1199
parisiensis	0.1155	0.023	0.1127	0.0275	0.094	0.0513	0.0603	0.0493	0.1107	0.1186	0.1071	0.039	0.0573	0.1001	0.1276	0.1114	0.1148	0.1261	0.1263	0.0583	0.1207
pneumophila (Phil-1)	0.1168	0.0843	0.1129	0.0843	0.0935	0.1004	0.0929	0.0857	0.1139	0.1108	0.1164	0.0879	0.0837	0.1035	0.1301	0.1104	0.1118	0.1219	0.1161	0.0862	0.1178
quateirensis	0.1226	0.0795	0.1087	0.081	0.0995	0.092	0.0844	0.0818	0.1139	0.1206	0.1206	0.0811	0.0853	0.1104	0.1336	0.1122	0.1154	0.1335	0.1186	0.0805	0.1227
quinlivanii	0.12	0.1175	0.0618	0.1145	0.1009	0.1263	0.1117	0.1161	0.1018	0.1045	0.1076	0.1156	0.1188	0.1046	0.1276	0.1004	0.1057	0.116	0.1317	0.1193	0.1084
rubrilucens	0.1225	0.1125	0.1022	0.1096	0.0944	0.1258	0.1157	0.1141	0.0339	0.0991	0.1001	0.1141	0.1123	0.0979	0.1306	0.1047	0.1062	0.1085	0.1167	0.1162	0.1065
sainthelens i-1	0.1176	0.0586	0.1152	0.0567	0.1015	0.0685	0.026	0.0656	0.1122	0.1151	0.1167	0.0643	0.0511	0.1103	0.1291	0.1183	0.1213	0.1271	0.1327	0.0324	0.1205
santicrucis	0.1225	0.0689	0.1206	0.0629	0.1111	0.083	0.0249	0.0763	0.1232	0.121	0.1231	0.072	0.0539	0.1124	0.1306	0.123	0.1288	0.1352	0.1404	0.0419	0.1271
shakespearei	0.1244	0.087	0.1111	0.085	0.1085	0.0916	0.0898	0.0832	0.1053	0.117	0.1185	0.0856	0.0953	0.1102	0.1296	0.1068	0.1203	0.13	0.1216	0.0911	0.1216
spiritensis	0.1298	0.1194	0.11	0.1201	0.1036	0.1305	0.1249	0.1191	0.0954	0.1054	0.1135	0.1176	0.1245	0.1021	0.1262	0.1049	0.1075	0.1217	0.1271	0.1165	0.1231
steigerwaltii	0.1202	0.036	0.1108	0.037	0.097	0.0443	0.0634	0.0395	0.1104	0.1114	0.1098	0.036	0.056	0.1048	0.1288	0.111	0.1135	0.1273	0.131	0.059	0.1219
tucsonensis	0.1144	0.0296	0.111	0.0261	0.0902	0.049	0.0581	0.05	0.1062	0.1121	0.107	0.0411	0.0561	0.1009	0.1224	0.1109	0.1107	0.1281	0.1242	0.0596	0.1211
wadsworthii	0.113	0.0435	0.1122	0.047	0.099	0.0586	0.0709	0.0578	0.1084	0.1191	0.1171	0.0496	0.0689	0.1048	0.1252	0.1101	0.1124	0.1221	0.1268	0.0718	0.1242
worsleiensis	0.1233	0.0819	0.1106	0.0848	0.1048	0.1022	0.0922	0.0896	0.1143	0.114	0.1229	0.0909	0.0887	0.1123	0.1369	0.1178	0.1112	0.1253	0.1165	0.0906	0.124
C. burnetii	0.2328	0.2276	0.2353	0.228	0.2328	0.2341	0.2366	0.2236	0.2381	0.2377	0.2447	0.2255	0.2293	0.2265	0.2386	0.2258	0.2267	0.2471	0.2308	0.2312	0.2426

Uncorrected ("p") distance matrix

Legionella 16S rRNA and *mip* sequence

Jukes-Cantor distance matrix

	micdadei	*moravica*	*nautarum*	*oakridgensis*	*parisiensis*	*pneumophila* (Phil-1)	*quateirensis*	*quinlivanii*	*rubrilucens*	*sainthelensi*-1	*santicrucis*	*shakespearei*	*spiritensis*	*steigerwaltii*	*tucsonensis*	*wadsworthii*	*worsleiensis*	*C. burnetii*
adelaidensis	0.1422	0.1274	0.1213	0.1355	0.1254	0.1269	0.1338	0.1307	0.1338	0.1279	0.1338	0.1361	0.1425	0.131	0.1242	0.1225	0.1347	0.2788
anisa	0.1361	0.0815	0.1285	0.1475	0.0233	0.0894	0.0841	0.1278	0.1219	0.0611	0.0723	0.0924	0.1301	0.0369	0.0302	0.0448	0.0867	0.2712
birminghamensis	0.1137	0.1122	0.1139	0.1279	0.1221	0.1223	0.1175	0.0645	0.1099	0.1251	0.1315	0.1203	0.119	0.1199	0.1201	0.1216	0.1196	0.2824
bozemanii-1	0.1337	0.0809	0.1244	0.1429	0.028	0.0894	0.0858	0.1242	0.1184	0.059	0.0657	0.0903	0.1309	0.0379	0.0265	0.0486	0.09	0.2718
brunensis	0.1133	0.1065	0.1019	0.1299	0.1004	0.0999	0.1067	0.1084	0.1009	0.109	0.1202	0.1172	0.1115	0.1039	0.0961	0.1062	0.1129	0.2787
cherrii	0.1491	0.0988	0.1348	0.1576	0.0532	0.1077	0.0982	0.1383	0.1377	0.0718	0.0879	0.0977	0.1433	0.0457	0.0507	0.061	0.1099	0.2806
cincinnatiensis	0.1408	0.0836	0.1276	0.1388	0.0628	0.0991	0.0895	0.1209	0.1257	0.0265	0.0253	0.0956	0.1367	0.0663	0.0604	0.0744	0.0984	0.2842
dumoffii	0.1351	0.0834	0.1263	0.1509	0.051	0.091	0.0866	0.1261	0.1238	0.0686	0.0805	0.0882	0.1297	0.0406	0.0518	0.0602	0.0954	0.2655
erythra	0.1153	0.1095	0.1154	0.1264	0.1198	0.1235	0.1236	0.1094	0.0347	0.1215	0.1346	0.1135	0.102	0.1194	0.1145	0.1171	0.124	0.2864
fairfieldensis	0.1137	0.1215	0.1086	0.1367	0.1291	0.1199	0.1315	0.1126	0.1063	0.1249	0.1319	0.1272	0.1136	0.1206	0.1214	0.1297	0.1236	0.2859
feeleii-1	0.1113	0.1239	0.1178	0.1469	0.1156	0.1265	0.1315	0.1161	0.1075	0.1268	0.1344	0.129	0.1231	0.1187	0.1154	0.1274	0.1343	0.2962
gormanii	0.1363	0.0861	0.1334	0.148	0.04	0.0935	0.0858	0.1256	0.1238	0.0672	0.0757	0.0909	0.128	0.0369	0.0423	0.0513	0.0969	0.2682
gratiana	0.1286	0.0904	0.1295	0.1464	0.0596	0.0888	0.0906	0.1293	0.1217	0.0529	0.0559	0.1019	0.1361	0.0582	0.0583	0.0723	0.0944	0.2736
hackeliae-1	0.1176	0.1091	0.1027	0.1306	0.1074	0.1113	0.1194	0.1126	0.1049	0.1193	0.1217	0.1192	0.1097	0.1129	0.1083	0.1129	0.1217	0.2696
israelensis	0.1497	0.1418	0.122	0.1422	0.1399	0.1428	0.1471	0.1398	0.1434	0.1417	0.1435	0.1422	0.1382	0.1413	0.1337	0.137	0.1511	0.2873
jamestowniensis	0.1158	0.1164	0.1067	0.1351	0.1206	0.1195	0.1215	0.1078	0.1127	0.1287	0.1344	0.1152	0.113	0.1201	0.12	0.119	0.1282	0.2687
jordanis	0.1143	0.1117	0.1082	0.1465	0.1246	0.1211	0.1253	0.1139	0.1145	0.1323	0.1413	0.1311	0.116	0.1231	0.1197	0.1218	0.1204	0.27
lansingensis	0.1319	0.1302	0.1302	0.1478	0.138	0.133	0.1471	0.126	0.1172	0.1392	0.1491	0.1428	0.1327	0.1395	0.1405	0.1333	0.1371	0.2997
londiniensis	0.1311	0.1174	0.1335	0.1327	0.1382	0.1262	0.1291	0.1448	0.1269	0.1461	0.1554	0.1327	0.1393	0.1439	0.1358	0.1389	0.1266	0.2758
longbeachae-1	0.1363	0.0812	0.1272	0.1436	0.0607	0.0915	0.0852	0.1299	0.1263	0.0331	0.0432	0.0971	0.1266	0.0615	0.0621	0.0754	0.0965	0.2764
maceachernii	0.0565	0.1261	0.091	0.1307	0.1316	0.1281	0.134	0.1171	0.1149	0.1313	0.1392	0.1327	0.1344	0.1331	0.1321	0.1358	0.1355	0.293
micdadei	-	0.1302	0.0985	0.1347	0.1382	0.1284	0.1401	0.1167	0.1167	0.1381	0.1588	0.1442	0.1244	0.1365	0.1386	0.1444	0.1442	0.2928
moravica	0.1196	-	0.1275	0.139	0.0891	0.0804	0.0292	0.1251	0.1133	0.0828	0.0943	0.0662	0.1178	0.0882	0.0901	0.0818	0.0585	0.2748
nautarum	0.0923	0.1173	-	0.1256	0.128	0.1218	0.1341	0.1245	0.1106	0.1275	0.1382	0.1286	0.1191	0.1277	0.1256	0.1263	0.1326	0.2816
oakridgensis	0.1233	0.1269	0.1156	-	0.1492	0.1412	0.1473	0.1405	0.1244	0.1391	0.1472	0.1381	0.1513	0.15	0.1436	0.148	0.1447	0.2861
parisiensis	0.1262	0.084	0.1177	0.1353	-	0.0914	0.0922	0.1285	0.125	0.0642	0.0732	0.0961	0.1322	0.0442	0.0285	0.0478	0.0965	0.2686
pneumophila (Phil-1)	0.118	0.0762	0.1124	0.1287	0.0861	-	0.0853	0.1367	0.1257	0.0972	0.1067	0.1014	0.1327	0.0893	0.0924	0.098	0.088	0.271
quateirensis	0.1278	0.0287	0.1228	0.1337	0.0868	0.0806	-	0.1297	0.1316	0.0848	0.0974	0.0683	0.1268	0.0902	0.0892	0.0872	0.0574	0.2796
quinlivanii	0.1081	0.1152	0.1147	0.1281	0.1181	0.1249	0.1191	-	0.1068	0.1273	0.132	0.1267	0.1206	0.1277	0.128	0.1315	0.1343	0.284
rubrilucens	0.1081	0.1052	0.1028	0.1146	0.1152	0.1158	0.1207	0.0995	-	0.125	0.1375	0.1144	0.0988	0.1237	0.1191	0.1216	0.1249	0.2791
sainthelens i-1	0.1262	0.0784	0.1173	0.1269	0.0615	0.0912	0.0802	0.1171	0.1151	-	0.0385	0.096	0.131	0.0628	0.0612	0.0731	0.0994	0.279
santicrucis	0.1431	0.0886	0.1262	0.1336	0.0697	0.0995	0.0913	0.1211	0.1256	0.0375	-	0.103	0.1477	0.0756	0.0697	0.085	0.1023	0.2879
shakespearei	0.1311	0.0634	0.1182	0.1261	0.0902	0.0948	0.0653	0.1165	0.1061	0.0901	0.0962	-	0.126	0.0953	0.0943	0.0916	0.0798	0.2859
spiritensis	0.1147	0.109	0.1101	0.137	0.1212	0.1217	0.1167	0.1114	0.0926	0.1202	0.134	0.116	-	0.1326	0.1334	0.1292	0.1308	0.2804
steigerwaltii	0.1248	0.0832	0.1174	0.136	0.0429	0.0842	0.085	0.1174	0.114	0.0602	0.0719	0.0895	0.1216	-	0.0379	0.0549	0.0995	0.2764
tucsonensis	0.1265	0.0849	0.1156	0.1307	0.028	0.0869	0.0841	0.1177	0.1102	0.0588	0.0665	0.0886	0.1223	0.037	-	0.0551	0.0998	0.279
wadsworthii	0.1313	0.0775	0.1163	0.1343	0.0463	0.0919	0.0823	0.1206	0.1122	0.0697	0.0803	0.0862	0.1187	0.053	0.0531	-	0.0977	0.2751
worsleiensis	0.1311	0.0563	0.1216	0.1316	0.0905	0.083	0.0553	0.1229	0.115	0.0931	0.0957	0.0757	0.12	0.0932	0.0934	0.0916	-	0.2767
C. burnetii	0.2424	0.2301	0.2348	0.2379	0.2258	0.2274	0.2334	0.2364	0.233	0.233	0.2391	0.2377	0.234	0.2312	0.233	0.2303	0.2314	-

Uncorrected ("p") distance matrix

Appendix 2. Distance matrices used for the neighbour joining and maximum liklihood phylogenetic analyses of the *L. anisa – L. worsleiensis* clade in Chapter Five. Distances are calculated on the combined *mip* and *mspA* sequence data. Upper matrix is the General time-reversible distances (Rodriguez, *et al.*, 1990) for each species pair. Lower matrix is the uncorrected "p" distance for each species pair.

Legionella (*anisa* - *worsleiensis*) clade
Combined *mspA* and *mip* sequence

General time-reversible distance matrix

	anisa	*bozemanii*-1	*cherrii*	*cincinnatiensis*	*gormanii*	*gratiana*	*moravica*	*parisiensis*	*quateirensis*	*santicrucis*	*shakespearei*	*steigerwaltii*	*sainthelensi*-1	*tucsonensis*	*wadsworthii*	*worsleiensis*	*longbeachae*-1	*dumoffii*	*pneumophila* (Phil-1)	*micdadei*
anisa	-	0.0562	0.14339	0.23718	0.10913	0.27135	0.37825	0.06595	0.32725	0.25708	0.35169	0.13281	0.24158	0.06181	0.18736	0.40972	0.26514	0.1808	0.56207	1.25763
bozemanii-1	0.04813	-	0.11512	0.23343	0.10918	0.24642	0.36224	0.06383	0.34762	0.26366	0.34333	0.11318	0.2642	0.04323	0.17984	0.42221	0.26934	0.1567	0.54463	1.35937
cherrii	0.1016	0.08556	-	0.22147	0.12906	0.23474	0.43873	0.15785	0.35217	0.2402	0.33431	0.06843	0.21871	0.11191	0.18919	0.53564	0.24219	0.13655	0.545	1.14342
cincinnatiensis	0.14082	0.13993	0.13458	-	0.23177	0.11536	0.41251	0.25505	0.34057	0.04704	0.42064	0.25112	0.06752	0.23415	0.29613	0.49951	0.06966	0.26208	0.49879	1.02774
gormanii	0.08405	0.08226	0.09479	0.14121	-	0.22101	0.39758	0.13629	0.32459	0.23099	0.39686	0.11653	0.25126	0.10802	0.20846	0.50324	0.26586	0.14348	0.52946	1.13536
gratiana	0.14973	0.14528	0.14082	0.08645	0.13588	-	0.43507	0.26497	0.34781	0.12287	0.4867	0.24737	0.13461	0.22992	0.33095	0.42834	0.14162	0.28505	0.48409	1.03104
moravica	0.1802	0.17661	0.18468	0.18292	0.18608	0.19005	-	0.45884	0.09157	0.48212	0.27262	0.40558	0.40017	0.40459	0.42127	0.20758	0.42442	0.38108	0.37017	1.11869
parisiensis	0.05526	0.05258	0.10695	0.14439	0.09659	0.14973	0.19181	-	0.4205	0.28406	0.42205	0.16159	0.27839	0.06786	0.20591	0.48529	0.29264	0.20976	0.70054	1.50588
quateirensis	0.17029	0.17386	0.17121	0.17385	0.17163	0.17653	0.06933	0.18902	-	0.3857	0.21071	0.34888	0.3641	0.35081	0.35356	0.17985	0.41246	0.35048	0.28094	0.95654
santicrucis	0.14349	0.14706	0.13993	0.041	0.13945	0.08824	0.19094	0.14973	0.18277	-	0.41258	0.25936	0.07685	0.25229	0.31951	0.55035	0.06844	0.31435	0.50266	1.42426
shakespearei	0.18096	0.17917	0.17206	0.189	0.19213	0.20147	0.14489	0.19611	0.13244	0.18989	-	0.34875	0.47206	0.35367	0.40055	0.30774	0.45875	0.3298	0.42294	1.21705
steigerwaltii	0.09299	0.08226	0.05721	0.14388	0.08668	0.14477	0.18516	0.10551	0.17521	0.14658	0.17962	-	0.24905	0.1074	0.19777	0.53265	0.25455	0.12594	0.50042	1.44158
sainthelensi-1	0.1426	0.14795	0.13191	0.05615	0.14659	0.09715	0.18202	0.15241	0.17832	0.06061	0.19525	0.14387	-	0.2483	0.30935	0.52534	0.06394	0.29944	0.48672	0.96727
tucsonensis	0.05258	0.03832	0.08556	0.14082	0.08315	0.13993	0.18554	0.05615	0.1774	0.14439	0.18185	0.08132	0.14528	-	0.19205	0.48985	0.26283	0.16224	0.61209	1.34656
wadsworthii	0.12302	0.11855	0.12567	0.16226	0.13324	0.17206	0.19289	0.12748	0.18311	0.1685	0.19467	0.12698	0.1667	0.12569	-	0.5466	0.32637	0.22574	0.66248	1.55458
worsleiensis	0.18908	0.19173	0.20602	0.20782	0.20294	0.19799	0.12711	0.20244	0.11911	0.21405	0.16533	0.20652	0.2096	0.20332	0.21778	-	0.58374	0.47935	0.36009	1.01082
longbeachae-1	0.14795	0.15062	0.13993	0.05704	0.14932	0.09893	0.18381	0.15152	0.18546	0.05615	0.19079	0.14659	0.05348	0.14884	0.16849	0.21494	-	0.30413	0.50177	0.99303
dumoffii	0.12041	0.10883	0.0981	0.15253	0.10112	0.157	0.18133	0.13024	0.17689	0.16414	0.17511	0.09212	0.15967	0.11149	0.14133	0.20267	0.16322	-	0.56796	1.65728
pneumophila (Phil-1)	0.20499	0.20677	0.20677	0.20232	0.20654	0.19786	0.17485	0.21569	0.1587	0.19964	0.18364	0.20204	0.19964	0.21301	0.2246	0.17924	0.2041	0.2104	-	1.16724
micdadei	0.27196	0.27463	0.27018	0.26752	0.27274	0.26841	0.26761	0.27997	0.25606	0.27822	0.26405	0.27364	0.26484	0.27909	0.28626	0.26315	0.26483	0.27471	0.26394	-

Uncorrected ("p") distance matrix

Appendix 3. The two publications reporting some of the results presented in this thesis

Ratcliff, R. M., S. C. Donnellan, J. A. Lanser, P. A. Manning, and M. W. Heuzenroeder. 1997. Interspecies sequence differences in the Mip protein from the genus *Legionella*: implications for function and evolutionary relatedness. Mol Microbiol. **25:**1149-1158.

Ratcliff, R. M., J. A. Lanser, P. A. Manning, and M. W. Heuzenroeder. 1998. Sequence-based classification scheme for the genus *Legionella* targeting the *mip* gene. J Clin Microbiol. **36:**1560-1567.

Ratclitf, R.M., Donnellan, S.C., Lanser, J.A., Manning, P.A. and Heuzenroeder, M.W. (1997). Interspecies sequence differences in the Mip protein from the genus Legionella: implications for function and evolutionary relatedness. *Molecular Microbiology, 25*(6), 1149-1158.

JOURNAL OF CLINICAL MICROBIOLOGY, June 1998, p. 1560–1567
0095-1137/98/$04.00+0

Vol. 36, No. 6

Sequence-Based Classification Scheme for the Genus *Legionella* Targeting the *mip* Gene

RODNEY M. RATCLIFF,[1]* JANICE A. LANSER,[1] PAUL A. MANNING,[2] AND MICHAEL W. HEUZENROEDER[1]

Infectious Diseases Laboratories, Institute of Medical and Veterinary Science, Adelaide, South Australia, 5000,[1] and Microbial Pathogenesis Unit, Department of Microbiology and Immunology, University of Adelaide, Adelaide, South Australia, 5005,[2] Australia

Received 15 September 1997/Returned for modification 30 December 1997/Accepted 17 March 1998

The identification and speciation of strains of *Legionella* is often difficult, and even the more successful chromatographic classification techniques have struggled to discriminate newly described species. A sequence-based genotypic classification scheme is reported, targeting approximately 700 nucleotide bases of the *mip* gene and utilizing gene amplification and direct amplicon sequencing. With the exception of *Legionella geestiana*, for which an amplicon was not produced, the scheme clearly and unambiguously discriminated among the remaining 39 *Legionella* species and correctly grouped 26 additional serogroup and reference strains within those species. Additionally, the genotypic classification of approximately 150 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 the latter classification. Strains thought to represent currently uncharacterized species were also found to be genotypically unique. The scheme is technically simple for a laboratory with even basic molecular capabilities and equipment, if access to a sequencing laboratory is available.

The genus *Legionella* comprises approximately 40 species, at least 7 of which have more than one serotype (3, 15, 31). Approximately half of the species have been associated with human disease (28). *Legionella*-like organisms isolated from clinical specimens, or from the environment during the course of an outbreak, need to be identified to elucidate the disease process and to identify the source. Legionellae have proved to be relatively unreactive when traditional biochemical tests are utilized, necessitating more complex identification methods (6, 7, 26, 41). Serologically based methods are widely used in clinical laboratories, but antigen cross-reactivity limits specificity and restricts their confident use to a few frequently isolated species (38). This is especially true for countries where legionellosis caused by species other than *L. pneumophila* is common (12). More complex classification schemes have been proposed (26, 38), the most successful being one based on the range and proportion of cellular fatty acids and ubiquinones (21, 22, 40, 43). As additional species have been characterized, this method has become less discriminating, since apparently unique patterns were proved to be shared by several species (43). The inclusion of hydroxylated fatty acids has improved discrimination, but it requires the analysis of both mono- and dihydroxylated fatty acids, and individual patterns are complex, making analysis difficult (21).

Gene sequence-based phylogenic (genotypic) schemes have become widely used for organisms which are difficult to classify, as more sequences have been determined and sequencing methods have become simpler, more widely available, and cost effective (11, 23, 24, 29, 32, 34). Genotypic schemes have the great advantage of being unaffected by colony age and growth conditions and, in contrast to chromatographic methods, are not subject to extraction and chromatographic conditions or constituent equipment. Additionally, because a gene sequence is essentially a long digital string, with each digit being one of only four nucleotides, genotypic schemes are less ambiguous and can utilize significantly more discriminatory data than phenotypic ones, and in a form that lends itself to widely available computer analysis software. Many genotypic schemes utilize variation in the 16S rRNA sequence (11, 23, 24, 32, 34), because of the ease with which regions can be amplified and sequenced with universal primers. The 16S rRNA sequences of *Legionella* species have been reported (18), as have the sequences of the *mip* gene (2, 12, 13, 31), which codes for an immunophilin of the FK506 binding protein (FKBP) class (14). This protein, which ranges in size from 232 to 251 amino acids, depending on the *Legionella* species (31), is an outer membrane protein important in the intracellular cycle of *Legionella*. While it is known to be involved with the survival of the bacterium immediately after uptake into phagocytic cells (9, 12, 28), its exact role is unclear. Additionally, analogs are found widely in both prokaryotes and eukaryotes and are likely to have a significant cellular role (14). Other gene sequences have been determined for *Legionella* (5, 17, 36), but only the rRNA sequences and the *mip* gene have been comprehensively determined for most species, an essential prerequisite for any gene to be the basis of a genotyping scheme. Ratcliff et al. (31) recently phylogenetically compared most *Legionella* species, using the species variation among both the 16S rRNA and *mip* genes, and found over twice the variation in the *mip* gene at the DNA level (56% of base sites) as in 16S rRNA (23% of base sites). A pairwise comparison of species reveals a *mip* gene variation of 3 to 31% (mean, 20%) between species pairs compared with 1 to 10% (mean, 6%) for 16S rRNA. For the *mip* gene, interspecies nucleotide variation occurred throughout the gene but especially within a hypervariable insert of up to 51 bases immediately adjacent to the region coding for the signal sequence, at redundant third codon sites, and in se-

* Corresponding author. Mailing address: Infectious Diseases Laboratories, Institute of Medical and Veterinary Science, P.O. Box 14, Rundle Mall, Adelaide, SA, 5000, Australia. Phone: 61 8 82223245. Fax: 61 8 82223543. E-mail: rod.ratcliff@imvs.sa.gov.au.

TABLE 1. Isolates used to test the robustness of the classification scheme

Legionella isolate[a]	Accession no.[b]	Phenotypic identification[c]	Sequence identification	No. of differences[d]	
				DNA	AA
L. anisa					
5 wild strains, LC3934	NA	*L. anisa*	*L. anisa*	0	0
IMVS-946	AF022312	*L. anisa*	*L. anisa*	18	1
L. birminghamensis					
IMVS-C4D5	NA	*L. birminghamensis*	*L. birminghamensis*	0	0
IMVS-K7H8	AF047743	*L. birminghamensis*	*L. birminghamensis*	1	0
LC2720	AF047744	*L. birminghamensis*	*L. birminghamensis*	41	1
L. bozemanii					
LC4348	NA	*L. bozemanii* serogroup 1	*L. bozemanii*	0	0
Serogroup 2 ATCC 35545, [IMVS-A5A1][c]	AF022308	NA, [*L. bozemanii* serogroup 2]	*L. bozemanii*	3	1
IMVS-A5F7, [IMVS-A8E7]	AF022309	*L. bozemanii* serogroup 1, [serogroup 2]	*L. bozemanii*	3	1
IMVS-A5I7, -K7B3, [LC2763]	AF022310	*L. bozemanii* serogroup 1 [or *L. parisiensis*]	*L. bozemanii*	6	1
IMVS-D2/7, -K7B4	AF022311	*L. bozemanii* serogroup 2	*L. bozemanii*	6	1
L. brunensis species E (IMVS-594)	AF022350	NC	*L. brunensis*	6	0
L. cherrii LC3664	NA	*L. cherrii*	*L. cherrii*	0	0
L. cincinnatiensis					
IMVS-C4B3	NA	*L. cincinnatiensis*	*L. cincinnatiensis*	0	0
IMVS-K5B7	AF022358	*L. cincinnatiensis*	*L. cincinnatiensis*	3	0
IMVS-K8D7	AF022359	*L. cincinnatiensis*	*L. cincinnatiensis*	1	0
IMVS-K8I2	AF047745	*L. cincinnatiensis*	*L. cincinnatiensis*	10	3
LC3936	AF047746	Cross-reactivity	*L. cincinnatiensis*	1	0
L. dumoffii IMVS-K7D6, -C7A3, LC0455[c]	AF022313	*L. dumoffii*	*L. dumoffii*	2	0
L. erythra					
IMVS-926, LC0709, LC3719	NA	*L. erythra*	*L. erythra*	0	0
LC1317[c]	AF047747	*L. erythra*	*L. erythra*	3	1
L. feeleii					
Serogroup 2 ATCC 35849, LC4210	AF022341	NA	*L. feeleii*	2	0
IMVS-913	AF023174	*L. feeleii*	*L. feeleii*	9	1
IMVS-853, -865	AF022340	*L. feeleii*	*L. feeleii*	11	1
Species J (IMVS-933)	AF022354	NC	*L. feeleii*	12	1
L. geestiana LC3644	NA	*L. geestiana*	NA	–[f]	–
L. gormanii LC0777C	AF047748	*L. tucsonensis*/*L. feeleii* serogroup 2	*L. gormanii*	26	4
L. hackeliae serogroup 2 ATCC 35999	NA	NA	*L. hackeliae*	0	0
L. jamestowniensis					
IMVS-724, -935	AF022339	*L. jamestowniensis*	*L. jamestowniensis*	2	0
IMVS-707, -708, -871, -945	AF022337	*L. jamestowniensis*	*L. jamestowniensis*	24	2
IMVS-857	AF022338	*L. jamestowniensis*	*L. jamestowniensis*	27	4
L. jordanis LC3940	NA	*L. jordanis*	*L. jordanis*	0	0
L. londiniensis					
Species C (IMVS-449), IMVS-755, -914, -967, LC4049	NA	*L. londiniensis*	*L. londiniensis*	0	0
IMVS-912	AF022346	*L. londiniensis*	*L. londiniensis*	1	0
L. longbeachae					
26 wild strains[e]	NA	*L. longbeachae* serogroup 1	*L. longbeachae*	0	0
Serogroup 2 ATCC 33484, [IMVS-C4E7]		NA, [*L. longbeachae* serogroup 2]	*L. longbeachae*	2	0
L. maceachernii					
IMVS-984 and IMVS-A4H7, LC4349	AF022315	*L. maceachernii*	*L. maceachernii*	7	0
IMVS-910, -943, -962	AF022314	*L. maceachernii*	*L. maceachernii*	6	0
L. micdadei					
10 wild strains, D4307, D4310, M097-017C1, LC0858	AF023175	*L. micdadei*	*L. micdadei*	1	0
D4309, D4363, D4534	AF047749	*L. micdadei*	*L. micdadei*	2	0
IMVS-K5D3	AF023176	*L. micdadei*	*L. micdadei*	2	0
L. oakridgensis LC3780	NA	*L. oakridgensis*	*L. oakridgensis*	0	0
L. parisiensis IMVS-916	NA	*L. parisiensis*	*L. parisiensis*	0	0
L. pneumophila					
Serogroup 1 Bellingham-1 ATCC 43111	AF022329	NA	*L. pneumophila*	9	1
Serogroup 1 Knoxville-1 ATCC 33153	AF022332	NA	*L. pneumophila*	2	0
Serogroup 1 OLDA ATCC 43109	AF022335	NA	*L. pneumophila*	0	0
Serogroup 1 Allentown ATCC 43106	AF022330	NA	*L. pneumophila*	2	0
Serogroup 1 Benidom 030 E ATCC 43108	AF022336	NA	*L. pneumophila*	9	1
Serogroup 1 Camperdown-1 ATCC 43113	AF022331	NA	*L. pneumophila*	0	0

Continued on next page

13, 31) and contained sufficient variation (3.6 to 30.5%) to uniquely identify each species.

To test the specificities of the primers, DNA extracted from the following organisms was tested: *Staphylococcus aureus*, *Streptococcus pneumoniae*, *Escherichia coli*, *Haemophilus influenzae*, *Salmonella* (multiple serovars), *Shigella flexneri*, *Proteus mirabilis*, *Aeromonas hydrophila*, *Pseudomonas aeruginosa*, *Yersinia enterocolitica*, *Campylobacter* (multiple species), *Helicobacter pylori*, *Corynebacterium diphtheriae*, *Mycobacterium abscessus*, *Vibrio parahaemolyticus*, and *Nocardia asteroides*. No amplification was detected, despite the presence of genes coding for Mip-like analogs in some species (19, 20, 30, 44).

Heat lysis proved adequate for extracting DNA for amplification, although boiling or steaming gave slightly more consistent results than microwave heating. Prolonged boiling or steaming for 15 min did not appear to degrade the DNA, as measured by the amount of amplicon produced by subsequent amplification.

Table 1 presents the results for additional reference strains, as well as the comparison of results from the wild strains. This includes both those strains which conform to type strains and those for which speciation cannot be determined, which may represent new species. For a few of these latter strains, amplification produced only limited product which sequenced poorly. Lowering the annealing temperature to 50°C produced more product, which in turn produced longer unambiguous sequence. Sequencing was also attempted after amplification at a below-optimal annealing temperature to produce multiple products of different sizes, and excellent sequence was still obtained with Legmip_fs but not with Legmip_r, as would be predicted. Isolate LC4381 was not able to be amplified, even with the annealing temperature reduced to 40°C.

Figure 1 is a UPGMA phylogenic dendrogram of the inter- and intraspecies similarities.

DISCUSSION

For a sequence-based classification scheme to be successful, the gene amplification needs to be specific for the genus but universal for the individual species and the subsequent interspecies sequence variation needs to be sufficient to discriminate clearly between species, even after allowing for the intraspecies sequence variation found among wild strains.

The primers reported here produced single amplicons of the correct size, and in sufficient quantity to enable accurate sequence determination, for all isolates tested, whether they were type strains or wild strains, with the exception of *L. geestiana* and one unusual, serologically nonreactive, nonidentifiable European wild strain, LC4381. Neither this last isolate nor the type strain or a European wild strain of *L. geestiana* produced an amplicon of the correct size, even after the annealing temperature was lowered to 40°C. The *mip* gene sequence for *L. geestiana* has not been published, and no sequence was determined with the various combinations of primers used by Ratcliff et al. to amplify and sequence this gene in 35 other *Legionella* species (31). However, a Mip analog has been detected in *L. geestiana* by using Mip-specific monoclonal antibody (16). Consequently, the lack of amplification for *L. geestiana*, and LC4381, is likely to be due to sequence variation in the sites targeted by the amplification primers. This was no impediment in our hands, as no locally derived wild strain tested failed to amplify. Similarly, other laboratories should have little problem, as such isolates are likely to be very rare.

Figure 1 shows that all of the remaining species could be easily discriminated and that the wild-strain isolates grouped within the same species as determined by serological, chromatographical, or molecular identification techniques, except for a few where the phenotypic identification was based solely on serology and cross-reactivity prevented unambiguous speciation. Many of these strains had been retained and stored at −70°C because of these nontypical reactions. For instance, LC2763, believed to be a *L. bozemanii* isolate, also showed strong cross-reactivity with *L. parisiensis*-specific antiserum. LC3936 reacted equally well with antisera specific for *L. cincinnatiensis*, *L. santicrucis*, and both serogroups of *L. sainthelensi* and *L. longbeachae*. LC4042 reacted weakly with only *L. spiritensis*-specific antiserum. However, this isolate was red autofluorescent, suggesting the isolate is likely to be related to either *L. erythra* or *L. rubrilucens*, the only red-autofluorescent species so far described. These results demonstrate how ambiguous serological identification can be for some species. The genotypic identification of other strains from the same laboratory, where the serological identification had been confirmed by ribotyping and/or DNA hybridization, was identical in every case.

The *Legionella* genomospecies clustered with *L. quinlivanii*, with which it has 69% DNA homology (4). With a difference of 3.6%, *L. bozemanii* and *L. tucsonensis* showed the least interspecies sequence variation. However, the four *L. bozemanii* wild strains and *L. bozemanii* serogroup 2, with an intraspecies variation of 1.3% or less, were easily discriminated from *L. tucsonensis*. Similarly, *L. cincinnatiensis* wild strains (maximum intraspecies variation, 1.6%) were easily discriminated from *L. santicrucis* strains (interspecies variation, 4.1%). Some species, such as *L. anisa*, *L. longbeachae*, *L. micdadei*, and *L. londiniensis*, for which several wild strains were tested, showed very little sequence difference, suggesting a single homogeneous clonal population, compared to *L. jamestowniensis*, *L. rubrilucens*, and *L. quinlivanii*, which appeared to demonstrate multiple divergent clonal populations. This is especially interesting given that the *L. longbeachae* strains were isolated on several continents whereas all but one of the *L. quinlivanii* isolates, including the type strain, came from one region within Australia. The monophyletic nature of *L. longbeachae* has been demonstrated by other workers (8). *L. jamestowniensis*, *L. rubrilucens*, and *L. quinlivanii* were generally quite divergent from their nearest neighbors, so the greater sequence variation among wild strains did not compromise the resolution of their speciation. This contrasts with chromatographic identification schemes, where ubiquinone and fatty acid profiles do not discriminate easily between some species, for example, *L. erythra* and *L. rubrilucens* (43). In fact, it was this lack of discrimination with such methods which was the motivation for this study.

LC0777C is a very interesting European strain. It reacted poorly serologically, and while there was moderate homology with *L. gormanii*, with which it was genotyped, there was also moderate homology with *L. tucsonensis* and *L. feeleii* serogroup 2. With 4.2% sequence variation from *L. gormanii*, LC0777C demonstrates more genotypic variation from *L. gormanii* than is typical for wild-strain variation in sister species. This is consistent with its reduced serological reactivity, and it may represent an additional serogroup, a new genotype clone, or even a new species.

However, the multiple serogroups of species where they have been determined could not be confidently discriminated by this scheme in every circumstance. For example, the sequences for *L. hackeliae* serogroups 1 and 2 are identical. Similarly, many of the serogroups of *L. pneumophila* are identical to each other, and as a species seem to mainly cluster into two closely related clonal populations, with serogroup 1 strains represented in both. *L. pneumophila* serogroup 5 and two

nonserogrouped wild strains fall outside these two clonal populations. These two strains may represent a new serogroup, since they are both serologically and genotypically different from other *L. pneumophila* strains. IMVS-D1/77, a serogroup 13 strain, grouped with the serogroup 13 type strain. For *L. bozemanii*, IMVS-A8E7, a serogroup 2 isolate, and IMVS-A5F7, a serogroup 1 isolate, both grouped with the serogroup 2 type strain. Two other serogroup 2 isolates, IMVS-D2/7 and IMVS-K7B4, produced similar sequences, but they were different from that of the serogroup 2 type strain. Similarly, some serogroup 1 strains clustered uniquely together. Consequently, while these latter genotype groups may be indicative of a particular serogroup, as may some of the unique *L. pneumophila* serogroup sequences, the differences are small, and it is possible that such differences may be within the normal clonal variation found in the wild-strain population independent of a specific serogroup. The testing of more strains of these species would be necessary to determine if these few differences are characteristic of specific serogroups.

In contrast, the differences in sequences from the two serogroups of *L. sainthelensi* do appear discriminatory, with two wild strains producing sequences identical to that of the serogroup 2 type strain. Strain LC4261 has been serologically confirmed as *L. sainthelensi* serogroup 2, although it also demonstrated serological cross-reactivity with *L. santicrusis* and, to a lesser extent, *L. cincinnatiensis* and *L. longbeachae*. Similarly, although the *L. longbeachae* serogroup 2 type strain differs from the serogroup 1 type strain by only 2 bases, the 26 serogroup 1 strains and one serogroup 2 strain all produced sequences identical to those of their type strains. Therefore, further testing of additional serogroup 2 wild strains may prove this small difference between the two serogroups to be definitive.

Wilkinson et al. (43) reported the existence of strains representing six uncharacterized species, designated species A to F. These strains were amplified, and the products were sequenced to determine how well the genotyping scheme would identify them. Species F is the type strain of *L. adelaidensis* (3). Species A and B give unique sequences, supporting the ubiquinone and fatty acid profile data indicating that they are indeed new species. Species C, D, and E cluster closely with *L. londiniensis*, *L. waltersii*, and *L. brunensis*, respectively. However, either these three species were not characterized or isolates were not available at the time of testing (4, 10, 42). A comparison of the ubiquinone and fatty acid profiles for species C, D, and E and *L. londiniensis*, *L. waltersii*, and *L. brunensis* confirmed that the scheme had grouped them correctly. The speciation of species C and E has been independently confirmed with DNA-DNA hybridization (2a). In addition to these uncharacterized species, other isolates, stored because they possessed unusual or unique ubiquinone and fatty acid profiles, were also tested. Isolates IMVS-911 and IMVS-960, designated species H, produced sequences identical to each other but unique from all characterized species. Similarly, IMVS-959, designated species I, produced a unique sequence. These three results are in complete agreement with the ubiquinone and fatty acid profiles, and these strains are likely to represent two new undescribed species. Species G isolates (IMVS-823 and IMVS-895) produced identical sequences but grouped moderately close to *L. shakespearei* (5.4% variation). These two isolates produce ubiquinone and fatty acid profiles similar to those of *L. shakespearei*, a species which had not been described at the time of their storage (39). The degree of difference from *L. shakespearei*, however, is greater than the intraspecies variation found in other species, so their true identity needs to be further elucidated. Species J (isolate IMVS-933) has grouped with *L. feeleii*. While the isolate produces a ubiquinone profile similar to that of *L. feeleii* (27), the fatty acid profile shows some differences. The significance of this result is unclear, and the identity of this isolate also needs to be further elucidated. Five European wild strains, similarly uncharacterized, were also tested. LC3043 and LC3044, serologically similar to each other, were found to also have sequences identical to each other and identical to those of the two species G strains. Interestingly they demonstrated weak serological homology only with *L. shakespearei*. The remaining three isolates, designated species L (LC1863), species M (LC4046), and species N (LC4048), were all found to have unique sequences, consistent with the phenotypic classification.

In conclusion, the scheme was able to unambiguously discriminate among 39 of 40 species and correctly group 26 additional serogroups or reference strains within those species. Additionally, 102 wild strains isolated within Australia and 50 wild strains from Europe, the United States, Singapore, Israel, and Kenya, including strains which are thought to represent currently uncharacterized species, were grouped consistently with their phenotypic identification. Two isolates grouped genotypically differently from their phenotypic classification, but it is probable that DNA hybridization would support the genotypic classification. There were no regional clonal variations detected which would indicate that laboratories in other countries would be troubled with ambiguous classification. The scheme is technically simple for a laboratory with even basic molecular capabilities and equipment, if access to a sequencing laboratory is available, especially given that heat lysis is quite adequate for extracting DNA suitable for amplification. Although all wild-strain amplification products were sequenced in both directions during this study, routinely only one direction would be necessary. While both sequencing primers performed extremely well, Legmip_fs would be the preferred sequencing primer, as it proved to be unaffected by nonspecific amplification.

There was no evidence of genetic recombination horizontally across species in the sequences from the 220 strains used in this study, and while they are still theoretically possible, such events would be unlikely to affect the classification of wild strains in practice. However, unusual results or critical isolates could be confirmed by other phenotypic methods or rRNA gene sequences.

ACKNOWLEDGMENTS

We thank Robert Benson and Tim Harrison for the provision of wild strains from the United States and Israel and from Europe, Singapore, and Kenya, respectively, and Norma Sangster for the provision of the remaining isolates used in this study. We also thank Mark Achtman for succinct manuscript advice and the Clive and Vera Ramaciotti Foundation for generous assistance in the purchase of equipment.

REFERENCES

1. **Achtman, M.** A phylogenetic perspective on molecular epidemiology. *In* M. Sussman (ed.), Molecular medical microbiology, in press. Academic Press, Inc., London, England.
2. **Bangsborg, J. M., N. P. Cianciotto, and P. Hindersson.** 1991. Nucleotide sequence analysis of the *Legionella micdadei mip* gene, encoding a 30-kilodalton analog of the *Legionella pneumophila* Mip protein. Infect. Immun. **59**:3836–3840.

2a. **Benson, R. (Centers for Disease Control and Prevention).** 1997. Personal communication.

3. **Benson, R. F., W. L. Thacker, J. A. Lanser, N. Sangster, W. R. Mayberry, and D. J. Brenner.** 1991. *Legionella adelaidensis*, a new species isolated from cooling tower water. J. Clin. Microbiol. **29**:1004–1006.
4. **Benson, R. F., W. L. Thacker, M. I. Daneshvar, and D. J. Brenner.** 1996. *Legionella waltersii* sp. nov. and an unnamed *Legionella* genomospecies isolated from water in Australia. Int. J. Syst. Bacteriol. **46**:631–634.

Appendix 4. (Two files located on CD in back sleeve) Sequence alignments of the data sets used for the phylogenetic analyses in Chapters Four and Five.