School of Biomedical Sciences
Faculty of Health Sciences

*Leptospirillum: A Study of the Nitrogen Fixing Capabilities*

Melissa Kaye Corbett

This thesis is presented for the Degree of
Doctor of Philosophy
of
Curtin University

December 2011
Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: .............................................

Date: .............................................
# Contents

Declaration ................................................. i

List of Figures ............................................. ix

List of Tables ................................................ x

Acknowledgements ......................................... xi

Abstract ..................................................... xii

Glossary ..................................................... xiv

1 Literature Review ........................................ 1

1.1 Introduction ........................................... 1

1.2 Biorefining .............................................. 1

1.2.1 History of biorefining ............................... 1

1.2.2 Modern biorefining .................................. 3

1.2.3 Factors affecting biorefining ...................... 3

1.2.4 Microbiology of biorefining ....................... 4

1.2.4.1 Microbial diversity of biorefining .......... 4

1.3 *Leptospirillum* ........................................ 5

1.3.1 Taxonomy of *Leptospirillum* ..................... 5

1.3.2 *Leptospirillum* metabolism ...................... 7

1.3.2.1 Energy generation ............................. 7

1.3.2.2 Carbon metabolism ............................ 9

1.3.2.3 Nitrogen metabolism .......................... 11

1.4 Biological nitrogen fixation .......................... 13

1.4.1 Microbial diversity of diazotrophy ............... 13

1.4.2 Biochemistry of BNF ............................... 14

1.4.2.1 Nitrogenase liability .......................... 16

1.4.3 Genetics of BNF .................................... 16

1.4.3.1 Structural *nif* genes ......................... 17

1.4.3.2 Assembly and biocluster formation genes .... 18
3 Nitrogen fixation gene regulation in *Leptospirillum* species

3.1 Introduction .................................................. 52
3.2 Materials and methods ........................................ 54
  3.2.1 Experimental design considerations ...................... 54
  3.2.2 Bacterial strains and growth media ........................ 54
  3.2.3 Experimental growth conditions ............................ 54
  3.2.4 Sampling regime ........................................ 55
  3.2.5 Isolating mRNA by proteinase K and acid phenol extraction . 55
  3.2.6 Removal of DNA from NA samples ........................ 55
  3.2.7 Confirmation of DNA removal .............................. 56
  3.2.8 Integrity analysis and quantification of isolated RNA .... 56
  3.2.9 Construction of cDNA by reverse transcription .......... 56
  3.2.10 Primers .................................................. 57
  3.2.11 Testing PCR primers ..................................... 57
  3.2.12 Quantitative Real Time PCR conditions .................. 58
  3.2.13 Data analysis for quantification of gene expression ...... 58
  3.2.14 Statistical analysis ..................................... 59
3.3 Results ......................................................... 60
  3.3.1 Extraction of RNA from *Leptospirillum* species ........ 60
  3.3.2 Expression of reference genes under test and control conditions ... 60
  3.3.3 Expression of target genes in *L. ferrodiazotrophum* .......... 65
  3.3.4 Expression of target genes in *L. ferrooxidans* ............ 66
  3.3.5 Expression of target genes in *L. ferriprophilum* ........... 68
  3.3.6 Comparison of target genes across *Leptospirillum* species ... 69
3.4 Discussion ....................................................... 71
  3.4.1 Effect of soluble nitrogen deprivation on nif mRNA expression ... 71
  3.4.2 Evaluation of the 16S rRNA and gyrB gene as reference genes .... 74
3.5 Conclusions ..................................................... 74

4 Proteomic analysis of whole cell extracts from *Leptospirillum* species grown in the absence of soluble nitrogen

4.1 Introduction .................................................... 76
4.2 Materials and methods ......................................... 78
  4.2.1 Bacterial strains and growth media ........................ 78
  4.2.2 Experimental growth conditions ............................ 78
  4.2.3 Isolating whole cellular proteins .......................... 78
  4.2.4 Assessing DNA contamination of protein extracts .......... 79
  4.2.5 Determination of protein concentration .................... 79
  4.2.6 Passive rehydration of IPG strips ......................... 79
4.2.7 First dimension Isoelectric Focusing (IEF) .......................... 79
4.2.8 Second dimension SDS-PAGE ....................................... 79
4.2.9 Protein visualisation with silver staining .......................... 80
4.2.10 Image analysis ....................................................... 81
4.2.11 Protein quantification with silver staining ......................... 81
4.2.12 Statistical analysis of proteomic data ............................. 81
4.2.13 Protein identification ................................................ 82
4.2.14 MALDI–TOF Mass Spectrometry .................................. 82
4.2.15 MALDI–TOF Mass Spectrometry data output ..................... 83
4.2.16 Analysis of Mass Spectrometry data ............................... 83
4.3 Results ........................................................................ 84
  4.3.1 Proteome changes in Leptospirillum species under soluble nitrogen deprivation .......................... 84
  4.3.2 L. ferrodiazotrophum proteomics ................................ 86
  4.3.3 L. ferriphilium proteomics .......................................... 89
  4.3.4 L. ferroxidans ......................................................... 92
  4.3.5 Comparative analysis of Molecular Weight and pl ................. 94
4.4 Discussion ..................................................................... 95
  4.4.1 Changes in L. ferrodiazotrophum protein expression .......... 95
  4.4.2 Changes in L. ferriphilium protein expression ................. 96
  4.4.3 Down regulated proteins of L. ferroxidans ...................... 97
  4.4.4 Significance of proteins identified in this study ............... 98
  4.4.5 Posttranslational modification of proteins ....................... 99
  4.4.6 Limitations of 2D-SDS PAGE proteomic analysis ............ 99
        4.4.6.1 Analytical Method ......................................... 100
4.5 Conclusions .................................................................. 101

5 Identification and structural analysis of the nif genes in Leptospirillum species 102
  5.1 Introduction .................................................................. 102
  5.2 Materials and Methods .................................................. 103
        5.2.1 Bacterial strains and growth media ........................... 103
        5.2.2 Bacterial cell harvesting ........................................ 103
        5.2.3 DNA preparation by proteinase K lysis and phenol extraction .................. 103
        5.2.4 PCR and sequencing primers .................................. 104
        5.2.5 Polymerase chain reaction (PCR) amplification ............ 104
        5.2.6 PCR clean up and sequencing preparation .................. 105
        5.2.7 Sequence analysis ............................................... 107
        5.2.8 Sequence alignments ............................................. 107
        5.2.9 Codon usage and determination ............................... 107
5.2.10 Theoretical protein modelling ................................. 107
5.3 Results ........................................................................ 108
  5.3.1 nifHDK genes in L. ferrodiazotrophum and L. ferrooxidans 108
  5.3.2 Identification of nifH, nifD and nifK in L. ferrhiphium DSM 14647 110
  5.3.3 Structures associated with the nifHDK operon in Leptospirillum 112
  5.3.4 Identifying conserved sequence regions in Leptospirillum nifHDK genes 116
     5.3.4.1 Analysis of nifH characteristics ............................. 117
     5.3.4.2 Analysis of nifD characteristics ............................. 117
     5.3.4.3 Analysis of nifK characteristics ............................. 117
5.4 Discussion .................................................................... 121
  5.4.1 Gene characteristics of nifH, D and K .......................... 121
  5.4.2 Signal structures .................................................... 123
  5.4.3 Amino acid structural assessment ............................... 124
5.5 Conclusions .................................................................. 128
6 Phylogenetic analysis of Leptospirillum species .................. 129
  6.1 Introduction .............................................................. 129
  6.2 Materials and methods ................................................ 131
     6.2.1 Leptospirillum sequences for Phylogenetic Analysis ........ 131
     6.2.2 PCR and sequencing primers .................................. 131
     6.2.3 Sequence selection for evolutionary comparison ............ 132
     6.2.4 Phylogenetic Analysis ............................................. 132
     6.2.5 Construction of Phylogenetic Trees ............................ 133
  6.3 Results ....................................................................... 135
     6.3.1 Leptospirillum phylogeny based on housekeeping gene sequences 135
        6.3.1.1 16S rRNA ....................................................... 135
        6.3.1.2 gyrB and rpoB .................................................. 137
        6.3.1.3 DNA concatenation of housekeeping genes ................. 138
     6.3.2 Leptospirillum nif gene phylogenies ........................... 140
        6.3.2.1 nifH analysis ..................................................... 140
        6.3.2.2 nifD analysis ..................................................... 144
        6.3.2.3 nifK analysis ..................................................... 148
        6.3.2.4 Concatenation of the NifHDK proteins ...................... 152
  6.4 Discussion .................................................................. 154
     6.4.1 Concatameric housekeeping phylogeny ........................ 154
     6.4.2 The nif phylogeny ................................................... 155
        6.4.2.1 The nif DNA phylogeny .................................... 156
        6.4.2.2 Nif protein phylogeny ....................................... 156
     6.4.3 The phylogenetic methods chosen ................................ 158
6.5 Conclusions .................................................. 158

7 General Discussion and Future Considerations 160
7.1 Final Conclusions and Outlook .......................... 165

Appendix A Supplementary Information - I 167
A.1 Quality assurance and standard curves .................. 167
A.2 Genbank accession numbers ................................. 174

Appendix B Supplementary Information - II 177
B.1 Computing .................................................... 177
B.2 Typesetting ................................................... 178

References ....................................................... 179

List of Figures

1.1 AMD in the Rio Tinto River ................................. 2
1.2 Electron micrograph of Leptospirillum ferrooxidans .......... 5
1.3 Phylogenetic positioning of Leptospirillum branching within the Nitrospira genus ........................................ 6
1.4 Metabolic pathways of Leptospirillum ........................ 8
1.5 Reductive Acetyl-CoA pathway of L. ferrooxidans ............. 10
1.6 Reverse TCA cycle ............................................ 10
1.7 Proposed regulatory model for nitrogen assimilation .......... 12
1.8 Nitrogenase action ............................................. 15
1.9 Organisation of structural nif genes in various diazotrophs .... 18
1.10 Phylogenetic nif grouping ..................................... 25

2.1 Growth patterns of At. caldus under soluble nitrogen deprivation ........ 34
2.2 Growth patterns of At. ferrooxidans .......................... 35
2.3 Growth patterns of Leptospirillum species .................... 36
2.4 Growth of L. ferrodiazotrophum under nitrogen starvation .... 37
2.5 Growth of L. ferriphilum under nitrogen starvation ............ 37
2.6 Growth of L. ferrooxidans under nitrogen starvation .......... 38
2.7 At. ferrooxidans Fe²⁺ oxidation rates ........................ 39
2.8 Leptospirillum Fe²⁺ oxidation rates .......................... 41
2.9  Averaged Fe$^{2+}$ oxidation rates cell$^{-1}$ h$^{-1}$ ................................. 43
2.10 Doubling events and Fe$^{2+}$ oxidation by *Leptospirillum* ...................... 44
2.11 Detection of nitrogenase activity via ethylene production .......................... 46

3.1  Livak Method formula ................................................................. 59
3.2  Pfaffi Method formula ................................................................. 59
3.3  RNA extracted from *Leptospirillum* species .................................... 60
3.4  Melt curve analysis of 16S rRNA qRT-PCR products .............................. 63
3.5  Melt curve analysis of *gyrB* qRT-PCR products ................................ 64
3.6  $\log_2$ transformed *L. ferrodiazotrophum* nif gene expression ................ 65
3.7  $\log_2$ transformed *L. ferrooxidans* nif gene expression ........................ 66
3.8  $\log_2$ transformed *L. ferriphilum* nif gene expression ........................ 68
3.9  Comparison of nif gene expression across *Leptospirillum* species .......... 70

4.1  2D SDS PAGE of *Leptospirillum* spp reference protein patterns ............. 85
4.2  *L. ferrodiazotrophum* spot map of protein abundance .......................... 87
4.3  *L. ferriphilum* spot map of protein abundance .................................. 90
4.4  *L. ferrooxidans* spot map of protein abundance ................................ 93
4.5  Comparative analysis of MW and pl ................................................ 94
4.6  Regulated proteins of *Leptospirillum* spp. ..................................... 98

5.1  Full length *nifHDK* PCR in two *Leptospirillum* species ...................... 108
5.2  Gradient PCR for *nifH* amplification in *L. ferriphilum* ...................... 110
5.3  PCR amplification of *nifH* in *L. ferriphilum* ................................ 111
5.4  PCR amplification of *nifD* and *nifK* in *L. ferriphilum* ..................... 111
5.5  Codon usage of *Leptospirillum* nif genes ..................................... 115
5.6  Inverted sequence string of the *nifH-nifD* intergenic spacer region ......... 116
5.7  Clustal W alignment of NifH sequences in *Leptospirillum* ..................... 118
5.8  Clustal W alignment of NifD sequences in *Leptospirillum* ..................... 119
5.9  Clustal W alignment of NifK sequences in *Leptospirillum* ..................... 120
5.10 *nifHDK* gene arrangement in *L. ferriphilum* .................................. 122
5.11 3D representation of conserved NifH regions .................................... 124
5.12 3D representation of conserved sites of NifD ................................... 126
5.13 3D representation of conserved sites of NifK ................................... 127

6.1  Gene organization and concatenation of housekeeping genes ................... 133
6.2  Nif gene organization and concatenation ......................................... 134
6.3  *Leptospirillum* 16S rRNA phylograms .......................................... 136
6.4  *Leptospirillum* *rpoB* and *gyrB* phylograms .................................. 138
6.5  Phylogenograms of concatenated sequences of 16S rRNA, *gyrB* and *rpoB*  139
List of Tables

1.1 Genes essential for nitrogen fixation ........................................ 17
1.2 Bacterial strains used in this study ............................................ 30
1.3 Calculations and Formulae ..................................................... 33
1.4 Exponential growth rates of *Leptospirillum* species ...................... 39
1.5 Doubling events of *Leptospirillum* species ................................ 44
3.1 PCR cycle parameters for RNA examination ................................. 56
3.2 Oligonucleotide primer sequences .......................................... 57
3.3 qRT-PCR cycle parameters .................................................. 58
3.4 Average 16S rRNA and gyrB C<T> values ................................ 61
3.5 T-test results on 16S and gyrB expression .................................. 62
3.6 Efficiency of housekeeping gene q-RTPCR ................................ 62
4.1 Mass Spectrometry Program Parameters .................................... 83
4.2 PMF identification of *L. ferrodiazotrophum* proteins .................... 88
4.3 PMF identification of *L. ferriphilum* proteins .............................. 91
4.4 PMF identification of *L. ferrooxidans* proteins ............................ 92
5.1 PCR cycle protocols for all *Leptospirillum* 16S rRNA ......................... 104
5.2 PCR cycle protocols for all *Leptospirillum nifH* amplification ............... 104
5.3 Primers for *Leptospirillum spp nifHDK* amplification ....................... 106
5.4 BLAST results for *L. ferrodiazotrophum* and *L. ferrooxidans nif* genes .... 109
5.5 BLAST results of *L. ferrihilum* gene fragments ................................ 112
5.6 Sequence characteristics of *Leptospirillum nif* genes .......................... 113
5.7 Shine-Delgarno/RBS sequences ..................................................... 115
5.8 Nucleotide sequence alignment ..................................................... 116

6.1 Degenerate gyrB and rpoB primer sequences used in this study ................. 131
6.2 PCR cycle parameters for gyrB and rpoB ........................................ 132
6.3 Housekeeping gene sequence identity ............................................. 135

A.1 Growth Media Composition Analysis .............................................. 167
A.2 Ferrous Standards for Iron oxidation ............................................ 168
A.3 *L. ferrodiazotrophum* C5 relative synonymous codon usage .................. 171
A.4 *L. ferrihilum* DSM 14467 relative synonymous codon usage .................. 172
A.5 *L. ferrooxidans* DSM 2705 relative synonymous codon usage ................ 173
A.6 Genbank Accession numbers of 'housekeeping' gene sequences ............... 174
A.7 DNA sequences Genbank Accession numbers ..................................... 174
A.8 Genbank Accession numbers of AA Nif sequences ................................ 175
A.9 Unidentified *Leptospirillum* sp proteines ..................................... 176

B.1 Commands for MrBayes 3.2 Phylogenetic analysis ................................ 177
B.2 Computer programs ................................................................. 178
Acknowledgements

I would sincerely like to first thank Associate Professor Elizabeth Watkin for giving me the opportunity to work in her laboratory. Her patience, support, enthusiasm and advice were greatly appreciated during the rough times of my career and personal life.

My gratitude is also extended to all the past and present members of the research group, I thank you all for your assistance and friendship throughout my years at Curtin University.

To Rob Steuart for teaching me the ways of 2D SDS-PAGE, your help put me on the right path. To Celia Smuts for her patience and invaluable assistance with most of the mass spectrometry work conducted in this thesis performed at the State Agricultural Biotechnology Centre. Thank you to Greg Cawthray and Professor Craig Atkins at the University of Western Australia for the use of the ARA equipment and expertise when it came to troubleshooting problems.

I also thank Dr. John Wetherall and Dr. Chris Bryan, for their assistance and helpful comments with the writing of this thesis of which I benefited from their knowledge and guidance.

Additionally, I would like thank the faculty members from the School of Biomedical Science, who provided invaluable knowledge and support as well as technical assistance throughout the course of this project.

Most importantly to my family especially my father and mother who have always been there for me with their kind support, patience and encouragement that help me in every step of my life. Thank you.

And last, but certainly not least, I thank Jason. I do not have enough room in this dissertation to begin to scratch the surface of what Jason has done for me. You have been there for me at each one of lifes turns, whether good or bad. Without you I would never have gotten this far. With love, I thank you.

This study has been supported by a Curtin University Postgraduate Award, and Scholarship from the Parker Cooperative Research Centre for Integrated Hydrometallurgy Solutions.
Abstract

To survive under conditions of soluble nitrogen-limitation, the ability of some prokaryotes to fix atmospheric nitrogen is advantageous. Members of the *Leptospirillum* genus are small, gram negative chemolithoautotrophs capable of growth in nutrient limited, extremely acidic, metal rich environments. Harboured within their genome, are the genes necessary for the regulation of nitrogen fixation. Three species of *Leptospirillum* were analysed in this project, *L. ferrooxidans*, *L. ferrihilum*, and *L. ferrodiazotrophum*. Each was exposed to soluble nitrogen deprivation and assessed for changes in physiology, genome expression and protein abundance.

Primary investigations into the effect of soluble nitrogen deprivation on *Leptospirillum* species growth demonstrated that cellular proliferation continued for all species under such conditions, with *L. ferrihilum* exhibiting the fastest generation times and iron oxidation rates of the three species analysed. In addition, all strains were subject to acetylene reduction assays to measure nitrogen fixation activity, with *L. ferrodiazotrophum* exhibiting maximum levels of acetylene reduction, 160 hours post exposure to soluble nitrogen deprivation.

Quantitative Real Time PCR analysis of *Leptospirillum* species mRNA following soluble nitrogen deprivation showed altered expression of a number of genes essential to nitrogen fixation. Expression of the *nifHDK* genes, crucial for formation of the nitrogenase enzyme, peaked after 96 hours of soluble nitrogen deprivation in *L. ferrodiazotrophum* and *L. ferrooxidans*. A noticeable increase in the level of *nifA* in *L. ferrihilum* was detected after 96 hours, remaining high up to 144 hours after soluble nitrogen deprivation.

As a consequence of detected fluctuations in the mRNA levels of nitrogen fixation genes, possible alterations to the proteome for each species were investigated. 2D-SDS PAGE combined with MALDI-TOF/MS was employed to assess and identify proteins from *Leptospirillum* whole cell protein extracts whose abundance changed following exposure to soluble nitrogen deprivation. These studies identified changes in a variety of proteins, most notably in pathways essential to maintaining metabolism and amino acid biosynthesis.

Structural analysis of the *nifHDK* genes within each of the species demonstrated strict conservation of amino acid structure in comparison to known functional diazotrophs, support-
Abstract

...ing theoretical conservation and functionality of this enzyme complex in all *Leptospirillum* species. Further analysis of these genes utilizing phylogenetic methods allowed for inferences to be made regarding the evolutionary history within the *Leptospirillum* genus. Additional *nifHDK* coding sequences from close phylogenetic and ecological related organisms were selected and with maximum-likelihood and Bayesian inference methods, *nifHDK* resulting trees showed congruent evolutionary history with the 16S rRNA genes for all species studied.

**Keywords:** *Leptospirillum*, nitrogen fixation, mRNA transcription, protein abundance, phylogenetics.
Glossary

GC  Gas Chromatography
DNA  deoxyribonucleic acid
IPG  Immobilized pH gradient
IEF  Isoelectric focusing
2D SDS-PAGE  Two dimensional sodium dodecyl sulfate Polyacrylamide Gel Electrophoresis
MALDI-TOF  Matrix-assisted laser desorption/ionization-time of flight
pI  Isoelectric point
MW  Molecular Weight
PMF  Peptide Mass Fingerprint
CHAPS  (3-[(3-Cholamidopropyl)-dimethylammonio]-1-propane sulfonate)
ARA  Acetylene reduction assay
PMSF  phenylmethanesulfonylfluoride
EDTA  Ethylenediaminetetraacetic acid
DTT  Dithiothreitol
UV  Ultraviolet
BSA  Bovine Serum Albumin
PCA  Principal Component Analysis
ACN  Acetonitrile
TFA  Trifluoroacetic Acid
SABC  State Agricultural Biotechnology Centre
CHCA  α-Cyano-4-hydroxycinnamic acid
AB  Applied Biosystems
kDa  kiloDalton
TEMED  Tetramethylethylenediamine
Min  minutes
s  seconds
h  hours
g  grams
mg  milligrams
mL  millilitre
mA  milliamps
V  volts
BSM  basal salts media
DSM  Deutsche Sammlung vonFur Mikroorganismen
PCR  Polymerase Chain Reaction
Glossary

DMSO  Dimethyl sulfoxide
DEPC  - Diethylpyrocarbonate
RT    Reverse transcriptase
NA    Nucleic Acid
CBB   Calvin-Benson-Bessel
PP    Posterior probabilities
BS    Bootstrap
IGS   Intergenic spacer
Γ     Gamma
I     Invariant sites
MP    Maximum parsimony
Chapter 1

Literature Review

1.1 Introduction

The ecosystem, both on land and in the water, depends heavily upon the activity of microorganisms, which play vital roles in recycling nutrients, such as nitrogen, carbon and sulphur. Prokaryotes represent by far the largest reservoir of genetic diversity on Earth, with a large proportion of them remaining uncharacterised (Rappe and Giovannoni 2003). Microorganisms have been exploited in biotechnological processes and play a fundamental role in modern life, from their inclusion in food, to the production of biofuels, and application such as biohydrometallurgy. As global technologies and civilizations have advanced, the demand for raw minerals such as copper, nickel and zinc has increased resulting in the depletion of high grade ores (Watling 2006). Consequently, to meet emerging demands, new technologies are being developed that capitalise on the natural action of micro-organisms.

1.2 Bioleaching

Bioleaching is the solubilisation of metals from mineral ores catalysed by micro-organisms, an action that occurs in many environments (Dopson et al. 2003). The use of commercial leaching of metal sulphides by microbial action, also referred to as “Biomining” has rapidly developed and is now a well established and incorporated technology within the mining industry.

1.2.1 History of bioleaching

Observations of microbial leaching were made as early as 162 A.D from sites in Cyprus, where in situ copper leaching and the concentration of cupriferous mine water resulted
in the deposition of crystallised copper sulphate (Constantinou 1992). More evidence of uncontrolled microbial leaching can also be seen in unmined mineralized areas, (Foster et al. 2007) and at abandoned metal mines where acidic waste water is pregnant with iron. The most notable example of this is the Rio Tinto (red-river), in Spain, where industrial mining for copper, gold and silver generated iron laden acidic waste waters, that drained into the river, resulting in a strong red pigment and extremely low pH (Davis Jr et al. 2000).

![Figure 1.1 - Acid mine drainage prevalent in the Rio Tinto River. Image: Courtesy of Anna Kaksonen, 2007](image)

The biological leaching of mineral bearing ores and the resultant generation of acid mine drainage (AMD) was first attributed to microbial action by Colmer and Hinkle (1947), who demonstrated that two microorganism were responsible. One, that oxidized ferrous iron to ferric sulphate named *Thiobacillus ferrooxidans* (reclassified as *Acidithiobacillus ferrooxidans*), and a second, *Thiobacillus thiooxidans* (reclassified as *Acidithiobacillus thiooxidans*) which is able to oxidise both iron, sulphur and sulphur based compounds resulting in the generation of sulphuric acid. This previously undocumented microbial behaviour explained the generation of such acidic waters from heavy metal mining of sulphide bearing ores, while bituminous coal mines generated near neutral or alkaline waters (Colmer, Temple, and Hinkle 1951).
1.2.2 Modern bioleaching

The application of bioleaching for industrial purposes began in the 1950’s (Brierley and Brierley 2001) and between 1950 and 1980 was utilized mainly for the recovery of copper and other metals from low grade ores (Acevedo 2002). Since then sophisticated technologies have developed, enabling microbial-driven operations to catalyse the oxidation of sulphide minerals for the subsequent extraction of valuable metals from ores; that would otherwise be unrefinable by traditional pyrometallurgical techniques (Rohwerder et al. 2003). Although certain metals can be extracted via this approach, not all minerals are susceptible to bioleaching (Rawlings 2002), and mineral ores that contain iron or a reduced form of sulphur will be more successful with bioleaching processes.

Bioleaching for the processing of mineral ores has several advantages over conventional physiochemical methods, including the use of naturally occurring components (microorganisms, water and air) with ambient pressure and temperatures (Lizama 2001). Unlike established smelting and refining processes for mineral extraction, less energy is required, reducing capital costs incurred with mineral recovery. Additionally, low-grade sulphide ores can be economically processed with no sulphur dioxide or other environmentally harmful gasses emitted (Okibe et al. 2003; Olson, Brierley, and Brierley 2003).

The replacement of all traditional smelting processes for mineral recovery by bioleaching is not viable as not all precious metals present in the ore body can be recovered, affecting the profitability of the operation. Furthermore, leaching rates of some operations are slow with, additional delivery of supplementary nutrients for high microbial growth and bioleaching rates costly (Zhou et al. 2009).

The application and effectiveness of metal bioleaching in an acidic environment can be influenced by the physical, chemical and biological factors in the system (Chen and Lin 2001).

1.2.3 Factors affecting bioleaching

There are a number of factors that will influence the rate at which bioleaching occurs and which micro-organisms will be present. A controlled acidic environment with a pH of 1.5 – 3.0 will permit the growth of a variety of micro-organisms as most metal ions will remain in solution (Bosecker 1997). However, continual fluctuations in pH will alter microbial community structures, resulting in variable leaching rates (Morin 2007). Adequate aeration, allowing for the exchange of CO₂ for growth and O₂ for proton turnover is necessary for high activity of most bioleaching microbes (Witne and Phillips 2001).

Within bioleaching systems, the mineral composition of the ore can vary, affecting the
growth of the microbial community (He et al. 2008). Different species of microorganisms have different sensitivities to the presence of heavy metal ions, with many leached ions fatally toxic to the microorganism (Nies 1999). Some microorganisms may develop a tolerance to heavy metals, either by constant exposure and resulting gene mutation, or by horizontal gene transfer (Martinez et al. 2006). Those organisms that have developed greater tolerance to heavy metal ions will be more successfully in leaching operations (Dopson et al. 2003).

1.2.4 Microbiology of bioleaching

Originally microorganisms utilized in bioleaching were indigenous to the mine site where the leaching occurred, and due to substrate limitations that exist within mining environments, it was previously thought that these environments would have extremely poor microbial diversity (Rohwerder et al. 2003). With the advancement of biomolecular techniques, research into the microbial ecology of extremely acidic leaching environments (natural and industrial) enabled the discovery of considerably diverse prokaryotes inhabiting the same niche (Johnson 2001). The majority of microorganisms chosen for bioleaching processes are extremely acidophilic chemolithoautotrophs, fixing CO₂ from the atmosphere for growth and obtaining energy from the oxidation of inorganic metal sulphides, and/or ferrous iron (Dopson and Lindstrom 2004).

1.2.4.1 Microbial diversity of bioleaching

The consortium of microorganisms commonly found in metal leaching populations are primarily determined by the temperature of the bioleaching process. Mesophilic microorganisms, those that thrive between 20 and 45 °C, are known to dominate bioleaching systems and are therefore the most highly researched group of organisms (Rawlings, D. E. 2005). Included in this group are At. ferrooxidans, At. thiooxidans, and At. caldus belonging to the gram negative γ-proteobacteria, and members of the Nitrospira genus, Leptospirillum ferrooxidans, L. ferrilphilum and L. ferrodiazotrophum. When temperatures approach 50 °C, or are moderately thermophilic, gram negative bacterial populations decrease, whilst gram positive bacteria including Sulfobacillus and Acidimicrobiium begin to dominate (Norris and Clark 1996), along with an archeal genus, Ferroplasma.

At present, the majority of microorganisms used in commercial bioleaching have been cultured from indigenous strains (Olson, Brierley, and Brierley 2003) with some selected for rapid growth on particular ores and concentrates (Pradhan et al. 2008). One shared feature of all bioleaching processes is that is that no attempt is made to maintain the sterility of
the operation. The highly acidophilic chemoautolithothrophic microorganisms involved create an environment that is inhospitable to other organisms (Rawlings, D. E. 2005), and those organisms that are less efficient at decomposing mineral ores to meet their energy requirements will be out competed. Over a period of time, this will result in a microbial population optimized for a leaching operation (Rawlings and Johnson 2007).

In bioleaching operations under 45 °C, two to four different microbial species will be present, including a dominate iron oxidizer and a sulphur oxidizer (Okibe et al. 2003). Due to the increased array and sensitivity of scientific techniques now available, examination of these mixed populations has shown that *Leptospirillum*, a strict iron oxidiser, outperforms the sulphur and iron oxidizer *At. ferrooxidans*, at a ratio of 2:1 (Coram and Rawlings 2002). Combined with its ability to leach minerals, and propensity to be highly abundant, the importance of understanding the *Leptospirillum* genus within bioleaching systems has increased.

### 1.3 *Leptospirillum*

The bacterium "*Leptospirillum*" was first described in 1972 by M. Markosyan and isolated from mine water of the Alaverda copper deposit in Armenia (Balashova et al. 1974). Phylogenetic studies performed on *Leptospirillum* established it as new bacterial genus. The name *Leptospirillum* is derived from its apparent shape. Leptos meaning thin and Speira a spiral.

![Figure 1.2 - Electron microscopic image of *L. ferrooxidans*](image)

**Figure 1.2** - Electron microscopic image of *L. ferrooxidans* (Source: Bangor University).

#### 1.3.1 Taxonomy of *Leptospirillum*

Genome studies of *Leptospirillum* species are still in their infancy, but examination of the 16S rRNA sequence, places these bacteria within the Nitrospira family (Johnson 2001). Members of the Nitrospira are a monophyletic but diverse group of organisms, widely distributed in many natural habitats, and play key roles in nitrogen cycling (Maixner et al. 2006).
As the majority of Nitrospira are uncultured (Lucker et al. 2010), and the available cultures difficult to maintain (Off, Alawi, and Spieck 2010), only a few studies have addressed their ecology and physiology. Grouping of *Leptospirillum* is determined by G+C content of the 16SrRNA sequence. Based on this 16S rRNA sequence, currently 3 species of *Leptospirillum* have been identified; *Leptospirillum ferrooxidans* (Group I), *Leptospirillum ferripilum* (Group II), and *Leptospirillum ferrodiazotrophum* (Group III) (Bond and Banfield 2001). A previously identified 4th *Leptospirillum* species designated *L. thermoferrooxidans*, capable of growth in thermophilic locations (>45 °C) was isolated from two separate locations; a hydrothermal spring on Kunashir Island, Russia, in 1992 and a bioleaching reactor in Australia circa 1994 (Hippe 2000) but has since been lost from laboratory cultivation (Rawlings 2002). *Leptospirillum rubarum*, a new species of *Leptospirillum* was described by Goltsman et al. 2009 but is yet to be formally named and recognised.

![Phylogenetic tree of Leptospirillum species](image)

**Figure 1.3** – Phylogenetic positioning of *Leptospirillum* branching within the Nitrospira genus. Inference analysis was performed using published 16S rRNA sequences from known organisms. Genbank accession numbers are indicated in parenthesis. The tree was constructed using MEGA 5.0 using GTR+G model. Percentages (as decimals) of 1000 bootstrap samplings supporting the shown tree are shown next to the nodes.

*Leptospirillum* species are found in metal rich acidic environments (pH 1.3-2.0), in particular, systems whereby iron-bearing minerals are exposed to oxygen and water, and the
temperature is mesophilic (25-40°C). They can be morphologically variable, but are recognised as small (0.3-0.9μm), gram negative, vibrioid to spiral shaped cells, with spirals consisting of 3 to 12 cells that allow for ease of movement through a fluid rich area (Krieg and Hylemon 1976). They possess a single polar flagellum and have been classified as obligately chemolithotrophic bacteria, utilizing CO₂ as a carbon source, with organic carbon sources demonstrated to have a detrimental effect on *Leptospirillum* growth (Sand et al. 1992). Energy requirements are met by the aerobic oxidation of ferrous iron to ferric, and under these conditions, this demonstrated survival in low pH regions combined with strict energy and nutrient requirements and tolerance to heavy metals has led to *Leptospirillum* species being referred to as ‘Extremophiles’ (Norris, Burton, and Foulis 1999).

### 1.3.2 *Leptospirillum* metabolism

#### 1.3.2.1 Energy generation

The physiology of *Leptospirillum* metabolism is homogeneous (Garcia-Moyano et al. 2008) and dissimilar to that of other chemolithoautotrophs, where heterogeneous metabolisms are prevalent. *Leptospirillum* species are not capable of sulphur oxidation and rely solely on the oxidation of ferrous iron as an electron donor to obtain energy, utilizing oxygen as an electron acceptor (Rawlings 2002).

\[ 2Fe^{2+} + \frac{1}{2}O_2 + 2H^+ \rightarrow 2Fe^{3+} + H_2O \]

Formula 1.1: Oxidation of ferrous iron by microbial action.

Oxygen acts as an electron acceptor and iron oxidation for the production of energy will only occur under aerobic conditions (Tyson et al. 2004) in acidic environments. Under aerobic and neutral pH conditions ferrous iron will spontaneously oxidise to ferric and this source of electron transfer isn’t made available to non-acidophiles, however some exceptions such as *Leptothrix ochracea* and *Gallionella ferruginea* can actively oxidize Fe²⁺ at a neutral pH (Katsoyiannis and Zouboulis 2004).

Oxidation of ferrous iron to ferric generates very little energy (Konhauser 1997), and by relying only on the oxidation of soluble Fe²⁺ to meet their energy requirement, *Leptospirillum* species require a high throughput of Fe²⁺ to obtain sufficient energy for growth (Blake et al. 1993). Consequently, growth of *Leptospirillum* populations will result in vast amounts of ferric iron precipitation (Rojas-Chapana and Tributsch 2004). A notable characteristic of *Leptospirillum* is that its proliferation isn’t inhibited by the high levels of ferric generated and as such, dominates its environmental habitat when levels of ferric begin to increase (Rawlings, Tributsch, and Hansford 1999).
The large quantities of iron required for energy aren’t transported through the cell membrane (Rawlings, D. E. 2005), but remain outside the cell and transport of the electron occurs via an enzymatic system composed of a series of cytochromes and quinone pools inside the cell membrane (Figure 1.4) (Merino, Andrews, and Asenjo 2010).

![Figure 1.4 - Metabolic pathways of *Leptospirillum* showing Fe²⁺ oxidation, NH₄⁺ assimilation and CO₂ incorporation. Image referenced from Tyson et al. 2004.](image)

These essential components to an electron transport chain include cytochrome cbb₃ heme copper terminal oxidases and cytochrome bd-type quinol oxidases which have a high affinity for oxygen (Pitcher, Brittain, and Watmough 2002). Studies have discovered that under aerobic respiration on iron, *Leptospirillum* species lack a rusticyanin molecule responsible for iron electron transport found in *At. ferrooxidans* and other acidophilic micro-organisms, and instead an acid stable cytochrome with a deep red colour is expressed (Blake et al. 1993; Ram et al. 2005).

Initially one red cytochrome was detected and its reduced form registered an absorbance peak of 579 nm and was thus named Cyt₅₇₉ (Hart et al. 1991). This Cyt₅₇₉ is unique to *Leptospirillum* and has been detected in all groups and is analogous to rusticyanin. It is highly abundant and localised to the extracellular membrane (Ram et al. 2005) where its enzymatic activity is indicative of an electron transfer protein, shuttling electrons resultant from Fe²⁺ oxidation to support critical metabolic functions (Singer et al. 2008). More recently, a second abundant membrane bound, acid stable, yellow-red cytochrome has been detected, that when reduced has an unique spectral signature peak at 572 nm, identified as Cyt₅₇₂ (Jeans et al. 2008). Cyt₅₇₂ oxidises Fe²⁺ at a low pH (0.95-3.4) more readily than Cyt₅₇₉ whose oxidation of Fe²⁺ isn’t thermodynamically favoured at a pH <3. This may be indicative of Cyt₅₇₂ shuttling electrons to Cyt₅₇₉ for critical steps in the electron transport.
chain.

Even though growth of Leptospirillum species is restricted to iron bearing mineral compounds, it hasn’t stopped them from being widespread throughout the ecosystem. Leptospirillum has been isolated from pyrite containing waste waters of uranium mines in Mexico (Norris and Kelly 1982) and sulphidic mines of Romania (Hallmann et al. 1992) to coal refuse heaps in South Carolina (Brofft, McArthur, and Shimkets 2002). The ability of Leptospirillum to tolerate the presence of some heavy metals also allows for growth on chalcopyrite and nickel laterites (Simate and Ndlovu 2008). Research has also shown that Leptospirillum can also survive on Fe$^{2+}$ from iron meteorites (Gonzalez-Toril et al. 2005).

1.3.2.2 Carbon metabolism

Autotrophic microorganisms obtain their carbon for cell mass synthesis from the carbon dioxide present in the atmosphere with CO$_2$ concentrations generally sufficient to avoid carbon limitation (Rawlings, D. E. 2005). Carbon fixation is essential for Leptospirillum proliferation, but the CO$_2$ fixation pathway it uses remains obscure. Earlier studies conducted on the carbon metabolism of Leptospirillum species stated that CO$_2$ fixation occurred via the Calvin-Benson-Bassham cycle (Coram and Rawlings 2002; Osorio et al. 2008), but more recent investigations have discovered that Leptospirillum group II and III lack the ribulose-5-phosphate kinase, an enzyme essential in the CBB cycle (Goltsman et al. 2009; Levican et al. 2008), indicating CO$_2$ fixation by a different mechanism.

In 2007, Victor Parro and colleagues determined that L. ferrooxidans (group I) assimilated CO$_2$ via the reductive acetyl-CoA pathway, due to the expression of the porGAB operon, encoding for a pyruvate flavodoxin/ferrodoxin oxidoreductase (por). In this instance, one CO$_2$ molecule is reduced to a methyl group and a second CO$_2$ molecule to carbonyl group by the enzyme CO dehydrogenase. The combining methyl and carbonyl groups give rise to acetyl-CoA which is reductively carboxylated to pyruvate by por, from which all other central metabolites can be formed (Hugler et al. 2005; Pezacka and Wood 1988). A metabolic model demonstrated that the stoichiometric chemistry of the reductive acetyl-CoA cycle was plausible, as was a reductive tricarboxylic acid cycle (rTCA) (Merino, Andrews, and Asenjo 2010), which is also reliant on the por enzyme.
Unlike *Leptospirillum* group I, genetic analysis of *Leptospirillum* group II (*L. ferrariphilum*) showed no active binding sites for CO dehydrogenase and Acetyl-CoA synthase required for the successful operation of the reductive acetyl-CoA cycle (Goltsman et al. 2009). On further analysis of the genome it became apparent that *Leptospirillum* group II species contained coding genes for all enzymes of the reverse TCA cycle (Levican et al. 2008), with four key enzymes present; ATP citrate lyase (ACL), fumarate reductase (FDR), 2-oxoglutarate ferredoxin oxidoreductase (OGOR) and pyruvate ferredoxin oxidoreductase (POR) (Hugler et al. 2003). *Leptospirillum* group III (*L. ferrodiazotrophum*) is genetically more closely aligned with group II than group I, and also has highly similar enzymes identified for the rTCA cycle.

Figure 1.5 – Reductive Acetyl-CoA pathway demonstrating the theoretical method of CO₂ incorporation by *L. ferrooxidans*. Image sourced from Hugler and Sievert 2011

Figure 1.6 – Reverse TCA cycle. Acetyl CoA is synthesized by the incorporation of 2 molecules of CO₂, the input of 8 H⁺ (in the form of NADH and/or FADH), and 2 ATP. The biochemical reactions involved, as well as the enzymes (red) catalysing the reactions, are depicted. Image sourced from (Hugler and Sievert 2011).

These finding are unusual because the reverse TCA cycle is commonly found in anaerobic
environments, as the POR and OGOR enzymes are oxygen sensitive (Yoon et al. 1996). Some microaerophilic autotrophs such as *Hydrogenobacter thermophilus* have functional rTCA pathways with isomeric forms of OGOR and POR being more oxygen tolerant (Yamamoto et al. 2006), as maybe the case for *Leptospirillum* species. Autotrophy via the rTCA cycle is theoretically probable considering the presence of these enzymes and the lack of those for the CBB cycle, even without supporting experimental data. It is worth noting that other members of the Nitrospira genus are known to use the rTCA cycle for carbon fixation Hugler and Sievert 2011; Lucker et al. 2010.

### 1.3.2.3 Nitrogen metabolism

Nitrogen is a constituent of nearly all biomolecules (Howard, J. B., and Rees 1996) and alongside carbon, hydrogen and oxygen, is imperative for life (Kasting and Siefert 2001). Although nitrogen is the single largest component of the Earth's atmosphere, in the form of dinitrogen (N\textsubscript{2}) it is inaccessible to the majority of plants and microorganisms because of the triple bond present between the two nitrogen atoms, rendering it extremely stable and unreactive (Davidson and Seitzinger 2006). Like the large majority of prokaryotes, *Leptospirillum* species readily assimilate nitrogen in the form of NH\textsubscript{4}+, with genes for regulating nitrogen assimilation found in all *Leptospirillum* species, including ammonia permeases (amtB), nitrogen regulatory protein PII (glnB), response regulator (ntrX) and signal transduction histidine kinase (ntrY)(Levican et al. 2008)(Figure 1.7).

\[
\text{NH}_4^+ + \alpha - \text{ketoglutarate} + \text{NAD}(P)H + H^+ \rightarrow \text{Glutamate} + \text{NAD}(P)^+ + H_2O
\]

**Formula 1.2: Bacterial nitrogen assimilation.**

Under conditions of nitrogen/ammonia limitation within a habitat, assimilation of nitrogenous material can be catalyzed by glutamine synthetase (Tyler 1978), most commonly utilized in enteric bacteria. An additional pathway, dependent on alanine dehydrogenase can permit ammonia assimilation and has been studied in autotrophic bacteria (Kenealy and Zeikus 1982), but is more often used in alanine catabolism or sporulation (Brown and Meers 1974).

Genomic analysis of the *Leptospirillum* genus has revealed that two species, *L. ferrooxidans* and *L. ferrooxidotrophum* contain genes necessary for nitrogen fixation, including structural nitrogenase proteins (nifHDKEN), regulatory proteins (ntrC, nifLA) and proteins coupling carbon and nitrogen metabolism (PII) (Tyson et al. 2005). To date, no structural *nif* operons for the nitrogenase enzyme have been identified in *L. ferrililum*, however the regulatory genes *nifLA* are present (Levican et al. 2008), as are biosynthesis genes (*nifSU*) (Garcia-Moyano et al. 2008). This suggests that *L. ferrililum* may have lost the *nif* operon over
time and that the *Leptospirillum* species may have obtained their *nif* operon via an ancient gene transfer event (Tyson et al. 2005).

*At. ferrooxidans* and *L. ferrooxidans* are the only mesophilic micro-organisms currently identified as being capable of fixing nitrogen in bioleaching environments (Norris, Murrell, and Hinson 1995). The capacity of *L. ferrodiazotrophum* and *L. ferrilium* to fix N$_2$ currently remains uncertain.

**Figure 1.7** – Proposed regulatory model for nitrogen assimilation within *At. ferrooxidans*. (Image reproduced from Levican et al. 2008). Red boxes and lines indicate proteins and their pathways activated during low conditions of nitrogen availability, whereas blue represents actions taken under high nitrogen concentrations. Signal transducer proteins are in orange, including GlnD, NtrB, GlnE. Principle effectors of nitrogen modulation (NH$_3$, N$_2$, glutamine and 2-oxoglutarate) are in green that trigger activation/inactivation of enzymes. Yellow circles represent proteins identified in some *Leptospirillum* species.
1.4 Biological nitrogen fixation

Biological nitrogen fixation (BNF) is restricted entirely to microorganisms and is the process by which N\textsubscript{2} is converted into nitrogenous compounds (Fay 1992) such as ammonia, ammonium or nitrate. Nitrogen fixation plays an critical role within the nitrogen cycle as it compensates for the loss of fixed nitrogen incurred by denitrification and replenishes the overall content of the biosphere (Martínez-Argudo, I., et al. 2005). Fixed nitrogen resultant from BNF is also more advantageous to the biosphere than abiotic fixation or industrial processing as it is less prone to leaching and volatilization as it is utilized \textit{in situ} (Cassman et al. 1999).

The ability to fix nitrogen from the atmosphere has evolved in many different environments (Vitousek and Howarth 1991), with more than 100 prokaryal and archaeal species reported to have this ability (Raymond et al. 2004). As the majority of micro-organisms are unable to metabolize dinitrogen for their survival needs they are reliant on the small subset of micro-organisms that can, called diazotrophs (Kneip et al. 2007).

1.4.1 Microbial diversity of diazotrophy

Due to the wide environmental distribution of diazotrophs, the ability to fix nitrogen is spread amongst all major taxonomic phylogenetic groups, including Cyanobacteria, Actinomycetes, and all subdivisions of the Proteobacteria. To date, Archaeal nitrogen fixing is restricted to the Methanogens. There is considerable biodiversity amongst the diazotrophs, with nitrogen fixation also found in a wide range of physiological states such as aerobic, anaerobic, heterotrophic, oxic, phototrophic and chemolithotrophic (Dixon and Kahn 2004).

In most environments, the capacity to fix nitrogen is only utilized by diazotrophs under conditions of nitrogen limitation and is stringently regulated by a complex system of nitrogen fixation (\textit{nif}) genes. Recent discoveries have identified \textit{nif} genes in a range of otherwise unrelated diazotrophs (Fani, Gallo, and Lio 2000). Phylogenetic trees plotted with \textit{nif} sequences from bacteria and archaea suggest that the \textit{nif} genes originated in a common ancestor of the two domains (Papineau and Mojzsis 2002). This wide distribution pattern of nitrogen fixing genes also lends support to the theory that the basic features of nitrogen fixation are strongly conserved (Dean and Jacobson 1991) and arose during prehistoric times (Zehr et al. 2003b) but have since diverged. The capacity of these diazotrophs to fix nitrogen relies on their ability to produce an energy expensive, and tightly regulated, nitrogenase enzyme system.
1.4.2 Biochemistry of BNF

BNF is an intricate process comprising a number of functional and regulatory gene products, accumulating in the production and activation of a nitrogenase protein complex (Zehr 1996), responsible for the biological hydrogenation of dinitrogen to ammonia. Nitrogenase enzyme complexes can be distinguished by the metal composition of their active site metallocluster (Zhao et al. 2006). Nitrogenase enzymes contain two constituent proteins, one; a large dinitrogenase \( \alpha_2\beta_2 \) heterotetramer component, referred to as the molybdenum-iron (MoFe) protein, contains the enzyme catalytic site. The Mo-Fe-sulphur clusters of the Mo-Fe protein are the actual binding and reduction sites of \( N_2 \). The second smaller dinitrogenase reductase \( \gamma_2 \) homodimer component known as the iron protein (Fe), functions as an ATP dependent electron donor to the large \( \alpha_2\beta_2 \) heterotetramer (Raymond et al. 2004), using at least two MgATP per election (Halbleib, Zhang, and Ludden 2000).

Even though the molybdenum containing nitrogenase is often the most prevalent nitrogenase identified in nature, there are two homologous alternative nitrogenases, that structurally have similar cofactors and subunit compositions (Rubio and Ludden 2005). Under conditions of molybdenum depletion, some microorganisms induce the synthesis of alternative nitrogenases containing vanadium iron (V-Fe \( N_2 \)ase) (Bishop, Hawkins, and Eady 1986) or iron-iron cofactors (Fe-N \( N_2 \)ase) (Pau, Mitchenall, and Robson 1989) to form the large \( \alpha_2\beta_2 \) heterotetramer component, substituting the Mo element in the metallocluster.

More recently, a fourth nitrogenase enzyme has been discovered in *Streptomyces thermoautotrophicus*, with a catalytic binding site composed of Mo-molybdopterin cytosine dinucleotide cofactor (Mo-MCD) (Frerichs-Deeken et al. 2003), rendering the nitrogenase subunit components distinctly dissimilar than the other three nitrogenase enzymes (Ribbe, Gadkari, and Meyer 1997). Of all the nitrogenase enzymes, the Mo-Fe nitrogenase is more specific and efficient at binding and reducing dinitrogen (Joerger, Wolfinger, and Bishop 1991) as the nitrogenase enzymes have comparatively different reaction kinetics and specificities (Burgess and Lowe 1996; Eady 1996).

Under ideal conditions, for successful dinitrogen reduction by Mo-Fe nitrogenase, three fundamental reactions must occur. Firstly, The reduction of the \( \gamma_2 \) homodimer Fe protein by ferredoxins or flavodoxins must occur, followed by transport of single electrons, one at a time, from the Fe protein to the large Mo-Fe heterotetramer protein in an MgATP hydrolysis dependent process. Finally, internal transfer of the electrons in the Mo-Fe protein to the active binding site completes the process (Seefeldt, Hoffman, and Dean 2009)(Figure 1.8).
Figure 1.8 – Electron transfer from the Fe protein to the Mo-Fe protein is coupled to the hydrolysis of MgATP, which is followed by the dissociated of the protein-protein complex. ATP hydrolysis, electron transfer and substrate reduction are the key steps for nitrogenase turnover. Creative commons license.

The electron transfer step is a critical function of the homodimer Fe protein because the Mo-Fe protein will not alone reduce $N_2$ in the absence of the Fe protein, despite the fact that the Mo-Fe protein component can be reduced by other electron donors (Cheng 2008). In all characterized nitrogenase systems, the Fe protein is the obligate electron donor. Substrate uptake will not occur unless the Mo-Fe cofactor is fully loaded with electron and proton equivalents present during turnover conditions (Burgess and Lowe 1996).

This reduction of dinitrogen is a demanding multi-electron redox process, and under ideal conditions, the overall stoichiometry of dinitrogen reduction is:

$$N_2 + 8H^+ + 8e^- + 16MgATP \rightarrow 2NH_3 + H_2 + 16MgADP + 16P$$

Formula 1.3: Nitrogenase reduction of dinitrogen.

The dissociation of the Mo-Fe protein complex is the rate limiting step in this reaction and as such, nitrogenase has a relatively slow turnover time of $5s^{-1}$ (Eady 1996). The slow enzymatic turnover places extreme physiological stress on the microorganism, as it requires diazotrophs to synthesize large quantities of nitrogenase (up to 20% of the total protein in the cells) in order to use dinitrogen as a sole nitrogen source. Both components of the
nitrogenase protein are also very oxygen liable and there is an obligation by the organism to protect the enzyme from damage (Dixon and Kahn 2004). Therefore the nitrogenase system is stringently regulated at the transcriptional level in response to the availability of fixed nitrogen (Raymond et al. 2004).

1.4.2.1 Nitrogenase liability

A large consortium of aerobic micro-organisms are capable of nitrogen fixation (Cheng 2008) and go to enormous lengths to protect the nitrogenase enzyme from oxygen attack. Of the aquatic heterocystous filamentous Cyanobacteria, 5 to 10% of cells undergo morphological differentiation into heterocysts, allowing oxygenic photosynthesis to occur in the vegetative cell and nitrogen fixation in the heterocyst (Reddy and Sherman 1993). Symbiotic root nodule bacterial species such as Rhizobia fix nitrogen under aerobic conditions as specialised leghaemoglobin molecules have a high affinity for oxygen (Preisig and Hennecke 1996) preventing attack on the nitrogenase enzyme. Azotobacter species are known to have one of the fastest metabolisms known to bacteria and utilise the action of respiratory protection, wasting carbon by consuming oxygen and therefore allowing the nitrogenase enzyme to be shielded (Liu and Wong 1995). As Leptospirillum species are free living diazotrophs, the mechanisms that these other prokaryotes employ to protect the nitrogenase enzyme from oxygen degradation are unlikely to be present. It is currently unknown as to how the Leptospirillum species that contain the genes for nitrogen fixation circumvent this issue and protect their nitrogenase enzyme from oxygen attack.

1.4.3 Genetics of BNF

As diazotrophy amongst prokaryotes is widely distributed, the number and arrangement of genes involved in BNF differs considerably. Nitrogenase function and expression is regulated by a complex system, incorporating a large number of different genes (Fischer 1994), found within one or several extensive, cotranscribed operons (Table 1.1). Core operons include structural genes (nifHDK), regulatory and activation genes (nifL/nifA), and other supplementary genes necessary for protein assembly and cluster biosynthesis (nifB, nifS, nifU, nifX) (Halbleib and Ludden 2000).
1.4.3.1 Structural nif genes

The nitrogenase two component protein is encoded for by a structural set of operons with the nifH gene responsible for the dinitrogenase reductase γ2 homodimer (Fe protein), nifD, the dinitrogenase heterodimer (α2 subunit) and nifK, the β2 subunit. In most diazotrophs, these three genes (nifHDK) form one contiguous transcriptional unit (Choo, Samian, and Najimudin 2003) and are likely to have two other genes, nifE and nifN adjacent, forming what is known as the nifHDKEN operon (Fani, Gallo, and Lio 2000). NifE and NifN share structural sequence homology with NifD and NifK (Brigle et al. 1987), resulting from a single duplication event of an ancestral operon (Fani, Gallo, and Lio 2000). However, NifEN acting as the α2β2 heterotetramer component of the nitrogenase enzyme lacks the ability to reduce atmospheric nitrogen (Hu et al. 2009), and instead is proposed to serve as a scaffold for Fe-MoCo synthesis (Roll et al. 1995), that later shifts the FeMo complex to the nifDK complex. Production of the alternative nitrogenase enzymes containing either vanadium or iron is reliant on vnfDK, vnfH and anfDK, anfH genes respectively, when Mo conditions are limiting (Bishop, Jarleniski, and Hetherington 1982).
Figure 1.9 – Organisation of structural nif genes in various diazotrophs. Gene organization adapted from Jacobson et al. 1989; Lee et al. 2000; Parro, V., and Moreno-Paz 2004. ORF = open reading frame. CHP=conserved hypothetical protein.

1.4.3.2 Assembly and biocluster formation genes

Unaided, the translational products of the nitrogenase structural genes (NifHDK) are inactive and immature (Dean, Bolin, and Zheng 1993). For the maturation of these gene products to occur, a consortium of nif genes with specific roles is necessary. For the maturation of the Fe protein ($\gamma_2$) is dependent on the products of nifH, nifM, nifU and nifS, while that of the heterodimer ($\alpha_2\beta_2$) requires the activation of 6 genes, nifE, nifN, nifV, nifQ and nifB for biosynthesis (Cheng 2008). Without the gene products of nifM, NifH is non-functional (Gavini, Tungtur, and Pulakat 2006), and and thus nitrogen fixation will not occur (Roberts et al. 1978). The nifS gene product is a cysteine desulfurase involved in the synthesis of the 4Fe-4S cluster of the Fe protein (NifH). NifU working synergistically with NifS, mobilizing iron and sulphur for nitrogenase specific iron-sulphur cluster formation (Zheng et al. 1993). Molybdenum metabolism and incorporation to the metallocluster is dependent on the NifQ protein (Imperial et al. 1984), often co-transcribed with nifB (Rodriguez-Quinones, Bosch, and Imperial 1993) ensuring metal cluster synthesis of a functional nitrogenase enzyme (Hoover et al. 1988).

1.4.3.3 Transcriptional regulatory genes

As the process of BNF requires a great deal of energy expenditure on behalf of the microorganism, it is highly regulated on several levels and by different factors (Dixon and Kahn 2004). Of all diazotrophs studied, control over BNF is regulated at the level of transcription (Halbleib and Ludden 2000), however, in some organisms it will also take place post-transcription (Ludden and Roberts 1989). The global nitrogen regulatory system (ntr) responds to the nitrogen status of the cell with gene products of the NtrBC two-component
system (Merrick and Edwards 1995) governing transcription of the \textit{nifA} gene. In response to low levels of cellular nitrogen, expression of NifA occurs, activating the transcription of the \textit{nifHDK} operon (Walmsley, Toukarian, and Kennedy 1994), permitting nitrogenase production. The NifA protein is responsible for transcription of all major \textit{nif} gene clusters and whose expression is strongly regulated by the NifL protein. In response to excess fixed nitrogen, NifL acts as the negative regulator of NifA and inhibits transcriptional activation by forming a NifL-NifA binary complex, therefore deactivating \textit{nifHDK} transcription. The two specific regulatory proteins NifL and NifA, are co-transcribed as the \textit{nifLA} operon (Martinez-Argudo, I., et al. 2005), however, the mechanism of this regulation varies according to the organism (Merrick and Edwards 1995).

In some diazotrophs such as \textit{Klebsiella pneumoniae}, transcription of \textit{nifA} is under the control of the \textit{ntrBC} gene products (Yousafzai, Buck, and Smith 1996), whilst in symbiotic nitrogen fixers such as \textit{Rhizobia}, \textit{nifA} expression is controlled by two genes, \textit{fixL} and \textit{fixJ}, sensitive to levels of oxygen (Gilles-Gonzalez, Ditta, and Helinski 1991). Within the \textit{Leptospirillum} genus, genes showing sequence homology to those necessary for the NtrBC regulatory system have been identified in \textit{L. ferrooxidans}, including \textit{ntrC-like} and \textit{glnB-like} genes as well as the specific transcriptional activator NifA (Parro, V., and Moreno-Paz 2004). However, in \textit{L. ferrodiazotrophum} and \textit{L. ferriphilum}, \textit{nifA} and \textit{nifL} genes have been discovered, but no homologs to \textit{ntrB} or \textit{ntrC} (Tyson et al. 2005). Nitrogenase transcriptional control mechanisms vary extensively between diazotrophs, even within the same genus and therefore gene regulation of this enzymatic system should be elucidated individually for each diazotroph (Halbleib and Ludden 2000).

### 1.5 Measuring biological nitrogen fixation

Quantification of BNF can be performed utilizing a number of different methods, however, measuring the amount of nitrogen fixed by a micro-organism is not only dependent on the available resources and the species in question, but also bacterial growth conditions and available nutrients, the oxygen concentration and how effectively nitrogen fixation is performed. The most common methods applied to assess BNF include the acetylene reduction assay (Dilworth 1966) and nitrogen isotope enrichment ($^{15}$N$_2$) (Montoya et al. 1996)

#### 1.5.1 $^{15}$N$_2$ isotope labelling

$^{15}$N$_2$ is a stable isotope of nitrogen that can be used as an extremely reliable marker for nitrogen fixation (Buckley et al. 2007). Methods applied are reliant on exposing the micro-organism of interest to an atmosphere enriched with $^{15}$N$_2$ and assessing the level of incor-
porated labelled nitrogen into the cell via mass spectrometry (Montoya et al. 1996). Analysis with $^{15}$N$_2$ is sensitive and definitive (Strandberg and Wilson 1968), but $^{15}$N$_2$ and the instrumentation necessary for subsequent examination can be expensive (Zahran 1999), limiting the number of samples that can be processed. A cheaper, but just as accurate method employed to measure BNF is the acetylene reduction assay (Hardarson and Atkins 2003).

1.5.2 Acetylene reduction assay (ARA)

The nitrogenase enzyme is not just specific for N$_2$, but also has the ability to catalyse reductions of a wide range of alternative substrates that contain C-C, C-N, N-N and N-O double and triple bonds such as: acetylene (C$_2$H$_2$), azide (N$_3^-$), cyanide (CN$^-$) and hydrazine (N$_2$H$_4$) (Hu et al. 2009).

The ability of nitrogenase to reduce acetylene to ethylene, was discovered by Dilworth (1966), and has since been used extensively as an indicator of nitrogen fixation. Acetylene has proven to be a particularly useful substrate in assessing nitrogenase activity because the reaction product ethylene, can be easily quantified by gas chromatography, even in minute amounts and no products other than ethylene have been identified from this reduction. Additionally, ethylene doesn't inhibit N$_2$ fixation, isn't reduced by nitrogenase (Kelly 1969), and is stable, which can be stored for analysis at a later period.

$$ C_2H_2 + 2H^+ + XATP + 2e^- \rightarrow C_2H_4 + ADP + XPi $$

Formula 1.4: Nitrogenase reduction of C$_2$H$_2$. Adapted from Danso 1995.

This ARA technique therefore indirectly measures BNF by estimating enzyme activity based on electron flux through nitrogenase, not the concentration of nitrogenase protein produced (Hardy, Burns, and Holsten 1978). Because ethylene and acetylene are both permeable to the bacterial envelope, nitrogenase activity of a potential diazotroph may be measured in vivo as well as in vitro by the acetylene reduction method (Hableib and Ludden 2000). It is therefore particularly apt for measuring nitrogenase activity (N fixation) at an instant in time, particularly that of free-living diazotrophs growing in culture medium. In the presence of an alternative non-molybdenum nitrogenase, reduction of acetylene will continue past ethylene, resulting in ethane, which can also be detected by gas chromatographic methods (Davis et al. 1996).
1.6 Determining diazotrophy by molecular methods

Diazotrophs undertake essential roles in mixed microbial habitats, providing non-diazotrophic organism with fixed nitrogen. Studies defining their diversity can help provide critical information needed to determine the productivity of an ecosystem; that is, which diazotrophs are the most active. The \textit{nifH} gene is highly conserved and the most phylogenetically studied gene in diazotrophs (Zehr et al. 2003a). It is one of the oldest functional genes (Rosado et al. 1998), and detection of it within a microorganisms genome, often signifies nitrogen fixation capabilities (Raymond et al. 2004). With advances in molecular techniques, novel ways to detect diazotrophic microbial activity and functionality have been developed. Techniques such as PCR and quantitative Real Time PCR can precisely analyse mRNA expression and provide information on activities within specific populations (Greer et al. 2001). Proteomic studies utilizing 2D SDS PAGE and Peptide Mass Fingerprinting (PMF) have also revealed direct genome functionality of various diazotrophs (MacLean, Finan, and Sadowsky 2007). Such techniques also permit phylogenetic (evolutionary) relationships between organisms to be examined (DeLong and Pace 2001).

1.6.1 Structural genome analysis

Multichromosomal genomic studies conducted on some diazotrophs have uncovered evidence regarding the conserved structure (Martinez-Aguilar et al. 2008) of the \textit{nifHDK} genes, with particular emphasis on the tendency towards nucleotide bias (Chen, Chen, and Johnson 1986; McEwan and Gatherer 1998). When codon usage frequencies are similar, genes clusters form (Bolhuis et al. 2010) and in diazotrophs such as \textit{Azotobacter}, \textit{Pseudomonas} and \textit{Frankia}, the \textit{nif} genes have a higher GC content compared to the rest of the genome (McEwan and Gatherer 1998; Morgan, Lundell, and Burgess 1988; Yan et al. 2008). The contrasting GC content and resulting amino acid bias often results in systemic variation in codon usage among the genes within a genome. This highlights the use of preferred codons to shape the role and function of the nitrogenase enzyme within an organism (Fouts et al. 2008). The relative codon bias of particular genes within microbial communities has been attributed to translational selection (Gouy and Gautier 1982) and demonstrated in \textit{nif} coding regions (McEwan and Gatherer 1998), where strict codon usage is necessary to ensure the functionality of the nitrogenase enzyme (Peters, Fisher, and Dean 1995).

All diazotrophs contain highly conserved cysteine residues within the nitrogenase structural protein (\textit{nifHDK}), two of which symmetrically co-ordinate the Fe protein (NifH) into its 3D structure (Hausinger and Howard 1983), with others involved in ligand binding of the 4Fe-4S cluster (Schlessman et al. 1998). Substitution of amino acids in these core positions is known to affect nitrogenase activity (Chang et al. 1988). Analysis of \textit{nifDK} products,
also reveals conserved cysteine and histidine residues essential for the co-ordination of the Fe-Mo cofactor to the Fe protein (Howard and Rees 1994). Surrounding the cysteine residue needed for Fe-Mo co-ordination (reference A. vinelandii Cys α275) of NifD, are 12 conserved residues. Further towards the N terminal end are a second set of nine conserved amino acids, centred on the His α442 reside of the NifD protein (Glazer and Kechrís 2009). Located between these highly conserved positions, an additional 50 amino acids can occasionally be found (Wang, Chen, and Johnson 1988), common to strict anaerobes and some archaea (Zehr et al. 2003a). Identification of these conserved residues within the NifH-D proteins can be used to classify and segregate diazotrophs into separate functional groups (Raymond et al. 2004; Schueler-Furman and Baker 2003). The nine-residue sequence (His α442) is entirely conserved in all Group I NifD sequences, whereas the presence or absence of the additional 50 amino acids allows for the unambiguous assignment of a NifD sequence to either Group I or Group II.

1.6.2 Analysis of mRNA transcription rates

Specialized molecular approaches including Northern blots, microarrays and RT-PCR for studying mRNA transcription rates have been used to assess the functional nitrogen fixing genes of many diazotrophs (Burgmann et al. 2003; Steunou et al. 2008), including bio-leaching microorganisms At. ferrooxidans (Li et al. 2011) and L. ferrooxidans (Parro and Moreno-Paz 2003). The foundation of these approaches is the extremely conserved nature of the genes involved, as nifH, nifD and nifK demonstrate high sequence identity in all diazotrophs (Hennecke et al. 1985). Since gene expression is a tightly controlled process (Grunberg-Manago 1999), detection of mRNA transcripts for a specified gene represents substantial evidence of the occurrence of a particular biological activity (Cho et al. 1998). As RNA is synthesized by actively growing cells (Neidhardt 1961), transcription of the nifHDK operon would be a good marker for nitrogen fixation as it is not constitutively expressed (Curatti et al. 2005) and is regulated in response to factors that control N₂ fixation. Expression of mRNA nif genes is therefore the key to identifying micro-organisms that are active in N₂ fixation (Zehr et al. 2003b).

1.6.3 Proteomic analysis

Transcription of mRNA does not always result in a functional protein, with some studies demonstrating a weak correlation between protein and mRNA abundance (Greenbaum et al. 2003; Nie, Wu, and Zhang 2006). To fully understand cell responses to external stimuli, both transcription and translation regulation should be studied (Jansen, Greenbaum, and Gerstein 2002). Investigation of various diazotrophic proteomes during periods of nitrogen
fixation with techniques such as 2D-SDS PAGE, MALDI-TOF, LC/MS have revealed significant differences in protein abundance (Resendis-Antonio et al. 2011; Selao, Nordlund, and Noren 2008). In species of *Bradyrhizobium japonicum*, the application of proteomic techniques revealed an elaborate, interconnected network of proteins for nitrogen and carbon metabolism, dependent on plant supplied metabolites (Sarma and Emerich 2005). Proteins associated with nitrogen fixation and assimilation were stimulated and detected under nitrogen fixing conditions for *Rhodospirillum rubrum* (Selao, Nordlund, and Noren 2008), consistent with expression of *nif* genes. Under nitrogen limiting conditions, proteomic studies conducted on *Gluconacetobacter diazotrophicus* detected proteins associated with cell wall biogenesis and nitrogen fixation accessory proteins (NifW and NifB) (Kruger et al. 2008). Proteomics performed on a mixed bioleaching community by Ram et al. 2005 detected low levels of *L. ferrodiatotrophicum* NifH protein as well as proteins involved in nitrogen regulation and ammonium uptake from *Leptospirillum* group II members, however, no proteins indicating nitrogen fixation by *L. ferroxidans* were isolated.

As diazotrophic proteomes can alter in response to their environment (Sur, Bothra, and Sen 2010), understanding the physical properties of the protein is essential to understand the functionality of the organism (Knight et al. 2004). By analysing patterns of protein expression under conditions in which all the detected proteins can be qualitatively and quantitatively studied (Wilkins et al. 1998), valuable information leading to improved understanding of metabolic pathways and physiological responses can be obtained.

### 1.6.4 Phylogeny

The evolution of nitrogen metabolic pathways is closely linked to the development of the atmosphere, with nitrogen composition of sediments being utilized to trace the origin of organic matter in the earth's crust (Papineau et al. 2005). Nitrogenase genes are highly conserved at both genetic and proteomic levels, across extensive phylogenetic lineages and among closely related organisms, reflecting the stringent structural requirements of the nitrogenase enzyme complex for proper catalytic functioning (Dixon and Kahn 2004). This conservation of the nitrogenase genes enables them to be uses as genetic markers for phylogenetic analysis.

Phylogenetic analysis of the *nifH* and *nifD* components of the nitrogenase protein together with their known homologs (*vnf, anf*), revealed that they separate into distinct, topologically consistent groups (Zehr et al. 2003a). Group I, forming the largest group of diazotrophs, encompasses the conventional Mo-Fe nitrogenases, principally comprised of proteobacterial and cyanobacterial phyla ( ). Predominantly anaerobic diazotrophs, such as *Clostridia* and *Spirochaeta* along with sulphate reducers, and some Archaea (*Methanococcus, Methanosarcina*) (Leigh 2000) are classified as group II (Raymond et al. 2004). Group II are
also reliant on the Mo-Fe metallocluster nitrogenase, however the structural composition is slightly different regarding genetic arrangement of the nif clusters, nucleotide sequences, and regulation (Thiel, Lyons, and Erker 1997).

Diazotrophic group III members are characterized by Mo-independent nitrogenases encoded for by the anf and vnf genes (Betancourt et al. 2008), excluding vnfH (Raymond et al. 2004). Nif genes demonstrating sequence identity to NifH and NifD that play no role in nitrogen fixation (Staples et al. 2007) found only in methanogens (Methanocaldococcus) and some anoxygenic photosynthetic bacteria, form group IV. Group V members are not diazotrophs, but chlorophyll and bacteriochlorophyll genes in these prokaryotes share sequence identity to the nif genes (BchL/BchX = NifH and BchNB, BchYZ = NifDK)(Fujita and Bauer 2000), possibly resultant from an ancient ancestral gene duplication event (Blankenship 2002).

Examination of the nifH gene structures from bioleaching diazotrophs At. ferrooxidans and some Leptospirillum species place them within group I, having the traditional Mo-Fe nitrogenase. However, analysis of diazotrophic grouping utilizing either the nifD or nifK genes as phylogenetic markers in Leptospirillum species has yet to be performed.

Since the distribution of the nif genes is not ubiquitous, studies conducted in conjunction with universal rRNA genes, aim to determine the true phylogeny of these organism (Henson et al. 2004; Zehr et al. 1997). Phylogenetic clustering based on the nifH genes, is known to produce topologies that are largely consistent with 16S rRNA based classification (Zehr et al. 2003a). However, not all species demonstrate congruent phylogenies when assessing evolution using the nifD and nifK genes (Hartmann and Barnum 2010; Kechris et al. 2006).
Figure 1.10 – Five phylogenetic groups elucidated in the text, shown on a concatenated phylogenetic tree composed of NifH and NifD homologs found in complete genomes. Adapted from Raymond et al. 2004

1.6.5 Evolution of nif genes

As nitrogen fixation is indiscriminately scattered throughout prokaryotic lineages (Gaby and Buckley 2011), the evolutionary history of diazotrophy is thought to be by both lateral transfer and vertical decent Henson et al. 2004. Multiple hypotheses have been put forward to explain the inheritance of such a highly conserved enzyme, including:

- Single occurrence of nitrogen fixation arising, followed by lateral transfer to many
different lineages (Normand and Bousquet 1989).

- Ubiquitous evolution of nitrogen fixation genes, with subsequent loss by a large number of organisms (Hirsch et al. 1995).

- Convergent evolution with nitrogen fixation arising multiple times.

If nitrogen fixation arose only once, with subsequence vertical descent responsible for the distribution of genes, comparison of the 16S rRNA and nif genes should reveal similar phylogenies, assuming mutations rates were similar (Burgmann 2003). However, if dissemination of diazotrophy is a result of lateral transfer, 16S rRNA and nif gene phylogenies constructed should deviate significantly (Dedysh, Ricke, and Liesack 2004).

### 1.7 Objectives

#### 1.7.1 Towards a better understanding of *Leptospirillum* diazotrophy

With the advance in molecular biology techniques, there has been an increased understanding of mixed environmental microbial populations that make up bioleaching operations. The discovery that *Leptospirillum* species dominate mineral processing bioreactors (Okibe et al. 2003), combined with its ability to leach minerals, has increased the importance of this genus over the past decade. Even though *Leptospirillum* spp have minimal nutrient requirements, commercial bioleaching processes often require the expensive addition of ammonium (for nitrogen) and phosphate containing compounds to enhance natural growth and increase the functionality of the system (Johnson, D. B. 2010). Upon depletion of essential nutrients, the influence of diazotrophs and their potential to contribute fixed nitrogen to these systems requires further examination.

A greater understanding of BNF within bioleaching operations, as performed by *Leptospirillum* is needed, as some *Leptospirillum* species have the capacity for nitrogen fixation and may possibly contribute to the global nitrogen input within bioleaching systems (Gonzalez-Toril et al., 2003). This current research has left questions regarding the ability of *Leptospirillum* species to fix atmospheric nitrogen for metabolic processes when under ammonium deficient conditions.

The objectives of this thesis were to assess whether nitrogen fixation by various *Leptospirillum* species was possible, and the extent to which it occurred utilizing modern molecular biology techniques. In particular:
1. Can all *Leptospirillum* species survive without soluble nitrogen, thus obtaining their nitrogenous needs via nitrogen fixation, and if so, is there a hierarchy based on their survival rates?

2. Do expression rates of key *nif* gene transcripts correlate with growth data, revealing any adaptations to certain environments?

3. Does an increase or decrease in *nif* gene transcription result in alterations to the proteome?

4. Are there any structural differences in the nitrogenase genes between the *Leptospirillum* spp, and how phylogenetically related are the genes?

These questions provided the theoretical framework for this thesis.

### 1.7.2 Research strategy and outline of the thesis

To gain more insight into the potential nitrogen fixation capability of the three *Leptospirillum* species, various techniques were employed. Initially, experiments covering cell proliferation, iron oxidation and acetylene reduction were conducted to appraise growth of all *Leptospirillum* species in the absence of soluble nitrogen. Few studies using molecular tools to assess mRNA expression of *Leptospirillum* spp have been published (Parro and Moreno-Paz 2003), but none on all three *Leptospirillum* species covering transcript abundance of various *nif* genes. Thus, this study employed Quantitative Real Time PCR assays to evaluate *nif* gene expression, as transcription of these genes would be a strong indicator of nitrogen fixation activity.

As mRNA transcripts often have rapid turnover rates (Deutscher 2006), it is also practical to measure abundance of the corresponding protein. Therefore, in this study a methodology was developed to extract whole cell proteins from all *Leptospirillum* species, allowing for the examination of proteins that differed in abundance in conditions when soluble nitrogen was absent. Quantitative changes observed in proteins produced by *Leptospirillum* spp between soluble nitrogen rich and poor environments could indicate metabolic adaptations to an altered environment. The identification of specific proteins that had altered abundance in response to the presence or absence of soluble nitrogen may provide insights into the pathways activated in *Leptospirillum* spp and identify possible functions.

Supporting proteomic analysis, examination and comparison of the structural components of the *nifHDK* genes between the *Leptospirillum* species to identify conserved and variable regions would help establish the functionality of the nitrogenase enzyme in this genus.
The final goal of this thesis was to compare and contrast the phylogeny predicted based on the 16S rRNA with that of the *nifHDK* genes. From this work, phylogenetic trees demonstrating species relatedness helped to establish the evolutionary history of these three *Leptospirillum* species.
Chapter 2

Physiological response of *Leptospirillum* species in the absence of soluble nitrogen

2.1 Introduction

The growth of all organisms depends on the availability of mineral nutrients, of which nitrogen is often the limiting nutrient in an ecosystem. Nitrogen stress in prokaryotes can affect aspects of cell size, respiration and growth rates, (Elliott et al. 1982) with a gradual loss of flagella and movement (Wei and Bauer 1998). Almost all bacteria will assimilate available nitrogen in the form of ammonia for use in biosynthetic pathways and generation of new cell mass, with a more limited set of bacteria capable of converting atmospheric nitrogen to ammonia.

BNF is the conversion of atmospheric dinitrogen to ammonia, catalysed by the nitrogenase enzyme. This system is restricted to prokaryotes, but widely distributed among a large percentage of known bacterial lineages. Within bioleaching environments, nutrient limiting conditions are common and therefore the ability of some micro-organisms to fix atmospheric nitrogen is of interest. Studies conducted on *At. ferrooxidans* and *L. ferrooxidans*, two bacteria known to dominate bioleaching operations, have demonstrated nitrogen fixation under depleted ammonium conditions (Parro and Moreno-Paz 2003). A more recently identified *Leptospirillum* species, *L. ferrodiazotrophum*, is known to possess the genes necessary for nitrogenase activity (Tyson et al. 2005), but its nitrogen fixing potential is yet to be assessed. To date, no nitrogenase encoding genes have been identified in *L. ferriphilum*, however, when devoid of a soluble nitrogen source it continues to proliferate (Garcia-Moyano et al. 2008).

In this study the ability of three *Leptospirillum* species to survive without soluble nitrogen was investigated. The rate of cell proliferation, as a measure of cell growth and the amount
of Fe$^{2+}$ oxidation, as a measure of cellular activity was used to determine whether the *Leptospirillum* species were capable of utilizing atmospheric nitrogen when deprived of a soluble nitrogen source. Nitrogenase not only reduces atmospheric nitrogen to ammonia, but also acetylene to ethylene (Dilworth 1966). The acetylene reduction assay therefore provided a useful means of measuring the nitrogen fixation capabilities of the *Leptospirillum* species, and allowed for comparisons to be made between these closely related species.

### 2.2 Materials and methods

#### 2.2.1 Bacterial strains and growth media

The bacterial strains used in this study as detailed in Table 2.1

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>Media</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Leptospirillum ferrooxidans</em> (DSM2705$^T$)</td>
<td>Modified 9K media L-1 FeSO$_4$.7 H$_2$O</td>
<td>Deutsche Sammlung Fur von Mikroorganismen (DSM)</td>
</tr>
<tr>
<td><em>Leptospirillum ferriphilum</em> (DSM14647$^T$)</td>
<td>Modified 9K media L-1 FeSO$_4$.7 H$_2$O</td>
<td>Deutsche Sammlung Fur von Mikroorganismen (DSM)</td>
</tr>
<tr>
<td><em>Leptospirillum ferro-diazotrophum</em> (C5)</td>
<td>Modified 9K media L-1 FeSO$_4$.7 H$_2$O</td>
<td>Donated by J. Banfield (U.C. Berkley, California)</td>
</tr>
<tr>
<td><em>Acidithiobacillus caldus</em> (DSM 8584$^T$)</td>
<td>BSM 5g L-1 of S0</td>
<td>Deutsche Sammlung Fur von Mikroorganismen (DSM)</td>
</tr>
<tr>
<td><em>Acidithiobacillus ferrooxidans</em> (DSM 14882$^T$)</td>
<td>BSM 13.9 g.L$^{-1}$ FeSO$_4$.7 H$_2$O</td>
<td>Deutsche Sammlung Fur von Mikroorganismen (DSM)</td>
</tr>
<tr>
<td><em>Azotobacter vinelandii</em></td>
<td>Burk’s N free liquid medium</td>
<td>Donated by N. Williams (Curtin University)</td>
</tr>
</tbody>
</table>

Stock cultures of *Leptospirillum* were aerobically cultivated in a modified 9K medium which contained (g L$^{-1}$): MgSO$_4$.7 H$_2$O, 0.4g; K$_2$HPO$_4$, 0.4g; (NH$_4$)$_2$SO$_4$, 0.4g. The final pH was adjusted to 1.8 with concentrated H$_2$SO$_4$. Ferrous sulphate (FeSO$_4$.7 H$_2$O), (final concentration 22g L$^{-1}$), was sterile filtered (0.2 μm, PALL corporation, hydrophilic polypropylene) and added post autoclaving. The medium was supplemented with trace salts μg L$^{-1}$: MgCl$_2$.2 H$_2$O, 62; ZnSO$_4$.7 H$_2$O, 68; CoCl$_2$.6 H$_2$O, 64; H$_3$BO$_3$, 30; Na$_2$MoO$_4$, 10; CuCl$_2$.2 H$_2$O, 66; NaVO$_3$, 30. Complete media analysis is detailed in Appendix A, table A.1.
At. *ferrooxidans* cultures were maintained in a basal salt media (BSM) which contained (g L⁻¹): MgSO₄·7H₂O, 0.25 g; K₂HPO₄, 0.25 g; and (NH₄)₂SO₄, 1.5g. The final pH was adjusted to 1.8 with concentrated H₂SO₄. Ferrous sulphate (FeSO₄·7H₂O), (final concentration 13.9 g L⁻¹), was sterile filtered (0.2 µm), and added post autoclaving and supplemented with trace salts as above. *At. caldus* cultures were grown in modified 9K media with the addition of 5 g L⁻¹ of S⁰ and supplemented with trace salts µg L⁻¹: MgCl₂·2H₂O, 62; ZnSO₄·7 H₂O, 68; CoCl₂·6H₂O, 64; H₃BO₃, 30; Na₂MoO₄, 10; CuCl₂·2H₂O, 66; NaVO₃, 30.

No ferrous sulphate added.

*Azotobacter vinelandii* cultures were maintained in Burk’s N free liquid medium, of which the composition was as follows (g L⁻¹): K₂HPO₄, 6.4; KH₂PO₄, 1.6; NaCl, 2.0; MgSO₄·7 H₂O, 2.0; CaSO₄·2 H₂O, 0.5; NaMoO₄·2 H₂O, 0.01; ferric citrate, 0.02; glucose, 10. The pH was adjusted to 7.2 using 0.1 N NaOH.

All cultures were incubated aerobically at 30 °C and were either stationary or gently shaking at 60 rpm.

### 2.2.2 Experimental growth conditions

Cultures in late exponential growth phase were harvested by centrifugation at 20,000 rpm in a Beckman centrifuge (Avanti J-E) for 40 min at 25 °C. Pellets were washed with 2 mL of acidified water, pH 1.8, to remove any traces of the initial growth medium that contained nitrogen then centrifuged at 15,300 rpm for 30 min at 25 °C and the supernatant removed. Cell pellets were resuspended in medium either containing soluble nitrogen or in the absence of soluble nitrogen.

Volumetric 250 mL conical flasks containing 50 mL of medium were inoculated at a concentration of 0.5 x 10⁷ cells ml⁻¹. Uninoculated abiotic controls zero N and 400 ppm N supplied as (NH₄)₂SO₄ were also included. Ammonium gas is known to have a high affinity for acidic solutions (Holden 1991). To limit the dissolution of NH₃ gas from the atmosphere in the acid culture media devoid of NH₄, a beaker of 100 ml of concentrated H₂SO₄ was placed with all experimental flasks, as minute traces of NH₃ gas have a higher affinity for concentrated H₂SO₄ than acidic growth media. This was to ensure that all nitrogenous sources in the nitrogen deficient cultures were obtained from atmospheric nitrogen fixation. *At. caldus* was employed as a control organism to test the functionality of the acid trap. Failure of *At. caldus* to grow in nitrogen deficient media would demonstrate that NH₃ gas had not dissolved in the growth media and was being drawn away by the acidic solution. Samples of 0.5 mL were taken every 24 hours over 7 days for cell counting, analysis of iron oxidation and to note changes in optical density, measured at 410 nm.
2.2.3 Cell growth

Cell numbers were determined by direct counting using a Helber counting chamber under phase contrast microscopy (Olympus CX41) at a magnification of 1000x. Each count was repeated three times.

2.2.4 Iron oxidation assay by Iron-dipyridyl complex formation

The amount of ferrous in the growth media was determined using the method described by (Wilson 1960). The standard curve is shown in Appendix A.

2.2.5 Measurement of acetylene reduction

Stock cultures of the three Leptospirillum species, At. ferrooxidans and At. caldus (500 mL, with a total of 3.5 x 10^{11} cells) all grown in media with soluble nitrogen were harvested by membrane filtration, washed with acidified sterile water pH 1.8 and centrifuged for 30 min at 15,300 rpm. Cell pellets were resuspended in 80 mL of modified 9K medium lacking (NH_{4})_{2}SO_{4} and 10 mL transferred to a sterile 25 mL conical flask and sealed with a SubaSeal (Sigma-Aldrich). A. vinelandii was subcultured into 25 mL flasks with fresh BURK’s N free media (ratio 1:5), with a final cell concentration of 2.5 x 10^{8} cells.

To each of the sealed cultures, 0.5 mL of acetylene (BOC, IR Grade 2.0) was injected. The gaseous head-space of the cultures was sampled (0.5 mL) in triplicate, at either 24, 72, 120 or 168 hours. Once the samples were taken the cultures were sacrificed and assessed for iron oxidation. Two cultures were examined at each time point for each condition.

Ethylene produced from the reduction of acetylene was measured using a Shimadzu GC-8A gas chromatograph fitted with a flame ionization detector. Glass column (100 cm x 2 mm) packed with Poropak R and run isothermally at 120 °C with high purity nitrogen as the carrier gas (60 cm^{3}min^{-1}). The detector and injector temperatures were maintained at 120°C. The system was calibrated by injecting 3.3 nmol/mL of ethylene with peak height detection, recorded on a Shimadzu C-R1A chromatopac. A standard curve was constructed (Appendix A). Although GC experiments are simple and accurate, peak heights detected below 150 were disregarded as they were considered to be below the sensitivity of the equipment.

2.2.6 Calculations and formulae

Calculating growth rate during the exponential phase-μmax (EXP) (#1) For the duration of the EXP, μ is constant and at a maximum as all limiting substrates are in excess. μmax is
Physiological response of *Leptospirillum*

indicative of how fast the culture is growing, representing the time it takes a single cell to double, so the following equation only applies during log growth.

Cell doubling times \((d_t)\) in hours were calculated from \(\mu_{\text{max}}\) \(\text{h}^{-1}\) values (#2).

The total average amount of \(\text{Fe}^{2+}\) that a cell consumed at any given time (#3) and the average amount of \(\text{Fe}^{2+}\) oxidized by each species over 168 hours, by the total number of cells recorded (#4).

Acetylene reduction values calculated from construction of a standard curve (#5).

**Table 2.2** – Calculations and formulae used for growth analysis and \(\text{Fe}^{2+}\) oxidation

<table>
<thead>
<tr>
<th>N° #</th>
<th>Product</th>
<th>Formula</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(\mu_{\text{max}}) (EXP)</td>
<td>(= \frac{\text{Ln}(C_2) - \text{Ln}(C_1)}{(T_2 - T_1)})</td>
<td>(\text{Ln} = \text{natural log})</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(C = \text{cell count})</td>
</tr>
<tr>
<td>2</td>
<td>(a_t)</td>
<td>(= \frac{1}{\mu_{\text{max}}})</td>
<td>(\mu_{\text{max}} = \text{EXP})</td>
</tr>
<tr>
<td>3</td>
<td>(\text{Fe}^{2+})ox cell (^{-1})h (^{-1})</td>
<td>(= \frac{(\text{Fe}_2 - \text{Fe}_1)}{\bar{x}(C_1 + C_2)}) (\frac{1}{(T_2 - T_1)})</td>
<td>(\text{Fe} = \text{nMol} \text{Fe}^{2+})ox</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(C = \text{cell count})</td>
</tr>
<tr>
<td>4</td>
<td>Average (\text{Fe}^{2+})ox cell (^{-1})h (^{-1})</td>
<td>(= \frac{\Sigma \text{Fe}^{2+}\text{ox}}{\Sigma \text{cells}})</td>
<td>(\Sigma = \text{sum of total})</td>
</tr>
<tr>
<td>5</td>
<td>Acetylene reduction</td>
<td>(y = 1820.9x - 214.4)</td>
<td>(y = \text{peak height}) (x = \text{ethylene conc})</td>
</tr>
</tbody>
</table>
2.3 Results

2.3.1 Analysis of cell proliferation data from *Leptospirillum* species grown in the absence of soluble nitrogen

In order to investigate the nitrogen fixation capabilities of *Leptospirillum* species, growth was assessed when deprived of a soluble nitrogen source (NH$_4^+$) and under stationary and shaking conditions.

*At. caldus* demonstrated a decrease in cell growth under shaking and stationary conditions when deprived of a soluble nitrogen source (Figure 2.1), with numbers falling 72 hours post inoculation.

![Figure 2.1](image)

*Figure 2.1* – The effect of soluble nitrogen on the growth of *At. caldus* (as measured by cell numbers). Cultures were incubated with either shaking (purple) or stationary (orange).

*At. ferrooxidans* cultures proliferated when deprived of soluble nitrogen, in both stationary and shaking cultures (Figure 2.2). Cell numbers were higher in the presence of nitrogen when the cultures were shaken. In the absence of nitrogen, the growth of *At ferrooxidans* was similar under both test conditions.
Physiological response of *Leptospirillum*

**Figure 2.2** – The effect of soluble nitrogen on the growth of *At. ferrooxidans* (as measured by cell numbers). Cultures were incubated shaking (purple) and stationary (orange).

In all *Leptospirillum* species (Figures 2.3 a, b and c), positive cell growth was demonstrated in all cultures deprived of soluble nitrogen (−NH₄) under shaking and stationary conditions. Despite all cultures being seeded with the same initial number of cells, the rate at which *Leptospirillum* proliferated differed amongst the species, with the greatest number of cells recorded in *L. ferritophilum* cultures.
Physiological response of *Leptospirillum*

![Graphs of cell growth over time for different species](image)

(a) *L. ferrodiazotrophum*

(b) *L. ferrophilum*

(c) *L. ferrooxidans*

**Figure 2.3** - The effect of soluble nitrogen on the growth of *Leptospirillum* spp. (as measured by cell numbers). Cultures were incubated shaking (purple) and stationary (orange).

### 2.3.2 Growth patterns identified in *Leptospirillum* spp. grown in the absence of soluble nitrogen

On closer inspection of the cell growth data collected on *Leptospirillum* species grown in the absence of soluble nitrogen, evidence of different growth patterns were observed. Regardless of the presence or absence of nitrogen, a long lag (λ) phase (72 hours) was recorded for *L. ferrodiazotrophum* (Figure 2.4). During period of exponential growth, different $\mu_{\text{max}}$
values were recorded for stationary and shaking cultures, with $\mu_{\text{max}}$ lasting 48 hours for stationary, whereas aerated cultures remained in exponential growth for 96 hours.

![Figure 2.4 - Growth of *L. ferrodiazotrophicum* in the absence of soluble nitrogen in shaking (purple) and stationary (orange) cultures.](image)

Cultures of *L. ferriphilum* grown in the absence of soluble nitrogen demonstrated an initial lag phase of 24 hours, followed by a slow non-exponential growth phase lasting 92 hours for both shaking and stationary experiments (Figure 2.5). Exponential growth phase ($\mu_{\text{max}}$) was observed, 120 hours post inoculation, with stationary cultures proliferating faster than those shaking.

![Figure 2.5 - Growth of *L. ferriphilum* in the absence of soluble nitrogen in shaking (purple) and stationary (orange) cultures.](image)
In the absence of soluble nitrogen *L. ferrooxidans* (Figure 2.6), with either the stationary or shaking incubation had two distinct lag phases and two exponential (\( \mu_{\text{max}} \)) growth phases were observed in both stationary and shaking experiments. Cultures lacking aeration experienced a longer second lag phase (48 hours) than those that were aerated (24 hours), but proliferation of stationary cultures was greater than those shaking. These growth patterns were noticeably absent in *L. ferrodiazotrophum* and *L. ferriphilum* (Figures 2.4 and 2.5 respectively).

**Figure 2.6** – Growth of *L. ferrooxidans* in the absence of soluble nitrogen in stationary (orange) and shaking (purple) cultures.

During periods of identified exponential growth, cultures of stationary *L. ferriphilum* had the fastest doubling rate (30 h) of all test cultures, similar to *L. ferriphilum* cultures with soluble nitrogen that were shaking (24 h) (Table 2.3). *L. ferrodiazotrophum* cultures had the slowest doubling time recorded, with both stationary and shaking taking close to 100 hours to double. *L. ferrooxidans* stationary cultures had a faster growth rate without soluble nitrogen (35 h) compared to those with aeration (41 h).
Table 2.3 – Average cell doubling times (h) during periods of exponential growth calculated from $\mu_{\text{max}}$ h^{-1} values of *Leptospirillum* species, under stationary and shaking conditions.

<table>
<thead>
<tr>
<th>Species</th>
<th>Aeration Regime</th>
<th>$+\text{NH}_4$</th>
<th>$-\text{NH}_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. ferrodiazotrophum</em></td>
<td>Shaking</td>
<td>100 ± 3</td>
<td>96 ± 2</td>
</tr>
<tr>
<td></td>
<td>Stationary</td>
<td>91 ± 8</td>
<td>98 ± 5</td>
</tr>
<tr>
<td><em>L. ferrphilum</em></td>
<td>Shaking</td>
<td>24 ± 2</td>
<td>30 ± 3</td>
</tr>
<tr>
<td></td>
<td>Stationary</td>
<td>34 ± 2</td>
<td>25 ± 2</td>
</tr>
<tr>
<td><em>L. ferrooxidans</em></td>
<td>Shaking</td>
<td>29 ± 2</td>
<td>41 ± 2</td>
</tr>
<tr>
<td></td>
<td>Stationary</td>
<td>43 ± 3</td>
<td>35 ± 4</td>
</tr>
</tbody>
</table>

2.3.3  Fe$^{2+}$ oxidation by *Leptospirillum* spp as a measure of cellular activity

To assess the ability of *Leptospirillum* to fix nitrogen, it was essential to monitor the rate at which the available energy substrate, Fe$^{2+}$, was consumed. *At. caldus* species are strict sulphur oxidizers and were not subjected to iron oxidation experiments.

*At. ferrooxidans* was able to oxidise Fe$^{2+}$ without soluble nitrogen (Figure 2.7), and as it was used as a control species in these experiments, no further analysis was done on the data following this result.

![Graph](image)

**Figure 2.7** – The effect of $\text{NH}_4^+$ on the proportion of Fe$^{2+}$ oxidized by *At. ferrooxidans*. Shaking (purple) and stationary (orange)

Up to 48 hours post inoculation, iron oxidation by *L. ferrodiazotrophum* without soluble nitrogen was similar to cultures supplemented with soluble nitrogen (Figure 2.8a) regardless of the aeration regime. Without soluble nitrogen and under stationary conditions, oxidation
Physiological response of *Leptospirillum*

of Fe$^{2+}$ surpassed shaking cultures at 96 hours (24.5% ± 3.5 compared to 18.5% ± 2.0) and was greater by 168 hours (47.4% ± 6.7 vs 29.2% ± 2.9). Oxidation of Fe$^{2+}$ by *L. ferrooxidans* was lower in the presence of soluble nitrogen, with only 24.8% ± 1.2 of available Fe$^{2+}$ being utilized by 168 hours (Figure 2.8), regardless of the aeration regime.

In the absence of soluble nitrogen and without shaking, *L. ferriphilum* cultures oxidised 88% ± 5.6 of the initial Fe$^{2+}$ substrate by 120 hours (Figure 2.8b), proceeding to 89% ± 4.2 by 168 hours. This result is equivalent to that obtained for Fe$^{2+}$ utilization of *L. ferriphilum* cultures with nitrogen and shaking (Figure 5.8) by 120 hours (85.7% ± 5.0). At 168 hours, aeration by shaking resulted in lower Fe$_{2+}$ oxidation for *L. ferriphilum* than stationary in the absence of N (70.6% ± 6.05 vs 88.6 ± 7.1).

By 168 hours, stationary cultures of *L. ferrooxidans* without soluble nitrogen oxidized a comparable amount of Fe$^{2+}$ substrate (48.3% ± 4.6) as seen in control cultures under the same conditions (52.66% ± 7.1). With aeration and without soluble nitrogen, only 25.03% ± 3.6 of Fe$_{2+}$ substrate was consumed by 168 hours, compared to 69.9% when soluble nitrogen was available (Figure 2.8c).

In stationary cultures lacking soluble nitrogen, higher iron oxidation rates were detected across all species when there was no aeration regime.
Figure 2.8 – The effect of NH$_4^+$ on Leptospirillum Fe$^{2+}$ oxidation rates. Shaking (purple) and stationary (orange). (a) L. ferrodiazotrophum, (b) L. ferrphilium and (c) L. ferrooxidans.

2.3.4 Rate of Fe$^{2+}$ oxidized by Leptospirillum spp. in the absence of soluble nitrogen

Each of the Leptospirillum species tested demonstrated different growth patterns and variable amounts of Fe$^{2+}$ oxidized over 168 hours. The amount of Fe$^{2+}$ consumed between
Physiological response of *Leptospirillum*

each sampling was calculated as nMol Fe$^{2+}$ cell$^{-1}$ h$^{-1}$. In all species over 168 hours, as the number of cells increased, the amount of Fe$^{2+}$ consumed per cell decreased in all species (Figure 2.9) regardless of the aeration regime.

After 24 hours, the consumption of Fe$^{2+}$ by *L. ferrodiazotrophum* was greater for stationary cultures than shaking cultures at each time point sampled. Between 72 and 96 hours, there was an increase in Fe$^{2+}$ oxidized per cell compared to the previous time point (Figure 2.9a) for both stationary and shaking cultures. Utilization of Fe$^{2+}$ cell$^{-1}$ h$^{-1}$ by *L. ferrodiazotrophum* stationary cultures was higher than shaking cultures for all following time periods. Consumption of Fe$^{2+}$ cell$^{-1}$ h$^{-1}$ by *L. ferriphilum* was initially lower than *L. ferrodiazotrophum* in both stationary and shaking cultures (Figure 2.9b). After 72 hours, an increase in Fe$^{2+}$ oxidized per cell was noted in stationary cultures (0.34 nMol ± 0.0064) compared to those shaking (0.14 nMol ± 0.0029). By 160 hours, utilization of Fe$^{2+}$ cell$^{-1}$ h$^{-1}$ was lower in stationary cultures than shaking. The decreased levels of Fe$^{2+}$ cell$^{-1}$ h$^{-1}$ in *L. ferriphilum* after 120 hours could be attributed to substrate limitation as >85% of iron had been oxidized by this time point. In *L. ferrooxidans*, stationary cultures consumed a higher amount of Fe$^{2+}$ cell$^{-1}$ h$^{-1}$ than those with aeration, until 96 hours (Figure 2.9c). Unlike *L. ferrodiazotrophum* and *L. ferriphilum* no increase in Fe$^{2+}$ consumption was recorded between 72-96 hours for *L. ferrooxidans*. During this time point, both stationary and shaking cultures demonstrated a lag in proliferation (Figure 2.6). After 96 hours there was no statistical difference between the amount of Fe$^{2+}$ consumed by stationary and shaking cultures for *L. ferrooxidans*. 
Figure 2.9 - Averaged Fe\textsuperscript{2+} oxidation rates cell\textsuperscript{-1} h\textsuperscript{-1} under shaking (purple) and stationary (orange) conditions without soluble nitrogen.

2.3.5 Doubling events and Fe\textsuperscript{2+} oxidation

By comparing the amount of times the culture doubled (Table 2.4) for each species under each growth condition, with the final amount of Fe\textsuperscript{2+} utilized, information regarding whether the energy liberated from Fe\textsuperscript{2+} oxidation was used for cell proliferation, or other metabolic functions could be obtained.
Table 2.4 – Number of doubling events calculated for Leptospirillum spp over 168 hours

<table>
<thead>
<tr>
<th>Species</th>
<th>Aeration Regime</th>
<th>+NH₄</th>
<th>-NH₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrodiazotrophum</td>
<td>Shaking</td>
<td>1.6 ± 0.17</td>
<td>1.8 ± 0.02</td>
</tr>
<tr>
<td></td>
<td>Stationary</td>
<td>1.7 ± 0.02</td>
<td>1.7 ± 0.27</td>
</tr>
<tr>
<td>L. ferrphilum</td>
<td>Shaking</td>
<td>6.6 ± 0.41</td>
<td>4.9 ± 0.51</td>
</tr>
<tr>
<td></td>
<td>Stationary</td>
<td>6.6 ± 0.30</td>
<td>5.7 ± 0.27</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td>Shaking</td>
<td>5.5 ± 0.33</td>
<td>5.0 ± 0.13</td>
</tr>
<tr>
<td></td>
<td>Stationary</td>
<td>3.9 ± 0.21</td>
<td>4.7 ± 0.27</td>
</tr>
</tbody>
</table>

With an aeration regime, similar amounts of Fe²⁺ were consumed by L. ferrphilum per doubling event in both stationary and shaking cultures, with or without soluble nitrogen (Figure 2.10). Shaking cultures of L. ferrooxidans without nitrogen utilized 0.237 ± 0.017 nMol Fe²⁺ per doubling event, compared to 0.178 ± 0.0048 nMol Fe²⁺ per doubling event whilst stationary, a 1.33 fold difference. With soluble nitrogen and without an aeration regime, oxidation of Fe²⁺ by L. ferrooxidans decreased to 0.096 ± 0.002 nMol Fe²⁺ per doubling event, a 1.85 fold decrease compared to the shaking cultures. With no aeration and no soluble nitrogen, L. ferrooxidans utilized 0.110 ± 0.013 nMol Fe²⁺ per doubling event, a 2.15 fold decrease compared to shaking cultures. Aerated cultures of L. ferrodiazotrophum with and without soluble nitrogen, utilized similar amounts of Fe²⁺ per doubling event. However, under stationary conditions, L. ferrodiazotrophum consumed more Fe²⁺ when cultured without soluble nitrogen (0.204 nMol ± 0.024) than in the presence of soluble nitrogen (0.168 ± 0.009)(a 1.21 fold increase).

![Graphs](a) Shaking  
![Graphs](b) Stationary

**Figure 2.10** – Amount of Fe²⁺ consumed per doubling event in by Leptospirillum species in shaking and stationary conditions.
2.3.6 Acetylene reduction to ethylene

The identification of nitrogenase activity was determined by analysing the atmospheric composition of flasks after incubation with acetylene. The integrated peak area of the ethylene produced by acetylene reduction was converted to a concentration by development of an ethylene standard curve. Nitrogenase activity as measured by acetylene reduction was detected in three of the five species tested, *A. vinelandii*, *L. ferrooxidans* and *L. ferrodiazotrophum*. At no time point were high enough levels of ethylene detected in cultures of *At. ferrooxidans*, and *L. ferriphilum* to indicate positive nitrogenase activity.

Of the two *Leptospirillum* species that successfully reduced acetylene, *L. ferrodiazotrophum* recorded the highest levels of ethylene (2.2 nmol mL\(^{-1}\)) under stationary conditions after 168 hours (Figure 2.11a). Detection of ethylene in shaking flasks was always lower than stationary flasks at all time points tested for *L. ferrodiazotrophum*.

Higher levels of ethylene were detected in stationary cultures of *L. ferrooxidans* cultures, than shaking, with the largest recording (1.77 nmol ml\(^{-1}\)) after 168 hours (Figure 2.11b). Ethylene was detected in shaking cultures of *A. vinelandii* after 24 hours (Figure 2.11c), with the greatest amount (3.50 nmol ml\(^{-1}\)) detected in flasks incubated for 168 hours. Cultures of *A. vinelandii* were more successful in reducing acetylene to ethylene than any of the *Leptospirillum* species with significantly less starting cell density (172x less cells).
Figure 2.11 – Amount of ethylene produced by *L. ferrodiazotrophum*, *L. ferrooxidans* and *A. vinelandii* cultures after incubation with acetylene gas.
2.4 Discussion

In an effort to determine if *Leptospirillum* species have the ability to survive without soluble nitrogen, possibly utilising nitrogen fixation pathways, a multi step experimental approach was undertaken. The evidence presented show that nitrogen fixation by *Leptospirillum* is highly likely, however extremely variable between the species.

2.4.1 Growth of *Leptospirillum* without soluble nitrogen

After exposing *Leptospirillum* species to growth conditions devoid of soluble nitrogen, continued cellular proliferation and iron oxidation was observed for all three species. This is highly suggestive of active molecular nitrogen utilization as the sole nitrogen source. All *Leptospirillum* species proliferated in the absence of soluble nitrogen with greater cell numbers recorded by 168 hours in stationary cultures, than those cultures with an aeration regime. It is apparent that a reduced distribution of oxygen throughout a stationary culture permitted more optimal growth for all three *Leptospirillum* when soluble nitrogen was unavailable. Unlike *L. ferrooxidans* and *L. ferrodiazotrophum*, continued growth of *L. ferriphilum* in media without soluble nitrogen was unexpected as it is assumed to contain no nitrogenase genes within its genome. Of the three species under examination, *L. ferriphilum* also demonstrated the greatest rate of proliferation and iron oxidation in nitrogen free media.

In 2008, Garcia-Moyano et al. demonstrated growth of *L. ferriphilum* up to 140 hours post subculture with nitrogen limitation, combined with continued iron oxidation. In this study, sustained cellular proliferation of *L. ferriphilum* was recorded up to 168 hours and beyond (data not shown). In this instance, ammonium ions absorbed from the atmosphere do not explain the growth of *L. ferriphilum*, as an acidic solution with greater affinity for atmospheric NH$_3$ was provided, preventing solubilisation in the test cultures; nor does scavenging of nitrogen from the inoculum, as trace amounts of N would have permitted the growth of *At. caldus*. Acidithiobacillus caldus was chosen as a negative control for nitrogen fixation, as its genome sequence reveals no homologs to any identified nitrogenase genes. Cessation of cell proliferation in the absence of soluble nitrogen confirmed nitrogen absence in the media constituents. The acquisition of nitrogenous compounds by nitrogen fixation for *L. ferriphilum* appears highly likely.

Of the three *Leptospirillum* species tested, *L. ferrodiazotrophum* experienced the longest initial lag phase of 72 hours post sub-culturing into nitrogen free media, compared to *L. ferrooxidans* and *L. ferriphilum*, both of which the lag lasted only 24 hours. Prior to subculturing for these experiments, when *Leptospirillum* cells were in late exponential phase, storage of appreciable amounts of NH$_3^{+}$ in the form of glutamine may have occurred due to
levels in the surrounding media decreasing. Under low concentrations of N, NH$_4^+$ incorporated into *Rhodospirillum rubrum* (a gram negative, nitrogen fixing bacterium) via glutamine synthetase forming glutamine (Nordlund et al. 1985), permitted glutamine to act as the amino acid donor for many cellular biosynthetic reactions, and thus acting as a storage reservoir for nitrogen (Graham et al. 1994). Storage of other elements such as iron and carbon by *Acidothiobacilli* and *Leptospirillia* during times of variable availability has been documented (Osorio et al. 2008), and therefore a viable possibility in regards to nitrogen storage.

The different growth pattern of *L. ferrodiazotrophum* in comparison to *L. ferrphilium* and *L. ferrooxidans* highlights a possible difference in metabolism and adaptation of the species to a new environment. For *L. ferrodiazotrophum*, increased sensitivity to the decreased cell to cell interaction and signalling mechanisms (Melke et al. 2010) resulting from the reduced cell number to volume ratio may have caused the longer lag phase, that did not affect either *L. ferrphilium* or *L. ferrooxidans*. It is also possible that cultivation of *L. ferrodiazotrophum* within the laboratory has not been optimized (Kaeberlein, Lewis, and Epstein 2002), as slow growth was also seen in control cultures with abundant soluble nitrogen and Fe$^{2+}$.

For all species without an aeration regime, evidence of biofilm formation was observed, and may have facilitated growth in the soluble nitrogen deprived environment, as biofilms assist in protecting the nitrogenase enzyme from atmospheric oxygen inactivation (Bazylinski et al. 2000; Flemming and Wingender 2010). Since diazotrophic nitrogen fixation is regulated at the transcriptional level in response to environmental oxygen and NH$_4^+$ levels, the lower dissolved oxygen content in the stationary cultures may have down regulated the inactivation of nitrogenase gene expression, and therefore allowed nitrogen fixation to occur. In cultures of *At. ferrooxidans* and *L. ferrooxidans*, as the oxygen content decreased, production of greater amounts of ethylene via acetylene reduction (Norris, Murrell, and Hinson 1995) were detected, indicating that the amount of nitrogenase produced was dependent on oxygen concentration. In this instance, the decreased aeration of stationary cultures, combined with the lack of soluble nitrogen, may have allowed for production of a greater amount of nitrogenase in comparison to the shaking cultures, resulting in higher cell proliferation rates.

### 2.4.2 Consumption of Fe$^{2+}$ under nitrogen depleted conditions

*Leptospirillum ferrphilium* and *L. ferrooxidans* oxidised Fe$^{2+}$ at a slower rate in the absence of soluble nitrogen than when cultures were provided with soluble nitrogen. However, in all *Leptospirillum* cultures grown in the absence of soluble nitrogen, Fe$^{2+}$ oxidation was more rapid in stationary cultures than in shaking cultures. As Fe$^{2+}$ oxidation and electron donation is responsible for maintaining the internal pH balance of the cell and therefore
subsequent generation of energy, it is likely that Fe$^{2+}$ electrons utilised for energy in N$_2$ fixation would affect the amount of Fe$^{2+}$ oxidised available for cell maintenance and growth. Analysis of the amount of Fe$^{2+}$ oxidised cell$^{-1}$ h$^{-1}$ by *Leptospirillum* at different time points over 168 hours demonstrated that during the first 24 hours after sub culturing, the largest amount of Fe$^{2+}$ was consumed per cell. For all species under observation, this indicated that the bacterial populations were adjusting to the new environment and various metabolic pathways required for adaptation and survival were being engaged.

In both shaking and stationary conditions, a surge in cell usage of Fe$^{2+}$ was detected in *L. ferrodiazotrophum* and *L. ferrihilum* cultures between 72 and 96 hours. This coincided with the start of the exponential growth phase for *L. ferrodiazotrophum*, and at which point consumption of Fe$^{2+}$ increased to meet energy requirements. Therefore necessary for the production of the nitrogenase enzyme, and proliferation in the absence of soluble nitrogen would be an augmented supply of energy. Increased Fe$^{2+}$ uptake at this time point by *L. ferrihilum*, is indicative of an increased energy demand, possibly for the maintenance of biosynthesis pathways already stimulated, as well as production of the nitrogenase enzyme.

### 2.4.2.1 Iron consumed per cell over 168 hours

Without soluble nitrogen, *L. ferrooxidans* and *L. ferrihilum* oxidized lower amounts of Fe$^{2+}$ per cell at any time point, compared to that of *L. ferrodiazotrophum* under both stationary and shaking conditions (Figure 2.9). This high consumption of Fe$^{2+}$ per cell by *L. ferrodiazotrophum* is not reflected in cell proliferation data as numbers remain low over 168 hours (cultures doubled 1.8x in shaking and 1.7x in stationary), whereas for *L. ferrihilum* and *L. ferrooxidans* substrate utilization per cell was low and the resultant cell numbers high. It has been observed for *Bacillus* strains that when substrate concentrations are sufficient, the correlation between ATP consumption and biomass creation is poor (Sundkvist, Gahan, and Sandstrom 2007), which from these results would be the case for *L. ferrodiazotrophum*. That is, even though consumption of Fe$^{2+}$ per cell was high, there was no corresponding increase in cell number.

At no point over 168 hours, in either shaking or stationary cultures did Fe$^{2+}$ become a limiting factor for *L. ferrodiazotrophum*. For *L. ferrodiazotrophum*, rather than converting the electrons liberated from Fe$^{2+}$ oxidation into energy solely for proliferation, the energy could be used in nitrogenase production or cell maintenance, including adjustment of the cell membrane potential, cell motility, production of flagella and controlling the internal pH balance.

Since final cell numbers were widely variable between the species, by comparing the total number of doubling events calculated with the final amount of Fe$^{2+}$ utilized, information regarding whether the energy liberated from Fe$^{2+}$ oxidation was used for cell proliferation, or
other metabolic functions could be revealed. It was apparent that without soluble nitrogen, more energy was required by all *Leptospirillum* species under both stationary and shaking conditions to proliferate. This data implies that adaptation to an environment depleted in soluble nitrogen altered the metabolic state of the cell in response to stress and highlights the possibility of nitrogen fixation in all species.

### 2.4.3 Evidence of nitrogenase activity via acetylene reduction

Although the ARA technique employed is not a direct measure of nitrogen fixation (Phillips 1980), in these experiments it provided an indication of which *Leptospirillum* species were potentially capable of fixing atmospheric nitrogen. Of the three *Leptospirillum* species assessed for nitrogenase activity, *L. ferroxidans*, and *L. ferrooxidans* were capable of acetylene reduction, when deprived of soluble nitrogen. *A. vinelandii* was used as a positive control for the GC system (Bishop, Hawkins, and Eady 1986) and under identical conditions, also reduced acetylene to ethylene. Only shaking cultures of *Azotobacter vinelandii* were examined for the production of ethylene as its nitrogenase enzyme is not completely inhibited by the presence of O$_2$ (Poole and Hill 1997) and agitation would assist growth. In comparison to the *Leptospirillum* species, *A. vinelandii* was more successfully in reducing acetylene to ethylene with significantly less cells. However, *A. vinelandii* cells are often 2-7 μm in size (Tilak, Pal, and Dey 2010) making them up to 14 times larger than *Leptospirillum* cells. They are also known to have one of the highest metabolic rates (Jurtshuk, Marcucci, and McQuitty 1975) of all bacteria, allowing faster consumption of oxygen and therefore less inhibition of the nitrogenase enzyme (Dworkin and Falkow 2006). To produce similar levels of ethylene as *A. vinelandii*, approximately 170 times as many *L. ferroxidans* and *L. ferrooxidans* cells were required, demonstrating that nitrogen fixation in these species is not as optimal as in *A. vinelandii*.

Both *L. ferroxidans* and *L. ferrooxidans* have been shown to contain genes for nitrogen fixation within their genome (Parro and Moreno-Paz 2003; Tyson et al. 2005), and a positive acetylene reduction result (Figure 2.11) demonstrates functionality of these genes. *L. ferrooxidans* demonstrated the greatest amount of acetylene reduction and combined with the high utilization of Fe$^{2+}$ per cell and lower cell yield could be indicative of elevated cellular metabolic activity, rather than cell proliferation, additionally supporting the theory that nitrogen fixation pathways were active. *L. ferroxidans* was also capable of acetylene reduction, albeit at lower rate.

Acetylene reduction was not detected for *At. ferroxidans* or *L. ferriphilum*. The failure to detect ethylene produced in *At. ferroxidans* cultures was unexpected due to continued proliferation and Fe$^{2+}$ oxidation (Figures 2.2 and 2.7) demonstrated when soluble nitrogen was absent. Nitrogen activity via acetylene reduction in *At. ferroxidans* was first reported
in 1971, (Mackintosh 1971) but subsequent research has failed to reproduce these results (Stevens and Tuovinen 1988). Decreasing the partial $O_2$ pressure aided in the detection of ethylene from *Thiobacillus ferrooxidans* (subsequently renamed *At. ferrooxidans*) (Norris, Murrell, and Hinson 1995) and perhaps the high $O_2$ concentration in this instance may be the reason for a negative result.

Similarly, *L. ferriphilum* failed to reduce acetylene to ethylene, with levels being below threshold detection, however, continued proliferation and oxidation of $Fe^{2+}$ in conditions with no soluble nitrogen occurred. As previous studies have shown that *At. ferrooxidans* can reduce acetylene, the lack of reduction by *L. ferriphilum* in this instance should not be considered definitive. In comparison to *L. ferrooxidans* and *L. ferrodiazotrophum*, the high $O_2$ environment may have negatively impacted on nitrogenase production in *L. ferriphilum*, or the amount produced may have been below the level of detection. By decreasing the partial $O_2$ concentration it may be possible to determine whether or not *L. ferriphilum* does in fact produce the nitrogenase enzyme, when in nitrogen deficient environments. It is also possible that the nitrogenase enzyme encoded by *L. ferriphilum* is in an isozyme form of V-nitrogenase or Fe-nitrogenase, which both reduce acetylene at decreased rates comparatively to Mo-nitrogenases (Masephol et al. 2002), making detection more difficult.

### 2.5 Conclusions

Under nitrogen stress, *Leptospirillum* species grew slower, and with visual examination (microscopic), were not as motile as control cultures (data not included). Consumption of $Fe^{2+}$ increased in all species without soluble nitrogen, but was not reflected in a proportional increase in cell biomass, implying use of the generated energy for other purposes, possibly nitrogen fixation.

The results of this study demonstrate that in the absence of soluble nitrogen, all three *Leptospirillum* species survived, and continued to proliferate, two of which, also reduced acetylene to ethylene. The difference in the amount of proliferation, $Fe^{2+}$ oxidation and varying abilities of reducing acetylene to ethylene demonstrate that the metabolism of *Leptospirillum* members is highly variable. Further studies combining these results with $^{15}N_2$ isotope dilution methods would improve the accuracy of these results and help to confirm the *Leptospirillum* genus as free living diazotrophs.
Chapter 3

Nitrogen fixation gene regulation in Leptospirillum species

As growth of all Leptospirillum species in the absence of soluble nitrogen was successfully demonstrated in Chapter 2, a change in the expression of genes regulating nitrogen fixation was of interest. In this chapter mRNA levels of various nif genes were analysed to identify changes occurring at the genomic level.

3.1 Introduction

Diazotrophs play a key role in global nitrogen cycling and studies assessing their nitrogen fixing capabilities provide important information for determining the productivity of an ecosystem. Within bioleaching operations, diazotrophs such as At ferrooxidans and L. ferrooxidans dominate microbial populations, and alterations in community dynamics resulting from environmental change can have a detrimental effect on mineral recovery (Brandl 2001). Leptospirillum species have a remarkable ability to cope with variable environmental stresses, including large increases in ferric iron as a result of Fe$^{2+}$ oxidation (Rawlings, Tributsch, and Hansford 1999) and changes in soluble nitrogen concentrations. These aerobic chemolithoautotrophs have a reported ability to reduce dinitrogen (Levican et al. 2008; Parro and Moreno-Paz 2003) to ammonium based compounds when levels of soluble nitrogen in their surrounding environment are depleted.

The gene regulation and mechanisms involved with nitrogen fixation have been examined in many prokaryotes (Dixon and Kahn 2004) and the overlap of genes induced by nitrogen fixation suggest extensive and complex cross talk of signalling pathways regulating respiration (Hurek et al. 1994), photosynthesis (Steunou et al. 2006) and metabolism. Investigating the physiological and metabolic changes that occur during nitrogen fixation can
be observed via quantitative expression changes in the pattern of gene transcription. Synthesis of the nitrogenase enzyme, essential for nitrogen fixation, is highly regulated at the level of mRNA transcription involving upwards of 80 different genes, controlling protein production and turnover.

Within *L. ferrodiazotrophum* and *L. ferrooxidans*, the genes encoding for the nitrogenase protein are organized in a contiguous arrangement (Parro, V., and Moreno-Paz 2004; Tyson et al. 2005), while to date, *nifHDK* genes have not been identified in *L. ferrirhizilum*. To support the function of the nitrogenase (*nifHDKEN*) enzyme, expression of the *nifs-nifu-hesB* operon is also important as *nifs* and *nifu* work synergistically to mobilize iron and sulphur for nitrogenase formation (Zheng et al. 1993) and is present in all *Leptospirillum* species (Garcia-Moyano et al. 2008; Parro and Moreno-Paz 2003; Tyson et al. 2005).

By measuring gene transcription rates (mRNA profiling), biological responses to various stimuli can be used to monitor and identify genes whose protein expression is regulated at the transcriptional level. Analysis of gene expression requires sensitive, precise and reproducible measurements for specified mRNA sequences. In order to detect and quantify gene expression from very small amounts of mRNA, real time quantitative PCR (q-RT-PCR) can be employed. Reverse transcription of mRNA followed by qRT-PCR is highly sensitive and allows for both the simultaneous amplification and quantification of initial amounts of genetic template in a PCR reaction (Gibson, Heid, and Williams 1996).

The objective of this study was to specifically examine whether soluble nitrogen deprivation would induce the activation of specific nitrogen fixation genes in *Leptospirillum* species. Comparative expression analysis (q-RT-PCR) of three genes from the *nifHDKEN* operon, *nifHDK* encoding for the Fe, α, β subunits of nitrogenase and three genes from the Fe-S cluster formation operon, *nifsU-hesB* were made between cultures grown in the presence of soluble nitrogen and cultures grown in the absence of soluble nitrogen at different time points of proliferation. Control genes for this study were the 16S rRNA gene and the Gyrase B gene (*gyrB*), which is necessary for regulating supercoiling of bacterial chromosomes during DNA replication (Hooper et al. 1989). Without knowing exact copy number, reference genes can be applied as standards to which test genes can be compared. Relative transcript abundance, in reference to the control genes was measured and reported as either up or down regulated.
3.2 Materials and methods

3.2.1 Experimental design considerations

Sensitivity, efficiency and reproducibility were important factors to consider when designing the mRNA assay. Therefore, optimization of the reagents used to perform the PCR was critical to obtaining reliable results.

- Target properties: Gene sequence lengths were between 100 bp and 250 bp to work optimally with SYBR green I (Bio-Rad handbook).
- Primer properties: Oligonucleotide lengths ranged between 15 and 30 bp with a GC content between 40-70%, and both primers had the same Tm (55-60 °C).
- The primers used in this study were designed using Primer3 and analysed for hairpins and primer dimers with Net Primer.
- mfold v3.5 was used to predict secondary structure interference at different annealing temperatures.

3.2.2 Bacterial strains and growth media

The bacterial strains of *L. ferrodiazotrophum*, *L. ferriphilum* and *L. ferrooxidans* used in this study as well as the growth media are detailed in Table 2.1.

3.2.3 Experimental growth conditions

Cultures in late exponential growth phase were harvested by centrifugation at 20,000 rpm in a Beckman centrifuge (Avanti J-E) for 40 minutes at 25 °C. Pellets were washed with 2 ml of acidified water, pH 1.8, to remove any traces of the initial growth media that contained nitrogen then centrifuged at 15,300 rpm, at 25 °C for 30 min and the supernatant removed. Cell pellets were resuspended in media either containing soluble nitrogen or in the absence of soluble nitrogen. Conical flasks (250 ml) containing 50 ml of media were inoculated at a concentration of 6.5 x 10^6 cells ml⁻¹. As experiments conducted in Chapter 2 demonstrated greater rates of acetylene reduction when cultures were stationary, only stationary cultures were assayed for RNA expression rates.
3.2.4 Sampling regime

The mRNA of the three *Leptospirillum* species was sampled at three time points; 72, 96 and 144 hours post inoculation in the absence of soluble nitrogen. Two replicates per gene, comprising three biological replicates in each reaction, were examined.

3.2.5 Isolating mRNA by proteinase K and acid phenol extraction

*Leptospirillum* cultures were harvested at the time periods of listed in section 3.2.4 by filtration with a 0.2 μm filter (PALL Corporation, hydrophilic polypropylene). The supernatant was discarded. Working on ice (4°C), bacterial pellets were resuspended in 500 μL of sterile 2x Buffer A (0.2 M NaCl pH 8.0, 0.2 M Tris, 2 mM sodium citrate, 1 mM CaCl₂, 50 mM EDTA pH 8.0), 20 μL of 20% w/v polyadenyllic acid sepharose (Sigma), 20 μL of 10% w/v sodium pyrophosphate and 30 μL lysozyme (1 mg/ml, Invitrogen) in listed order. Samples were mixed by inversion and incubated at 37°C for 20 minutes. To each tube, 10 μL of 20% SDS and 60 μL of Proteinase K (27 mg/ml) was added, mixed by inversion and incubated at 50°C for 30 minutes. A further 200 μL of 20% SDS was added to the tubes along with 500 μL of Phenol: Chloroform: Isoamyl Alcohol (25:24:1) pH 4.2 (Sigma-Aldrich). Tubes were vortexed for 10 seconds and centrifuged at 15,300 rpm at 4°C for 5 min. The over layer was removed and transferred to a new tube, followed by adding an equal volume of Phenol: Chloroform: Isoamyl Alcohol (25:24:1) pH 4.2. Tubes were mixed by inversion and centrifuged at 15,300 rpm at 4°C for 3 min. The top layer was extracted, placed in a new tube and an equal volume of 30% polyethylene glycol 9000 (PEG-9000)/1.6 M NaCl added. Tubes were placed at -20°C for 2 hours. After incubation, tubes were gently inverted and centrifuged at 15,300 rpm at 4°C for 20 min to isolate a nucleic acid pellet. The supernatant was discarded and the pellet washed with 4°C, DEPC treated 70% ethanol and centrifuged at 15,300 rpm at 4°C for 5 min. A repeat wash of DEPC treated 70% ethanol was performed removing all traces of polyethylene glycol. Tubes were centrifuged at 15,300 rpm at 4°C for 3 min. The pellet was allowed to air dry before being resuspended in 30 μL of TE buffer (10 mM Tris-Cl, pH 8.0, 1 mM EDTA, pH 8.0). Nucleic acids were stored at -80°C in 10 μL aliquots.

3.2.6 Removal of DNA from NA samples

DNA contamination was removed leaving only RNA transcripts by treatment with DNase I enzyme (Promega). To 10 μL of the NA sample, 1 μL of DNase I was added along with 2 μL of 10X DNase Buffer I and 7 μL of ddH₂O. The reaction was incubated for 30 minutes at 37°C. DNase I was inactivated by the addition of 0.5 μL 0.1 M EDTA pH 8.0.
3.2.7 Confirmation of DNA removal

Samples of extracted RNA were used as a template nucleic acid in a PCR with 16S rRNA primers. All PCR reagents except primers were from Invitrogen. The 12.5 µL PCR contained 2 µL 10 X PCR buffer, 1.0 U DNA Taq polymerase, 2.5 mM MgCl₂, 0.2 mM dNTPs, 0.4 µM forward and reverse primers, 2% DMSO, 50 ng NA template and ultrapure water. The PCR was performed on an Alpha Metrix G-Storm. PCR cycle parameters are listed in Table 3.1.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>No. repeats</th>
<th>Temp (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
<td>1x</td>
<td>94</td>
<td>5 min</td>
</tr>
<tr>
<td>Cycle 2</td>
<td>25x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td>94</td>
<td>30 sec</td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td>56</td>
<td>30 sec</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td>72</td>
<td>1 min</td>
</tr>
<tr>
<td>Cycle 3</td>
<td>1x</td>
<td>72</td>
<td>10 min</td>
</tr>
<tr>
<td>Cycle 4</td>
<td>1x</td>
<td>10</td>
<td>∞</td>
</tr>
</tbody>
</table>

Table 3.1 – PCR cycle parameters testing for DNA contamination of RNA samples.

PCR products were analysed by 1% TBE, pH 8.0 (44 mM Tris-Cl, 44 mM Boric Acid, 10 mM EDTA pH 8.0) agarose gel electrophoresis, post stained with Ethidium Bromide and visually assessed for fragments. Failure of the PCR to generate product deemed samples to be free of contaminating DNA. Genomic DNA was used as a positive control for the PCR reaction.

3.2.8 Integrity analysis and quantification of isolated RNA

Visual assessment of extracted RNA integrity was assessed on a 1.5% TBE agarose gel with SYBR Green I staining. Secondary analysis of integrity was performed on a Nanodrop 1000, and an OD₂₆₀/₂₈₀ ratio between 1.9 and 2.1 confirming purity of the RNA sample.

3.2.9 Construction of cDNA by reverse transcription

Construction of cDNA was made from RNA samples treated with DNase I using the Qiagen QuantiTech® Reverse Transcription kit. Reactions were performed on ice. In a sterile microfuge tube, 14 µL of RNA sample was combined with 1 µL of RT primer mix, 1 µL of Quantscript Reverse Transcriptase and 4 µL of 5X RT buffer. Incubation of the reaction occurred at 42°C for 15 minutes followed by 3 minutes at 95°C to inactivate the reverse transcriptase. Aliquots of the cDNA were stored at -20°C until use.
3.2.10 Primers

Primers (Table 3.2) were designed to amplify selected fragments from genes of interest including housekeeping genes (16S rRNA and gyrB) nitrogenase encoding genes (*nifH*, *D*, *K*) and nitrogenase biosynthesis genes (*nifS*, *U*, *hesB*) and *nifA*. The 16S rRNA 27F primer is from Lane (1991) and 402R from Zammit et al. (2008). All other primers were designed in this study against sequences found in the NCBI database for these genes of interest.

<table>
<thead>
<tr>
<th>Species</th>
<th>Target</th>
<th>Primer names</th>
<th>Forward (5'-3')</th>
<th>Reverse</th>
<th>Product size (bp)</th>
<th>Tm ° C</th>
<th>Gen Bank Acc#</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrodiazotrophum</td>
<td><em>nifH</em></td>
<td>FdHF and FdHR</td>
<td>GCTGAAGCCGAATGAGGGATT</td>
<td>GCATGCAAAAGCCCAGAGCA</td>
<td>~210</td>
<td>56</td>
<td>CH003548</td>
</tr>
<tr>
<td></td>
<td><em>nifD</em></td>
<td>FdDF and FdDR</td>
<td>GGAGGGTTCCATCATGAGTCA</td>
<td>AAGGACATGATCCATATCA</td>
<td>~210</td>
<td>56</td>
<td>CH003548</td>
</tr>
<tr>
<td></td>
<td><em>nifK</em></td>
<td>FdKF and FdHR</td>
<td>CGGACATGAGGGACACCCCAAG</td>
<td>GTTTTTCGACCTGGGTTTT</td>
<td>~200</td>
<td>57</td>
<td>CH003548</td>
</tr>
<tr>
<td></td>
<td><em>nifS</em></td>
<td>FdSF and FdSR</td>
<td>GTCACAGATGGATCCGGTCTA</td>
<td>AATCTGGCCATCAAGGGCGT</td>
<td>~250</td>
<td>56</td>
<td>CH003604</td>
</tr>
<tr>
<td></td>
<td><em>nifU</em></td>
<td>FdUF and FdHR</td>
<td>GGAGAATATCATGGCATACAG</td>
<td>GCGATGACGGGATCATTGAA</td>
<td>~160</td>
<td>56</td>
<td>CH003604</td>
</tr>
<tr>
<td></td>
<td>hesB</td>
<td>FdhesF and FdhesR</td>
<td>CCGTCTCAAGAAGGTCCAGA</td>
<td>GGTTGAGCCGTCAAGGTAAA</td>
<td>~250</td>
<td>56</td>
<td>CH003604</td>
</tr>
<tr>
<td>L. ferriphilum</td>
<td><em>nifS</em></td>
<td>FpSF and FpSR</td>
<td>GAGGAGTCCTTCATGGATCGA</td>
<td>ACCTGGCCATCAAAGGGCGT</td>
<td>~240</td>
<td>56</td>
<td>DS995259.1</td>
</tr>
<tr>
<td></td>
<td><em>nifU</em></td>
<td>FpUF and FpUR</td>
<td>ATCGAAGAGGCCAAGTTCA</td>
<td>AAAGCCAACGTCTGAAGGAGG</td>
<td>~170</td>
<td>56</td>
<td>DS995259.1</td>
</tr>
<tr>
<td></td>
<td>hesB</td>
<td>FphesF and FphesR</td>
<td>CCGATTTCGAAATGTACGAA</td>
<td>CGCGCCGTCTAACGTTTTACTT</td>
<td>~200</td>
<td>56</td>
<td>DS995259.1</td>
</tr>
<tr>
<td></td>
<td><em>nifA</em></td>
<td>FpAF and FpAR</td>
<td>GAGAACCGCGTTTTCTGAAC</td>
<td>AGTTTGTCGAAGAGCGGAAA</td>
<td>~240</td>
<td>56</td>
<td>DS995259.1</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td><em>nifH</em></td>
<td>FxFH and FdHR</td>
<td>CCTGAGCGAGATGGGGAAGAA</td>
<td>GTTATCACCTCGATCAACTT</td>
<td>~250</td>
<td>55</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td><em>nifD</em></td>
<td>FxDF and FxDR</td>
<td>CGGAGGGTTCCATGAGCAACA</td>
<td>GATCCACATCAGCCACGGT</td>
<td>~250</td>
<td>56</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td><em>nifK</em></td>
<td>FxFK and FxKR</td>
<td>GGAACAGACAATGAGCCAGAAT</td>
<td>GAGTGGACCAAAACGGAAGA</td>
<td>~150</td>
<td>56</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td><em>nifS</em></td>
<td>FxFS and FxSR</td>
<td>TTGAAGCTCAAGATGTACCT</td>
<td>TCTTTCTGGGCAAGTTCTC</td>
<td>~200</td>
<td>56</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td><em>nifU</em></td>
<td>FxFU and FxUR</td>
<td>GGAACAGACAGGGACAGACAG</td>
<td>TAGCTTTGCGATGACCT</td>
<td>~200</td>
<td>56</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td>hesB</td>
<td>FxhesF and FxhesR</td>
<td>GGGCGAGTCTTTCTGAAAG</td>
<td>ATCAGGCGAGAAGTTGAAA</td>
<td>~220</td>
<td>56</td>
<td>AY204398</td>
</tr>
<tr>
<td>All species</td>
<td>16S rRNA</td>
<td>27F and 402R</td>
<td>AGAGTTTGATCAGGCTCAGA</td>
<td>GTTGGCTGGGTCAGGGTT</td>
<td>~370</td>
<td>56</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>gyrB</td>
<td>UgyrBF and UgyrBR</td>
<td>TGGATCCCTCAAGGCTCAGA</td>
<td>TGGATCCCTGCGAGCTGTT</td>
<td>~210</td>
<td>56</td>
<td>/</td>
</tr>
</tbody>
</table>

Table 3.2 - Oligonucleotide primer sequences for RT-PCR gene amplification. Tm temperatures were determined from NetPrimer and optimized PCRs. Genbank accession numbers represent sequences on which primers were modelled.

3.2.11 Testing PCR primers

Conventional PCRs were performed to test the specificity of the primers. All PCR reagents except primers were from Invitrogen. The 12.5 μL PCR reaction contained 1x PCR buffer, 1.0 U DNA Taq polymerase, 2.5 mM MgCl₂, 200 μM dNTPs, 0.4 μM forward and reverse primers, 2% DMSO and ultrapure water. The PCR was performed on an Alpha metix G-Storm. PCR products were analysed by 1% Tris-Borate EDTA pH 8.0 (44 mM Tris-Cl, 44 mM Boric Acid, 10 mM EDTA pH 8.0) agarose gel electrophoresis, post stained with Ethidium Bromide and visually assessed for fragments. Identity of all fragments was verified by sequencing and BLAST analysis to confirm amplification of the target gene.
3.2.12 Quantitative Real Time PCR conditions

Quantitative Real Time PCR was carried out by a Bio-Rad iQ5 Thermocycler. Every 25 μL reaction contained 2.5 μL 10X PCR buffer, 0.5 U Platinum Taq DNA polymerase (Invitrogen), 2.5 mM MgCl₂, 400 μM dNTPs (Invitrogen), 0.4 μM forward and reverse primers, SYBR® green (1:10,000 Bio-Rad) and ultrapure H₂O. Standard curves for all housekeeping and target genes were constructed from serial dilutions of cDNA ranging from 10⁻² to 10⁻⁶. Master mixes for all samples were prepared in triplicate (3 x 25 μL) and run on 96 well PCR plates (Axygen), sealed with microfilm (Bio-Rad). Reaction components were exactly the same for all target and housekeeping genes with an appropriate volume of cDNA for each gene determined during qPCR optimization.

The annealing temperature (Tₘ) of each of the reactions was determined theoretically by using NetPrimer (PremierBiosoft), and experimentally confirmed by performing a PCR with visualization of products on an agarose gel. Specific Tₘs for all target genes are listed in Table 3.2. The PCR cycle parameters used to amplify all genes is listed in Table 3.3. A melt curve analysis from SYBR green PCR products was performed to confirm amplification of only one product (single peak) among the replicated and serially diluted samples.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>N° repeats</th>
<th>Temp (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
<td>1x</td>
<td>94</td>
<td>7 min</td>
</tr>
<tr>
<td>Cycle 2</td>
<td>40x</td>
<td>94</td>
<td>30 sec</td>
</tr>
<tr>
<td>Step 1</td>
<td></td>
<td>Tm</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td></td>
<td>72</td>
<td>30 sec</td>
</tr>
<tr>
<td>Step 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cycle 3</td>
<td>1x</td>
<td>72</td>
<td>5 min</td>
</tr>
<tr>
<td>Cycle 4</td>
<td>61x</td>
<td>65-95°*</td>
<td>10 sec</td>
</tr>
<tr>
<td>Cycle 5</td>
<td>1x</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

*Increase set point temperature by 0.5°C/ read

Table 3.3 – qRT-PCR cycle parameters including data collection point and melt curve analysis.

3.2.13 Data analysis for quantification of gene expression

Calculation of the threshold cycle values (Cₚ) for every sample were performed using the Bio-Rad iQ5 real time detection software. Data was acquired once per cycle at the same temperature (Tₘ) in order to monitor amplicon yield. From a standard curve with log input amount and Cₚ values for each gene, the input amount for each sample was calculated for target gene using a standard curve. Relative quantification using the Livak Method was employed, where expression of the control sample is equal to 1.
\[ \Delta C_T(-N) = C_T(nif-N) - C_T(housekeeping-N) \]
\[ \Delta C_T(+N) = C_T(nif+N) - C_T(housekeeping+N) \]
\[ \Delta \Delta C_T = \Delta C_T(-N) - \Delta C_T(+N) \]
\[ 2^{\Delta \Delta C_T} = \text{Normalized expression ratio} \]

**Figure 3.1** - Livak Method formula employed for normalisation of expression data (Bio-Rad Real-time PCR, applications guide, 2006)

The relative gene expression rate of target genes was compared with the reference genes (16S rRNA, gyrB) generated by the standard curve and expression fold change determined. Where amplification efficiency of the target and reference genes was similar but not equal, the Pfaffl Method (Pfaffl 2001) was employed for calculations, and the actual amplification efficiency value (E, calculated from the slope of the standard curve) was substituted. Where amplification efficiency (E) differed between target and reference genes, the formula is as follows.

\[ E = 10^{-(1/\text{slope})} \]

\[ \text{Ratio} = \frac{(E_{nif})^{\Delta C_T}_{nif(control-test)}}{(E_{housekeeping})^{\Delta C_T}_{housekeeping(control-test)}} \]

**Figure 3.2** - Pfaffl method formula for when amplification efficiencies of target and reference gene are not similar

### 3.2.14 Statistical analysis

All statistical analyses were performed with SPSS v19 and data were expressed as mean and S.E.M. Data was Log₂ transformed in order to ensure homogeneity of variance. A T-test was used to compare the expression of test and control genes of 16SrRNA and gyrB. A probability level of \( P < 0.05 \) was considered statistically significant. Gene expression data were analysed by 1-way ANOVA with a Tukey’s HSD test (Hsu 1996) to assess the effects of time, with and without soluble nitrogen.
3.3 Results

3.3.1 Extraction of RNA from *Leptospirillum* species

Extraction of RNA from all *Leptospirillum* species under control and test conditions was successful. The 16 and 23S bands are visible. No contaminating genomic DNA was identified.

![Image of gel electrophoresis](image)

*Figure 3.3 – Leptospirillum spp. on a 1% TBE agarose gel.*

The quality of the primers and specificity to the target gene were assessed by PCR and products were visualised by agarose gel electrophoresis. This confirmed that each primer set produced a single fragment of the expected size. Resultant sequencing and BLAST analysis revealed that the PCR products matched the target sequence of interest, proving correct amplification of primer sets.

3.3.2 Expression of reference genes under test and control conditions

Expression of both the 16S rRNA and *gyrB* genes were detected at measurable levels in all *Leptospirillum* species, at all time points sampled and under both test and control conditions. Table 3.4 demonstrates averaged threshold detection (*C*<sub>T</sub>) values recorded under test and control conditions for both genes. A calibration curve from the *C*<sub>T</sub> values of the reference genes (16S rRNA and *gyrB*) were first constructed from test and control samples using template dilutions. Examination of the standard curves produced by the 16S rRNA
and gyrB genes demonstrated high correlation coefficients and similar slopes with single melting peaks showing no evidence of artefact such as primer dimer.

Table 3.4 – Average 16S rRNA and gyrB C\textsubscript{T} values from serial cDNA template dilutions of control and test cultures. Averages were constructed from C\textsubscript{T} values of cDNA serial dilutions; 10\textsuperscript{-1}, 10\textsuperscript{-2}, 10\textsuperscript{-3}, 10\textsuperscript{-4} and 10\textsuperscript{-5} recorded for each sample.

(a) 16S rRNA

<table>
<thead>
<tr>
<th></th>
<th>L. ferrodiazotrophum</th>
<th>L. ferrilium</th>
<th>L. ferrooxidans</th>
</tr>
</thead>
<tbody>
<tr>
<td>+N -N</td>
<td>+N -N</td>
<td>+N -N</td>
<td></td>
</tr>
<tr>
<td>10\textsuperscript{-2}</td>
<td>20.23 20.152</td>
<td>20.683 19.949</td>
<td>20.650 20.318</td>
</tr>
<tr>
<td>10\textsuperscript{-3}</td>
<td>24.162 23.522</td>
<td>24.041 23.319</td>
<td>24.162 23.522</td>
</tr>
<tr>
<td>10\textsuperscript{-4}</td>
<td>27.653 27.004</td>
<td>27.211 26.847</td>
<td>27.450 26.765</td>
</tr>
<tr>
<td>10\textsuperscript{-5}</td>
<td>31.375 30.690</td>
<td>30.270 30.900</td>
<td>30.678 30.008</td>
</tr>
</tbody>
</table>

(b) gyrB

<table>
<thead>
<tr>
<th></th>
<th>L. ferrodiazotrophum</th>
<th>L. ferrilium</th>
<th>L. ferrooxidans</th>
</tr>
</thead>
<tbody>
<tr>
<td>+N -N</td>
<td>+N -N</td>
<td>+N -N</td>
<td></td>
</tr>
<tr>
<td>10\textsuperscript{-3}</td>
<td>23.655 23.595</td>
<td>23.820 24.081</td>
<td>23.655 23.595</td>
</tr>
<tr>
<td>10\textsuperscript{-4}</td>
<td>27.076 26.830</td>
<td>27.423 27.381</td>
<td>26.770 27.088</td>
</tr>
<tr>
<td>10\textsuperscript{-5}</td>
<td>30.374 30.163</td>
<td>30.585 30.652</td>
<td>29.982 30.482</td>
</tr>
</tbody>
</table>

For the reference genes to be used as relative constants, the expression rate of the 16S rRNA and gyrB genes must be similar under both conditions tested; that is the C\textsubscript{T} value of 16S +N ≈ 16S –N, ± 5% and C\textsubscript{T} of gyrB +N ≈ gyrB –N, ± 5%. Accepting a 95% confidence interval, the expression of reference genes in the test and the control samples were statistically similar (two-tailed t-test) for cDNA values dilutions of 10\textsuperscript{-1}, 10\textsuperscript{-2}, 10\textsuperscript{-3}, 10\textsuperscript{-4} and 10\textsuperscript{-5} failed the hypothesis and were excluded from all reference calculations.
Nitrogen fixation gene regulation

Table 3.5 – Resultant p values from two-tailed t-test determining similarity between housekeeping gene expression with serially diluted cDNA. P values >0.05 indicate no significant difference between the control and test Ct values recorded for 16S rRNA and gyrB.

<table>
<thead>
<tr>
<th></th>
<th>L. ferrodiazotrophum</th>
<th>L. ferriphilum</th>
<th>L. ferrooxidans</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>16S rRNA</td>
<td>gyrB</td>
<td>16S rRNA</td>
</tr>
<tr>
<td>10^{-1}</td>
<td>0.551</td>
<td>0.518</td>
<td>0.237</td>
</tr>
<tr>
<td>10^{-2}</td>
<td>0.483</td>
<td>0.960</td>
<td>0.195</td>
</tr>
<tr>
<td>10^{-3}</td>
<td>0.060</td>
<td>0.748</td>
<td>0.085</td>
</tr>
<tr>
<td>10^{-4}</td>
<td>&lt;0.01</td>
<td>0.246</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td>10^{-5}</td>
<td>&lt;0.01</td>
<td>0.190</td>
<td>0.01</td>
</tr>
</tbody>
</table>

To ensure that the fluorescence detected was entirely produced by the desired DNA fragment, a dissociation curve of the PCR product was performed. Melt curve analysis of the 16S rRNA and gyrB genes amplified showed that the fluorescent signal obtained for the primer sets chosen originated from a single product only. Melting temperatures (Figures 3.4 and 3.5) for L. ferrodiazotrophum 16S rRNA was 91 °C and gyrB, 90.5 °C, L. ferriphilum 16S rRNA was 91 °C and gyrB 88 °C and L. ferrooxidans 16S rRNA, 90.5 °C and gyrB, 88 °C.

The amplification efficiencies (E) for the 16S rRNA and gyrB in Leptospirillum control and test samples ranged between 93% to 109%. Correlation coefficient calculated from the serial diluted samples ranged between 0.985 and 0.998 (Table 3.6).

Table 3.6 – Efficiency and sensitivity of 16S rRNA and gyrB quantitative RT-PCR assays under control and test conditions

<table>
<thead>
<tr>
<th>Target Species</th>
<th>Gene and Condition</th>
<th>Correlation coefficient</th>
<th>Amplification Efficiency</th>
<th>Slope</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrodiazotrophum</td>
<td>16S rRNA +N</td>
<td>0.998</td>
<td>98.9%</td>
<td>-3.54</td>
</tr>
<tr>
<td></td>
<td>16S rRNA -N</td>
<td>0.998</td>
<td>94.5%</td>
<td>-3.46</td>
</tr>
<tr>
<td></td>
<td>gyrB +N</td>
<td>0.997</td>
<td>100%</td>
<td>-3.30</td>
</tr>
<tr>
<td></td>
<td>gyrB -N</td>
<td>0.994</td>
<td>98.7%</td>
<td>-3.35</td>
</tr>
<tr>
<td>L. ferriphilum</td>
<td>16S rRNA +N</td>
<td>0.996</td>
<td>95%</td>
<td>-3.44</td>
</tr>
<tr>
<td></td>
<td>16S rRNA -N</td>
<td>0.998</td>
<td>100%</td>
<td>-3.30</td>
</tr>
<tr>
<td></td>
<td>gyrB +N</td>
<td>0.995</td>
<td>109%</td>
<td>-3.11</td>
</tr>
<tr>
<td></td>
<td>gyrB -N</td>
<td>0.995</td>
<td>98.5%</td>
<td>-3.36</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td>16S rRNA +N</td>
<td>0.99</td>
<td>93%</td>
<td>-3.50</td>
</tr>
<tr>
<td></td>
<td>16S rRNA -N</td>
<td>0.997</td>
<td>100%</td>
<td>-3.50</td>
</tr>
<tr>
<td></td>
<td>gyrB +N</td>
<td>0.985</td>
<td>97.7%</td>
<td>-3.37</td>
</tr>
<tr>
<td></td>
<td>gyrB -N</td>
<td>0.994</td>
<td>94.5%</td>
<td>-3.60</td>
</tr>
</tbody>
</table>
Figure 3.4 - Melt-curve analysis of 16S rRNA qRT-PCR products from *Leptospirillum* samples. Melt curves were performed at the end of SYBR Green qPCRs to check for primer-dimer or non-specific product formation. These are plots of dI/dT from one qRT-PCR run; the peaks indicate the $T_m$ of the amplification products.
Figure 3.5 - Melt-curve analysis of *gyrB* qRT-PCR products from *Leptospirillum* samples. Melt curves were performed at the end of SYBR Green qPCRs to check for primer-dimer or non-specific product formation. These are plots of dI/dT from one qRT-PCR run; the peaks indicate the $T_m$ of the amplification products.
3.3.3 Expression of target genes in *L. ferrodiastrophum*

As the amplification efficiency differed between the target gene and the reference gene, the Pfaffl Method was used to calculate the expression ratio for each target gene. To compare the changes in *nif* gene expression between the genes, data normalized to the relative constant (16S rRNA and *gyrB* average) was log₂ transformed and an average of all the replicate samples obtained (Figure 3.6).

![Figure 3.6 - Differences in expression of the *nif* genes relative to normalized data. Ratios of target genes were obtained from 16SrRNA and *gyrB*. Mean values ± S.E.M are presented (n = 12).](image)

There was no statistically significant difference between the expression of the *nifH, D* and *K* genes at any time point sampled, however a difference in single gene expression was noted with time. At 72 hours post exposure to the absence of soluble nitrogen, *nifH* expression was detected, similarly *nifD* and *nifK*. Incubation for a following 24 hours under these conditions significantly increased *nifH* expression, demonstrating a statistical difference occurring with time (p < 0.001). Significant increases were also recorded for *nifD* (p < 0.01) and *nifK* (p < 0.01) expression at 96 hours compared to 72 hours. A slight increase in *nifD*K expression was noted after 144 hours, but this escalation was not significantly different than results recorded at 96 hours (p = 0.194).

Expression of the *nifs* gene was similar to *nifD*K at 72 hours, as was *nifU*. Expression of *hesB* was lowest at 72 hours, and significantly less than *nifs* (p = 0.05), but not *nifU* (p = 0.21). The difference in expression between *nifs* and *hesB* was not seen at either 96 or 144 hours post exposure to soluble nitrogen absence. Expression of *nifs-U* and *hesB*
genes increased over time, with *nifU* and *hesB* levels significantly greater after 96 hours (p = <0.05, and <0.01 respectively) than at 72 hours, whereas the *nifs* increase was not significant. A further increase in *nifs-U* and *hesB* expression was recorded at 144 hours, of which only *hesB* was statistically different (p = < 0.05) than the value recorded at 96 hours.

### 3.3.4 Expression of target genes in *L. ferrooxidans*

As the amplification efficiency of all experiments was similar, the Livak Method equation was used to obtain a relative expression ratio of the *nif* genes compared to the housekeeping genes. To compare the changes in *nif* gene expression between the genes, data normalized to the relative constant (16S rRNA and *gyrB* average) was log₂ transformed and an average of all the replicate samples obtained (Figure 3.7).

![Graph](image)

**Figure 3.7** - Differences in expression of the *nif* genes relative to normalized data. Ratios of target genes were obtained from 16SrRNA and *gyrB*. Mean values ± S.E.M are presented (n = 12).

Expression of each of the *nifH*, and *nifD* genes did not differ significantly over time (p = 0.8005), at all time points sampled. Noticeably different was *nifK*, demonstrating a statistically significant increase in expression at 144 hours compared to expression at 72 hours (p = 0.011). After 144 hours exposed to the absence of soluble nitrogen, *nifH* and *nifK* expression appeared to differ, but this variance was not statistically significant (p = 0.09).

The *nifs* gene demonstrated different levels of expression at each time point sampled.
Expression of *nifS* was detected at 72 hours and had decreased by 96 hours, but this change was not statistically significant (*p* = 0.23). Comparatively, by 144 hours, expression had increased, and was statistically different than the values recorded at 72 hours (*p*<0.05) and 96 hours (*p*<0.01).

Expression of *nifU* and *hesB* at 72 hours was lower than *nifS*, however the difference was not statistically significant (*p* = 0.346). After 96 hours, expression of *nifU* and *hesB* remained constant, with changes in expression not being statistically different than those recorded at 72 hours for either gene (*p* = 0.639). The decrease in *nifS* at 96 hours meant that the increased expression of *hesB* at this time point was significantly different (*p* = 0.043).

During the second exponential growth phase (144 h), expression of all the *nifS-U-hesB* genes increased, depicting values similar to those recorded with the *nifHDK* genes. The increase in *nifS-U-hesB* expression after 144 hours was significantly different than the values recorded at 96 hours (*p* = <0.05).
3.3.5 Expression of target genes in \textit{L. ferriphilum}

As amplification efficiency for the test and reference genes were similar, the Livak Method was used to obtain an expression ratio for each gene. Data normalized to the relative constant (16S rRNA and \textit{gyrB} average) by \(\log_2\) transformation allowed comparison of the expression rates between the genes (Figure 3.8).

![Figure 3.8 - Differences in expression of the \textit{nif} genes relative to normalized data. Ratios of target genes were obtained from 16SrRNA and \textit{gyrB}. Mean values \pm S.E.M are presented (\(n = 12\)).](image)

Following 72 hours post exposure to the absence of soluble nitrogen, expression of \textit{nifA} was recorded at significantly higher levels than \textit{nifS} (\(p = 0.02\)), \textit{nifU} (\(p < 0.01\)) and \textit{hesB} (\(p = 0.009\)). There was no statistical difference between the expression of the \textit{nifs-U-hesB} genes at 72 hours. After 96 h, expression of the \textit{nifA} gene increased, demonstrating a statistical difference related to time (\(p < 0.01\)). This change in expression was also mirrored by \textit{nifS}, \textit{nifU} and \textit{hesB} genes, demonstrating a statistically significant increase (\(p < 0.05\)) in values recorded after 96 hours. At this time point, \textit{nifS} expression was greater than \textit{hesB} (\(p = 0.01\)), but not \textit{nifU} (\(p = 0.08\)). There was no difference in gene expression between \textit{nifU} and \textit{hesB} after 96 hours.

By 144 hours, expression of \textit{nifA} remained high and did not return to levels recorded at 72 hours. Expression of \textit{nifS} continued unchanged after 144 hours, remaining at the same level seen after 96 hours. An additional rise in \textit{nifU} and \textit{hesB} expression was recorded after 144 hours, however, only the increase in \textit{hesB} expression was statistically significant (\(p < 0.01\)).
3.3.6 Comparison of target genes across *Leptospirillum* species

Comparison of Log2 transformed expression ratios from all *Leptospirillum* species genes demonstrated differences in expression rates of the same genes when exposed to soluble nitrogen deficiency (Figure 3.9). Expression of *nifS* was greatest in *L. ferrophilum* after 96 hours, significantly more than *L. ferrodiazotrophum* (*p* = <0.05) and *L. ferrooxidans* (*p* = <0.01) at the same time point. No statistical difference in *nifS* expression between the three species was detected at either 72 or 144 hours. Values of *nifU* expression were statistically lower in *L. ferrooxidans* at 72 and 96 hours than *L. ferrophilum* and *L. ferrodiazotrophum* (*p* = <0.05 for both), but not significantly different after 144 hours. After 72 hours, *hesB* expression by *L. ferrophilum* was greater than *L. ferrooxidans* (*p* = <0.05), but not *L. ferrodiazotrophum*. The difference in *hesB* expression between *L. ferrooxidans* and *L. ferrodiazotrophum* was not significant at any time point. A further change in *hesB* expression by 96 hours demonstrated a significant increase in *L. ferrophilum* *hesB* mRNA compared to both *L. ferrodiazotrophum* (*p* = <0.01) and *L. ferrooxidans* (*p* = <0.05). At this time point, *L. ferrodiazotrophum* *hesB* expression was also significantly more than *L. ferrooxidans* (*p* = <0.05). By 144 hours, *hesB* expression by *L. ferrophilum* was significantly greater than *L. ferrooxidans* (*p* = <0.05), but not *L. ferrodiazotrophum*. There was no statistical difference between *L. ferrodiazotrophum* and *L. ferrooxidans* *hesB* expression at 144 hours.

There was no statistical difference in the expression of *nifH, D* or *K* genes between *L. ferrooxidans* and *L. ferrodiazotrophum* at 72 hours. After 96 and 144 hours, levels of *nifH* were statistically greater in *L. ferrodiazotrophum* than *L. ferrooxidans* (*p* = <0.01, <0.05). The increase in *nifD* expression by *L. ferrodiazotrophum* at 96 h was greater than *L. ferrooxidans* (*p* = <0.01), but not so after 144 hours (*p* = 0.24). The expression of *nifK* was not statistically different between the two species until 144 post exposure to absent soluble nitrogen, where *L. ferrodiazotrophum* demonstrated greater expression than *L. ferrooxidans* (*p* = 0.01).
Figure 3.9 – Comparison of nif gene expression across Leptospirillum species
3.4 Discussion

In this study, quantitative real time PCR was used to monitor expression of nitrogen fixing genes of *Leptospirillum* species grown in the absence of nitrogen. There was an increase in the relative expression of the *nif* genes in all *Leptospirillum* species, most notably during periods of exponential growth. From the observed changes in gene expression we can conclude that there was a transcriptional response to the growth of the *Leptospirillum* species the absence of nitrogen, with the increased *nif* mRNA transcripts indicating of probable activation of nitrogen fixation pathways.

3.4.1 Effect of soluble nitrogen deprivation on *nif* mRNA expression

Increased expression of the *nifH-D-K* genes in *L. ferrooxidans* was detected after 72 hours growth in the absence of soluble nitrogen, at which point exponential growth had yet to begin after sub-culturing (chapter 2). Sampling of mRNA transcripts during the identified exponential growth phase (96 h) showed a significant rise in *nifHDK* expression with similar levels detected when sampling was repeated at 144 hours. Even though the mRNA transcript levels were similar at 96 and 144 hours, nitrogenase activity, via acetylene reduction was greatest at 144 hours. Similar phenomena have been described in *Klebsiella pneumonia* where nitrogenase activity did not appear until 60 minutes after mRNA initiation (Collmer and Lamborg 1979). In the absence of soluble nitrogen, *nif* mRNA stability increases (Kahn, Hawkins, and Eady 1982) and there is also a trend towards longer mRNA half lives when growth rates are slow. The slow growth of *L. ferrooxidans* without ammonium may have resulted in slower mRNA degradation and turnover, explaining why *nif* mRNA abundance showed little variation between 96 and 144 hours. The increase in *nifHDK* transcription from 72 to 96 hours before reaching the final levels recorded at 144 hours may also have served to accelerate protein changes (Lee et al. 2011) necessary for nitrogen fixation.

Increased expression of the *nifHDK* genes in cultures of *L. ferrooxidans* 3.2 under ammonium starvation was demonstrated by Parro and Moreno-Paz (2003), with microarray analysis showing a 4.9 fold increase at 300 hours as well as changes in the regulation of the *nifSU-hesB* genes. The RT-PCR analysis in this study confirmed these results, and demonstrated that by 144 hours, average expression of *nifHDK* in *L. ferrooxidans* was 5.39 (±0.47) times greater than in cultures with soluble nitrogen (non-log2 transformed). Unlike *L. ferrooxidatum*, when expression of *nifH*, *D* and *K* genes was low during the non exponential growth phase at 72 h, the corresponding *L. ferrooxidans* non-exponential growth phase (recorded at 96 h) demonstrated an increase in *nifHDK* expression, which remained constant even after the exponential growth phase recommenced at 144 h. This may high-
light a possible difference between the two species regarding the mechanisms of nitrogen management and proliferation. However, like *L. ferrooxidans*, the small variation in expression between 96 and 144 hours could also be attributed to slow mRNA degradation and turnover. The higher expression rate of the *nifHDK* genes in *L. ferrooxidans* at 96 h may also explain the greater rate of acetylene reduction recorded in Chapter 2, in comparison to *L. ferrooxidans*.

Unlike *L. ferrooxidans* and *L. ferrodiazotrophum*, to date, no homologous *nifHDK* genes have been identified in *L. ferriphilum*. Nonetheless, it continued to proliferate when grown in the absence of soluble nitrogen. As the *nifA* gene is known to encode for the NifA protein – responsible for transcription of nitrogen fixation operons (Martínez-Argudo, L., et al. 2005), permitting induction of other secondary *nif* genes (*nifs-U-hesB*), expression levels of *nifA* were monitored in *L. ferriphilum* in an attempt to determine patterns of nitrogen fixation in this species. In many diazotrophs, the highly specific *nifLA* regulatory system governs nitrogen fixing bacteria at the transcription level in response to the fixed nitrogen status of the environment (Little et al. 2002).

The detection of increased *nifA* transcript abundance in *L. ferriphilum* at 96 h in the absence of soluble nitrogen indicates the release of the NifA activator protein from NifL, allowing for transcriptional activation of the *nif* gene promoters. Over time, a noticeable increase in the expression of the *nifs-U-hesB* genes followed the increase in *nifA*, peaking at 144 hours, at which point expression of the *nifA* had stabilized. Under nitrogen fixing conditions, the transcript abundance of *nifA* increased in species of *Azoarcus* (Egener et al. 2002), *Pseudomonas* (Yan et al. 2010) and *Rhzobium etli* (Salazar et al. 2010), resulting in the modulation of *nifH* transcripts in response to low ammonium levels. These results suggest that similar regulatory mechanisms controlling the expression of nitrogen fixation genes could be present in *L. ferriphilum* allowing for continued survival in areas of depleted soluble nitrogen.

After exposure of all *Leptospirillum* species to the absence of soluble nitrogen, an increased abundance in the *nifs-U-hesB* gene levels was detected. In *L. ferrodiazotrophum*, the expression of *nifs-U* genes increased significantly from 72 to 96 hours, plateauing by 144 hours. This gene expression mirrored the trend seen with *nifHDK* gene expression in *L. ferrodiazotrophum* at the same time period. It appears that at the start of the *L. ferrodiazotrophum* exponential growth phase (96 h), significant increases in nitrogen fixation gene expression are necessary for proliferation in the absence of soluble nitrogen. Similar to *L. ferrodiazotrophum*, *L. ferriphilum* also demonstrated a significant increase in *nifs-U* gene transcription after 96 h, remaining high up to 144 hours. As the *nifs* and *nifu* gene products (NifS, NifU) are required to mobilize and transport sulphur for the structural nitrogenase (Dixon and Kahn 2004), essential for full activation of the nitrogenase component proteins (Zheng et al. 1993), increased transcription of the *nifs-U* genes in these two spe-
cies strongly suggests that nitrogen fixation was taking place.

A slight decrease in nifS gene expression was detected in L. ferrooxidans at 96 hours, but no change in nifU levels were detected. Unlike L. ferrodiazotrophum and L. ferriphilum, L. ferrooxidans had a lag phase in growth at 96 hours. Even though there was a slight increase in nifHDK expression, the lower nifS expression at this time point in L. ferrooxidans may have impacted on the ability to form a functional nitrogenase enzyme, affecting cellular proliferation mechanics. After 144 hours and during the exponential growth phase, nifS-U transcription in L. ferrooxidans reached similar levels to those expressed by L. ferriphilum and L. ferrodiazotrophum. For all Leptospirillum spp it appears that nif mRNA synthesis occurs early in the exponential growth phase, with total RNA synthesis at maximum during mid-exponential growth (144 h). This demonstrates a time period necessary for the synthesis and construction of all the components crucial for the formation of an active nitrogen fixing system, similar to that seen with Klebsiella pneumonia (Kaluza and Hennecke 1981). The increased expression of the nifS-U genes by 144 hours may also be a result of the need to physiologically stabilize the nitrogenase enzyme in an aerobic environment (Dean and Jacobson 1991) as aerobic inactivation of these genes is known to inhibit growth of diazotrophs (Dean, Bolin, and Zheng 1993).

Expression of hesB is known to occur only under conditions of nitrogen fixation (Huang et al. 1999) and in this study, levels of hesB increased significantly over time, in each Leptospirillum species, with maximum expression detected at 144 hours. In L. ferrodiazotrophum at 72 hours and L. ferrooxidans at 96 hours, there was a significant difference in the levels of hesB and nifS expressed by each species, which was unexpected as the nifS-U-hesB genes are organized in a continuous fashion and should have also demonstrated the same expression level. In other nitrogen fixing species, hesB shows homology to iscA\textsuperscript{Nif} in A. vinelandii (Johnson et al. 2005) and iscN in Rhizobium etli (Dombrecht et al. 2002). Differences in the expression rate between the iscA\textsuperscript{Nif}-nifU-S genes organized in a contiguous fashion has been documented in A. vinelandii under conditions of nitrogen fixation, where iscA\textsuperscript{Nif} expression increased 23.7 fold, whilst nifS only by 8.0 fold (Hamilton et al. 2011). The different abundance of these co-transcribed genes in Leptospirillum suggests regulation by transcriptional attenuation (Naville and Gautheret 2009) in response to the availability of soluble nitrogen, stored nitrogen (chapter 2) and oxygen exposure, or processing resulting in different segmental stabilities for expressed genes (Meinken et al. 2003). Also possible is the presence of more than one gene promoter (Yun and Szalay 1984), resulting in multiple transcripts, where mRNA is processed as either one primary transcript or multiple primary transcripts (Dominic, Chen, and Zehr 1998). Additionally, different mRNA stability between the genes of the operon may also explain the dissimilar expression levels detected. Full genomic sequencing of the Leptospirillum species would assist in determining if this is the case.
3.4.2 Evaluation of the 16S rRNA and gyrB gene as reference genes

As there is no complete genome sequence for any *Leptospirillum* species the number of gene copies for either of the reference genes or test nitrogen fixation genes remains unknown. Therefore the choice of reference genes for qRT-PCR normalization is crucial for accurate comparisons of gene expression. Historically, the ribosomal 16S ribosomal RNA gene is used, but as it is abundantly expressed and known to be present in multiple copies, it can yield very small C\textsuperscript{T} values, that can be undesirable when the target genes have low expression values. To prevent large errors arising from the use of a single gene for normalization, it has been demonstrated that at least two housekeeping genes should be used for internal standards (Thellin et al. 1999). Even housekeeping genes can be expressed at different levels under different conditions, especially in bacteria, where expression rates are directly proportional to growth rate (Avison 2007).

In this study, the 16S rRNA and *gyrB* genes were selected based on the fact that their functions were unrelated, minimizing the possibility of co-regulation expression bias (Garcia-Vallejo et al. 2004). Expression correlations between the test (-N) and control (+N) conditions reliably generated quantitative results at template dilutions of 10\textsuperscript{1} to 10\textsuperscript{3} with both reference genes. The lack of expression correlation between the 16S rRNA and *gyrB* genes at the lower template concentrations could be attributed to less than optimal PCR efficiency or pipetting error. As housekeeping gene-averaged experiments are subject to systematic errors in interpretation (Warren et al. 2006), this rigorous examination of each C\textsubscript{r} value permitted identification of qualitative trends with confidence.

3.5 Conclusions

The purpose of this study was to develop and evaluate a quantitative real time PCR method for characterising mRNA transcription rates in *Leptospirillum* species. Given the constitutive expression, moderate abundance and high RT-PCR efficiencies, the 16S rRNA and *gyrB* genes proved to be suitable candidates to act as reference genes, as fluctuations in expression between control and test conditions was minimal. Stimulation of the *nif* genes in all the *Leptospirillum* species was evident after growth in the absence of soluble nitrogen, indicating activation of nitrogen fixation pathways.

The identification of homologous *nifH, D and K* genes in *L. ferrihilum*, taken together with the increased abundance of *nifA, nifs, nifu* and *hesB* demonstrated in this study, would strengthen the hypothesis that nitrogen fixation in this species is probable. Consequently, these results confirm previous studies describing *L. ferrooxidans* as a diazotroph and endorse *L. ferrodiazotrophum* as a second functional diazotroph in the *Leptospirillum* genus.
In conclusion, the real time PCR technology was reliable, reproducible and accurate in assessing the nitrogen fixing capabilities of the *Leptospirillum* species studied.
Chapter 4

Proteomic analysis of whole cell extracts from *Leptospirillum* species grown in the absence of soluble nitrogen

*The increase in mRNA expression of nif genes in all Leptospirillum species demonstrated a response to changes in environmental conditions. The work in this chapter aimed to determine whether the altered genomic expression was further reflected in the proteome.*

4.1 Introduction

*Leptospirillum* species are known to dominate microorganism populations inhabiting bio-leaching environments under temperate conditions. Identification of *Leptospirillum* genomes from these habitats (Tyson et al. 2004) has provided valuable information regarding population composition, but genomic sequences alone are not sufficient to reveal the specific molecular functions of all the genes. Genomic information does not identify how *Leptospirillum* species may alter their metabolism in response to changes in environmental conditions. By investigating protein expression a more accurate view of *Leptospirillum* functionality can be determined, as it is important to establish how differences in metabolic potentials between these lineages allow for distinct niches to be occupied. Identification of any potentially adaptive traits to environmental stress in closely related organisms is also significant from an evolutionary perspective. Proteomic analysis is possible even if the genome sequences are not fully available and annotated, as homologous sequences from genetically related microorganisms can be used for elucidation of function.

The purpose of this analysis was to identify proteins that were differently expressed under nitrogen limitation for the three *Leptospirillum* species and as such, attempt to identify proteins involved in nitrogen fixation to support growth data and acetylene reduction results from Chapter 2, and up-regulation of nitrogen fixation genes discerned in Chapter 3. In order to identify those proteins that were differentially regulated or modified in conditions
of soluble nitrogen deprivation, two-dimensional Polyacrylamide Gel Electrophoresis (2D PAGE) was employed to assess whole cell protein extracts. 2D PAGE allowed for the separation of proteins according to charge (pl) by isoelectric focusing (IEF) in the first dimension and by size (MW) with SDS PAGE in the second dimension. The high reproducibility of 2D PAGE was particularly valuable for multiple sample comparisons as it directly correlated the changes at the peptide level to individual proteins detected. The identification of the differently expressed or modified proteins was achieved by Mass Spectrometric analysis of digested peptides. The quantitative comparison of proteomic changes with statistical confidence was made possible by analysis of replicate samples, and these experiments offered insight into pathways that were activated, deactivated or altered under conditions of nitrogen deficiency.
4.2 Materials and methods

4.2.1 Bacterial strains and growth media

The bacterial strains of *L. ferrodiazotrophum*, *L. ferriphilum* and *L. ferroxidans* used in this study as well as the growth media are detailed in Table 2.1 from section 2.1.1 - Bacterial strains and growth media.

4.2.2 Experimental growth conditions

Cultures in exponential growth phase (determined from cell counts) were harvested by centrifugation at 20,000 rpm in a Beckman centrifuge (Avanti J-E) for 40 min at 25°C. Pellets were washed with 2 mL of acidified water, pH 1.8, to remove any traces of the initial growth media that contained nitrogen then centrifuged at 15,300 rpm for 30 min at 25°C and the supernatant removed. Cell pellets were resuspended in media either containing soluble nitrogen or without soluble nitrogen. Conical flasks (1 L) containing 200 ml of media were inoculated at a cell concentration of 3.8 x 10^6. Flasks were incubated at 30°C with no aeration.

4.2.3 Isolating whole cellular proteins

*Leptospirillum* cultures were collected by filtration with a 0.2 µL filter (PALL Corporation, hydrophilic polypropylene). The supernatant was discarded and cells adherent to the filter were removed by washing with 2 mL of acidified sterile water, pH 1.8, and transferred to a sterile 2 mL tube. Cells were centrifuged at 15,300 rpm at 4°C for 30 min, the supernatant discarded and the cell pellet resuspended in 2 mL of cold acidified sterile water. Tubes were placed on ice for 10 min to allow Fe^{3+} to precipitate, the cell suspension then transferred to a new tube and centrifuged at 15,300 rpm at 4°C for 20 min. The supernatant was discarded and the cell pellet resuspended in 100 µL of lysis buffer (2% CHAPS, 1 µL of 100 mg/mL Lysozyme (Invitrogen)). Cell pellets remained at room temperature for 20 min. The tubes were heated at 95°C for 10 minutes and briefly chilled on ice. To the lysis solution, 1 µL of PMSF, 400 µL of rehydration buffer (7 M Urea, 2 M Thiourea, 4% CHAPS, 0.8% Biolyte pH 3-10, 1 mM EDTA, 40 mM Tris pH 7.4, 50 mM DTT) and 1x Complete protease inhibitor cocktail (Roche), was added. Tubes were inverted manually to mix proteins with rehydration buffer. Protein samples were centrifuged at 15,300 rpm at 4°C for 10 min to remove contaminating cellular debris. The supernatant containing the protein was transferred to a new 1.5 mL tube. Proteins were stored at -80°C.
4.2.4  Assessing DNA contamination of protein extracts

A 5 µL sample of the protein extract was mixed with 2 µL of gel loading dye (1x) and loaded onto a 1% agarose TBE gel. The gel was run at 100 V for 40 minutes and stained with Ethidium bromide for 10 minutes, then analysed under UV light. Protein samples with no nucleic acid contamination were deemed of suitable quality for subsequent experiments.

4.2.5  Determination of protein concentration

Concentration of the protein sample was determined by the Bradford Protein Assay, using a BSA standard (Spector 1977). Two different dilutions of the protein sample, 1:5 and 1:10 were prepared and protein concentrations determined by comparison to the calibration curve.

4.2.6  Passive rehydration of IPG strips

Protein samples were mixed with rehydration buffer so that the concentration of protein applied to each IPG (Immobilized pH gradient strip) was 80 µg with a total volume of 250 µL. Protein samples were placed in rehydration wells and 17 cm, pl 4-7 IPG strips, gel side down, were overlaid. Strips and proteins remained at room temperature for 60 min to initiate rehydration. Mineral oil (Bio-Rad) was overlaid on the strip to prevent evaporation of the sample and the rehydration tray was covered with a lid and left overnight.

4.2.7  First dimension Isoelectric Focusing (IEF)

Proteins were electrophoretically separated in the first dimension by a Bio-Rad Protean IEF cell. Electrode wicks (Bio-Rad) were placed over the wire electrodes and 10 µL of 10 mM DTT applied to each wick. The IPG strip was placed on top of the electrode wicks. Strips were covered with mineral oil. A default temperature of 20 °C was selected. IEF was performed using a rapid gradient protocol of 15 min at 250 V, 3 hours at 10,000 V and a final step up to 40,000 Volt-Hours. The current was limited to a maximum of 50 µA per IPG strip. After focusing, IPG strips were removed from the focusing tray and drained of mineral oil. Strips were transferred to a clean rehydration tray whereby they were prepared for 2D SDS PAGE.

4.2.8  Second dimension SDS-PAGE

SDS-Polyacrylamide gels were cast using a Bio-Rad Protean II XL cell, with 1 mm spacers and 16 cm x 20 cm glass plates. A 10% gel was prepared with Acrylamide/bis (30%T,
2.67%, 1.5 M Tris-Cl pH 8.8, 10% SDS, 10% APS and TEMED. Gels were over-layed with water-saturated 2-butanol, left to polymerise for 45 minutes and then the overlay rinsed off with distilled water. Gels were stored at 4°C for 16-24 hours prior to use.

Equilibration buffer I (6 M Urea, 0.375 M Tris-Cl pH 8.8, 2% SDS, 20% glycerol 2% w/v DTT) was added to focused IPG strips and shaken for 10 min at room temperature. Used equilibrium buffer I was removed and replaced with equilibration buffer II (6 M Urea, 0.375 M Tris-Cl pH 8.8, 2% SDS, 20% glycerol 2.5% w/v iodoacetamide), and returned to the shaker for a further 10 min. Equilibrium buffer II was removed by decanting. IPG strips were washed with 1X Cathode buffer (0.2 M Glycine, 25 mM Tris-Cl, 0.1% w/v SDS), to remove traces of equilibrium II buffer. IPG strip were laid onto the gels. SeeBlue Plus 2 molecular weight marker (Invitrogen) was loaded on to a piece of filter paper and placed on top of the gel, next to the IPG strip. IPG strips were covered with melted overlay agarose (BIO-RAD ReadyPrep™) and allowed to solidify for 5 min. Gels were locked into place on the central cooling core and placed in the lower buffer chamber. Cathode buffer was loaded into the upper chamber while the lower chamber was filled with Anode buffer (1.9 M Tris-Cl pH 8.8). Electrophoresis was performed in two steps. An initial step of 16 mA/gel for 30 min and a second step of 25 mA/gel for 5 hours. Gels were removed from the tank and transferred to container with seal-able lid for staining.

4.2.9 Protein visualisation with silver staining.

Fixing solution (12% glacial acetic acid, 50% ethanol, 0.05% formalin) was applied to the gel and left overnight. Fixative solution was removed and gels washed with 20% ethanol for 20 min. After washing twice, the wash solution was discarded and sensitizing solution (0.02% w/v sodium thiosulfate) added to cover the gel. Gels were incubated for 5 min with shaking and washed twice with ultrapure water for 1 min. The water was discarded and cold (4°C) silver staining solution (0.2% w/v silver nitrate, 0.076% formalin) applied and incubated with shaking for 20 min. The silver stain was removed and gels rinsed twice with water for 20-30 sec. Gels were quickly rinsed (10 s) with developer solution (0.5 M sodium carbonate, 0.0025% sodium thiosulfate, 0.05% formalin) and used developer solution discarded. Fresh developer solution was added and the protein image developed over 2-5 min until desired intensity was reached. Developer solution was removed and terminating solution (35 mM Na₂EDTA) added. Terminating solution remained on the gel for 10 min and then discarded. Gels were washed with distilled water and stored in cling film at 4°C.
4.2.10 Image analysis

Silver stained 2D SDS-PAGE gels were scanned using a Perkin Elmer ProScan Imager. Densitometric image analysis for all 2D gels was performed using Progenesis Same Spots v1.0 software (Nonlinear Dynamics, UK). Quality control was performed for each image, verifying resolution (16-bit) and flagging saturated spots. Saturated spots were excluded from analysis. To enable the same protein pattern on each gel (100% matching strategy, for spot detection) a composite image from three gels under the same test conditions was created through visual examination and computer aided alignment. For each *Leptospirillum* species, the control experiment with soluble nitrogen was the reference gel. All proteins were detected and edited in reference to this gel. Subsequently, the spot pattern was transferred to all gels in the experiment. All matches or mismatches were visually inspected for accuracy and further management of spots was completed manually.

4.2.11 Protein quantification with silver staining

Protein quantification by relative integrated optical density (volume) was used for estimation of protein amount. The amount of protein present in a spot was the area of the spot multiplied by the pixel density. All spot volumes were normalised to the total protein amount detected, with each spot expressed relative to the total volume of all spots on the image.

\[ NV = \left( \frac{V_n}{V_t} \right) \times S \]

Where:

- \( NV \) = total spot volume method normalised volume of spot \( n \)
- \( V_n \) = volume of spot \( n \)
- \( V_t \) = total volume of all spots in that gel \( (V_1 + V_2 + \ldots + V_n, \text{ where } n \) = total number of spots)\n- \( S \) = user-defined scaling value (default = 100)

Experimental pI and Mw values were estimated with the Mw marker and Progenesis Same Spots. Formula from Nonlinear Dynamics, technical note - Normalisation methods for 2D gel image analysis using Progenesis.

4.2.12 Statistical analysis of proteomic data

Two analytical methods were employed for comparison of gels with control cell extracts to gels with cell extracts without soluble nitrogen: a multivariate pattern analysis method
to compare the whole proteome pattern and a univariate spot-to-spot analysis method to obtain detailed information on single spots that changed following soluble nitrogen deprivation. Principle component analysis (PCA) method was applied to compare protein patterns and carried out using Progenesis Same Spots. Proteins of different abundance were analysed for significance using a Student’s t-Test. Only spots demonstrating a significant change (p<0.05) and at least a 2-fold difference in abundance were considered to have changed in abundance.

4.2.13 Protein identification

Spots identified by Progenesis Same Spots analysis were excised manually and collected in 1.5 mL tubes. Gel pieces were subjected to destaining with 1% w/v potassium ferricyanide and 1.6% w/v sodium thiosulphate, followed by rinsing with ultra pure water. Destained gel slices were treated with multiple washes of 100 mM ammonium bicarbonate followed by washes of 50% acetonitrile (ACN) for dehydration. Slices were freeze dried to remove excess liquid and once dry, rehydrated with the addition of 100 μL of 100 mM ammonium bicarbonate containing 1 μg of trypsin (Roche). Peptide digestion was performed at 37 °C overnight. Peptide extracts were pooled and then freeze dried. Dried peptide fragments were rehydrated in 10 μL of 0.1% trifluoroacetic acid (TFA) and stored at −80 °C.

4.2.14 MALDI–TOF Mass Spectrometry

MALDI-TOF mass spectrometry was performed using a Voyager DE-Pro (Applied Biosystems, USA). A solution of α-cyano-4-hydroxycinnamic acid matrix (CHCA) was prepared fresh to a working concentration of 10 mg mL⁻¹ in H₂O/ACN/TFA (50:50:0.1, v/v/v), and 9 μL of the CHCA was mixed with 1 μL of the trypsin digest. A 2 μL sample of the matrix and peptide solution was deposited onto a stainless steel AB 100 spot MALDI target plate and allowed to air dry at room temperature. After complete drying, a further 2 μL of the sample was added and allowed to dry. Samples were then ready for analysis. Parameters for Mass Spectrometry are listed in Table 4.1, with a mass range of 500-3000 Da selected, as the CHCA matrix obscures peaks less than 500 Da
4.2.15 MALDI–TOF Mass Spectrometry data output

The peptide mass peak list generated was analysed using Data Explorer 4.8 (Applied Biosystems, USA). Peak lists generated were corrected for baseline drift and noise reduction.

4.2.16 Analysis of Mass Spectrometry data

The MASCOT search engine (MatrixScience) was used to identify peptides, restricted to the NCBI non-redundant database and from bacterial sources only. Searches were performed under the parameters of fixed carbamidomethylation of cysteines, variable oxidation of methionines, a mass tolerance of 0.1 Da and two missed trypsin cleavage. Confirmation of calculated pl and Mw from MASCOT results were performed by the ExPASy Proteomics Server with Monoisotopic resolution. Protein identifications were considered reliable when the MASCOT score was > 71 (MASCOT score was calculated as $-10 \times \log P$, where P is the probability that the observed match is a random event). This is the lowest score indicated by the program as significant ($p < 0.05$) and indicated by the probability of incorrect protein identification.
4.3 Results

4.3.1 Proteome changes in *Leptospirillum* species under soluble nitrogen deprivation

Reference protein patterns (+NH₄ conditions) for the three *Leptospirillum* species are seen in Figure 4.1 a, b and c, with three replicate gels used to create a spot map for each species and condition being tested. All identical and matchless proteins were given a unique identifier (spot #) by Progenesis Same Spots. For each *Leptospirillum* species, gels were compared between conditions of soluble nitrogen and deficient soluble nitrogen in an attempt to discern changes in *Leptospirillum* protein expression brought about by nitrogen stress. The analysis across the replicates of protein maps revealed qualitative (presence/absence) differences. Univariate spot to spot analyses were performed in order to obtain information about changes in abundance of single proteins. Proteins detected in only nitrogen deficient conditions were nominated as increased in abundance compared to controls, whilst those only discovered in control cultures were designated as decreased in abundance under nitrogen deficient conditions.
Figure 4.1 – 2D SDS PAGE of reference protein patterns (controls) for *L. ferrodiazotrophum*, *L. ferrphilum* and *L. feroxidans* without displayed positions of significantly differed spots.
4.3.2  *L. ferrodiazotrophum* proteomics

Progenesis Same Spots initially detected >3600 resolved proteins (n=6) in *L. ferrodiazotrophum* gels. After all detected proteins were edited and correctly matched, 46 proteins demonstrated different abundance levels between tested conditions (Figure 4.2) and were submitted for further statistical analysis. Of these 46 proteins, 27 met the statistically requirements (>2 fold abundance, p <0.05) with 19 proteins having increased abundance under soluble nitrogen deficiency and 8 with decreased abundance. Proteins notably different in abundance that met statistical requirements were selected for processing and identification by mass spectrometry (MS). Of the 27 proteins whose abundance had changed, 9 could not be visually isolated on the gel for excision and subsequent analysis by MS. From the remaining 18 proteins, identification of 9 were successful (Table 4.2). Peptide identification by Mascot returned results that were not statistically significant (score >71) for 9 digested proteins.

The identities of the 9 unique *L. ferrodiazotrophum* proteins determined by Mascot analysis of the PMF are detailed in Table 4.2, with extensive peptide coverage ranging from 34 - 61%. Homologous proteins were identified in other genera, the majority of which were acidophilic bacteria and not discredited as an incorrect match. In most cases, similar experimental and theoretical MW/pl values were obtained, even with spots where homology was obtained with a different organism.
Figure 4.2 – *L. ferrodiazotrophum* protein spot map generated by Progenesis Same Spots showing the initial 46 proteins detected as having different abundance levels between conditions of nitrogen depletion and soluble nitrogen abundance. The spots highlighted exhibited a detectable appearance in nitrogen deficient conditions common to at least 3 replicate experiments with yellow spots, demonstrating decreased abundance and green spots, increased. Identities of these highlighted spots are listed in Table 4.2.
Table 4.2 – PMF Identification of *L. ferrodiazotrophum* proteins exhibiting a significant change in abundance in nitrogen deficient cultures. The coloured regions of the table correspond to the 2D gel spots shown in Figure 4.2, with yellow indicating spots decreased abundance and green, increased (>2 fold change).

<table>
<thead>
<tr>
<th>Spot #</th>
<th>Genbank Accession</th>
<th>Homologous protein (microorganism)</th>
<th>Sequence Coverage</th>
<th>MOWSE score*</th>
<th>Theoretical Mw(Da)/pi</th>
<th>Experimental Mw(Da)/pi</th>
<th>E value</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>1008</td>
<td>gi</td>
<td>31790361</td>
<td>Putative 2-methylthioadenine synthetase (uncultured Acidobacteria bacterium)</td>
<td>45%</td>
<td>82</td>
<td>50,236/5.19</td>
<td>~64,000/5.2</td>
<td>0.052</td>
</tr>
<tr>
<td>1899</td>
<td>gi</td>
<td>251770823</td>
<td>Protein of unknown function (<em>L. ferrodiazotrophum</em>)</td>
<td>44%</td>
<td>82</td>
<td>36,631/5.74</td>
<td>~45,000/6.0</td>
<td>0.075</td>
</tr>
<tr>
<td>2939</td>
<td>gi</td>
<td>251771825</td>
<td>NUDIX hydrolase (<em>L. ferrodiazotrophum</em>)</td>
<td>39%</td>
<td>92</td>
<td>18,816/4.98</td>
<td>~17,000/5.0</td>
<td>0.07</td>
</tr>
<tr>
<td>1348</td>
<td>gi</td>
<td>206602946</td>
<td>Anthranilate synthase component I (<em>Leptospirillum</em> sp. Group II)</td>
<td>54%</td>
<td>100</td>
<td>49,064/6.15</td>
<td>~55,000/6.5</td>
<td>0.00074</td>
</tr>
<tr>
<td>2452</td>
<td>gi</td>
<td>308048382</td>
<td>Two component transcription regulator, winged helix (<em>Ferrimonas balearica</em> DSM 9799)</td>
<td>55%</td>
<td>79</td>
<td>27,273/5.82</td>
<td>~36,000/5.7</td>
<td>4.00E-13</td>
</tr>
<tr>
<td>2764</td>
<td>gi</td>
<td>206603867</td>
<td>Imidazolglycerol-phosphate dehydratase (<em>Leptospirillum</em> sp. Group II)</td>
<td>61%</td>
<td>104</td>
<td>23,340/6.36</td>
<td>~24,000/6.2</td>
<td>0.0068</td>
</tr>
<tr>
<td>2815</td>
<td>gi</td>
<td>198282587</td>
<td>Cpr/Fnr family transcriptional regulator (<em>At. ferrooxidans</em>)</td>
<td>43%</td>
<td>84</td>
<td>26,933/5.75</td>
<td>~22,000/6.0</td>
<td>4.00E-08</td>
</tr>
<tr>
<td>3167</td>
<td>gi</td>
<td>251772007</td>
<td>Putative reiske iron-sulfur family protein (<em>L. ferrodiazotrophum</em>)</td>
<td>48%</td>
<td>80.5</td>
<td>10,887/5.5</td>
<td>~11,000/4.9</td>
<td>1.00E-17</td>
</tr>
<tr>
<td>3560</td>
<td>gi</td>
<td>251773021</td>
<td>Enolase (<em>L. ferrodiazotrophum</em>)</td>
<td>39%</td>
<td>78</td>
<td>45,493/4.91</td>
<td>~36,000/4.5</td>
<td>1.40E-02</td>
</tr>
</tbody>
</table>

*Scores greater than 71 are significant (p<0.05)*
4.3.3 L. ferriphilum proteomics

Of the 2000 individual L. ferriphilum proteins detected by Progenesis Same Spots, 44 demonstrated different abundance levels between the conditions tested (Figure 4.3). Statistic evaluation of these selected spots revealed that levels of 29 of these proteins were significantly different, with 13 increasing in abundance, and 16 decreasing in abundance. Of the 29 proteins selected, eight could not be visually isolated on the gel for excision. From the remaining 21 proteins, identification of seven was successful. Peptide identification by Mascot returned results that were not statistically significant (score > 71) for 14 digested proteins.

Peptide mass fingerprint sequence coverage ranged from 26 - 44% for the seven proteins successfully identified (Table 4.3). Of the seven proteins, two were matched directly to L. ferriphilum and another two were homologous to other Leptospirillum species (L. ferrooxidans and L. rubarum), with the remaining three matching proteins from other micro-organisms.
Figure 4.3 – *L. ferriphilum* protein spot map generated by Progenesis Same Spots showing the initial 44 proteins detected as having differential expression rates between conditions of nitrogen depletion and soluble nitrogen abundance. The spots highlighted exhibited a detectable appearance in nitrogen deficient conditions common to at least 3 replicate experiments with yellow spots, decreased abundance and green spots, increased abundance. Identities of these highlighted spots are listed in Table 4.3.
Table 4.3 - PMF Identification of *L. ferriphilum* proteins exhibiting a significant change in abundance in nitrogen deficient cultures. The coloured regions of the table correspond to those 2D gel spots shown in Figure 4.3 with yellow indicating spots decreased in abundance and green, increased (>2 fold change).

<table>
<thead>
<tr>
<th>Spot #</th>
<th>Genbank Accession</th>
<th>Homologous protein (microorganism)</th>
<th>Sequence Coverage</th>
<th>MOWSE score*</th>
<th>Theoretical Mw(Da)/pI</th>
<th>Experimental Mw(Da)/pI</th>
<th>E value</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>594</td>
<td>gi</td>
<td>206603508</td>
<td>putative SAM-dependent methyltransferase (<em>L. ferriphilum</em>)</td>
<td>26%</td>
<td>86</td>
<td>88,495/5.23</td>
<td>~85,000/5.4</td>
<td>0.02</td>
</tr>
<tr>
<td>977</td>
<td>gi</td>
<td>206603034</td>
<td>Transposase (<em>L. ferriphilum</em>)</td>
<td>42%</td>
<td>95</td>
<td>45,748/6.99</td>
<td>~50,000/6.5</td>
<td>0.0004</td>
</tr>
<tr>
<td>1569</td>
<td>gi</td>
<td>291280117</td>
<td>L-fuculose phosphate aldolase (<em>Deferribacter desulfuricans SSM1</em>)</td>
<td>38%</td>
<td>73</td>
<td>20,782/5.85</td>
<td>~22,000/6.0</td>
<td>1.00E-06</td>
</tr>
<tr>
<td>1066</td>
<td>gi</td>
<td>78043811</td>
<td>3-isopropylmalate dehydratase large subunit family protein (<em>Carboxydothermus hydrogenoformans Z-2901</em>)</td>
<td>44%</td>
<td>91</td>
<td>44,244/5.81</td>
<td>~49,000/5.9</td>
<td>0.59</td>
</tr>
<tr>
<td>1176</td>
<td>gi</td>
<td>317151889</td>
<td>formate dehydrogenase accessory protein (<em>Desulfovibrio aespoeensis Aspo-2</em>)</td>
<td>38%</td>
<td>71</td>
<td>32,051/5.77</td>
<td>~38,000/5.7</td>
<td>0.55</td>
</tr>
<tr>
<td>1251</td>
<td>gi</td>
<td>124541846</td>
<td>Dihydroorotate dehydrogenase (<em>L. rubarum</em>)</td>
<td>39%</td>
<td>70</td>
<td>38,575/5.89</td>
<td>~38,000/5.7</td>
<td>0.0031</td>
</tr>
<tr>
<td>1499</td>
<td>gi</td>
<td>31747807</td>
<td>Lfe194p1 (<em>L. ferrooxidans</em>)</td>
<td>27%</td>
<td>191</td>
<td>11,978/4.66</td>
<td>~25,000/4.9</td>
<td>7.01E-14</td>
</tr>
</tbody>
</table>

*Scores greater than 71 are significant (p<0.05)
4.3.4 L. ferrooxidans

More than 1600 proteins exhibited by L. ferrooxidans were edited and matched (Figure 4.4), with a noticeable difference between 26 individual spots. Of these detected 26 spots, only nine proteins were statistically significant in their expression levels, with five demonstrating increased abundance and four with decreased abundance. Not all spots were visible for excising and only six were processed by PMF. Of the six spots analysed, only four protein identities were obtained for L. ferrooxidans, as two were below the confidence level of identification (<71).

All identified proteins demonstrated decrease abundance under nitrogen deficient conditions. The identities of the four unique L. ferrooxidans proteins are detailed in Table 4.4, with peptide coverage ranging from 39 - 54%. A homologous protein was identified in L. ferrooxidans and another in Leptospirillum group II species, with the remaining two found in other genera. Experimental and theoretical MW/pl values were compared and have a high correlation even with spots where homology was obtained with a different organism.

Table 4.4 – PMF Identification of L. ferrooxidans proteins exhibiting a significant change in abundance in nitrogen deficient cultures. The coloured region of the table correspond to the 2D gel spots shown in Figure 4.4 with yellow indicating spots representing proteins of decreased abundance (>2 fold change).

<table>
<thead>
<tr>
<th>Spot #</th>
<th>Accession</th>
<th>Homologous protein (microorganism)</th>
<th>Sequence Coverage</th>
<th>MOWSE score*</th>
<th>Theoretical Mw(Da)/pI</th>
<th>Experimental Mw(Da)/pI</th>
<th>E value</th>
<th>Fold Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>704</td>
<td>gi</td>
<td>116750906</td>
<td>OmpA family protein (Syntrophobacter fumaroxidans MPOB)</td>
<td>39%</td>
<td>99</td>
<td>58,039/6.16</td>
<td>~43,000/5.58</td>
<td>0.059</td>
</tr>
<tr>
<td>1120</td>
<td>gi</td>
<td>262194926</td>
<td>D-isomer specific 2-hydroxyacid dehydrogenase NAD-binding protein (Haliangium ochraceum DSM)</td>
<td>54%</td>
<td>77</td>
<td>32,057/6.46</td>
<td>~27,000/5.92</td>
<td>0.23</td>
</tr>
<tr>
<td>1262</td>
<td>gi</td>
<td>206602195</td>
<td>Carbohydrate kinase family protein (Leptospirillum group II)</td>
<td>38%</td>
<td>123</td>
<td>55,361/6.05</td>
<td>~35,000/5.6</td>
<td>5.22E-05</td>
</tr>
<tr>
<td>1362</td>
<td>gi</td>
<td>206602046</td>
<td>LexA repressor (L. ferrooxidans)</td>
<td>42%</td>
<td>71</td>
<td>24,856/9.55</td>
<td>~20,000/5.2</td>
<td>0.00037</td>
</tr>
</tbody>
</table>

*Scores greater than 71 are significant (p<0.05)
Figure 4.4 — *L. ferrooxidans* protein spot map generated by Progenesis Same Spots showing the initial 26 proteins detected as having differential expression rates between conditions of nitrogen depletion and soluble nitrogen abundance. The spots highlighted exhibited a detectable appearance in nitrogen deficient conditions common to at least 3 replicate experiments with yellow spots, decreased abundance, green spots, increased abundance. Identities of the 4 down regulated (yellow) spots are listed in Table 4.4. Identification of the two proteins with increased abundance (#507, 804) was not significant.
4.3.5 Comparative analysis of Molecular Weight and pl

The pl and molecular weight (kDa) of all identified proteins regardless of species or experimental condition were estimated, as mentioned in the method 4.2.11. The estimated experimental values were comparatively analysed against the theoretical value of pl and molecular weight of the identified proteins. Both pl and molecular weight correlation plots were created (Figure 4.5).

![Correlation plots](image)

**Figure 4.5** - Correlation between calculated (theoretical) and experimental (a) pl values of all spots and correlation between calculated (theoretical) and experimental (b) Mw values of all spots.

Comparative analysis of all molecular weights showed a stronger correlation ($r^2=0.902$) than pl correlation ($r^2=0.672$) of successfully identified proteins. Of all proteins identified, 90% showed a difference of less than 10 kDa between the experimental and the calculated.
4.4 Discussion

The study presented here demonstrated that statistically significant changes in protein abundance were detectable between the different species of *Leptospirillum* when placed under conditions of nitrogen deprivation. Although PMF identification of these proteins proved only partially successful, the 2D SDS PAGE analysis provided novel insight into the proteomic differences (and by extension genetic differences) presented under the conditions examined.

4.4.1 Changes in *L. ferrodiazotrophum* protein expression

A small segment of the *L. ferrodiazotrophum* proteome was identified using PMF, with nine proteins ascribed a specific function. Proteins of increased abundance matching *L. ferrodiazotrophum* included a 2.1 fold increase of Enolase (#5360), an enzyme crucial to glycolytic and gluconeogenic pathways, generally highly conserved amongst all organisms (Dutow et al. 2010). Not only is it essential for the basic catalytic functions of metabolism but in other bacterial species, such as *E. coli*, it is responsible for regulating mRNA stability in response to metabolic stress (Morita et al. 2004) as well as interacting with proteins involved with transcription and translation (Commichau et al. 2009). In addition to Enolase, increased abundance of a putative Reiske iron sulphur family protein (#3167), in particular a Reiske non-heme oxygenase was detected. It is responsible for the metabolism of aromatic compounds and resultant storage of electrons form a soluble electron transport chain (Ferraro, Gakhar, and Ramaswamy 2005). This is unusual as *Leptospirillum* species fix atmospheric carbon and are usually inhibited by the presence of organic carbon molecules. Golsman et al. in 2009 discovered a putative extradiol gene (LigB) and a carboxymethyl butenolidease enzyme, both of which are involved in aromatic compound metabolism in *Leptospirillum group II* and *L. ferrodiazotrophum*, indicating that manipulation of aromatic compounds by *Leptospirillum* species is possible.

The increase in abundance of enolase has also been demonstrated in other bacteria during periods of environmental stress (Sharma et al. 2006) and in prokaryotes capable of symbiotic nitrogen fixation such as *Bradyrhizobium japonicum* (Oehrle et al. 2008). Species of *B. japonicum* exhibit acetylene reduction with increased levels of enolase in extracts of soybean nodule cytosol during periods of nitrogen fixation. In Chapter 2, *L. ferrodiazotrophum* demonstrated positive acetylene reduction and high levels of Fe$^{2+}$ oxidation per cell in cultures devoid of soluble nitrogen. With transduction of these liberated electrons by an electron transport protein (Reiske iron sulphur), together with increased enolase production, the maintenance and support of nitrogen fixation by this *Leptospirillum* species appears highly probable.
Proteins identified in species other than *Leptospirillum* included a membrane oxidase complex imidazoleglycerol-phosphate dehydratase (#2764, *Leptospirillum sp Group II*), essential in microbial de novo synthesis of histidine, which is necessary for the co-ordination of the Rieske iron sulphur cluster (Cammack, Gay, and Shergill 1999). Detection and response of *L. ferrooxidans* to changes in the environment could be under the control of a two winged component signal transduction system (#2452, *Ferrimonas balearica*), also reliant on histidine for phosphorylation, via histidine kinase action.

Proteins that demonstrated decreased abundance and were homologous to *L. ferrooxidans* included a Nudix hydrolase (#2939), indicating a change in cellular metabolism but not isolating the exact metabolic function that was altered. The number of hydrolase genes varies between 0 to 30 in prokaryotes (McLennan 2006), and any number of cellular metabolic or homoeostatic processes may have been altered whilst under nitrogen stress. Other proteins identified include the Anthranilate synthase component 1 (#1348, *Leptospirillum. sp Group II*), an enzyme necessary in the biosynthesis of tryptophan utilizing ammonia and chorismate. With no soluble nitrogen available, it is unsurprising that the abundance of this protein had decreased.

### 4.4.2 Changes in *L. ferriphilum* protein expression

The majority of *L. ferriphilum* proteins identified were related to metabolism. Increased abundance of the large subunit of 3-isopropylmalate dehydratase (#1066) was recorded, necessary for the biosynthesis of leucine (Tamakoshi, Yamagishi, and Oshima 1998), as well as dihydroorotate dehydrogenase, an enzyme needed for the de novo biosynthesis of pyrimidine. A formate dehydrogenase accessory protein, utilized in either exporting mature formate dehydrogenase complexes to the periplasmic space, or maturation of the formate dehydrogenase complex (Mandrand-Berthelot et al. 1988) also increased in abundance under conditions of soluble nitrogen deficiency.

Transposases (spot #977) are responsible for DNA reorganization and are potentially deleterious to the host, therefore the lowered abundance of this enzyme by *L. ferriphilum* may have been required to keep control over the destructive multiplication of the mobile element (Horak and Kivisaar 1999) during periods of soluble nitrogen depletion. Methylation of proteins, small molecules or nucleic acids by SAM-dependent methyltransferase (#549) decreased under the tested conditions and resultantly, changes to protein trafficking, signal transduction, metabolism and biosynthesis may have been affected (Martin and McMillan 2002). Changes to metabolism would have also been brought about by the lowered level of an adolase protein (#1569), which is essential in catalysing the central steps of carbohydrate metabolism.
Although *L. ferriphilium* failed to reduce acetylene in previous experiments, increased abundance (5.2 fold) of dihydroorotate dehydrogenase, essential in the biosynthesis of pyrimidine (Rowland et al. 1998) was detected. In *Geobacter sulfurreducens*, expression of dihydroorotate dehydrogenase mRNA transcription increased 2.32 fold whilst fixing nitrogen (Methe et al. 2005), and the lowered levels of a fusulose adolase protein was also demonstrated in *Rhodobacter sphaeroides*, a known nitrogen fixing bacteria (Arai, Roh, and Kaplan 2008) under aerobic conditions. With the continued proliferation of *L. ferriphilium* under soluble nitrogen deficiency detailed in chapter 2, it is possible that in a reduced oxygen environment, nitrogen fixation proteins could be revealed.

### 4.4.3 Down regulated proteins of *L. ferrooxidans*

Only proteins that decreased in abundance were successfully uncovered in *L. ferrooxidans* cultures, two of which were involved in metabolism. Carbohydrate kinase family protein (#1262, *Leptospirillum* sp Group II) responsible for changes in either signalling or priming compounds for metabolism (Deutscher, Francke, and Postma 2006) and D-isomer specific 2-hydroxyacid dehydrogenase (#1120) utilized in the production of pyruvate (Taguchi and Ohta 1991). It is possible that the metabolic rate of *L. ferrooxidans* slowed due to deficiencies in soluble nitrogen, resulting in the decreased expression of these two enzymes. Environmental stresses imposed by depleted soluble nitrogen could have also triggered the decrease in LexA repressor protein (#1362) abundance, as it is responsible for negatively regulating SOS genes involved with DNA repair and mutagenesis under normal conditions. By reducing the LexA repressor protein, any damages accrued during nitrogen fixation could be repaired. Decreased expression of a putative membrane porin, OmpA (#704) could have affected the integrity of cell membrane, resulting in an aberrant exchange of molecules between the cell and the outside environment. Porin control by OmpA also plays an important role in the oxidation of Fe$^{2+}$ (Kamimura et al. 2011) and a decrease in porin activity may help to explain the slow growth and iron oxidation rates recorded under nitrogen deficient conditions (Chapter 2).

Identification of *L. ferrooxidans* specific peptides was made difficult due to only 26 proteins detected by Progenesis Same Spots as having altered in abundance, demonstrating a greater than 2 fold change. As the NCBInr database only has 309 *L. ferrooxidans* proteins recorded, of which 29 have an identified function with the remaining 280 being hypothetical proteins or open reading frames (ORF); the number of proteins detected by 2D gel analysis in this experiment meant that identification of *L. ferrooxidans* proteins was less than optimal and more reliant on homologous protein signatures from other microorganisms.
4.4.4 Significance of proteins identified in this study

The successful identification of 20 proteins revealed that, under conditions of soluble nitrogen deprivation, protein expression was altered. No specific proteins or pathways involved with nitrogen fixation were isolated from any of the three *Leptospirillum* studied. However, as carbon and nitrogen metabolisms are closely linked (Fisher and Sonenschein 1991), the regulation of metabolic pathways across all species changed, in response to the enforced nitrogen stress. Those proteins whose abundance increased include the *de novo* synthesis of nucleic and amino acids, glycolysis, signalling and energy transport (Figure 4.6a). The proteins identified with decreased abundance were heavily involved in cellular metabolism and membrane channel control (Figure 4.6b). It is possible that the proteins needed for nitrogen fixation, function and regulation were not highly expressed, had a high turnover rate or were not represented as abundant proteins, hence their lack of recognition in these experiments.

![Figure 4.6](image)

(a) Increased abundance

(b) Decreased abundance

*Figure 4.6* – Distribution of biological processes related to proteins whereby abundance (a) increased and (b) decreased in *Leptospirillum* sp. (>2 fold) under soluble nitrogen depletion.
Proteomic Analysis of *Leptospirillum*

No *Leptospirillum* species has had its entire genome sequenced, and as such protein identification by database screening is problematic. In these experiments, the different proteomic changes brought about by environmental stresses can only be attributed to nitrogen deprivation, and not as a result of nitrogen fixation.

### 4.4.5 Posttranslational modification of proteins

The discrepancies observed between the pl and Mw of proteins from the experimental to the theoretical (Figure 4.5) could be a result of posttranslational modifications of *Leptospirillum* proteins that occurred *in vivo*, as theoretical proteomic translations do not take these into account. Of all proteins identified (n=20), 55% had a greater experimental Mw than was theoretically derived. Phosphorylation, dephosphorylation or glycosylation are the likely modifiers responsible for producing these observed discrepancies, whereby a rapid change in the protein molecular weight occurred due to cross-linking of smaller proteins with larger ones. These posttranslational modifications would explain the pl shift away from the theoretical. Although 55% of proteins were larger than expected, only 41% of these (25% of total) were identified as belonging to the *Leptospirillum* genus. Factoring this in, protein identification based on homology with sequences from different species should be interpreted with care and as such, the experimental Mw and pl values obtained should be given more latitude when compared to theoretical values. Proteins of lower experimental molecular weight than theoretically expected (total of 25%) could be due to proteolysis (*in vivo* or during sample processing) or as products of alternative mRNA gene transcription. Artefactual modifications that may have occurred as a result of sample preparation and processing of proteins were marginally accounted for when searching the peptide database, such as carbamidomethylation of cysteines and variable oxidation of methionine, but other unknown alterations to protein structures that occurred may not be reflected when searching the database for homologous proteins within the *Leptospirillum* genus.

### 4.4.6 Limitations of 2D-SDS PAGE proteomic analysis

Silver staining was used as the stain of choice as it can detect proteins of low abundance with concentrations as small as 1 ng. Due to the slow growth rate of the *Leptospirillum* species (detailed in chapter 2), it was difficult to obtain large amounts of protein for 2D gel analysis. Despite the pooling of multiple gel pieces, the lack of sufficient quantities of protein present in the excised spots, resulted in the inability to obtain significant protein identifications for all selected spots. This is because at least five peptide masses need to be matched to the protein and 15% of the protein sequence needs to be covered for an unambiguous identification (Mann, Hendrickson, and Pandey 2001). Of the 46 proteins examined, 26 in total did not meet this standard, and were discarded from further analysis.
To correct for factors that result in experimental variation, normalisation was required in all proteomics experiments to calibrate the data between different sample runs. When multiple gels were run, there were a number of spots present on some gels, but absent from others. Meticulous examination of multiple gels was important to determine genuine differences between the samples. Relative protein volume variations can range from 16 to 39% for replicates that are from the same sample, and from 22 to 55% for replicated experimental samples (Molloy et al. 2003). Therefore, to be statistically significant and considering the semi-quantitative nature of silver staining, changes in protein volume must be at least two-fold different.

Of the all the proteins resolved and identified (n=20), the magnitude change in protein expression detected ranged from 2.0 to 5.8, with an average fold difference of 2.87 ± 0.23. In a cell, the concentrations of different proteins can span 5–6 orders of magnitude (Barrett, Brophy, and Hamilton 2005), and detection of these proteins with silver stain is only possible up to four orders of magnitude. For these experiments, detection by silver staining can be deemed reliable as 90% of the protein identified expressed less than four-fold magnitude change.

4.4.6.1 Analytical Method

Statistical analyses were carried out on the whole data sets to filter out interesting protein spots. In this study, two statistical approaches were applied for data analysis, based on univariate and multivariate methods. The multivariate method (PCA) applied by Progenesis Same Spots utilised all the proteome data simultaneously to uncover patterns in expression change, and in this instance was used just to support the univariate results (data not shown). The univariate method successfully detected changes in the expression of individual proteins, with the Student’s t-Test identifying proteins whose expression from conditions of soluble nitrogen to conditions without soluble nitrogen altered significantly. Next to the significance cut-off (5%) employed by the Student’s t-Test, a threshold in fold change (>2) of expression rate of the proteins in relation to control conditions was applied to create the final set of proteins selected for mass spectrometry analysis. The initial analysis performed by Progenesis Same Spots software isolated proteins that did not adhere to these criteria. For *L. ferrooxidans*, 46 proteins were identified as having a difference in expression between test and control subjects, but follow up calculations performed reduced this number to 27. In these cases, it was therefore important to manually evaluate all results before confirming a change in protein expression.
4.5 Conclusions

The combination of 2D-SDS PAGE and MALDI-TOF mass spectrometry analysis to compare changes in protein expression has allowed for the detailed, reproducible, and sensitive characterization of *Leptospirillum* species under soluble nitrogen deficiency. In this study it is clear that the changes in the proteomic profile of the *Leptospirillum* species were brought about due to the presence or absence of soluble nitrogen in the media. The changes in protein abundance detected were primarily associated with roles essential to metabolism, which in turn had a cascading effect, resulting in modifications to other pathways, including altering production of stress and signalling proteins.

This characterization of protein expression and diversity of *Leptospirillum* species has increased the understanding of how *Leptospirillum* species function in soluble nitrogen depleted environments and has shed light on the complex interactions between a variety of metabolic pathways. The discovery of proteins expressed and regulated in the presence and absence of nitrogen revealed that there is great variation within the genus, and understanding the physiological adaptations taking place may help understand the molecular bases of *Leptospirillum*, as a free living diazotroph, but further research is required to determine whether or not proteins involved in nitrogen fixation are involved.
Chapter 5

Identification and structural analysis of the \textit{nif} genes in \textit{Leptospirillum} species

Previous chapters have demonstrated both genomic and proteomic changes of the \textit{Leptospirillum} species activity in the absence of soluble nitrogen. This chapter examination of the nitrogenase enzyme structure at the genomic level was the focus of this chapter.

5.1 Introduction

\textit{Leptospirillum} species inhabit a variety of acidic environments where the concentrations of nitrogenous compounds can be extremely low. Thus, the ability to utilize atmospheric dinitrogen as a nitrogen source is an advantageous trait for \textit{Leptospirillum} species to possess.

Within diazotrophic microorganisms, irrespective of the environment, the nitrogenase enzyme shows a high degree of conservation in structure, function and amino acid sequence (Dean, Bolin, and Zheng 1993). The \textit{nifH} gene, encoding for component II of the nitrogenase enzyme is traditionally used to identify micro-organisms that may have the ability to fix atmospheric nitrogen. The \textit{nifH} gene, along with \textit{nifD} and \textit{nifK}, (encoding for component I of the nitrogenase) are often contiguous within a DNA region, organized as a single transcriptional unit (\textit{nifHDK}) (Choo, Samian, and Najimudin 2003). Strong conservation of these genes reflects the structural requirements of the nitrogenase proteins for catalytic functions (Schaechter and Lederberg 2004); however the degree of linkage and arrangement of the specific and associated \textit{nif} genes can vary considerably (Dean and Jacobson 1991).

The presence of these three genes within a prokaryotic genome is often a strong indicator of nitrogen fixation ability. Genome analysis of \textit{L. ferrooxidans} and \textit{L. ferrediazotrophum}
shows the presence of these genes (Parro and Moreno-Paz 2003; Tyson et al. 2005), although none have been identified in \textit{L. ferrilunum}. The presence of functional \textit{nifHDK} genes in \textit{L. ferrilunum} would support the theory that the entire \textit{Leptospirillum} genus is diazotrophic and capable of surviving periods of soluble nitrogen depletion.

This study sought to determine the presence of \textit{nifH}, \textit{nifD} and \textit{nifK} genes in \textit{L. ferrilunum}, as well as confirm the presence of the \textit{nifHKD} genes in \textit{L. ferrodiazotrophum C5} and \textit{L. ferrooxidans DSM 2705}.

### 5.2 Materials and Methods

#### 5.2.1 Bacterial strains and growth media

The bacterial strains of \textit{L. ferrodiazotrophum}, \textit{L. ferrilunum} and \textit{L. ferrooxidans} used in this study as well as the growth media are detailed in Table 2.1.

#### 5.2.2 Bacterial cell harvesting

Conical flasks (250 mL) containing 50 mL of bacterial suspension were inoculated at a concentration of \(1.5 \times 10^6\) cells mL\(^{-1}\) and incubated at 30°C with aeration. Cells were harvested after reaching stationary phase, pelleted by centrifugation at 20,000 rpm in a Beckman centrifuge (Avanti J-E) for 40 minutes. The pellets were washed with 2 mL of acidified water, pH 1.8, and transferred to a 2 mL tube. Tubes were centrifuged again at 15,300 rpm for 30 minutes at 4°C and the supernatant removed.

#### 5.2.3 DNA preparation by proteinase K lysis and phenol extraction

DNA was extracted as detailed in section 3.2.5 with the following exceptions:

1. Phenol:chloroform:isoamyl alcohol (25:24:1) (Sigma-Aldrich) was pH 6.7.
2. DNA was precipitated with 0.1 volume of 3 M sodium acetate (pH 5.2) and 1 volume of 4°C isopropanol.
3. Tubes were placed at -80°C for 2 hours to precipitate DNA.
4. DNA was stored at -20°C in 5 µL aliquots.
5.2.4 PCR and sequencing primers

Four genes were targeted for PCR amplification and DNA sequencing. These were the 16S rRNA gene, *nifH*, *nifD*, and *nifK*. Universal primers were used for 16S rRNA (Lane, 1991). Species specific primers were designed using Primer 3 (Rozen and Skaletsky 2000) for *nifH*, *nifD* and *nifK* based on sequences in the GenBank public database for *L. ferrooxidans* and *L. ferrodiazotrophum*. Degenerate primers for *nifH* and *nifD* for *L. ferriphilum* were taken from previous studies and modified based on their identity to *L. ferrooxidans* and *L. ferrodiazotrophum* sequences. Primers for *nifK* in *L. ferriphilum* were designed based on sequence homology between the *nifK* gene of *L. ferrooxidans* and *L. ferrodiazotrophum*. Primers to cover intergenic spacer regions between *nifH* and *nifD*, and *nifD* and *nifK* for *L. ferriphilum* were designed manually from sequence data accrued in this study. All oligonucleotide primers were manufactured by GeneWorks (Australia). Dry lyophilized nucleotide pellets were reconstituted to a stock concentration of 100 μM, and further diluted to a working concentration of 20 μM. All primer solutions were stored at -20 °C.

5.2.5 Polymerase chain reaction (PCR) amplification

PCR amplification was conducted on an Alpha Metrix G-Storm GS1 thermal cycler using a 25 μL reaction volume. The PCR conditions for each species and gene are listed in Table 5.1. Each PCR mixture (25 μL) contained 2.5 μL 10X PCR buffer (Promega), 2.5 mM MgCl₂, 0.2 mM dNTPs, 0.4 μM of each primer, 1U Taq DNA polymerase (Promega), 2% DMSO, 100 ng DNA template and ddH₂O. Negative controls were run by substituting PCR grade water for template DNA.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>N° repeats</th>
<th>Temp (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
<td>1x</td>
<td>95</td>
<td>5min</td>
</tr>
<tr>
<td>Cycle 2</td>
<td>30x</td>
<td>95</td>
<td>30s</td>
</tr>
<tr>
<td>Step 1</td>
<td>55</td>
<td>30s</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>72</td>
<td>90s</td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td>1x</td>
<td>72</td>
<td>10min</td>
</tr>
<tr>
<td>Cycle 4</td>
<td>1x</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.1 – PCR cycle protocols for all *Leptospirillum* 16S rRNA

<table>
<thead>
<tr>
<th>Protocol</th>
<th>N° repeats</th>
<th>Temp (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
<td>1x</td>
<td>95</td>
<td>5min</td>
</tr>
<tr>
<td>Cycle 2</td>
<td>30x</td>
<td>95</td>
<td>30s</td>
</tr>
<tr>
<td>Step 1</td>
<td>56</td>
<td>30s</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>72</td>
<td>45s</td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td>1x</td>
<td>72</td>
<td>10min</td>
</tr>
<tr>
<td>Cycle 4</td>
<td>1x</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.2 – PCR cycle protocols for all *Leptospirillum nifH* amplification
PCR cycle conditions for \textit{nifD} and \textit{nifK} are the same as for \textit{nifH} but with an increased extension time of 90 seconds. PCR products were visualised as described in results section 3.1

5.2.6 PCR clean up and sequencing preparation

PCR products were column purified using the Qiagen PCR clean up kit following the manufacturer's instructions. Purified PCR products were eluted in 30 µL of TE buffer, pH 8.0 and stored at -20°C. The concentration of PCR products was determined by Nanodrop analysis, and diluted to a working concentration between 10-40 ng/µL for sequencing reactions. Each 10 µL sequencing reaction contained 5.75 µL ddH$_2$O, 0.5 µL of BigDye® terminator and 1.75 µL of 5x sequence buffer, both from Applied Biosystems, with 1 µL 20 mM forward or reverse primer, and 1 µL of PCR template. Sequencing PCR's were performed on a Alpha Metrix G-Storm GS1 thermal cycler with the cycle conditions following the manufacturer's instructions.
## Table 5.3 – Primers used in this study

<table>
<thead>
<tr>
<th>Species</th>
<th>Target</th>
<th>Primer names</th>
<th>Forward (5’-3’)</th>
<th>Reverse</th>
<th>Product size (bp)</th>
<th>Reference</th>
<th>Gen Bank accession (sequences used for primer design)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L. ferrodiazotrophum</strong></td>
<td>nifH</td>
<td>FdnifH F and FdnifH R</td>
<td>GCTGAAGCCGAATGGAGGATT</td>
<td>CGTCGCTTGATCCACCTTCCATT</td>
<td>~880</td>
<td>This study</td>
<td>CH003548</td>
</tr>
<tr>
<td></td>
<td>nifD</td>
<td>FdnifD F and FdnifD R</td>
<td>GAGGGTTCACATCATGTCGTA</td>
<td>TCTGAGCCGGGAAAAGACAAAT</td>
<td>~1460</td>
<td>This study</td>
<td>CH003548</td>
</tr>
<tr>
<td></td>
<td>nifK</td>
<td>FdnifK F and FdnifK R</td>
<td>CCGAATAGCGCCAGAACCACAGG</td>
<td>ACATCATCGCTAGCCACTCT</td>
<td>~1560</td>
<td>This study</td>
<td>CH003548</td>
</tr>
<tr>
<td></td>
<td>nifH (Set 1)</td>
<td>Za-nifH F and Za-nifH R</td>
<td>TGGGAYCCNAARGCNGA</td>
<td>ANDGCCATATCAYTCNCC</td>
<td>~460</td>
<td>Zani et al. 2000</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>nifH (Set 2)</td>
<td>MS-nifH F and MSnifH R</td>
<td>GCIWITAYGGNAARGGGNGG</td>
<td>TCIGGIGARATGATGGC</td>
<td>~470</td>
<td>Marusina et al. 2001</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>nifH (Set 3)</td>
<td>FpnifH F and FpnifH R</td>
<td>TAYGNNARGGNGGSHATGGY</td>
<td>ATRTRTTNGCNCGRTAVBBGCG</td>
<td>~850</td>
<td>This study</td>
<td>CH003548, AY204398</td>
</tr>
<tr>
<td></td>
<td>nifD</td>
<td>FpnifD F and FpnifD R</td>
<td>CGYCATTCCGACGGAGGT</td>
<td>TCCANGARTGCATCTGDSGA</td>
<td>~1200</td>
<td>Dedys et al. 2004 (Modified)</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>nifK</td>
<td>FpnifK F and FpnifK R</td>
<td>GGGAGATACCGACAATGGAC</td>
<td>GGAACATGGTRTTCTTGTGCT</td>
<td>~1500</td>
<td>This study</td>
<td>CH003548, AY204397</td>
</tr>
<tr>
<td></td>
<td>nifH-D IGS</td>
<td>Fp-HFor and Fp-Drev</td>
<td>CGACCCCGATCGACATATGGA</td>
<td>ATTCTCGGGTTAATCCTGGAG</td>
<td>~250</td>
<td>This study</td>
<td>JN390678</td>
</tr>
<tr>
<td></td>
<td>nifD-K IGS</td>
<td>Fp-Dfor and Fp-KRev</td>
<td>TCTGAGTTGCTCCAGG</td>
<td>GTGYTTGGGRTTCTGCTGCT</td>
<td>~130</td>
<td>This study</td>
<td>JN390678</td>
</tr>
<tr>
<td><strong>L. ferriphilum</strong></td>
<td>nifH</td>
<td>FdnifH F and FdnifH R</td>
<td>CCACTGAGCGAGATGGGAAAGAAA</td>
<td>CCATCGCCTAGTGCTGCCCTTCTTCT</td>
<td>~850</td>
<td>This study</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td>nifD</td>
<td>FdnifD F and FdnifD R</td>
<td>CAGGAGGATTTCATGACGAGA</td>
<td>CACCTTTGAGGCGGCGCGAGAAAA</td>
<td>~1430</td>
<td>This study</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td>nifK</td>
<td>FdnifK F and Fx nifK R</td>
<td>GGAACACAGACAATGGGCCAGAAT</td>
<td>AAACCCCGGAGAGGATTTCTGCTC</td>
<td>~1500</td>
<td>This study</td>
<td>AY204397/AY204398</td>
</tr>
<tr>
<td><strong>L. ferrooxidans</strong></td>
<td>nifH</td>
<td>FxnifH F and FxnifH R</td>
<td>CCACTGAGCGAGATGGGAAAGAAA</td>
<td>CCATCGCCTAGTGCTGCCCTTCTTCT</td>
<td>~850</td>
<td>This study</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td>nifD</td>
<td>FxnifD F and FxnifD R</td>
<td>CAGGAGGATTTCATGACGAGA</td>
<td>CACCTTTGAGGCGGCGCGAGAAAA</td>
<td>~1430</td>
<td>This study</td>
<td>AY204398</td>
</tr>
<tr>
<td></td>
<td>nifK</td>
<td>FxnifK F and Fx nifK R</td>
<td>GGAACACAGACAATGGGCCAGAAT</td>
<td>AAACCCCGGAGAGGATTTCTGCTC</td>
<td>~1500</td>
<td>This study</td>
<td>AY204397/AY204398</td>
</tr>
<tr>
<td><strong>All species</strong></td>
<td>16S rRNA</td>
<td>Uni-27F and Uni 1492R</td>
<td>AGAGTTGATCMTGGCTCAG</td>
<td>ACCTTGTTACGACTT</td>
<td>~1479</td>
<td>Lane, 1991</td>
<td>/</td>
</tr>
<tr>
<td></td>
<td>16S rRNA</td>
<td>800F and 920R</td>
<td>GTAGTCCAGGCGGCTAAGCA</td>
<td>CCCGCTCAATCTCCTTGA</td>
<td></td>
<td>Yanagi, M et al. 1993</td>
<td>/</td>
</tr>
</tbody>
</table>

Abbreviations used: Y = T, C; R = A, G; M = A, C; W = A, T; S = G, C; B = T, G, C; D = A, T, G; H = A, T, C; N = G, A, T, C; I = Inosine.
5.2.7 Sequence analysis

Raw sequencing data and quality was assessed using 4Peaks v1.7. Manual modifications to sequences were made after visual analysis of ABI graphical output if base pair conflicts were detected. Corrected files were imported to Geneious Pro 3.7.0 (Drummond et al. 2010) and consensus sequences of *Leptospirillum* gene fragments from forward and reverse priming generated to ensure entire gene coverage. Nucleotide sequences were converted to peptide strings using the genetic universal code. Sequences identities (nucleotide and protein) were confirmed by comparison to known sequences in the EMBL and NCBI databases, and similarity percentages checked using BLAST (Basic Local Alignment Search Tool) software (Altschul et al. 1990).

5.2.8 Sequence alignments

Within MEGA 5.05 (Tamura et al. 2011), multiple sequence alignments for each gene were generated by CLUSTAL W (Thompson, Higgins, and Gibson 1994). Translation products were aligned using the PAM amino acid substitution matrix (Dayhoff, Schwartz, and Orcutt 1978).

5.2.9 Codon usage and determination

Using EMBOSS 6.4.0, codon usage, GC content and 3rd codon GC content were determined using the following packages: Cusp, Chips and Wobble (Rice, Longden, and Bleasby 2000).

5.2.10 Theoretical protein modelling

Conserved amino acid domains were identified and located using CDD (Marchker-Bauer 2011) and 3D protein models viewed with Cn3D v4.3 software. Amino acid locations highlighted on protein models were from the resolved crystal structure of *A. vinelandii* *nifH* (PDB: 1NIP_A), *nifD* (PDB: 2AFK_A) and *nifK* (PDB: 2AFK_B).
5.3 Results

5.3.1 \textit{nifHDK} genes in \textit{L. ferrodiazotrophum} and \textit{L. ferrooxidans}

To determine whether \textit{L. ferrodiazotrophum} C5 and \textit{L. ferrooxidans} DSM 2705 contained the complete \textit{nifHDK} genes as previously reported from similar strains (\textit{L. ferrodiazotrophum} UBA1 and \textit{L. ferrooxidans} 3.2), PCRs with species specific primers for these gene regions were conducted. The results of these PCRs can be seen in Figure 5.1. All specific primer sets successfully amplified a product of the correct size. No other contaminating products were detected in any reaction. The \textit{nifA} gene was also amplified in \textit{L. ferrooxidans}, but no further analysis done.

![Figure 5.1](image.png)

\textbf{Figure 5.1} - PCR products of \textit{nifHDK} (a) \textit{L. ferrodiazotrophum} C5 and (b) \textit{L. ferrooxidans} DSM 2705 run on a 1% agarose TBE gel. Each primer set has two replicates. L = Hyperladder II (Bioline). The expected product size in both species for \textit{nifH} was \textasciitilde985bp, \textit{nifD}: \textasciitilde1464bp, \textit{nifK}: \textasciitilde1560bp and \textit{nifA}: \textasciitilde1600bp.

Sequence analysis of amplified gene regions and BLAST search of the nucleotide and translated theoretical proteins in the NCBI (nr) databases returned high sequence identity results for \textit{L. ferrooxidans} and \textit{L. ferrodiazotrophum} (Table 5.4). PCR products from \textit{L. ferrooxidans} DSM 2705 demonstrated sequence identity of 99\% to \textit{L. ferrooxidans} sp 3.2. BLASTn analysis of \textit{L. ferrodiazotrophum} sequences demonstrated high sequence identity to the precise genes of interest - \textit{nifH}, D and K but were unsuccessful in matching them to the correct species. BLASTp searches with translated theoretical proteins returned results with greater sequence identity to both the correct species and target gene for \textit{L. ferrooxidans} and \textit{L. ferrodiazotrophum} than nucleotide sequence search results.
Table 5.4 – BLASTn results (a) for nucleotide sequences and BLASTp results (b) for amino acid sequences producing significant alignments in *L. ferrooxidans* and *L. ferrodiazotrophum*. Genbank Accession results with the highest homology are displayed.

(a) BLASTn

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query coverage</th>
<th>E value</th>
<th>Max identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrooxidans</td>
<td>nifH</td>
<td>Leptospirillum ferrooxidans nifHDK operon, partial sequence</td>
<td>1615</td>
<td>1615</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AV204398.1</td>
<td>Leptospirillum ferrooxidans NifH (nifH) and NifD (nifD) genes, complete cds; and NifK (nifK) gene, partial cds</td>
<td>1615</td>
<td>1615</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td>nifD</td>
<td>Leptospirillum ferrooxidans NifH (nifH) and NifD (nifD) genes, complete cds; and NifK (nifK) gene, partial cds</td>
<td>2643</td>
<td>2643</td>
<td>100%</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>AV204398.1</td>
<td>Leptospirillum ferrooxidans nifHDK operon, partial sequence</td>
<td>2639</td>
<td>2639</td>
<td>99%</td>
<td>0</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td>nifK</td>
<td>Leptospirillum ferrooxidans NifH (nifH) and NifD (nifD) genes, complete cds; and NifK (nifK) gene, partial cds</td>
<td>558</td>
<td>558</td>
<td>100%</td>
<td>6.00E-156</td>
</tr>
</tbody>
</table>

(b) BLASTp

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query coverage</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrooxidans</td>
<td>NifH</td>
<td>Nitrogenase iron protein [Leptospirillum ferrooxidans]</td>
<td>593</td>
<td>593</td>
<td>99%</td>
</tr>
<tr>
<td></td>
<td>EES53484.1</td>
<td>Nitrogenase iron protein [NifH] [Leptospirillum ferrodiazotrophum]</td>
<td>574</td>
<td>574</td>
<td>99%</td>
</tr>
<tr>
<td></td>
<td>AAO38343.1</td>
<td>NifH [Leptospirillum ferrooxidans]</td>
<td>540</td>
<td>540</td>
<td>90%</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td>NifD</td>
<td>Nitrogenase molybdenuin-iron protein alpha chain component I [Leptospirillum ferrooxidans]</td>
<td>987</td>
<td>987</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>EES53485.1</td>
<td>Nitrogenase, molybdenuin-iron protein alpha chain (NifD) [Leptospirillum ferrodiazotrophum]</td>
<td>947</td>
<td>947</td>
<td>100%</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td>NifK</td>
<td>Nitrogenase, molybdenuin-iron protein beta chain (NifK) [Leptospirillum ferrodiazotrophum]</td>
<td>890</td>
<td>890</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>YP_001523955.1</td>
<td>Nitrogenase molybdenuin-iron protein beta chain [Azorhizobium caulinodans ORS 571] &gt;dbj</td>
<td>BAF87037.1</td>
<td></td>
<td>749</td>
</tr>
<tr>
<td></td>
<td>YP_001207332.1</td>
<td>Nitrogenase molybdenuin-iron protein beta chain, nifK [Bradyrhizobium sp. ORS278] &gt;emb</td>
<td>CAL79113.1</td>
<td></td>
<td>748</td>
</tr>
</tbody>
</table>

109
5.3.2 Identification of \textit{nifH}, \textit{nifD} and \textit{nifK} in \textit{L. ferrphilum DSM 14647}.

To assess whether \textit{L. ferrphilum} had a copy of the \textit{nifH} gene, 4 primer sets were tested on genomic DNA. The first set of degenerate \textit{nifH} primers (Za\text{-}nifH F and Za\text{-}nifH R) were developed by Zani et al. (2000). Adhering to PCR reagent conditions and cycle parameters from the study, no \textit{nifH} gene product was amplified. Altering Mg$^{2+}$ concentrations, denaturation temperatures or extension times failed to produce a PCR product of ~800bp in size as was expected.

A second set of degenerate primers (MS\text{-}nifH F and MSnifH R) from the study of Marusina et al. (2001) also failed to yield a \textit{nifH} product of approximately ~470 bp in size.

The third degenerate primer set for \textit{nifH} was developed for this study (FpnifH F and FpnifH R) based on the \textit{nifH} sequences of \textit{L. ferrooxidans} and \textit{L. ferrodiazotrophum}. No product of 800bp was identified at an annealing temperature of 56 °C which successfully amplified the \textit{nifH} gene in \textit{L. ferrodiazotrophum} and \textit{L. ferrooxidans}. Adjusting the temperature did not result in a product of the correct size as can be seen in Figure 5.2. Primer set 3 resulted in non-specific amplification.

![Figure 5.2](image)

\textit{Figure 5.2} - Agarose gel of a gradient PCR of possible \textit{nifH} products in \textit{L. ferrphilum} using the third primer set for \textit{nifH}. Multiple PCR products were detected in temperatures ranging from 54-58 °C. No amplification of any product was detected at the top end of the gradient at 59 °C. No evidence of primer dimer is apparent.

A fourth set of degenerate primers (LG\text{-}nifHF and R) developed by Laguerre et al. (2001) for \textit{nifH} identification, successfully amplified a single PCR product of approximately 800bp (expected 780 bp) from \textit{L. ferrphilum} genomic DNA (Figure 5.3).
Modified degenerate primers for *nifD* from Dedysh, Ricke, and Liesack (2004) were successful in amplifying a single PCR product at an annealing temperature of 56°C. Primers constructed for this study to amplify *nifK* were designed based on sequence identity of the *L. ferrooxidans* and *L. ferrooxidans* *nifK* sequences. These primers successfully amplified a gene fragment of the expected size (Figure 5.4).

Sequenced PCR product analysis and BLASTn/BLASTp searches in the NCBI (nr) databases indicated high sequence identity to known bacterial and archaeal *nifH, nifD* and *nifK* sequences (Table 5.5) with low E values.
Table 5.5 – BLASTn results (a) for nucleotide sequences and BLASTp results (b) for amino acid sequences producing significant alignments in *L. ferrphilum*. Genbank Accession results with the greatest sequence identity to the *nif* genes are displayed.

(a) BLAST(n)

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query coverage</th>
<th>E value</th>
<th>Max ident</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrphilum</td>
<td>nifH</td>
<td>Leptospirillum ferrooxidans nifHDK operon, partial sequence</td>
<td>846</td>
<td>846</td>
<td>97%</td>
<td>0</td>
</tr>
<tr>
<td>AF547999.1</td>
<td>Leptospirillum ferrooxidans NifH (nifH) and NifD (nifD) genes, complete cds; and NifK (nifK) gene, partial cds</td>
<td>846</td>
<td>846</td>
<td>97%</td>
<td>0</td>
<td>84%</td>
</tr>
<tr>
<td>AV204398.1</td>
<td>Leptospirillum ferrooxidans nifHDK operon, partial sequence</td>
<td>1402</td>
<td>1402</td>
<td>99%</td>
<td>0</td>
<td>84%</td>
</tr>
<tr>
<td>AV204398.1</td>
<td>Leptospirillum ferrooxidans NifH (nifH) and NifD (nifD) genes, complete cds; and NifK (nifK) gene, partial cds</td>
<td>1402</td>
<td>1402</td>
<td>99%</td>
<td>0</td>
<td>84%</td>
</tr>
<tr>
<td>L. ferriphilum</td>
<td>nifD</td>
<td>Leptospirillum ferrooxidans NifH (nifH) and NifD (nifD) genes, complete cds; and NifK (nifK) gene, partial cds</td>
<td>758</td>
<td>758</td>
<td>99%</td>
<td>0</td>
</tr>
<tr>
<td>AV204397.1</td>
<td>Leptospirillum ferrooxidans NifE (nifE) gene, complete cds; and NifK gene, partial cds</td>
<td>516</td>
<td>516</td>
<td>85%</td>
<td>2.00E-142</td>
<td>74%</td>
</tr>
</tbody>
</table>

(b) BLAST(p)

<table>
<thead>
<tr>
<th>Accession</th>
<th>Description</th>
<th>Max score</th>
<th>Total score</th>
<th>Query coverage</th>
<th>E value</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrphilum</td>
<td>nifH</td>
<td>Nitrogenase iron protein [Leptospirillum ferrooxidans]</td>
<td>534</td>
<td>534</td>
<td>99%</td>
</tr>
<tr>
<td>AAO12257.1</td>
<td>Nitrogenase iron protein [NifH] [Leptospirillum ferrodiazotrophum]</td>
<td>533</td>
<td>533</td>
<td>99%</td>
<td>0</td>
</tr>
<tr>
<td>EES53484.1</td>
<td>Nitrogenase, molybdenum-iron protein alpha chain (NifD) [Leptospirillum ferrodiazotrophum]</td>
<td>942</td>
<td>942</td>
<td>99%</td>
<td>0</td>
</tr>
<tr>
<td>EES53485.1</td>
<td>Nitrogenase, molybdenum-iron protein alpha chain (NifD) [Leptospirillum ferrodiazotrophum]</td>
<td>942</td>
<td>942</td>
<td>99%</td>
<td>0</td>
</tr>
<tr>
<td>AAO38344.1</td>
<td>Nitrogenase, molybdenum-iron protein alpha chain (NifD) [Leptospirillum ferrodiazotrophum]</td>
<td>984</td>
<td>984</td>
<td>99%</td>
<td>0</td>
</tr>
<tr>
<td>EES53486.1</td>
<td>Nitrogenase, molybdenum-iron protein beta chain (NifK) [Leptospirillum ferrodiazotrophum]</td>
<td>838</td>
<td>838</td>
<td>99%</td>
<td>0</td>
</tr>
<tr>
<td>YP_001523955.1</td>
<td>Nitrogenase, molybdenum-iron protein beta chain (Azorhizobium caulinodans ORS 571)</td>
<td>838</td>
<td>838</td>
<td>99%</td>
<td>0</td>
</tr>
</tbody>
</table>

Based on the continuous fashion in which the *nifH, D* and *K* genes are arranged in *L. ferrodiazotrophum* and *L. ferrooxidans*, a primer walking strategy with specific primers (Table 5.3) was designed on the flanking regions of *nifH, D* and *K* amplified the missing intergenic spacer regions, with the *nifH-nifD* intergenic region spanning 175 bp and *nifD-nifK* intergenic region, 72 bp in length.

5.3.3 Structures associated with the *nifHDK* operon in *Leptospirillum*

Fundamental characteristics of the *nifH, nifD* and *nifK* *Leptospirillum* sequences are in Table 5.6, demonstrating conserved sequence length between the species. GC% content for all genes was slightly above 50%; however at the third codon, GC% composition increased for all genes, in all species. To determine the presence of synonymous codon usage bias, the effective number of codons used in a gene (Nc) was calculated (with a maximum value
of 61 = no codon bias, and 20 = extreme bias) (Fuglsang 2006).

**Table 5.6** – Sequence characteristics of the *Leptospirillum nifH, D and K* genes, and theoretical molecular weight of translated amino acids.

<table>
<thead>
<tr>
<th>Species</th>
<th>NifH</th>
<th>DNA length (bp)</th>
<th>MW(Da)</th>
<th>Overall %GC content</th>
<th>%GC 3rd codon</th>
<th>Nc</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. ferrodiazotrophum</em></td>
<td></td>
<td>876</td>
<td>31358</td>
<td>59</td>
<td>78</td>
<td>39.36</td>
</tr>
<tr>
<td><em>L. ferrphilum</em></td>
<td></td>
<td>876</td>
<td>31503</td>
<td>59.3</td>
<td>76</td>
<td>35.688</td>
</tr>
<tr>
<td><em>L. ferrooxidans</em></td>
<td></td>
<td>876</td>
<td>31533</td>
<td>55</td>
<td>67</td>
<td>42.514</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>NifD</th>
<th>DNA length (bp)</th>
<th>MW(Da)</th>
<th>Overall %GC content</th>
<th>%GC 3rd codon</th>
<th>Nc</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. ferrodiazotrophum</em></td>
<td></td>
<td>1431</td>
<td>54026</td>
<td>55.8</td>
<td>78</td>
<td>40.225</td>
</tr>
<tr>
<td><em>L. ferrphilum</em></td>
<td></td>
<td>1431</td>
<td>54102</td>
<td>57.5</td>
<td>82</td>
<td>37.143</td>
</tr>
<tr>
<td><em>L. ferrooxidans</em></td>
<td></td>
<td>1431</td>
<td>53884</td>
<td>53.7</td>
<td>74</td>
<td>41.269</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>NifK</th>
<th>DNA length (bp)</th>
<th>MW(Da)</th>
<th>Overall %GC content</th>
<th>%GC 3rd codon</th>
<th>Nc</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>L. ferrodiazotrophum</em></td>
<td></td>
<td>1539</td>
<td>57581</td>
<td>57.5</td>
<td>78</td>
<td>39.197</td>
</tr>
<tr>
<td><em>L. ferrphilum</em></td>
<td></td>
<td>1539</td>
<td>57692</td>
<td>58.3</td>
<td>79</td>
<td>39.903</td>
</tr>
<tr>
<td><em>L. ferrooxidans</em></td>
<td></td>
<td>1539</td>
<td>57647</td>
<td>53.8</td>
<td>70</td>
<td>45.657</td>
</tr>
</tbody>
</table>

Further analysis of codon usage for each of the genes is illustrated in Figure 5.5, with codon usage frequencies (as calculated by the Cusp package) remarkably similar for each of the *nif* genes analysed. The following codons – UUA (L), CUA (L), AUA (I), GUA (V), AGU (S) and UGG (W) were absent from the *nifH* gene for all species. The lack of UGG is notable, as tryptophan is only encoded for by one codon. Showing similar codon bias (Nc) in all species for *nifD* as was seen in *nifH*, a large number of codons were absent, however only two, UUA and GUA were common to all three species. Analogous to *nifH* and *nifD*, *nifK* was also deficient in UUA and GUA, as well as GUA codons (Complete codon usage break down for each gene is in Appendix A, Table A.3, A.4 and A.5).
nif gene structural analysis
**nif** gene structural analysis

**Figure 5.5** – Frequency of codon usage in *nifH*, *nifD* and *nifK* genes of *Leptospirillum* (a) *ferrodiazotrophum*, (b) *ferrophilum* and (c) *ferrooxidans*

For the initiation of translation, a Shine-Dalgarno sequence (AGGAGG) was found immediately upstream of the *nifH* gene in *L. ferrodiazotrophum* and *L. ferrooxidans* -14 and -13bp from the initiation codon respectively (Table 5.7). No sequence data upstream of the initiation codon was available for *L. ferrophilum*. In all *Leptospirillum* species, putative ribosomal binding sites (RBS) were found upstream of the start codons for both *nifD* and *nifK* genes.

**Table 5.7** – Shine-Delgramno/RBS sequences found prior to *nifH*, *nifD* and *nifK*. RBS are highlighted in bold. Initiation codon is underlined.

<table>
<thead>
<tr>
<th>Loci</th>
<th>Species</th>
<th>Shine-Delgramno/ RBS</th>
</tr>
</thead>
<tbody>
<tr>
<td>nifH</td>
<td><em>L. ferrodiazotrophum</em></td>
<td>CAGGAGGA GAACACTATG</td>
</tr>
<tr>
<td></td>
<td><em>L. ferrophilum</em></td>
<td>unknown</td>
</tr>
<tr>
<td></td>
<td><em>L. ferrooxidans</em></td>
<td>CAGGAGGA AAACATTATG</td>
</tr>
<tr>
<td>nifD</td>
<td><em>L. ferrodiazotrophum</em></td>
<td>CGGAGG GTTCCATCATG</td>
</tr>
<tr>
<td></td>
<td><em>L. ferrophilum</em></td>
<td>CGGAGG GTTCCATATG</td>
</tr>
<tr>
<td></td>
<td><em>L. ferrooxidans</em></td>
<td>CGGAGG GTTCCATAG</td>
</tr>
<tr>
<td>nifK</td>
<td><em>L. ferrodiazotrophum</em></td>
<td>AGAGGATACGCAAACTG</td>
</tr>
<tr>
<td></td>
<td><em>L. ferrophilum</em></td>
<td>GGGAGA TACCGACAAATG</td>
</tr>
<tr>
<td></td>
<td><em>L. ferrooxidans</em></td>
<td>GCAGGA ACAGACAAATG</td>
</tr>
</tbody>
</table>

Within the *nifH-nifD* intergenic regions of all *Leptospirillum* species, an inverted sequence string was found (Figure 5.6). Even though there were hypothetical RBS, no -24 to -12 pro-
motermotor sequences were found in nifH-nifD intergenic spacer region for any Leptospirillum species. Modelling of the DNA sequence into the most stable secondary structure depicts a theoretical transcription terminator. No direct or inverted sequence repeats were found in the nifD-nifK intergenic region of any Leptospirillum species, nor was there a -24 to -12 promoter motif.

![Diagram of secondary structure](image)

(a) *L. ferrodiazotrophum*  
(b) *L. ferrihilum*  
(c) *L. ferrooxidans*

**Figure 5.6** – Inverted sequence string (theoretical transcription terminators) found in the nifH-nifD intergenic spacer region of *Leptospirillum* spp. Shown as DNA sequences folded into the most stable secondary structure based on thermodynamic values (ΔG) by mfold (Zuker 2003).

### 5.3.4 Identifying conserved sequence regions in *Leptospirillum* nifHDK genes

Analysis of aligned *Leptospirillum* sequences is in Table 5.8. Nucleotide sequence identity was greatest with the 16S rRNA and showed minimal variation between the three species examined, revealing the absence of any insertions or deletions.

**Table 5.8** – Percentage identity between aligned nucleotide sequences of *Leptospirillum* species. One housekeeping gene, 16S rRNA, and three nitrogenase encoding genes; nifH, nifD and nifK.

<table>
<thead>
<tr>
<th>Loci</th>
<th>N° aligned sites (bp)</th>
<th>N° conserved sites</th>
<th>N° variable sites</th>
<th>% Sequence identity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA AA</td>
<td>DNA AA</td>
<td>DNA AA</td>
<td>DNA AA</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>1349 -</td>
<td>1221 -</td>
<td>172 -</td>
<td>90.51 -</td>
</tr>
<tr>
<td>NifH</td>
<td>876 292</td>
<td>678 258</td>
<td>198 16</td>
<td>77.40  88.36</td>
</tr>
<tr>
<td>NifD</td>
<td>1431 477</td>
<td>1106 439</td>
<td>325 36</td>
<td>77.29  92.03</td>
</tr>
<tr>
<td>NifK</td>
<td>1539 512</td>
<td>1113 418</td>
<td>426 94</td>
<td>72.32  81.64</td>
</tr>
</tbody>
</table>

Sequence identity of the selected nif genes and theoretical translated amino acids was generally high amongst the three species. Conserved nucleotide sequence identity of the
**nif** gene structural analysis

*nifH* gene was 77.4%, but increased to 88.3% when in a peptide string. The *nifD* gene was similarly conserved as *nifH* at the nucleotide level (77.3%), but as a protein (NifD) showed higher sequence identity than NifH (92.03%). Both as a nucleotide and peptide string, sequence identity of the *nifK* gene (72.32%) and subsequent NifK protein (81.64%), showed the lowest amount of sequence identity between species of the three *nif* genes analysed.

Deduced amino acid sequences of *Leptospirillum* species NifH, NifD and NifK were aligned against the functional Nif peptides of two other known diazotrophs, *Azotobacter vinelandii* and *Acidithiobacillus ferrooxidans* for further analysis of their structural composition.

### 5.3.4.1 Analysis of *nifH* characteristics

Conserved characters in the aligned peptide sequences of *Leptospirillum* NifH with *At. ferrooxidans* and *A. vinelandii* equalled 69.5% compared to the corresponding *nifH* nucleotide sequence (42.9%). The multiple sequence alignment of NifH (Figure 5.7) identified conserved and functional regions in all *Leptospirillum* species, including a Walker A motif (A), nucleotide switch one (1) and switch (2) regions, as well as nucleotide and ion binding sites.

### 5.3.4.2 Analysis of *nifD* characteristics

Conserved characters in the aligned peptide sequences of *Leptospirillum* NifD with *At. ferrooxidans* and *A. vinelandii* equalled 68.3% compared to the corresponding *nifD* nucleotide sequence (56.9%). The multiple sequence alignment of NifD (Figure 5.8) identified conserved and functional regions in all *Leptospirillum* species notably, a twelve-residue sequence surrounding Cys-288 (LNVLHCYRSMNY) and a nine-residue sequence centred on His-455 (FRQMHSWDY).

### 5.3.4.3 Analysis of *nifK* characteristics

Multiple sequence alignment of the NifK proteins from *Leptospirillum* are shown in Figure 5.9 detailing conserved and functional regions on the β component of the nitrogenase protein. All *Leptospirillum* species NifK sequences contained 4 cysteine residues, of which three are strongly conserved; Cys 69, 94 and 152 (Figure 5.13) interacting with the Fe$_8$S$_7$ heterogen. The cysteine residues were flanked by conserved residues on both the N and C terminal sides. Conservation of amino acid homology was greatest in the N terminal third of the sequence alignment. Numerous α/β subunit interaction sites and MoFe protein β subunit Fe protein contacts were also noted.
Figure 5.7 – Clustal W alignment of deduced amino acid sequences of the nifH proteins from *Leptospirillum* species with known diazotrophic NifH sequences. Sequence alignment showing conserved sequence signatures in this protein commonly shared by closely related diazotrophs. Amino acids are designated with one-letter abbreviations. Gaps (indicated by hyphens) were introduced for maximal alignment. Conserved AA are indicated by stars above the alignment. Colonos and periods denote similar AA and acceptable substitutions.
Figure 5.8 – Clustal W alignment of deduced amino acid sequences of the NifD proteins from *Leptospirillum* species. Conserved regions in and around the crucial FeMo-Co and P cluster ligands are highlighted. Amino acids are designated with one-letter abbreviations. Gaps (indicated by hyphens) were introduced for maximal alignment. Conserved AA are indicated by stars above the alignment. Colons and periods denote similar AA and acceptable substitutions.
Figure 5.9 – Clustal W alignment of deduced amino acid sequences of the NifK proteins from *Leptospirillum* species. Conserved regions in and around the crucial FeMo-Co and P cluster ligands are highlighted. Amino acids are designated with one-letter abbreviations. Gaps (indicated by hyphens) were introduced for maximal alignment. Conserved AA are indicated by stars above the alignment. Colons and periods denote similar AA and acceptable substitutions. The numbers below the alignment are in reference to the protein sequence of *A. vinelandii*. 
5.4 Discussion

For a microbe to be deemed capable of nitrogen fixation, genes for the production of the nitrogenase enzyme must be present (Streicher, Gurney, and Valentine 1971). This study demonstrates that the three *Leptospirillum* species each contain the structural genes necessary for nitrogenase production. Discovery of *L. ferrihilum* nifHDK genes is novel as previous studies (Tyson et al. 2005) have remarked that group II *Leptospirillum* species have no nitrogenase protein genes. The *nifH*, encoding for the Fe component of the nitrogenase enzyme demonstrated a high degree of conserved structure and function across the three species as did the *nifD* and *nifK* genes responsible for forming the Fe-Mo protein. Conservation of these gene structures appears to be maintained to give a functional enzymatic complex. Taken together, in all three *Leptospirillum* spp, functionality of these three genes, and therefore production of the nitrogenase enzyme for nitrogen fixation by appears highly probable.

5.4.1 Gene characteristics of *nifH, D and K*

The observations recorded in Chapter 2 which demonstrated continued proliferation and Fe$^{2+}$ oxidation by *L. ferrihilum* species without soluble nitrogen was of interest, as it was thought to be incapable of nitrogen fixation due to the supposed lack of structural genes for the nitrogenase enzyme. Possible theories to explain this phenomenon included minute traces of dissolved atmospheric NH$_3$ in the media. However, a more acidic solution with greater NH$_3$ affinity was included in the experiments to prevent this from occurring. The lack of growth by *At. caldus* in cultures minus soluble nitrogen also demonstrated that atmospheric NH$_3$ in the media was not providing enough nitrogen to permit proliferation. It is therefore unlikely that this was the reason for continued proliferation of all *Leptospirillum* species. A second proposition where *L. ferrihilum* contained the nifHDK genes but with sequence identity significantly different from *L. ferrooxidans* and *L. ferrodiazotrophum* suggested that the application of more degenerate priming sequences could possibly isolate the missing genes.

In this study, the *nifH* gene in *L. ferrihilum* was successfully isolated which then lead to the identification of the *nifD* and *nifK* genes. Contig arrangement of the *nifH, nifD* and *nifK* sequences with the *nifH-nifD* and *nifD-nifK* intergenic spacer regions successfully confirmed that the nifHDK genes in *L. ferrihilum* were arranged in a continuous fashion, resulting in a genomic fragment of 4093 bp that was similar to that of *L. ferrodiazotrophum* and *L. ferrooxidans*.
The *nifH* gene in all *Leptospirillum* species were highly conserved at the amino acid level (Table 5.8), however, the degeneracy of the genetic code resulted in significant variability at the nucleotide level.

Nucleotide composition and codon usage for the *nifH*, *nifD* and *nifK* genes in all the *Leptospirillum* species demonstrated in Table 5.6, highlight the preference for G and C residues at the third codon (67-82%), however, the overall G+C content for each of the genes was lower (53-59%). Comparison of the third base GC content with codon usage for the individual genes demonstrated a correlation; as GC content increased, the effective number of codons decreased. This asymmetric codon usage was noted in all *Leptospirillum* species with a number of possible codons not used for any of the three *nif* genes. The amino acids commonly affected included leucine, isoleucine, serine, histidine, valine and tryptophan. The absence of tryptophan residues in the *nifH* genes is characteristic of most diazotrophs (Souillard and Sibold 1986), demonstrating conservation of this attribute in *Leptospirillum* species.

Histidine, encoded for by CUA was also found to be absent from the *nifH* codons in *R. trifolii* (Mathur and Tuli 1991) and *Anabaena* sp (Mazur and Chui 1982). The codons UUA for leucine and GUA for valine were never present within the *nifH, nifD* or *nifK* genes for any *Leptospirillum* species. The lack of UUA is not unexpected as it is classified as a rare codon, found much less frequently in some organisms than other codons representing the same amino acid, leucine (Chater and Chandra 2008; Leskiw et al. 1991). Rare codons are known to reduce the translational efficiency rate (Proshkin et al. 2010) and bias against these codons is noted in highly expressed genes (Bulmer 1991). Therefore the absence of this codon from the *Leptospirillum nif* genes indicates the potential for optimal transcription with little dampening of the efficiency rate. Presuming that the apparent codon bias within the *nif* structural genes is reflected in the relative abundance of the corresponding tRNA species (Jacobson et al. 1989), the preference for codons ending in a G or C could signal that these codons are optimal for translation.

As conserved codon usage is known to lead to more accelerated protein synthesis (Elf et al. 2003), and the use of optimal codons indicating high gene expression (Stenico, Lloyd, and Sharp 1994), it is likely that the *Leptospirillum* genes for nitrogenase are ideally encoded for translation, resulting in consistent and controlled production of the nitrogenase protein.
5.4.2 Signal structures

Previous research by Parro and Moreno-Paz in 2003 identified a nifH sequence from micro-array data in *L. ferrooxidans* 3.2 (Genbank Accession AY204398), whose coding sequence was 99% identical to the *L. ferrooxidans* DSM2705 nifH amplified in this study, with the exception of a Y nucleotide at position +21 downstream of the initiator codon, which in this study was resolved as a C nucleotide. Analysis of upstream sequences of the nif genes demonstrated similarities and special features associated with the functionality of the genes. Annotation of the *L. ferrooxidans* 3.2 nif operon fragment demonstrated a -24 to -12 nif promoter (CTGGC-N₁-TTGC) (Beynon et al. 1983) -225 bp upstream of the nifH start codon.

While this study only sequenced up to 20 bases upstream of the initiator codon, analysis of the genomic scaffold (Genbank Accession GQ693862) on which a nifH sequence in *L. ferrodiazotrophum* UBA1 was identified, homologous to the *L. ferrodiazotrophum* C5 nifH sequence from this study, is the presence of a σ⁵⁴ promoter (CTGGC-N₁-TTGC) -227 bp upstream of the nifH initiator codon. It is highly probable that the nif promoters identified from *L. ferrooxidans* 3.2 and *L. ferrodiazotrophum* UBA1 are also present in *L. ferrooxidans* DSM 2705 and *L. ferrodiazotrophum* C5 and are responsible for the initiation of nifHDK transcription.

The intergenic region between the nifH-nifD is longer (110-175 bp) than the region separating the coding sequences of nifD and nifK (62-72 bp) for all *Leptospirillum* species supporting the theory of another regulatory function such as ribosome binding prior to nifD translation (Holland et al. 1987). The homology of these RBS sequences found prior to the initiator codons may also contribute to the coordinated expression of the separate nitrogenase subunits. Within the nifH-nifD intergenic spacer region of all *Leptospirillum*, was the presence of an inverted sequence repeat, rich in cytosine-guanine base pairs. This sequence string may form a stem loop, acting as a theoretical transcription terminator, however, they lack traditional adenine residues flanking the 3’ end, frequently associated with intrinsic terminator mechanisms (Wilson and Hippel 1995). In this instance the stable stem loop structure may protect mRNA from exonucleolytic degradation from the 3’ end (Willison, Pierrard, and Hubner 1993) or alter translation efficiency by reducing ribosomal scanning rates (Yang et al. 2004).

Evidence of these stem loops and RBS within the nifH-D and nifD-K intergenic regions has also been demonstrated in *A. vinelandii* (Brigle et al. 1987), *Rhodobacter capsulatus* (Willison, Pierrard, and Hubner 1993) and *Burkholderia* (Minerdi et al. 2001) and assigned putative roles of transcription terminators and intramolecular processors. The production of separate mRNA transcripts may be responsible for the different expression levels of nifH and nifDK detected in the nitrogen fixation experiments conducted in Chapter 3.
5.4.3 Amino acid structural assessment

The nitrogenase enzyme has conserved structural and mechanistic features (Rees and Howard 2000) that make it identifiable amongst other enzymes also involved in ATP binding. A Walker A motif, indicative of ATP/GTP binding found in many nucleotide binding proteins was recognized in NifH of all Leptospirillum species. As nitrogenase enzyme activity requires a large amount of energy (ATP) expenditure on behalf of the cell (Burns 1969), identification of this motif in the Fe component of the nitrogenase is unsurprising. Completing the motif was the presence of two signature lysine residues (GKGGiGKS) (Figure 5.11) identifying it as a member of the ‘deviant Walker A subgroup’ (Lutkenhaus and Sundaramoorthy 2003). The initial C-terminal lysine (Lys13) of the Walker motif is crucial for ATP binding nucleotide proteins, whereas the second unique conserved lysine is assumed to contact the α and γ phosphate of the nucleotide bound to the opposite nifH monomer (Lutkenhaus and Sundaramoorthy 2003). Along with the conserved Walker A motif were two protein chains, namely switch I and II, that perform important roles in the communication between the nucleotide binding site of NifH and the two cysteine ligands of the 4Fe-4S cluster (Ryle and Seefeldt 1996).

![Figure 5.11](image) – 3D representation of conserved areas of the NifH protein in Leptospirillum spp. modelled on the crystal structure of A. vinelandii. Regions highlighted include the Walker A motif (magenta), Switch 1 and 2, Fe₄S₄ and two cysteine binding sites.

From the theoretical amino acid translation of nifH, the structures essential for binding of MgATP to the iron component I (NifH), and resultant protein confirmation change are
present in all *Leptospirillum* species studied, theoretically permitting NifH docking to component II (NifDK), forming the nitrogenase protein. Prior to this result, it was assumed that *L. ferrilphilum* had no structural genes for nitrogenase production and therefore no capacity for nitrogen fixation. Discovery of the nifH gene, and apparent conserved functionality, mirroring *L. ferrooxidans* and *L. ferrodiazotrophum*, with no structural deletions or insertions, is highly suggestive that the activity of this enzyme (and subsequent nitrogen fixation) was responsible for the continued growth of *L. ferrilphilum* in media with no soluble nitrogen.

Similarly to other diazotrophs, all *Leptospirillum* NifD sequences demonstrated conservation of five cysteine residues (positions 59, 85, 151, 181 and 272 of the unaligned sequences) proposed to be responsible for ligating and coordinating the P cluster (Kim and Rees 1996) necessary for electron shuttling and substrate reduction (Hu et al. 2005). Associated with the cysteine residues were four aspartyl residues, α-Asp158, α-Asp159, β-Asp160, and β-Asp161 found in all *Leptospirillum* NifD(α) and NifK(β) sequences, acknowledged to also be completely conserved in known diazotrophs (Dean, Bolin, and Zheng 1993). The conserved aspartate residues are connected to the P cluster ligands (conserved αCys151 and βCys153) by short α helices (Cantwell 1998), that may function as a conduit, permitting signalling between the nitrogenase protein surface and metallocluster (Peters, Fisher, and Dean 1995). Preservation of these residues within all *Leptospirillum* species is indicative of retained significant catalytic activity for nitrogen fixation (Figure 5.12).
Figure 5.12 – NifD conserved sites modelled on the nitrogenase-α-component of *A.vinelandii* resolved crystal structure (PDB: 2AFK_A). Alpha helices are represented by green tubes and beta sheets by tan planks. Green and blue wire indicate protein secondary structure. Coloured balls represent amino acids identified in all *Leptospirillum spp* as conserved and homologous to the NifD structure from *A. vinelandii*. Conserved cysteine residues (blue), ligand cofactor interaction sites (pink), and P cluster conserved residues (purple) interacting with the Fe₆S₇ heterogen.

The highly conserved residues (LNVLH[CYRSMNY]) surrounding the Cys272 (unaligned) and His439 (FRQMHSWDY) sequence regions of the *Leptospirillum* NifD protein that ligate the FeMo cofactor to the α-component of the nitrogenase protein (Glazer and Kecharis 2009) are unique identifiers which place the *Leptospirillum* species within Group I diazotrophic prokaryotes with *A. vinelandii* and *Burkholderia* (Kecharis et al. 2006). Unlike *Clostridium pasteurianum*, *Leptospirillum* species do not contain an insert of 50 residues at the carboxyl terminal half of NifD, as seen in the multiple sequence alignment. The absence of these residues, also cements their position as Group I diazotrophs (Raymond et al. 2004).

NifK was less structurally conserved than NifH and NifD (82% vs 88 and 92%), but three conserved cysteine residues (Cys70, 95 and 153), essential for P cluster binding were confirmed to be present. Substitutions of these residues by different amino acids can decrease or completely eliminate nitrogenase activity (Kim and Rees 1996). Two other conserved cysteine residues were identified at positions 113 and 415, but analysis of this residue on
a 3D model (Figure 5.13) shows no obvious interactive role with the P cluster.

![Figure 5.13](image_url) – NifK conserved sites modelled on the nitrogenase-β-component of *A. vinelandii* resolved crystal structure (PDB:2AFK_B). Alpha helices are represented by green tubes and beta sheets by tan planks. Green and blue wire indicate protein secondary structure. Coloured balls represent amino acids identified in all *Leptospirillum* spp as conserved and homologous to the NifD structure from *A. vinelandii*. Conserved cysteine residues (blue) and P cluster conserved residues interacting with the Fe₅S₇ heterogen are in purple.

Along with the conserved cysteine residues were numerous β/α subunit interaction sites and MoFe protein β subunit Fe protein contacts (highlighted in the protein sequence alignment) also common to other diazotrophs. Considering the highly conserved regions in the α and β subunits are responsible for the major folding properties of the polypeptides (Holland et al. 1987), theoretically the strength of the conserved regions of the homodimeric NifH, combined with the identified FeMo cofactor binding sites and P cluster cysteine residues of the tetrameric NifD and NifK proteins, a functional nitrogenase enzyme in all *Leptospirillum* species of this study appears highly probable.

The lack of acetylene reduction activity by *L. ferrphilium* in Chapter 2 does not appear to be the result of the lack of structural nif genes, as evidenced by this analysis. The nifHDK genes in *L. ferrphilium* demonstrate similar sequence conservation analogous to those found in *L. ferrodiazotrophum* and *L. ferrooxidans*, representing no structural defects or mutations, that would render any of the individual genes inactive. Active nitrogenase production, and therefore nitrogen fixation by *L. ferrphilium* would explain the continual growth observed in soluble nitrogen deficient conditions as well as the increase in transcription of the nifS-U genes recorded in Chapter 3, necessary for synthesis of the 4Fe-4S cluster of the NifH protein.
5.5 Conclusions

Sequencing and analysis of the \emph{nifH}, \emph{D} and \emph{K} genes in all \emph{Leptospirillum} species identified highly conserved regions in each gene, indicating that all the \emph{Leptospirillum} members assessed in this study belong to the Group I nitrogenase cluster (typical Mo-Fe nitrogenase), alongside members from proteobacteria and cyanobacteria phyla. Structural analysis of the \emph{Leptospirillum} NifHDK encoded proteins revealed conserved amino acids necessary for the assembly of a functional molybdenum nitrogenase with no apparent deletions or mutations that would affect activity of this enzyme.

In environments where soluble nitrogen deficiencies can prevent microbial proliferation and community expansion, \emph{Leptospirillum} are in a unique position to dominate these habitats. Despite \emph{At. ferrooxidans} having a similar capacity for nitrogen fixation as \emph{Leptospirillum} and an ability to survive in low pH conditions, it does not tolerate high levels of ferric in the environment. Therefore, \emph{Leptospirillum} species ability to tolerate low pH, high levels of ferric ions combined with the presence of these \emph{nif} genes, may enable them to survive periods of soluble nitrogen shortages, out-competing other microbes and eventually dominating microbial populations.
Chapter 6

Phylogenetic analysis of *Leptospirillum* species

*Previous chapters have demonstrated that three species of Leptospirillum contain functional genes necessary for the production of the nitrogenase enzyme complex. This chapter explores the evolutionary relatedness of the three species based on the phylogeny of their 16S rRNA and nif genes.*

6.1 Introduction

Bacterial systematics is the study of diversity of organisms and their relationships and comprises of classification, nomenclature, and identification. Bacteria have long been classified by their microscopic morphological features (coccis, spirals, short and long rods) (Cohn 1872), as well as by cell wall staining (Gram 1884). Complementing the morphological descriptions, bacteria are now also identified and described according to key physiological traits, such as respiration, metabolism, photosynthesis and nitrogen fixation (Boucher et al. 2003). With the advent of bacterial identification and taxonomic classification based on the stable 16S rRNA subunit (Woese et al. 1985), the hierarchical evolutionary relatedness between organisms can also be derived.

Phylogenetic identification and understanding of evolutionary relatedness between organisms, is based upon similarities and differences in their physical and/or genetic characteristics. By using molecular data such as DNA or protein sequences, or distinct molecular markers, phylogenetic analysis allows for the inference or estimation of these evolutionary relationships (Felsenstein 1985). The 16S rRNA gene is the most common tool for phylogenetic characterization of micro-organisms as signatures in their structure provide consistent means for distinguishing species (Woese 1987). Due to the conserved nature of this gene, it shows limited variation for members of closely related taxa (Fox, Wisotzkey, and Jurtshuk
1992), but it is also prone to homologous recombination, resulting in problems for phylogenetic estimation using tree reconstruction methods (Wang and Zhang 2000).

Even though the 16S rRNA gene is highly conserved, and used to discern bacterial identification, it does not provide definitive information on biological, physiological or ecological function of the microorganism studied. Consequently, when dealing with closely related species, a single gene should not be used to infer evolutionary relationships.

Alternatives to the 16S rRNA gene sequence include protein coding genes (rpoB, gyrB, nifD, recA, atpD), which provide additional insight into relationships between organisms that share similar functions (Holmes, Nevin, and Lovley 2004; Richert, Brambilla, and Stackebrandt 2007). Nevertheless, protein coding phylogenograms frequently contradict rRNA phylogenies because of lateral gene transfer events (Doolittle 1999) that occur between species. By combining different datasets (the 16S rRNA gene with protein coding genes) a superior picture of the true evolutionary history may be resolved.

The capacity to utilize dinitrogen as a nitrogen source is an important phenotypic trait, with the nifH gene often used as an evolutionary marker in phylogenetic reconstructions (Raymond et al. 2004; Zehr et al. 2003a) due to its highly conserved function. The discovery of the full length nifHDK genes in all Leptospirillum species (Chapter 5) has allowed these protein encoding genes to be used for phylogenetic studies, elucidating the evolutionary lineage amongst the species and reconstruction of genealogical ties to other known diazotrophs.

Along with the 16S rRNA, two other constitutively expressed ‘house keeping’ genes, DNA gyrase subunit beta (gyrB, responsible for regulating supercoiling of bacterial chromosomes during genome replication) and DNA directed RNA polymerase beta (rpoB, necessary for the transcription of DNA to RNA) were selected and identified in all Leptospirillum species. Concatenation of the NifHDK protein encoding genes as well as the 16S rRNA with gyrB and rpoB allowed for a more rigorous investigation of the genotypes.

This study offers a comparison of the Leptospirillum 16S rRNA phylogeny with the phylogenies based on two essential housekeeping genes and three different structural genes of the nitrogenase enzyme.
6.2 Materials and methods

6.2.1 *Leptospirillum* sequences for Phylogenetic Analysis

The *Leptospirillum* species 16S rRNA, *nifH*, *nifD* and *nifK* sequences used for phylogenetic analysis were generated in Chapter 5, following the protocols listed in sections 5.2.1 to 5.2.6.

6.2.2 PCR and sequencing primers

Growth and subsequent DNA extraction from *Leptospirillum* species was performed according to protocols listed in sections 5.2.1 to 5.2.3. Two additional housekeeping genes were targeted for PCR amplification and DNA sequencing. Degenerate primers for partial sequence amplification of *gyrB* and *rpoB* were designed from BLAST alignments of homologous sequences found in the NCBI(nr) database. The *gyrB* and *rpoB* oligonucleotide primers were manufactured by GeneWorks (Australia), with dry lyophilized nucleotide pellets reconstituted to a stock concentration of 100 μM, and further diluted to a working concentration of 20 μM. All primer solutions were stored at -20 °C.

<table>
<thead>
<tr>
<th>Species</th>
<th>Target</th>
<th>Primer names</th>
<th>Forward (5'-3')</th>
<th>Reverse (3'-5')</th>
<th>Product size (bp)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Leptospirillum</em></td>
<td><em>gyrB</em></td>
<td>UnigyrB-F and UnigyrB-R</td>
<td>GCGAGCCRTCNACRTCNGC</td>
<td>TGACGGGAATCATCAGTGTC</td>
<td>~1000</td>
<td>This study</td>
</tr>
<tr>
<td><em>Leptospirillum</em></td>
<td><em>rpoB</em></td>
<td>UnirpoB-F and UnirpoB-R</td>
<td>GGAGGgARAGGWTCTGCTCA</td>
<td>GGAGGgARAGGWTCTGCTCA</td>
<td>~780</td>
<td>This study</td>
</tr>
</tbody>
</table>

PCR amplification was conducted on an Alpha Metrix G-Storm GS1 thermal cycler using a 25 μL reaction volume. The optimized PCR conditions for each gene are listed in Tables 6.2 a and b. Each PCR mixture contained ddH₂O, 2 mM 10X PCR buffer (Promega), 2.5 mM MgCl₂, 0.2 mM dNTPs, 3.2 μM of each primer, 1U Taq DNA polymerase (Promega), 2% DMSO and 100 ng DNA template. Negative controls were run by substituting PCR grade water for template DNA.
Phylogenetic analysis of *Leptospirillum* species

Table 6.2 – PCR cycle parameters for *gyrB* and *rpoB*.

<table>
<thead>
<tr>
<th>Protocol</th>
<th>N° repeats</th>
<th>Temp (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
<td>1x</td>
<td>95</td>
<td>5min</td>
</tr>
<tr>
<td>Cycle 2</td>
<td>30x</td>
<td>95</td>
<td>30s</td>
</tr>
<tr>
<td>Step 1</td>
<td>95</td>
<td>30s</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>60</td>
<td>60s</td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td>72</td>
<td>120s</td>
<td></td>
</tr>
<tr>
<td>Cycle 3</td>
<td>1x</td>
<td>72</td>
<td>7min</td>
</tr>
<tr>
<td>Cycle 4</td>
<td>1x</td>
<td>10</td>
<td>x</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protocol</th>
<th>N° repeats</th>
<th>Temp (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cycle 1</td>
<td>1x</td>
<td>95</td>
<td>5min</td>
</tr>
<tr>
<td>Cycle 2</td>
<td>30x</td>
<td>95</td>
<td>30s</td>
</tr>
<tr>
<td>Step 1</td>
<td>95</td>
<td>30s</td>
<td></td>
</tr>
<tr>
<td>Step 2</td>
<td>57</td>
<td>30s</td>
<td></td>
</tr>
<tr>
<td>Step 3</td>
<td>72</td>
<td>90s</td>
<td></td>
</tr>
<tr>
<td>Cycle 3</td>
<td>1x</td>
<td>72</td>
<td>10min</td>
</tr>
<tr>
<td>Cycle 4</td>
<td>1x</td>
<td>10</td>
<td>x</td>
</tr>
</tbody>
</table>

PCR clean up, sequencing, and confirmation of target identity were performed following protocols listed in sections 5.2.5 and 5.2.6.

6.2.3 Sequence selection for evolutionary comparison

All *Leptospirillum* species nucleotide and translated amino acid sequences were submitted to the NCBI(nr) database for BLAST analysis and 10 to 15 sequences with the highest sequence identity were selected for evolutionary comparison. Micro-organisms selected and Genbank accession numbers for multiple sequence alignment and phylogenetic analysis are detailed in Tables A.6, A.7 and A.8 of Appendix A. The out-group for housekeeping gene analysis (16S rRNA, *gyrB* and *rpoB*) tree rooting was *Thermodesulfovibrio yellowstonii DSM 1147*, obtained from the complete genome sequence. *Clostridium pasteurianum* was selected as the out-group for the *nif* genes.

6.2.4 Phylogenetic Analysis

Multiple sequence alignments for each gene were generated using CLUSTAL W (Thompson, Higgins, and Gibson 1994) within MEGA 5.05 (Tamura et al. 2011). Translated peptides were aligned using the PAM amino acid substitution matrix (Dayhoff, Schwartz, and Orcutt 1978). Following alignment, obvious errors were manually corrected. Sequences were truncated to ensure all analysed gene sequences were of the same length, preferentially maintaining *Leptospirillum* sequence length. Sequence alignments for both nucleotides and proteins were imported into Model Generator v0.85 (Keane et al. 2006) to assess the most appropriate data substitution model for inferring phylogenetic relatedness. From log likelihood (-lnL) values, Akaike Information Criteria and Bayesian Information Criteria scores the most appropriate model was selected. Concatameric sequences analyses were constructed manually with head to tail arrangement after individual gene alignments were performed (Figure 6.1 and 6.2).
6.2.5 Construction of Phylogenetic Trees

Phylogenetic relationships for comparison of evolutionary distances and branching order were performed by the construction of maximum likelihood (ML) and Bayesian inference (BI) trees. Neighbour Joining trees were used for a quick evaluation of tree shape and taxa position, but not used for phylogenetic inference. The commands used for running analyses are presented in Appendix B. ML phylogenetic trees were constructed using MEGA 5.05, with 1000 bootstrap replicates and optimal substitution models selected by Model Generator, for the appropriate gene alignment. Bootstrap values are given as a percentage. Branch lengths were scaled according to the number of base substitutions in both ML and BI trees.

To complement the ML analyses, Bayesian Inference of phylogeny with Markov chain Monte Carlo sampling was performed using MrBayes 3.2 (Ronquist and Huelsenbeck 2003). Following the author’s recommendations, each gene studied was assigned a best fitting substitution model, as determined by Model Generator. For all data sets, one cold chain and three heated chains were run simultaneously for 1 million generations with pseudo-sampling every 1000 with a random starting tree. Each analysis was replicated twice. Evidence for convergence of the different chains was obtained by examining the correlation between the posterior probabilities of the individual clades in pairwise comparisons among runs, and by comparing the generation plots for overall model likelihood (-lnL). The first 250 generations (25%) were removed from analysis as ‘burn-in’. Then number of generations chosen was sufficient to ensure the standard deviation of split frequencies was less than 0.01. After burn-in, all runs were checked to ensure a similar mean and variance of model likelihood and parameter values. Bayesian posterior probability (PP) values were estimated on the 50% majority rule consensus of the remaining trees. Tree files were read using FigTree v1.3.1.

![Figure 6.1 - Gene organization and concatenation of housekeeping genes for phylogenetic analysis. Position of the partial selection of gyrB and rpoB genes](image-url)
Figure 6.2 - Nif gene organization and concatenation. Intergenic spacer regions were not included in the concatenated sequence used for tree construction.
6.3 Results

Greatest sequence identity between the *Leptospirillum* species was recorded with the 16S rRNA, however, it is lower than expected as 16S rRNA is often highly conserved between species (>97%). Sequence identity for *gyrB* and *rpoB* was significantly lower than the 16S rRNA.

<table>
<thead>
<tr>
<th>Loc</th>
<th>Nº aligned sites (bp)</th>
<th>Nº conserved sites</th>
<th>Nº variable sites</th>
<th>% Sequence identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>16S rRNA</td>
<td>1349</td>
<td>1221</td>
<td>172</td>
<td>90.51</td>
</tr>
<tr>
<td><em>gyrB</em></td>
<td>1050</td>
<td>723</td>
<td>327</td>
<td>68.86</td>
</tr>
<tr>
<td><em>rpoB</em></td>
<td>781</td>
<td>483</td>
<td>297</td>
<td>61.84</td>
</tr>
</tbody>
</table>

6.3.1 *Leptospirillum* phylogeny based on housekeeping gene sequences

6.3.1.1 16S rRNA

Successful amplification of the 16S rRNA gene was performed in Chapter 5 and sequences used for phylogenetic analysis. Sequences for other known *Leptospirillum* 16S rRNA genes were obtained from Genbank. The sequence alignment contained 9 taxa and was 1349 bp in length. No significant insertion or major deletions were identified. The model of evolution selected under Maximum Likelihood was General Time Reversible (GTR) + Γ. The same data were analysed under Bayesian analysis with the same model for one million generations. The analyses were rooted with a close member of the nitrospira group, *T. yellowstonii DSM 1147*.

The 16S rRNA tree topologies Figure 6.3 generated with ML and BI phylogenetic models were in general agreement when assigning genomic groups based on sequence clustering, with the three type species forming distinct clades. Both models identified *L. ferrodiazotrophum* as the most recently divergent species supported with high BS values (100) and PP (100). When attempting to elucidate the evolutionary pathway of each of the species, ML and BI topologies are not congruent. ML phylogenies inferred divergence of *L. ferrodiazotrophum* prior to *L. ferrooxidans* and *L. ferrilfihum*, whereas BI depicted evolution of *L. ferrooxidans* (PP - 100) adrift of *L. ferrodiazotrophum* and *L. ferrilfihum*, although supported with low BS values (54)
Figure 6.3 – Phylogenetic reconstruction of *Leptospirillum* 16S rRNA genes by Maximum Likelihood (ML) and Bayesian Inference (BI). Evolutionary history was inferred by ML and BI methods based on the GTR model (Waddell and Steel 1997). Numbers at nodes on the ML tree depict bootstrap percentages after 1000 replications (Felsenstein 1985). The Bayesian posterior probabilities (as a percentage) are shown next to the nodes. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. *T. yellowstonii* was selected as an out-group.
6.3.1.2 \textit{gyrB} and \textit{rpoB}

Partial \textit{gyrB} and \textit{rpoB} sequences were successfully amplified for all \textit{Leptospirillum} species with sequence information subsequently deposited in Genbank. Alignment of the \textit{gyrB} sequence was 1050 bp in length and alignment of the \textit{rpoB} sequence was 781 bp in length. Housekeeping gene phylogenies were performed with only sequences from this study as to determine the evolutionary age between these three \textit{Leptospirillum} species. Additionally, the large number of \textit{gyrB} and \textit{rpoB} sequences in the NCBI database are from environmental sampling and clones, therefore their identification can not be authenticated. The model of DNA evolution selected for both \textit{gyrB} and \textit{rpoB} under ML and BI was GTR+\Gamma and the resulting phylogenograms are depicted in Figure 6.4. BI analyses were run for one million generations, and the 50\% consensus tree shown. The out-group selected was \textit{T. yellowstonii}.

Evolutionary age of the species differ between the \textit{gyrB} and \textit{rpoB} phylogenograms, but all methods support the divergence of \textit{L. ferrodiazotrophum} separately from \textit{L. ferrooxidans} and \textit{L. ferriphilum}, as was depicted in the ML tree of the 16S rRNA gene (Figure 6.3a). Confidence in the divergence was strongest when modelled on the \textit{gyrB} gene, supported by high PP and BS values.
Figure 6.4 - Phylogenetic reconstruction of *Leptospirillum* species *gyrB* and *rpoB* genes by ML and BI. Numbers at nodes on the ML tree depict bootstrap percentages after 1000 replications. The Bayesian posterior probabilities (as a percentage) are shown next to the nodes. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. *T. yellowstonii* was selected as an out-group.

### 6.3.1.3 DNA concatenation of housekeeping genes

The aligned 16S rRNA, *gyrB* and *rpoB* DNA sequences were concatenated into a single dataset, 3,180 nucleotides in length. The model of DNA evolution selected under ML and BI was GTR+Γ and the resulting phylograms are depicted in Figure 6.5. BI analyses were run for one million generations, and the 50% consensus tree shown. The out-group selected was *T. yellowstonii*.

With high PP and BS values, both ML and BI phylogenies constructed from the concatenomic analyses support the phylogenetic trees resolved when *gyrB* and *rpoB* genes were examined separately. The divergence of *L. ferrodiazotrophum* from a common ancestor prior to *L. ferrooxidans* and *L. ferriphilum* is consistent regardless of the model of evolution chosen (ML or BI), or the house keeping gene selected (*gyrB* or *rpoB*). Evolutionary distance of *L. ferrodiazotrophum* from the last common ancestor was greatest with the 16S rRNA, and *gyrB* genes, but not when assessed with *rpoB* alone.
Figure 6.5 - Phylogenetic reconstruction of concatenated housekeeping genes by ML and BI. Numbers at nodes on the ML tree depict bootstrap percentages after 1000 replications. The Bayesian posterior probabilities (as a percentage) are shown next to the nodes. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. *T. yellowstonii* was selected as an out-group.
6.3.2 *Leptospirillum nif* gene phylogenies

Greatest *nif* gene sequence identity between the *Leptospirillum* species was recorded with the *nifH*, followed by *nifK* and finally *nifD* (Chapter 5). The trend differs when nucleotide sequences were translated to the corresponding theoretical amino acid peptide, and NifD is the most conserved sequence between the species. Determination of inheritance patterns of the *nif* genes within the *Leptospirillum* genus, was performed with *nifHDK* genes from known diazotrophs, that demonstrated the greatest sequence identity to the *Leptospirillum* species *nif* genes.

6.3.2.1 *nifH* analysis

The alignment of *nifH* DNA included 11 taxa and was truncated to a consensus length of 828 bp. The translated protein sequence alignment (NifH) was 275 AA in length following truncation and demonstrated in Figure 6.6. The model of DNA evolution selected under ML was GTR+Γ+I. The same data were analysed under Bayesian analysis using the same model for one million generations. Consensus trees are shown in Figure 6.7. The model of protein evolution selected under ML was WAG+Γ (Whelan and Goldman 2001) and also applied to Bayesian analysis with one million generations (Figure 6.8). *Cl. pasteurainum* was chosen as an out-group as it had the lowest AA sequence identity to the *Leptospirillum* species of all the NifH sequences selected.

The *nifH* DNA and AA data, when analysed with ML and BI generated phylogenetic trees that were generally similar (Figure 6.7 and 6.8). All *Leptospirillum* species were monophyletic, forming a single cluster. Resolution of the *Leptospirillum* clade was not congruent between the DNA and AA phylograms, as the order of *nifH* divergence differed between the species.

DNA phylograms depicting *nifH* evolution (Figure 6.7) demonstrate the divergence of *L. ferriphilum* prior to the separation of *L. ferrooxidans* and *L. ferrodiazotrophum*. However, as the NifH protein, *L. ferrodiazotrophum* divergence occurs prior to *L. ferrooxidans* and *L. ferriphilum*, with *L. ferriphilum* the most recently divergent species (Figure 6.8). The individual branching of *L. ferrooxidans* and *L. ferriphilum* from *L. ferrodiazotrophum* was not supported with high bootstrap values (52) by ML, although resolved slightly stronger with BI (86).

The placement of the *Leptospirillum* genus was congruent between ML and BI trees using both DNA and AA sequences in regards to its position amongst the other prokayrotes, as the closest evolutionary ancestors belong to γ-proteobacteria, in particular *Acidithiobacillus*, of which *Leptospirillum* species share a common environment. Resolution of the *Leptospirillum* clades had greater support with Bayesian PP than with BS values from ML regardless of the type of sequence analysed.
Burkholderia xenovorans
Methylocystis sp ATCC49242
Bradyrhizobium japonicum
At. ferrivorans
At. ferrooxidans
L. ferriphilum
L. ferrooxidans
L. ferrodiazotrophum
Methylacidiphilum fumariolicum
Azotobacter vinelandii
Clostridium pasteurianum

The numbers below the alignment are in reference to the protein sequence of A. vinelandii.
Figure 6.7 – Phylogenetic reconstruction of the nifH gene inferred by using the ML and BI methods based on GTR+Γ+I model. The ML bootstrap consensus tree was inferred from 1000 replicates with the percentage of replicate trees in which the associated taxa clustered together are shown next to the branches. The 50% consensus rule tree for BI is depicted. The Bayesian PP (as a percentage) are shown next to the nodes. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. *Cl. pasteurianum* was selected as an out-group.
Figure 6.8 – Phylogenetic reconstruction of the NifH protein was inferred by ML and BI. The ML bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed. The 50% consensus rule tree for BI is depicted. The Bayesian PP (as a percentage) are shown next to the nodes. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. *Cl. pasteurianum* was selected as an out-group.
6.3.2.2 *nifD* analysis

The alignment of *nifD* DNA had 11 taxa and was 1449 bp in length including insertions for multiple alignment. The translated protein sequence was 527 AA in length after truncation depicted in Figure 6.9. The model of DNA evolution selected under ML was GTR+Γ. The same data were analysed under Bayesian analysis using the same model for one million generations. The consensus trees are shown in Figure 6.10.

The model of protein evolution selected under ML was WAG+Γ+F and was also applied to Bayesian analysis with one million generations (Figure 6.11). *Cl. pasteurianum* was chosen as an out-group for both DNA and AA.

Resolution of the *Leptospirillum* clade was not congruent between the DNA and AA phylogenograms, as the order of *nifD* inheritance differed between the species. As seen with the *nifH* phylogenograms, *nifD* DNA phylogenograms resolved *L. ferriphilum* divergence as a separate event prior to *L. ferrokioxidans* and *L. ferrodiazotrophum* separation, supported by high PP values (100), but with low BS values (61). Branch lengths constructed from *nifD* genes demonstrated that the relative amount of character change was similar for all *Leptospirillum* species.

*NifD* (AA) phylogenograms depict divergence of *L. ferrodiazotrophum* prior to *L. ferrokioxidans* and *L. ferriphilum*, as was previously demonstrated with the NifH protein. Branching of *L. ferrokioxidans* and *L. ferriphilum* after *L. ferrodiazotrophum* divergence was not supported with reliable BS values (55)(Figure 6.11a), although higher with PP (91) (Figure 6.11b).

Placement of the *Leptospirillum* *nifD* gene amongst other prokaryotes demonstrated distance from the γ-proteobacteria genus *Acidithiobacillus*, of which *nifH* showed close ancestry (Figure 6.7a). Comparatively, the *nifD* phylogeny depicted a close evolutionary history to *A. vinelandii* (as did *nifH*). The BI NifD phylogram (Figure 6.11b) demonstrates a close relationship of the *Leptospirillum* clade with *B. xenovorans* and members from the *Acidithiobacillus* genus, with *A. vinelandii* divergence occurring prior (PP=100). Evolution of *Leptospirillum* after the divergence of *A. vinelandii* is also depicted under ML, but now separate from the other β, and γ-proteobacteria. Bootstrapping values generated from the WAG+F+Γ analysis were very low for all branching events bar the division of *Acidithiobacillus* and may be attributed to the highly conserved nature of the NifD protein.
Figure 6.9 – Truncated NiF alignment. Clustal W alignment of deduced amino acid sequences of the NiF proteins from *Leptospirillum* species with other diazotrophic NiF sequences. Amino acids are designated with one-letter abbreviations. Gaps (indicated by hyphens) were introduced for maximal alignment. Conserved AA are indicated by stars above the alignment. Colon and periods denote similar AA and acceptable substitutions. The numbers below the alignment are in reference to the protein sequence of *A. vinelandii*.
(a) Maximum Likelihood

(b) Bayesian Inference

Figure 6.10 – Phylogenetic reconstruction of the *Leptospirillum* *nifD* genes was inferred by ML and BI. The ML bootstrap consensus tree values inferred from 1000 replicates are shown next to the branches. The 50% consensus rule tree for BI is depicted. The Bayesian PP (as a percentage) are shown next to the nodes. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. *Cl. pasteuriunum* was selected as an out-group.
Figure 6.11 – Phylogenetic reconstruction of the NifD by ML and BI. The ML bootstrap consensus tree values inferred from 1000 replicates are shown next to the branches. The 50% consensus rule tree for BI is depicted. The Bayesian PP (as a percentage) are shown next to the nodes. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Cl. pasteurianum was selected as an out-group.
6.3.2.3 *nifK* analysis

The alignment of *nifK* DNA had 11 taxa and contained 1583 coding positions. The aligned translated protein sequence contained 481 positions after truncation, depicted in Figure 6.12. The model of DNA evolution selected under ML was GTR+Γ. The same data were analysed under Bayesian analysis using the same model for one million generations (Figure 6.13). The model of protein evolution selected under ML was WAG+Γ and was also applied to Bayesian analysis with one million generations (Figure 6.14). *Cl. pasteurianum* was chosen as an out-group for both DNA and AA analysis.

The *nifK* DNA and NifK protein sequences when analysed with ML and BI resolved congruent evolutionary pathways, unlike the *nifH* and *nifD* generated phylogenetic trees. With both the DNA and AA sequences, phylograms established *L. ferrodiazotrophum* divergence prior to *L. ferrooxidans* and *L. ferrihilum*, supported by high BS and PP values. Of the three *Leptospirillum* species, *L. ferrooxidans* has undergone the most evolutionary change.
Figure 6.12 – Truncated NifK alignment. Clustal W alignment of deduced amino acid sequences of the NifK proteins from *Leptospirillum* species with other diazotrophic NifK sequences. Amino acids are designated with one-letter abbreviations. Gaps (indicated by hyphens) were introduced for maximal alignment. Conserved AA are indicated by stars above the alignment. Colons and periods denote similar AA and acceptable substitutions. The numbers below the alignment are in reference to the protein sequence of *A. vinelandii*.
Figure 6.13 – Phylogenetic reconstruction of *Leptospirillum nifK* by ML and BI. The ML bootstrap consensus tree values inferred from 1000 replicates are shown next to the branches. The 50% consensus rule tree for BI is depicted. The Bayesian PP (as a percentage) are shown next to the nodes. The trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Trees were rooted on *Cl. pasteurianum* as an out-group.
Figure 6.14 – Phylogenetic reconstruction of *Leptospirillum* NifK by ML and BI. The ML bootstrap consensus tree values inferred from 1000 replicates are shown next to the branches. The 50% consensus rule tree for BI is depicted. The Bayesian PP (as a percentage) are shown next to the nodes. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Trees were rooted on *Cl. pasteurianum* as an out-group.
6.3.2.4 Concatenation of the NifHDK proteins

The aligned protein sequences - NifH, NifD and NifK from 11 taxa were concatenated into a single dataset, with a total 1,333 coding positions. The model of evolution selected under ML and BI was WAG+Γ and the resulting phylograms are depicted in Figure 6.15. BI analyses were run for one million generations, and the 50% consensus tree shown. The out-group selected was *Cl. pasteurianum*.

With high PP and BS values, both ML and BI phylogenies constructed from the concatameric analyses demonstrated congruent phylogenetic trees. The divergence of *L. ferrodiazotrophum* from a common ancestor prior to *L. ferrooxidans* and *L. ferrihilum* is consistent. Concatenation of the protein encoding genes demonstrates that the evolutionary distance of *L. ferrooxidans* from the last common ancestor was greatest of all *Leptospirillum* species studied.
Figure 6.15 – Phylogenetic reconstruction of *Leptospirillum* concatenated NifHDK sequences by ML and BI. The ML bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analysed and the percentage of replicate trees in which the associated taxa clustered together are shown next to the branches. The 50% consensus rule tree for BI. The Bayesian PP (as a percentage) are shown next to the nodes. Trees are drawn to scale, with branch lengths measured in the number of substitutions per site. Trees were rooted on *Cl. pasteurianum* as an out-group.
6.4 Discussion

Phylogenetic inference, as an approach to classifying and clarifying prokaryotic relationships is often based on the comparative analysis of the 16S rRNA sequences (Lane et al. 1985). Concatenation of the 16S rRNA gene with two other housekeeping genes, \textit{gyrB} and \textit{rpoB} successfully resolved congruent evolutionary pathways for the \textit{Leptospirillum} species, demonstrating divergence of \textit{L. ferrodiazotrophum} prior to \textit{L. ferrooxidans} and \textit{L. ferrphilum}. Mirroring this result, were phylogenetic trees constructed from amino acid sequences of the NifHDK genes. Examination of these tree topologies has allowed for evolutionary inferences of nitrogenase gene acquisition by \textit{Leptospirillum} species to be postulated.

6.4.1 Concatameric housekeeping phylogeny

Based on the 16S rRNA sequence, the evolutionary history of the \textit{Leptospirillum} species was inferred, determining that \textit{L. ferrodiazotrophum} was the most divergent species of the \textit{Leptospirillum} genera, with successful segregation of the \textit{Leptospirillum} species into three separate groups. The topology of the 16S rRNA dendogram derived from the ML parameters (Figure 6.3) was in agreement with the phylograms published by Tyson et al. (2005), where \textit{L. ferrodiazotrophum} diverged separately from \textit{L. ferrphilum} and \textit{L. ferrooxidans}.

Both ML and BI depict \textit{L. ferrodiazotrophum} as the most divergent species of the \textit{Leptospirillum} genus, but the relationship between it and the other two \textit{Leptospirillum} species remains unclear due to the lack of congruence between the evolutionary modelling methods. Having limited informative positions within a gene is known to increase stochastic noise (Brinkmann et al. 2005), often resulting in inadequately resolved phylogenies. This may help explain the inconsistent and inadequately supported evolutionary relatedness of the \textit{Leptospirillum} 16S rRNA. The high sequence conservation of the 16S rRNA gene could have rendered the homologous nucleotide positions phylogenetically uninformative, contributing to the inability to obtain congruent accurate phylogenetic inferences.

As the phylograms constructed with 16S rRNA gene information did not produce consistent trees, analysis using partial \textit{gyrB} and \textit{rpoB} sequences was conducted in order to resolve these incongruencies. Both these genes have been proposed as potential biomarkers to overcome the high conservation of the 16S rRNA gene, as taxonomic resolution of the \textit{rpoB} gene is three times greater than that of the 16S rRNA for known bacteria, including \textit{Vibrio}, \textit{Bacillus} and \textit{Pseudomonas} (Ki, Zhang, and Qian 2009). As they are chromosomally borne, neither are transmitted horizontally (Yamamoto and Harayama 1995).

Even though the \textit{gyrB} and \textit{rpoB} genes were less conserved than the 16S rRNA (Table 6.3),
consistent evolutionary phylograms were obtained with each of these sequences, demonstrating congruence with the 16S rRNA phylogram constructed using ML methods. Concatenation of the *gyrB* and *rpoB* genes sequences with the 16S rRNA gene resulted in a longer sequence, which is known to help reduce variance in the distance estimates (Gadagkar, Rosenberg, and Kumar 2005). This concatenation enhanced the resolution of the inferred 16S rRNA tree topology, supported with high BS and PP values, indicating that *L. ferrodiazotrophum* is the most ancient species in the genus, followed by *L. ferriphilum* and *L. ferrooxidans*.

Evolutionary age of the three species was not consistent between the housekeeping genes, as the *rpoB* gene depicted less evolutionary divergence of *L. ferrodiazotrophum* than both the 16S rRNA and *gyrB* genes. Only half (~45%) of the *rpoB* gene was amplified in this study and could have contributed to the inconsistency of this particular tree structure. Concatenation of the housekeeping genes was again able to resolve this discrepancy, and discern that *L. ferrodiazotrophum* was the most highly divergent member of the *Leptospirillum* species.

Out-groups are traditionally nominated based on their close but dissimilar relationship to the species of interest (Phillipe and Laurent 1998) as they share a recent common ancestor. For this study *T. yellowstonii* was selected as an out-group as it is positioned within the Nitrospira family along with the ‘in-group’ *Leptospirillum*, but is deficient in genes for nitrogen fixation. In order to prevent long branch attraction (LBA) artefacts from misleading phylogenetic inferences, optimal out-group selection is known to be important (Holton and Pisani 2010). In this instance, the selection of *T. yellowstonii* for 16S rRNA analysis may have been less than optimal and in fact been too closely related to resolve a clear evolutionary pattern, compromising the taxa of primary interest (Stackebrandt and Ludwig 1994). Conversely, use of the *T. yellowstonii gyrB* and *rpoB* genes were successful as out-group sequences, resolving consistent evolutionary pathways in reference to all the *Leptospirillum* species, regardless of the substitution model selected.

### 6.4.2 The *nif* phylogeny

To ensure the stoichiometry of the nitrogenase reaction, *nifHDK* genes are often arranged in a contiguous fashion, for co-transcription (Fani, Gallo, and Lio 2000). This synteny was confirmed in all *Leptospirillum* species (Chapter 5) and therefore assumed that evolution of these three genes would have occurred at the same rate, and be subject to the same evolutionary pressures. However, phylograms constructed with molybdenum based nitrogenase genes, *nifH, nifD* and *nifK* were not congruent, suggesting a complex evolutionary history combining vertical inheritance with possible lateral gene transfer events. The potential variable rates of inheritance and evolution experience by these three genes may have introduced a source of bias, subsequently affecting the determination of the correct
evolutionary pathway.

6.4.2.1 The nif DNA phylogeny

With nifH and nifD genes, the evolutionary age of *L. ferrodiazotrophum* and *L. ferrooxidans* appeared similar, diverging separately from *L. ferrihilum*, whereas evolution of the nifK gene demonstrated divergence of *L. ferrodiazotrophum* prior to *L. ferrooxidans* and *L. ferrihilum*. Assessment of all nif genes separately, demonstrated that phylogeny of nifK was congruent with evolutionary pathway depicted from the housekeeping genes. Comparatively, nifH and nifD gene phylogenies differed from the concatenated 16S-gyrB-rpoB sequence phylograms, highlighting potential differences between evolutionary pressures acting on essential ‘informative’ genes and those ‘operational’ genes expressed depending on environmental conditions. Lateral transfer of genetic material is less likely with complex informational genes (ribosomal, housekeeping) (Kechriss et al. 2006) than role specific proteins, with other studies confirming that phylogenetic reconstructions conducted with nitrogen fixation genes were not in agreement with 16S rRNA phylograms (Kechriss et al. 2006). The highly conserved nature of the nifH and nifD genes in comparison to nifK (Raymond et al. 2004) could also explain the discrepancies observed in the phylogeny, where mutational drift of the nifK gene may have overridden the translational efficiency that keeps mutations rates in codon regions low (Sur, Sen, and Bothra 2007), resulting in more phylogenetically informative nucleotide positions.

Alterations to the translation efficiency may also be responsible for nucleotide substitutions, resulting in the observed incongruent positioning between the nif phylograms. As nucleotide frequencies are known to vary significantly between organisms for a large number of genes (Galtier, Tourasse, and Gouy 1999), diverging nucleotide frequencies between the *Leptospirillum* species could have contributed to the phylogenetic discrepancies documented.

6.4.2.2 Nif protein phylogeny

The translated amino acid sequences of nifHDK genes indicated a congruent evolutionary history when examined separately and as a NifHDK concatamer sequence, demonstrating *L. ferrodiazotrophum* divergence from a common ancestor, prior to *L. ferrooxidans* and *L. ferrihilum*. Application of the amino acid sequences in place of nucleotides counteracts any substitution bias or saturation effects occurring at the third codon position that can hinder phylogenetic resolution (Hashimoto et al. 1995; Mindell 1996). With the *Leptospirillum* nif genes, multiple substitutions at the third codon that did not result in a non-synonymous mutation, may have obscured the actual phylogenetic relationship.
Thus, phylogeny utilizing the translated amino acids better reflected the actual evolutionary route of the nifHDK genes within the _Leptospirillum_ genus. The high congruency between the three trees suggests that even the shorter NifH sequence (275 AA) provided abundant phylogenetic information for the reconstruction of an accurate Nif evolutionary pathway.

The amino acid phylogenies depicting _L. ferrodiazotrophum_ divergence separate from a common ancestor prior to the other _Leptospirillum_ members were supported by high PP and BS values. Housekeeping phylograms are in agreement with this evolutionary pathway. Although the _Leptospirillum_ species were grouped together with high confidence, the apparent evolutionary age of _L. ferrooxidans_ and _L. ferriphilum_ changed depending on the Nif sequence analysed. As demonstrated in Chapter 5, there are additional RBS upstream of the _nifD_ and _nifK_ genes, indicating possible acquisition and/or evolution of these genes occurring separately. Nevertheless, the high sequence identity of all the Nif proteins and resulting congruent phylogenograms indicates a functional constraint on the structure of these proteins within each _Leptospirillum_ species.

HGT of entire operons is possible (Price, Dehal, and Arkin 2008), allowing for the rapid gain of new functionality. However, the complex cross talk of genes necessary for nitrogen fixation to occur means that just the acquisition of the structural _nifHDK_ genes in _Leptospirillum_ species would not result in successful nitrogen fixation. As growth for all species was demonstrated in the absence of nitrogen (Chapter 2), as well as increased mRNA transcription of various _nif_ genes (Chapter 3), vertical decent of these metabolic genes appears more likely than acquisition via HGT. However, the differences in nucleotide sequences between the genes could be the result of genetic material exchange with micro-organisms sharing their habitat or mutations during DNA replication.

The concatenated NifHDK sequence demonstrated a close evolutionary relationship with _Methylacidiphilum fumariolicum_ (Figure 6.15), an aerobic, prokaryl species with the ability to use methane as a carbon source. HGT of these genes between the _Leptospirillum_ species appears probable, as the exchange of metabolic genes commonly occurs between organisms cohabiting the same environmental niche. Selective pressure from an environment depleted in soluble nitrogen, may have forced the _Leptospirillum_ species to retain these functional nitrogen fixation genes.

These phylogenetic differences among the _nifHDK_ and housekeeping trees for _Leptospirillum_ species may be attributed to a combination of factors including, varied rate of evolution between the genes, selective adaptation to the environment, mutation, or possibly computational artefacts. Permutation of any of these effects is known to influence tree topographies (Hartmann and Barnum 2010). The small number of _Leptospirillum_ sequences available for phylogenetic analysis may have also hindered this study, hampering resolution of the evolutionary pathway. These slight discrepancies between the _nif_ phylogenograms may
also be explained by the possibility of a gene duplication event, whereby more than one copy of the nifHDK genes is present. Gene duplication has occurred in the evolution of nif gene sequences, with some prokaryotes such as A. vinelandii possessing multiple copies (Robson, Woodley, and Jones 1986). As nifD and nifK genes are known to be products of an ancient gene duplication event (Fani, Gallo, and Lio 2000), potentially the primers used in this study only amplified one of the nif alleles, resulting in the incongruence seen between the phylogenetic trees. By sequencing the Leptospirillum species genomes, the number of nifHDK genes carried could be confirmed.

6.4.3 The phylogenetic methods chosen

Both ML and BI methods attempt to construct the ‘exact’ tree—one that reflects evolutionary pathway of the organism. No phylogenetic inferences were made using Maximum Parsimony (MP), as it is the most sensitive method to LBA artefacts (Philippe et al. 2005), that obscure the true evolutionary history when species evolve at a rapid rate (Bergsten 2005). Probabilistic methods, specifically ML and BI demonstrate greater robustness than MP with less LBA (Ho and Jermiin 2004; Huelsenbeck 1995). These methods are also more suitable to handle real substitution patterns than MP methods, and help reduce the impact of inconsistent phylogeny caused by suboptimal model selection (Steel 2005). However, the phylogenies constructed from ML and BI weren’t always entirely consistent, raising the question of the ‘true’ phylogenetic position of the Leptospirillum strains.

The differences between the methods of analysis of exactly the same data can partially be explained by different biases and assumptions of the algorithms used, although, the lack of resolution between these models could also be attributed to the fact that even complex models can not replicate all aspects of biological reality (Felsenstein 1985; Guindon and Gascuel 2003; Kühner and Felsenstein 1994). Incongruence may have also arisen due to the reliability of the sequence alignment, regions of the sequences that were selectively retained or excluded, differences in the evolutionary rates among species, or even as a result of bias in the taxa sampling, as sample taxa were selected based on their homology to the nifH gene in Leptospirillum. The influence of these aspects on branching order of the species in phylogenetic trees cannot be predicted.

6.5 Conclusions

The conserved and variable regions within each of the three nif genes allowed for them to be used as a phylogenetic markers, clearly resolving their evolutionary relationships in regards to each other, and distinctly separating the Leptospirillum species from other diazotrophs. Inferred Leptospirillum phylogenetic trees demonstrated distinct topologies when
analysed with housekeeping and metabolic based genes. The accuracy and use of the *gyrB* and *rpoB* genes to supplement the 16S rRNA phylogenetic analysis clearly resolved *Leptospirillum* evolution, demonstrating *L. ferrodiazotrophicum* to be the most ancient of the three species examined. Phylogenetic analysis with the NifHDK protein encoding genes were in agreement with this statement and produced more accurate phylograms than those inferred from the nucleotide sequences. Of all three *Leptospirillum* species, divergence of *L. ferrooxidans* appears most recent.

As the 16S rRNA and concatenated NifHDK sequences in the *Leptospirillum* species demonstrated congruent evolutionary pathways, it is likely that the *nif* genes are of ancient origin and their retention and maintenance by the *Leptospirillum* species should be credited to positive selective pressure encountered in the environment, where their use will enable survival.
Chapter 7

General Discussion and Future Considerations

Nitrogen is a vital element for bacterial life, and necessary for the production of biomolecules such as proteins, nucleic acids and other cellular components. Assimilation of available nitrogen by bacteria is commonly in the form of NH$_4^+$; however, a small subset of prokaryotes can reduce atmospheric dinitrogen to NH$_4^+$. This process referred to as BNF, described in Chapter 1 is catalysed by the nitrogenase enzyme complex, and requires extensive the cross talk of more than 80 metabolic regulatory genes (Dos Santos and Dean 2011). The ability of a microorganism to perform BNF in a nutrient-limited ecosystem is a beneficial trait, allowing survival when soluble nitrogen levels have been exhausted.

*Leptospirillum* species inhabit nutrient poor environments. As such the capacity of *Leptospirillum* species to function under nitrogen limiting conditions was of interest and provided the framework for this thesis. Primarily, investigations into the impact of growth media lacking soluble nitrogen on *Leptospirillum* species' survivability were performed, to determine whether or not mechanisms for nitrogen fixation were engaged. Previous studies reported that *L. ferrooxidans* demonstrated properties associated with nitrogen fixation (Norris, Murrell, and Hinson 1995; Parro and Moreno-Paz 2003) when levels of soluble nitrogen in the surrounding environment were depleted. However, examination of the underlying processes had not been studied. Additionally, as the description of *L. ferriphilum* is fairly recent (Coram and Rawlings 2002) as is *L. ferrodiazotrophum* (Tyson et al. 2004), detailed studies regarding their nitrogen fixation capabilities have yet to be undertaken.

With advances in molecular biology and bioinformatics methods, a variety of techniques are now available to assess diazotrophy in microorganisms. Through analysis of growth studies, gene expression, structural homology modelling and proteomics, the effect of soluble nitrogen deprivation on *Leptospirillum* species was assessed in an effort to elucidate whether or not nitrogen fixation was occurring.

Continued growth of all *Leptospirillum* species in the absence of soluble nitrogen was con-
firmed by observations of cell proliferation and Fe\(^{2+}\) oxidation. Growth of *L. ferrooxidans* and *L. ferrodiazotrophum* in these conditions was expected (previous studies Parro and Moreno-Paz 2003; Tyson et al. 2005). Of all *Leptospirillum* species analysed, *L. ferriphilum* demonstrated the fastest growth, and Fe\(^{2+}\) oxidation rates when soluble nitrogen was absent. This was interesting as *L. ferriphilum* was assumed to not contain for the structural nitrogenase enzyme (Garcia-Moyano et al. 2008; Levican et al. 2008; Tyson et al. 2004). However, this study demonstrated the presence of these genes providing an explanation for the continued survival and growth of *L. ferriphilum*. Exposure of all three *Leptospirillum* species to conditions without soluble nitrogen resulted in the continued Fe\(^{2+}\) oxidation in non-shaking conditions, without the generation of increased cell mass. As utilization of the nitrogenase enzyme for N\(_2\) fixation is energetically expensive, this increase in Fe\(^{2+}\) oxidation could possibly be attributed to the engagement of this pathway, where successfully reduced molecular N was incorporated to maintain essential biosynthesis processes, rather than for proliferation.

The nitrogenase protein complex is composed of two components, genetically determined by three genes, *nifH, nifD* and *nifK*, linked in a contiguous operon (Raymond et al. 2004). DNA sequences encoding these genes are highly conserved amongst nitrogen fixing organisms (Dixon and Kahn 2004), and are often used as markers to identify diazotrophic potential. The *nifHDK* genes in a strain of *L. ferrooxidans* (*L. ferrooxidans* sp 3.2) have been previously identified (Parro and Moreno-Paz 2003). This study confirmed their presence in another strain of *L. ferrooxidans*, *L. ferrooxidans* DSM 2705\(^T\). The *L. ferrodiazotrophum* C5 analysed in this study demonstrated homology of the *nifHDK* genes to that of *L. ferrodiazotrophum UBA1* reported by Tyson et al. 2005. The presence of this operon in two species of *Leptospirillum* enabled the search towards uncovering the same genes in *L. ferriphilum*. Ultimately, the first three genes of the *nif* operon encoding for the nitrogenase enzyme, *nifHDK*, were identified. Analysis of the 4093 long base pair fragment revealed high sequence similarity to those nitrogenase genes from *L. ferrooxidans* and *L. ferrodiazotrophum*. Further evaluations of the *nifHDK* sequences demonstrated that the nitrogenase protein encoded is a Mo-dependent enzyme, and not an alternative form reliant on vanadium (Lee, Hu, and Ribbe 2010) or iron (Davis et al. 1996) for the metallocluster. As the Mo-dependent nitrogenase is deemed the most effective enzyme (Seefeldt, Hoffman, and Dean 2009) for nitrogen fixation (excluding temperatures below 10 °C Miller and Eady 1988), this appears to be the optimal nitrogenase enzyme for all *Leptospirillum* species considering their temperate environment.

In Chapter 3, it was theorized that the lack of acetylene reduction by cultures of *L. ferriphilum* that demonstrated positive proliferation in ammonium deficient conditions could be attributed to the presence of alternative structural nitrogenase genes (*vnfHDK/afnHDK*). The identification of the *nifHDK* genes in *L. ferriphilum* depicted in Chapter 5 demonstrated
a lack of an additional vnfG or anfG gene positioned between the nifD and nifK genes. This gene is common to all bacteria with alternate nitrogenase enzymes such as A. vinelandii (Bishop, Jarlenski, and Hetherington 1982), Rhodopseudomonas palustris (Oda and Harwood 2005) and Clostridium pasteurianum, and therefore the absence of a vnfG/anfG gene means the production of a secondary nitrogenase enzyme in L. ferrophilum is highly unlikely.

Conservation of the nifHDK genes among all three Leptospirillum species studied in comparison to sequences from known diazotrophs was high. Structural analysis of these genes identified no deletions or insertions that would result in a non-functional nitrogenase enzyme. Therefore, the continued proliferation of all Leptospirillum species recorded in the absence of soluble nitrogen is likely to be due to nitrogenase activity, enabling nitrogen fixation, with subsequent assimilation of fixed N₂. The retention of these structural nitrogenase genes may provide a potential survival mechanisms for when fluctuation in levels of soluble N are encountered, allowing for the maintenance of homoeostasis – enabling a selective advantage in nutrient poor environments.

Identification of an organism is often initially performed by analysis of its rRNA genes (Woese 1987). This however does not provide information regarding the function and activity in the environment. Profiling of Leptospirillum mRNA transcription by q-RT-PCR methods identified changes to numerous nif gene expression levels when exposed to a lack of soluble nitrogen. The manifestation of nifH-D-K mRNA is a good indicator of activity as they are known to have stringent transcriptional regulatory mechanisms (Storz and Hengge-Aronis 2000) preventing their activation when levels of NH₃⁺ are not limiting. Exposure of L. ferrooxidans and L. ferrodiazotrophum to conditions lacking soluble nitrogen resulted in the increased expression of these three candidate genes responsible for the production of the nitrogenase enzyme. The increased expression of these genes in the two Leptospirillum species mentioned is highly indicative of nitrogenase enzyme production, resulting in the engagement of nitrogen fixation machinery to supply the cell with a nitrogenous source. These findings corroborate the continued proliferation, Fe²⁺ oxidation and acetylene reduction demonstrated in Chapter 2, indicative of survival in a nitrogen depleted environment via nitrogen fixation. As no mutations or deletions were detected in the detailed structural analysis of these genes in L. ferrooxidans and L. ferrodiazotrophum, functional nitrogenase activity in environments depleted in soluble nitrogen appears credible.

As the nifHDK genes of L. ferrophilum were identified late in this study, time constraints did not permit examination of mRNA transcription rates by q-RT-PCR. In an effort to define why continued proliferation of L. ferrophilum occurred in the absence of soluble nitrogen, but with no demonstrable acetylene reduction activity, previously identified genes essential to permitting (nifA) and assisting nitrogen fixation (nifS-U-hesB) were examined. The detection of increased mRNA transcription of each of these genes when soluble nitrogen was absent
supports the theory of nitrogen fixation by this species. Future studies conducted on transcription of the \( \text{nifHDK} \) genes would allow for greater confidence in declaring diazotrophy in this species of \( \text{Leptospirillum} \).

The increased transcriptional mRNA expression of all the \( \text{Leptospirillum nif} \) genes clearly indicates a functional response to the absence of soluble nitrogen, and supports the concept of diazotrophy in all three species. This study also identified the \( \text{nifA} \) gene in \( \text{L. ferrooxidans DSM 2705} \) (Genbank Accession JN390681), and assessments of whole shotgun sequence data in the NCBI database also located it in \( \text{L. ferrodiazotrophum UBA1} \) (EES52452.1). Future mRNA profiling of these genes would provide additional insight into the changes occurring at the level of transcription, as \( \text{nifA} \) is responsible for activation of the \( \text{nifH} \) promoter, permitting \( \text{nifH} \) transcription for production of the Fe nitrogenase component. As nitrogen fixation requires the activation of many genes, examination of other cellular transformations would highlight the numerous modifications to metabolic and physiological processes that occur in \( \text{Leptospirillum} \) species under conditions of absent soluble nitrogen.

Even though transcription of mRNA is often used as an indicator of phenotypic expression (Gordon et al. 2009), it is only the transient step necessary to make a functional protein. Identification of proteins produced can highlight function and metabolic activity (Wilmes, Wexler, and Bond 2008) in an environment. Therefore by assessing protein abundance of \( \text{Leptospirillum} \) species subject to the absence of soluble nitrogen, representation of the metabolic changes that occurred could be determined. The proteomic approaches used to analyse whole cell extracted proteins from \( \text{Leptospirillum} \) species (chapter 4) demonstrated that well separated protein patterns were achieved for all species under test and control conditions. Common to all \( \text{Leptospirillum} \) species, the absence of soluble nitrogen resulted in alterations to protein abundance. In \( \text{L. ferriphilum} \) and \( \text{L. ferrodiazotrophum} \), the absence of soluble nitrogen increased the abundance of certain metabolic proteins including those responsible for de novo synthesis of nucleic and amino acids, glycolysis, signalling and energy transport. Substantial changes in the relative concentration of proteins from \( \text{L. ferrooxidans} \) only identified those whose abundance decreased during exposure to absent soluble nitrogen, particularly membrane channel and stress response proteins.

The relative abundance of the proteins responsible for forming the nitrogenase enzyme complex (NifH, D and K) are principal candidates for modification as they catalyse the reduction of dinitrogen into ammonia (Ow et al. 2007). Examination and MS identification of proteins selected whose abundance had changed >2 fold, did not identify specific nitrogen fixation proteins in any \( \text{Leptospirillum} \) species. It is possible that at the time of sampling, abundance of these particular proteins had not sufficiently increased, and weren't detected computationally as being of particular significance. Turnover of nitrogenase components may also explain lack of detection in these species, where continued proliferation during exponential growth and reduction of acetylene in two of the three \( \text{Leptospirillum spp.} \) pos-
itive presumed nitrogen fixation.

Proteomic observations made in this study provided a greater understanding of the metabolic changes occurring in *Leptospirillum* spp, resulting from soluble nitrogen deprivation. In each species, different ancillary pathways/processes affected and possibly engaged were highlighted, during what appeared to be, molecular nitrogen fixation. A further, more detailed proteomic analysis of all *Leptospirillum* species under these conditions would permit proteomic profiling of the complex biological systems involved, piecing together the various regulatory processes responsible for continued *Leptospirillum* spp growth in the absence of soluble nitrogen. This would provide detailed knowledge regarding proteins of low abundance and potentially identify those involved with nitrogen fixation.

Supporting the changes observed in mRNA transcription and protein abundance were assessments of *Leptospirillum* species’ ability to reduce acetylene. Diazotrophy has been confirmed in various bacterial species by measuring BNF using the acetylene reduction assay (Bishop et al. 1986; Dalton and Whittenbury 1976). The application of this technique in this study confirmed data previously recorded for *L. ferrooxidans* (Norris, Murrell, and Hinson 1995; Sato et al. 2009), which demonstrated acetylene reduction. *L. ferrodiazotrophum* was shown to reduce acetylene at greater rates than recorded with *L. ferrooxidans*. The lack of acetylene reduction activity by *L. ferriphilum* was unexpected due to the continued proliferation and Fe$^{2+}$ oxidation in the absence of soluble nitrogen. As evaluation of the assay specificity was confirmed by *A. vinelandii* reduction of acetylene to ethylene in the presence of oxygen, it is likely that the levels produced of ethylene produced by *L. ferriphilum* were too low and below the level of detection of the system. The discrepancies in proliferation, iron oxidation and acetylene reduction abilities highlights a metabolic difference between these three *Leptospirillum* species when exposed to an environment devoid of soluble nitrogen. As these experiments were conducted in pure cultures, analysis of *Leptospirillum* behaviour in mixed populations under these conditions would be of great interest and possibly highlight any synergistic relationships that occur between organisms occupying the same environmental niche.

Phylogenetic analysis of individual *Leptospirillum nif* genes demonstrated incongruent evolutionary pathways when analysed as DNA sequences. In these instances it is likely that base pair substitutions occurring at the third codon (as was demonstrated by the high GC% content in Chapter 5) masked the true evolutionary route of *nif* inheritance. Nonetheless, assessment of evolution via *nif* protein sequence phylogeny was additionally supported when protein coding genes were concatenated, resolving a clear evolutionary path of *nif* inheritance in *Leptospirillum*. The most accurate evolutionary pathway of the *nif* genes demonstrated that *L. ferrodiazotrophum* is the most ancient species of the genus, whereas *L. ferriphilum* and *L. ferrooxidans* diverged more recently.
General Discussion and Future Considerations

Whether or not nitrogen fixation evolved via vertical descent or horizontal gene transfer in these \textit{Leptospirillum} species still remains unknown. Both the \textit{nifH} and \textit{nifD} genes were highly conserved and as such, were useful for determining \textit{Leptospirillum} diazotrophic grouping based on sequence identity (group I). However, this high level of conservation could also be responsible for obscuring the true evolutionary pathway, limiting the phylogenetic resolution that was obtained. As the \textit{nifHDK} genes are organized in a contiguous fashion in all three \textit{Leptospirillum} species studied – structurally conserved for optimal transcription of the nitrogenase enzyme (Wang, Chen, and Johnson 1988), it was assumed that inheritance of each gene would be identical. Yet, the presence of intact RBS upstream of each of the genes start codons, a theoretical transcriptional terminator as well as the differing rates of mRNA transcription recorded meant that there was a possibility that the inheritance of these genes could have occurred at different stages of \textit{Leptospirillum} evolution.

As the phylogeny depicted with the concatenated Nif sequences was congruent with the 16S rRNA, \textit{Leptospirillum nif} genes appear to have evolved with the 16S rRNA and are unlikely to be a recent acquisition from neighbouring diazotrophs. Exchange of genetic material with \textit{Acidithiobacillus} or \textit{Methylacidiphilum}, of which the \textit{Leptospirillum} species share a joint habitat appears possible, as do transcriptional mutations, however these alterations and possible patterns of variable inheritance did not result in functional changes to the encoded proteins.

Unlike some diazotrophs harbouring multiple copies of \textit{nifH} that result in the exhibition of different phylogenetic relationships, this study only identified one \textit{nifH} gene in the \textit{Leptospirillum} species, thus helping reduce uncertainty of the phylograms depicted. Interestingly, the less structurally conserved \textit{nifK} gene demonstrated congruent evolution with the 16S rRNA, at both the DNA and AA level. This decreased conservation may have allowed for more informative variable sites of each of the three species. It is possible that the \textit{nifK} gene could be used as a marker to identify and differentiate \textit{Leptospirillum} species. Further research identifying more \textit{nifHDK} genes from other strains of these \textit{Leptospirillum} type species could aid in improving the phylogenetic resolution of these protein encoding genes.

### 7.1 Final Conclusions and Outlook

The purpose of this study was to determine the response of \textit{Leptospirillum} species when subject to conditions depleted of soluble nitrogen, possibly identifying processes involved in nitrogen fixation. The most significant outcome of this work was the identification of the \textit{nifHDK} genes in \textit{L. ferrilphilum} and observations of its continued ability to survive in environments lacking in soluble nitrogen. The sustained proliferation of all three \textit{Leptospirillum} species studied, combined with the structural analysis of these genes, provides evidence
of functional nitrogen fixation in each of these species. The examination of other *Leptospirillum* strains for these nitrogen fixing genes and studies regarding the ability to survive in nitrogen depleted environments would assist in determining whether or not the genus as a whole is capable of fixing nitrogen. The origins of the *nif* genes in these bacteria can not be categorically determined, however, the evidence presented here suggests that the genes were not a recent acquisition as evolutionary inheritance is congruent with that of the 16S rRNA, suggesting selective environment pressures retained the *nif* genes. As no *Leptospirillum* species has had its entire genome sequenced, advances in this regard would be advantageous, and assist in the development of new proteomic and genomic assays to evaluate nitrogen fixation processes.

Collectively, the work presented in this thesis provides new insights into the nature of three *Leptospirillum* species, with the possibility of nitrogen fixation helping to explain their success and ability to dominate microbial communities in nutrient restrictive environments.
Appendix A

Supplementary Information - I

A.1 Quality assurance and standard curves

<table>
<thead>
<tr>
<th>SAMPLE NAME</th>
<th>9K Media (+NH₄)₂SO₄</th>
<th>9K Media (-NH₄)₂SO₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>SERIAL #</td>
<td>80672</td>
<td>80673</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ELEMENT</th>
<th>9K Media (+NH₄)₂SO₄</th>
<th>9K Media (-NH₄)₂SO₄</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH₄N (mg/L)</td>
<td>59.12</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>NO₃N (mg/L)</td>
<td>0.1</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>S (mg/L)</td>
<td>808</td>
<td>794</td>
</tr>
<tr>
<td>P (mg/L)</td>
<td>110.3</td>
<td>109.1</td>
</tr>
<tr>
<td>K (mg/L)</td>
<td>113.2</td>
<td>110.1</td>
</tr>
<tr>
<td>Na (mg/L)</td>
<td>2.41</td>
<td>2.16</td>
</tr>
<tr>
<td>Ca (mg/L)</td>
<td>35.54</td>
<td>39.77</td>
</tr>
<tr>
<td>Mg (mg/L)</td>
<td>25.27</td>
<td>22.78</td>
</tr>
<tr>
<td>Cu (mg/L)</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>Zn (mg/L)</td>
<td>1.6</td>
<td>1.51</td>
</tr>
<tr>
<td>Mn (mg/L)</td>
<td>&lt; 0.05</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Fe (mg/L)</td>
<td>0.33</td>
<td>0.4</td>
</tr>
<tr>
<td>B (mg/L)</td>
<td>0.12</td>
<td>&lt; 0.05</td>
</tr>
<tr>
<td>Cl (mg/L)</td>
<td>59.9</td>
<td>67</td>
</tr>
<tr>
<td>pH</td>
<td>1.67</td>
<td>1.66</td>
</tr>
<tr>
<td>COND. (dS/m)</td>
<td>10.16</td>
<td>11.16</td>
</tr>
</tbody>
</table>
Table A.2 – Ferrous Standards for Iron oxidation

<table>
<thead>
<tr>
<th>Concentration (ppm) of Ferrous Standard</th>
<th>Nanodrop reading</th>
<th>for 0.1mm pathlength</th>
<th>Calculated Ferrous (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>0.0086</td>
<td>0.086</td>
<td>25.66</td>
</tr>
<tr>
<td>50</td>
<td>0.0151</td>
<td>0.151</td>
<td>49.09</td>
</tr>
<tr>
<td>80</td>
<td>0.0229</td>
<td>0.229</td>
<td>77.21</td>
</tr>
<tr>
<td>100</td>
<td>0.0302</td>
<td>0.302</td>
<td>103.52</td>
</tr>
<tr>
<td>200</td>
<td>0.0566</td>
<td>0.566</td>
<td>198.68</td>
</tr>
<tr>
<td>500</td>
<td>0.138</td>
<td>1.38</td>
<td>492.09</td>
</tr>
<tr>
<td>1000</td>
<td>0.28</td>
<td>2.8</td>
<td>1003.93</td>
</tr>
</tbody>
</table>

Figure A.1 – Standard curve for Fe$^{2+}$ oxidation
Figure A.2 - Calibration curve for ARA generated by plotting peak heights against known ethylene concentrations.
Figure A.3 – *Leptospirillum* spp, non-Log2 transformed mRNA replicate data, normalised to 16S rRNA and *gyrB* gene expression. Fold change range from red (decreased) to green (increased).
### Table A.3 - *L. ferrodiazotrophum* C5 relative synonymous codon usage breakdown for *nifH*, *nifD* and *nifK*. Corresponds to Figure 5.5a

#### (a) *nifH*

<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>3</td>
<td>0.86</td>
<td>UCU(S)</td>
<td>0</td>
<td>0</td>
<td>UAU(Y)</td>
<td>4</td>
<td>0.89</td>
<td>UGU(C)</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>4</td>
<td>1.14</td>
<td>UCC(S)</td>
<td>11</td>
<td>4.71</td>
<td>UAC(Y)</td>
<td>5</td>
<td>1.11</td>
<td>UGC(C)</td>
<td>4</td>
<td>1.6</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>0</td>
<td>0</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
<td>UGA(*)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>3</td>
<td>0.82</td>
<td>UCG(S)</td>
<td>0</td>
<td>0</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
<td>UGG(W)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>3</td>
<td>0.82</td>
<td>CUC(P)</td>
<td>0</td>
<td>0</td>
<td>CAU(H)</td>
<td>2</td>
<td>1.33</td>
<td>CGU(R)</td>
<td>2</td>
<td>0.92</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>7</td>
<td>1.91</td>
<td>CCC(P)</td>
<td>4</td>
<td>2</td>
<td>CAC(H)</td>
<td>1</td>
<td>0.67</td>
<td>CGG(R)</td>
<td>5</td>
<td>2.31</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>0</td>
<td>0</td>
<td>CCA(P)</td>
<td>0</td>
<td>0</td>
<td>CAA(Q)</td>
<td>1</td>
<td>0.2</td>
<td>CGA(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>9</td>
<td>2.45</td>
<td>CCG(P)</td>
<td>4</td>
<td>2</td>
<td>CAG(Q)</td>
<td>9</td>
<td>1.9</td>
<td>CGG(R)</td>
<td>4</td>
<td>1.85</td>
</tr>
<tr>
<td>AUU(I)</td>
<td>0</td>
<td>0</td>
<td>AAC(T)</td>
<td>0</td>
<td>0</td>
<td>AAA(K)</td>
<td>5</td>
<td>3.07</td>
<td>AGA(R)</td>
<td>2</td>
<td>0.92</td>
</tr>
<tr>
<td>AUC(I)</td>
<td>16</td>
<td>2.29</td>
<td>ACC(T)</td>
<td>7</td>
<td>1.87</td>
<td>AAC(N)</td>
<td>8</td>
<td>1.33</td>
<td>AGC(S)</td>
<td>3</td>
<td>1.29</td>
</tr>
<tr>
<td>AUA(I)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>0</td>
<td>0</td>
<td>AAA(K)</td>
<td>3</td>
<td>0.35</td>
<td>AGA(R)</td>
<td>2</td>
<td>0.92</td>
</tr>
<tr>
<td>AUG(M)</td>
<td>12</td>
<td>1</td>
<td>ACC(T)</td>
<td>8</td>
<td>2.13</td>
<td>AAC(N)</td>
<td>14</td>
<td>1.65</td>
<td>AGG(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GUC(V)</td>
<td>2</td>
<td>0.35</td>
<td>GCC(A)</td>
<td>2</td>
<td>0.28</td>
<td>GAL(D)</td>
<td>8</td>
<td>1.0</td>
<td>GGU(G)</td>
<td>3</td>
<td>0.4</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>10</td>
<td>1.74</td>
<td>GCG(A)</td>
<td>22</td>
<td>3.03</td>
<td>GAC(D)</td>
<td>8</td>
<td>1.0</td>
<td>GGC(G)</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>0</td>
<td>0</td>
<td>GCA(A)</td>
<td>1</td>
<td>0.14</td>
<td>GAA(E)</td>
<td>7</td>
<td>0.56</td>
<td>GGA(G)</td>
<td>10</td>
<td>1.33</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>11</td>
<td>1.91</td>
<td>GCG(A)</td>
<td>4</td>
<td>0.55</td>
<td>GAG(E)</td>
<td>18</td>
<td>1.44</td>
<td>GGG(G)</td>
<td>2</td>
<td>0.27</td>
</tr>
</tbody>
</table>

#### (b) *nifD*

<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>8</td>
<td>0.76</td>
<td>UCU(S)</td>
<td>1</td>
<td>0.26</td>
<td>UAU(Y)</td>
<td>8</td>
<td>0.64</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>13</td>
<td>1.24</td>
<td>UCC(S)</td>
<td>10</td>
<td>2.61</td>
<td>UAC(Y)</td>
<td>17</td>
<td>1.36</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>1</td>
<td>0.26</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>1</td>
<td>0.29</td>
<td>UCG(S)</td>
<td>7</td>
<td>1.83</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>4</td>
<td>1.14</td>
<td>CCC(P)</td>
<td>0</td>
<td>0</td>
<td>CAU(H)</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>8</td>
<td>2.29</td>
<td>CCA(P)</td>
<td>10</td>
<td>2.11</td>
<td>CAC(H)</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>0</td>
<td>0</td>
<td>CCA(P)</td>
<td>0</td>
<td>0</td>
<td>CAO(Q)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>8</td>
<td>2.29</td>
<td>CCG(P)</td>
<td>9</td>
<td>1.89</td>
<td>CAG(Q)</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>AUU(I)</td>
<td>5</td>
<td>0.38</td>
<td>ACC(T)</td>
<td>0</td>
<td>0</td>
<td>AAA(K)</td>
<td>4</td>
<td>0.5</td>
</tr>
<tr>
<td>AUC(I)</td>
<td>35</td>
<td>2.63</td>
<td>ACC(T)</td>
<td>15</td>
<td>2.73</td>
<td>AAC(N)</td>
<td>12</td>
<td>1.5</td>
</tr>
<tr>
<td>AUA(I)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>0</td>
<td>0</td>
<td>AAA(K)</td>
<td>14</td>
<td>0.52</td>
</tr>
<tr>
<td>AUG(M)</td>
<td>15</td>
<td>1</td>
<td>ACC(T)</td>
<td>7</td>
<td>1.27</td>
<td>AGG(K)</td>
<td>31</td>
<td>1.48</td>
</tr>
<tr>
<td>GUC(V)</td>
<td>7</td>
<td>0.8</td>
<td>GCC(A)</td>
<td>2</td>
<td>0.31</td>
<td>GAL(D)</td>
<td>6</td>
<td>0.4</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>19</td>
<td>2.17</td>
<td>GCC(A)</td>
<td>14</td>
<td>2.15</td>
<td>GAC(D)</td>
<td>24</td>
<td>1.6</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>0</td>
<td>0</td>
<td>GCC(A)</td>
<td>2</td>
<td>0.31</td>
<td>GAA(E)</td>
<td>15</td>
<td>0.81</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>9</td>
<td>1.03</td>
<td>GCG(A)</td>
<td>8</td>
<td>1.23</td>
<td>GAG(E)</td>
<td>22</td>
<td>1.19</td>
</tr>
</tbody>
</table>

#### (c) *nifK*

<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>9</td>
<td>0.84</td>
<td>UCU(S)</td>
<td>2</td>
<td>0.4</td>
<td>UAU(Y)</td>
<td>7</td>
<td>0.87</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>19</td>
<td>1.36</td>
<td>UCC(S)</td>
<td>11</td>
<td>2.2</td>
<td>UAC(Y)</td>
<td>14</td>
<td>1.33</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>1</td>
<td>0.2</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>2</td>
<td>0.28</td>
<td>UCG(S)</td>
<td>6</td>
<td>1.2</td>
<td>UAG(*)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>8</td>
<td>1.12</td>
<td>CUC(P)</td>
<td>2</td>
<td>0.32</td>
<td>CAU(H)</td>
<td>9</td>
<td>0.96</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>20</td>
<td>2.79</td>
<td>CCC(P)</td>
<td>12</td>
<td>1.92</td>
<td>CAC(H)</td>
<td>10</td>
<td>1.05</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>0</td>
<td>0</td>
<td>CCA(P)</td>
<td>1</td>
<td>0.16</td>
<td>CAO(Q)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>13</td>
<td>1.81</td>
<td>CCG(P)</td>
<td>10</td>
<td>1.6</td>
<td>CAG(Q)</td>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>AUU(I)</td>
<td>3</td>
<td>0.39</td>
<td>ACC(T)</td>
<td>1</td>
<td>0.13</td>
<td>AAA(K)</td>
<td>10</td>
<td>0.83</td>
</tr>
<tr>
<td>AUC(I)</td>
<td>20</td>
<td>2.61</td>
<td>ACC(T)</td>
<td>21</td>
<td>2.8</td>
<td>AAC(N)</td>
<td>14</td>
<td>1.17</td>
</tr>
<tr>
<td>AUA(I)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>3</td>
<td>0.4</td>
<td>AAA(K)</td>
<td>2</td>
<td>0.13</td>
</tr>
<tr>
<td>AUG(M)</td>
<td>19</td>
<td>1</td>
<td>ACC(T)</td>
<td>5</td>
<td>0.67</td>
<td>AGA(R)</td>
<td>28</td>
<td>1.87</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>7</td>
<td>0.85</td>
<td>GCC(A)</td>
<td>3</td>
<td>0.32</td>
<td>GAL(D)</td>
<td>9</td>
<td>0.64</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>15</td>
<td>1.82</td>
<td>GCC(A)</td>
<td>22</td>
<td>2.38</td>
<td>GAC(D)</td>
<td>19</td>
<td>1.36</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>0</td>
<td>0</td>
<td>GCC(A)</td>
<td>3</td>
<td>0.32</td>
<td>GAA(E)</td>
<td>14</td>
<td>0.72</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>11</td>
<td>1.33</td>
<td>GCG(A)</td>
<td>9</td>
<td>0.97</td>
<td>GAG(E)</td>
<td>25</td>
<td>1.28</td>
</tr>
</tbody>
</table>

---

**171**
<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>1</td>
<td>0.29</td>
<td>UCU(S)</td>
<td>1</td>
<td>0.5</td>
<td>UAU(Y)</td>
<td>2</td>
<td>0.44</td>
<td>UGU(C)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>6</td>
<td>1.71</td>
<td>UCG(S)</td>
<td>7</td>
<td>3.5</td>
<td>UAG(Y)</td>
<td>7</td>
<td>1.56</td>
<td>UGC(C)</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>0</td>
<td>0</td>
<td>UAA(*)</td>
<td>0</td>
<td>0</td>
<td>UGA(*)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>0</td>
<td>0</td>
<td>UC(G)</td>
<td>4</td>
<td>2</td>
<td>UAC(*)</td>
<td>0</td>
<td>0</td>
<td>UGG(W)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUU(L)</td>
<td>3</td>
<td>0.75</td>
<td>CCG(P)</td>
<td>1</td>
<td>0.5</td>
<td>CAU(H)</td>
<td>0</td>
<td>0</td>
<td>CGU(R)</td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>6</td>
<td>1.5</td>
<td>CCC(P)</td>
<td>2</td>
<td>1</td>
<td>CAC(H)</td>
<td>3</td>
<td>2</td>
<td>CGC(R)</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>0</td>
<td>0</td>
<td>CC(A)</td>
<td>0</td>
<td>0</td>
<td>CAA(Q)</td>
<td>0</td>
<td>0</td>
<td>CGA(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>15</td>
<td>3.75</td>
<td>CCG(P)</td>
<td>5</td>
<td>2.5</td>
<td>CAG(Q)</td>
<td>9</td>
<td>2</td>
<td>CGG(R)</td>
<td>7</td>
<td>2.8</td>
</tr>
<tr>
<td>AUG(L)</td>
<td>2</td>
<td>0.33</td>
<td>ACA(T)</td>
<td>2</td>
<td>0.44</td>
<td>AAU(N)</td>
<td>3</td>
<td>0.46</td>
<td>AGU(S)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AUU(U)</td>
<td>16</td>
<td>2.67</td>
<td>ACC(T)</td>
<td>10</td>
<td>2.22</td>
<td>AAC(N)</td>
<td>10</td>
<td>1.54</td>
<td>AGC(S)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AAU(L)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>1</td>
<td>0.22</td>
<td>AAA(K)</td>
<td>6</td>
<td>0.75</td>
<td>AGA(R)</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>AUG(L)</td>
<td>11</td>
<td>1</td>
<td>ACC(T)</td>
<td>5</td>
<td>1.11</td>
<td>AAG(K)</td>
<td>10</td>
<td>1.25</td>
<td>AGG(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GUA(V)</td>
<td>2</td>
<td>0.33</td>
<td>GCU(A)</td>
<td>2</td>
<td>0.28</td>
<td>GAU(D)</td>
<td>6</td>
<td>0.86</td>
<td>GGU(G)</td>
<td>5</td>
<td>0.69</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>11</td>
<td>1.83</td>
<td>GCC(A)</td>
<td>20</td>
<td>2.76</td>
<td>GAC(D)</td>
<td>8</td>
<td>1.14</td>
<td>GGC(G)</td>
<td>15</td>
<td>2.07</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>0</td>
<td>0</td>
<td>GCA(A)</td>
<td>1</td>
<td>0.14</td>
<td>GAA(E)</td>
<td>21</td>
<td>1.56</td>
<td>GGA(G)</td>
<td>7</td>
<td>0.97</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>11</td>
<td>1.83</td>
<td>GCG(A)</td>
<td>6</td>
<td>0.83</td>
<td>GAG(E)</td>
<td>6</td>
<td>0.44</td>
<td>GGG(G)</td>
<td>2</td>
<td>0.28</td>
</tr>
</tbody>
</table>

(a) nifH

<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>3</td>
<td>0.29</td>
<td>UCU(S)</td>
<td>1</td>
<td>0.24</td>
<td>UAU(Y)</td>
<td>7</td>
<td>0.58</td>
<td>UGU(C)</td>
<td>1</td>
<td>0.33</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>18</td>
<td>1.71</td>
<td>UCG(S)</td>
<td>10</td>
<td>2.4</td>
<td>UAG(Y)</td>
<td>17</td>
<td>1.42</td>
<td>UGC(C)</td>
<td>5</td>
<td>1.67</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>0</td>
<td>0</td>
<td>UAA(*)</td>
<td>0</td>
<td>0</td>
<td>UGA(*)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>0</td>
<td>0</td>
<td>UC(G)</td>
<td>11</td>
<td>2.64</td>
<td>UAC(*)</td>
<td>0</td>
<td>0</td>
<td>UGG(W)</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>CUU(L)</td>
<td>2</td>
<td>0.57</td>
<td>CCG(P)</td>
<td>2</td>
<td>0.44</td>
<td>CAU(H)</td>
<td>5</td>
<td>0.77</td>
<td>CGU(R)</td>
<td>2</td>
<td>0.5</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>9</td>
<td>2.57</td>
<td>CCC(P)</td>
<td>7</td>
<td>1.56</td>
<td>CAC(H)</td>
<td>8</td>
<td>1.23</td>
<td>CGC(R)</td>
<td>7</td>
<td>1.75</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>0</td>
<td>0</td>
<td>CC(A)</td>
<td>0</td>
<td>0</td>
<td>CAA(Q)</td>
<td>0</td>
<td>0</td>
<td>CGA(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>10</td>
<td>2.86</td>
<td>CCG(P)</td>
<td>9</td>
<td>2</td>
<td>CAG(Q)</td>
<td>11</td>
<td>2</td>
<td>CGG(R)</td>
<td>15</td>
<td>3.75</td>
</tr>
<tr>
<td>AUG(L)</td>
<td>5</td>
<td>0.38</td>
<td>ACA(T)</td>
<td>1</td>
<td>0.19</td>
<td>AAU(N)</td>
<td>2</td>
<td>0.24</td>
<td>AGU(S)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AUU(U)</td>
<td>35</td>
<td>2.63</td>
<td>ACC(T)</td>
<td>11</td>
<td>2.1</td>
<td>AAC(N)</td>
<td>15</td>
<td>1.76</td>
<td>AGC(S)</td>
<td>3</td>
<td>0.72</td>
</tr>
<tr>
<td>AAU(L)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>0</td>
<td>0</td>
<td>AAA(K)</td>
<td>9</td>
<td>0.46</td>
<td>AGA(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AUG(M)</td>
<td>14</td>
<td>1.1</td>
<td>ACC(T)</td>
<td>9</td>
<td>1.71</td>
<td>AAG(K)</td>
<td>30</td>
<td>1.54</td>
<td>AGG(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>GUA(V)</td>
<td>2</td>
<td>0.22</td>
<td>GCU(A)</td>
<td>1</td>
<td>0.16</td>
<td>GAU(D)</td>
<td>5</td>
<td>0.3</td>
<td>GGU(G)</td>
<td>3</td>
<td>0.27</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>22</td>
<td>2.44</td>
<td>GCC(A)</td>
<td>10</td>
<td>1.6</td>
<td>GAC(D)</td>
<td>28</td>
<td>1.7</td>
<td>GGC(G)</td>
<td>17</td>
<td>1.55</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>0</td>
<td>0</td>
<td>GCA(A)</td>
<td>3</td>
<td>0.48</td>
<td>GAA(E)</td>
<td>20</td>
<td>1.11</td>
<td>GGA(G)</td>
<td>12</td>
<td>1.09</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>12</td>
<td>1.33</td>
<td>GCG(A)</td>
<td>11</td>
<td>1.76</td>
<td>GAG(E)</td>
<td>16</td>
<td>0.89</td>
<td>GGG(G)</td>
<td>12</td>
<td>1.09</td>
</tr>
</tbody>
</table>

(b) nifD

<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>4</td>
<td>0.28</td>
<td>UCU(S)</td>
<td>4</td>
<td>0.83</td>
<td>UAU(Y)</td>
<td>9</td>
<td>0.9</td>
<td>UGU(C)</td>
<td>1</td>
<td>0.4</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>25</td>
<td>1.72</td>
<td>UCG(S)</td>
<td>8</td>
<td>1.66</td>
<td>UAG(Y)</td>
<td>11</td>
<td>1.1</td>
<td>UGC(C)</td>
<td>4</td>
<td>1.6</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>1</td>
<td>0.21</td>
<td>UAA(*)</td>
<td>0</td>
<td>0</td>
<td>UGA(*)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>3</td>
<td>0.43</td>
<td>UC(G)</td>
<td>6</td>
<td>1.24</td>
<td>UAC(*)</td>
<td>1</td>
<td>3</td>
<td>UGG(W)</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>CUU(L)</td>
<td>3</td>
<td>0.43</td>
<td>CCG(P)</td>
<td>1</td>
<td>0.17</td>
<td>CAU(H)</td>
<td>11</td>
<td>1.16</td>
<td>CGU(R)</td>
<td>1</td>
<td>0.23</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>14</td>
<td>2</td>
<td>CCC(P)</td>
<td>7</td>
<td>1.17</td>
<td>CAC(H)</td>
<td>8</td>
<td>0.84</td>
<td>CGC(R)</td>
<td>9</td>
<td>2.08</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>1</td>
<td>0.14</td>
<td>CC(A)</td>
<td>1</td>
<td>0.17</td>
<td>CAA(Q)</td>
<td>0</td>
<td>0</td>
<td>CGA(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>21</td>
<td>3</td>
<td>CCG(P)</td>
<td>15</td>
<td>2.5</td>
<td>CAG(Q)</td>
<td>10</td>
<td>2</td>
<td>CGG(R)</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>AUG(L)</td>
<td>3</td>
<td>0.41</td>
<td>ACA(T)</td>
<td>2</td>
<td>0.24</td>
<td>AAU(N)</td>
<td>11</td>
<td>1.05</td>
<td>AGU(S)</td>
<td>1</td>
<td>0.21</td>
</tr>
<tr>
<td>AUU(U)</td>
<td>19</td>
<td>2.59</td>
<td>ACC(T)</td>
<td>21</td>
<td>2.55</td>
<td>AAC(N)</td>
<td>10</td>
<td>0.95</td>
<td>AGC(S)</td>
<td>9</td>
<td>1.86</td>
</tr>
<tr>
<td>AAU(L)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>0</td>
<td>0</td>
<td>AAA(K)</td>
<td>8</td>
<td>0.57</td>
<td>AGA(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AUG(M)</td>
<td>17</td>
<td>1</td>
<td>ACC(T)</td>
<td>10</td>
<td>1.21</td>
<td>AAG(K)</td>
<td>20</td>
<td>1.43</td>
<td>AGG(R)</td>
<td>3</td>
<td>0.69</td>
</tr>
<tr>
<td>GUA(V)</td>
<td>5</td>
<td>0.67</td>
<td>GCU(A)</td>
<td>3</td>
<td>0.32</td>
<td>GAU(D)</td>
<td>4</td>
<td>0.28</td>
<td>GGU(G)</td>
<td>6</td>
<td>0.6</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>16</td>
<td>2.13</td>
<td>GCC(A)</td>
<td>23</td>
<td>2.42</td>
<td>GAC(D)</td>
<td>25</td>
<td>1.72</td>
<td>GGC(G)</td>
<td>13</td>
<td>1.3</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>0</td>
<td>0</td>
<td>GCA(A)</td>
<td>0</td>
<td>0</td>
<td>GAA(E)</td>
<td>17</td>
<td>0.89</td>
<td>GGA(G)</td>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>9</td>
<td>1.2</td>
<td>GCG(A)</td>
<td>12</td>
<td>1.26</td>
<td>GAG(E)</td>
<td>21</td>
<td>1.11</td>
<td>GGG(G)</td>
<td>11</td>
<td>1.1</td>
</tr>
</tbody>
</table>

(c) nifK

Table A.4 - *L. ferriphilum DSM 14647* relative synonymous codon usage breakdown for *nifH*, *nifD* and *nifK*. Corresponds to Figure 5.5b
### Table A.5 - *L. ferrooxidans* DSM 2705 relative synonymous codon usage breakdown for nifH, nifD and nifK. Corresponds to Figure 5.5c

<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>0</td>
<td>0</td>
<td>UCU(S)</td>
<td>0</td>
<td>0</td>
<td>UAU(Y)</td>
<td>3</td>
<td>0.67</td>
<td>UGU(C)</td>
<td>2</td>
<td>0.8</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>7</td>
<td>2</td>
<td>UCC(S)</td>
<td>8</td>
<td>3.2</td>
<td>UAC(Y)</td>
<td>6</td>
<td>1.33</td>
<td>UGC(C)</td>
<td>3</td>
<td>1.2</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>2</td>
<td>0.8</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
<td>UGA(*)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>4</td>
<td>1</td>
<td>UCG(S)</td>
<td>2</td>
<td>0.8</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
<td>UGG(W)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>6</td>
<td>1.5</td>
<td>CCU(P)</td>
<td>2</td>
<td>0.5</td>
<td>CAU(H)</td>
<td>1</td>
<td>0.67</td>
<td>CGU(R)</td>
<td>2</td>
<td>0.86</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>7</td>
<td>1.75</td>
<td>CCC(P)</td>
<td>3</td>
<td>1.5</td>
<td>CAC(H)</td>
<td>2</td>
<td>1.33</td>
<td>CGC(R)</td>
<td>9</td>
<td>3.86</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>0</td>
<td>0</td>
<td>CCA(P)</td>
<td>2</td>
<td>1</td>
<td>CA(A)</td>
<td>0</td>
<td>0</td>
<td>CGA(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>7</td>
<td>1.75</td>
<td>CGG(P)</td>
<td>2</td>
<td>1</td>
<td>CAG(Q)</td>
<td>9</td>
<td>2</td>
<td>CGG(R)</td>
<td>1</td>
<td>0.43</td>
</tr>
<tr>
<td>AAU(I)</td>
<td>3</td>
<td>0.39</td>
<td>ACA(T)</td>
<td>1</td>
<td>0.27</td>
<td>AA(U)</td>
<td>3</td>
<td>0.5</td>
<td>AGU(S)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AUG(L)</td>
<td>20</td>
<td>2.61</td>
<td>ACC(T)</td>
<td>9</td>
<td>2.4</td>
<td>AAC(N)</td>
<td>9</td>
<td>1.5</td>
<td>AGG(S)</td>
<td>3</td>
<td>1.2</td>
</tr>
<tr>
<td>AAU(I)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>4</td>
<td>1.07</td>
<td>AAA(K)</td>
<td>10</td>
<td>1.18</td>
<td>AGA(R)</td>
<td>1</td>
<td>0.43</td>
</tr>
<tr>
<td>AUG(M)</td>
<td>11</td>
<td>1</td>
<td>ACC(T)</td>
<td>1</td>
<td>0.27</td>
<td>AAC(K)</td>
<td>7</td>
<td>0.82</td>
<td>AGG(R)</td>
<td>1</td>
<td>0.43</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>7</td>
<td>1.33</td>
<td>GCA(A)</td>
<td>3</td>
<td>0.44</td>
<td>GAU(D)</td>
<td>9</td>
<td>1.06</td>
<td>GGU(G)</td>
<td>7</td>
<td>0.97</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>11</td>
<td>2.1</td>
<td>GCC(A)</td>
<td>16</td>
<td>2.37</td>
<td>GAC(D)</td>
<td>8</td>
<td>0.94</td>
<td>GGC(G)</td>
<td>16</td>
<td>2.21</td>
</tr>
<tr>
<td>GUA(V)</td>
<td>0</td>
<td>0</td>
<td>GCA(A)</td>
<td>4</td>
<td>0.59</td>
<td>GA(A)</td>
<td>18</td>
<td>1.44</td>
<td>GGA(G)</td>
<td>5</td>
<td>0.69</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>3</td>
<td>0.57</td>
<td>GCG(A)</td>
<td>4</td>
<td>0.59</td>
<td>GAG(E)</td>
<td>7</td>
<td>0.56</td>
<td>GGG(G)</td>
<td>1</td>
<td>0.14</td>
</tr>
</tbody>
</table>

(a) nifH

<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>6</td>
<td>0.57</td>
<td>UCU(S)</td>
<td>3</td>
<td>0.72</td>
<td>UAU(Y)</td>
<td>10</td>
<td>0.83</td>
<td>UGU(C)</td>
<td>2</td>
<td>0.67</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>15</td>
<td>1.43</td>
<td>UCC(S)</td>
<td>13</td>
<td>3.12</td>
<td>UAC(Y)</td>
<td>14</td>
<td>1.17</td>
<td>UGC(C)</td>
<td>4</td>
<td>1.33</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>2</td>
<td>0.48</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
<td>UGA(*)</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>2</td>
<td>0.57</td>
<td>UCG(S)</td>
<td>2</td>
<td>0.48</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
<td>UGG(W)</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>5</td>
<td>1.43</td>
<td>CCU(P)</td>
<td>2</td>
<td>0.4</td>
<td>CAU(H)</td>
<td>3</td>
<td>0.46</td>
<td>CGU(R)</td>
<td>2</td>
<td>0.63</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>8</td>
<td>2.29</td>
<td>CCC(P)</td>
<td>12</td>
<td>2.4</td>
<td>CAC(H)</td>
<td>10</td>
<td>1.54</td>
<td>CGC(R)</td>
<td>8</td>
<td>2.53</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>1</td>
<td>0.29</td>
<td>CCA(P)</td>
<td>2</td>
<td>0.4</td>
<td>CA(A)</td>
<td>1</td>
<td>0.18</td>
<td>CGA(R)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>5</td>
<td>1.43</td>
<td>CGG(P)</td>
<td>4</td>
<td>0.8</td>
<td>CAG(Q)</td>
<td>10</td>
<td>1.82</td>
<td>CGG(R)</td>
<td>4</td>
<td>1.26</td>
</tr>
<tr>
<td>AAU(I)</td>
<td>5</td>
<td>0.36</td>
<td>ACA(T)</td>
<td>0</td>
<td>0</td>
<td>AA(U)</td>
<td>4</td>
<td>0.42</td>
<td>AGU(S)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AUG(L)</td>
<td>36</td>
<td>2.57</td>
<td>ACC(T)</td>
<td>15</td>
<td>2.5</td>
<td>AAC(N)</td>
<td>15</td>
<td>1.58</td>
<td>AGG(S)</td>
<td>5</td>
<td>1.2</td>
</tr>
<tr>
<td>AAU(I)</td>
<td>1</td>
<td>0.07</td>
<td>ACA(T)</td>
<td>3</td>
<td>0.5</td>
<td>AAA(K)</td>
<td>11</td>
<td>0.52</td>
<td>AGA(R)</td>
<td>2</td>
<td>0.63</td>
</tr>
<tr>
<td>AUG(M)</td>
<td>14</td>
<td>1</td>
<td>ACC(T)</td>
<td>6</td>
<td>1</td>
<td>AAC(K)</td>
<td>31</td>
<td>1.48</td>
<td>AGG(R)</td>
<td>3</td>
<td>0.95</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>4</td>
<td>0.46</td>
<td>GCA(A)</td>
<td>0</td>
<td>0</td>
<td>GAU(D)</td>
<td>7</td>
<td>0.47</td>
<td>GGU(G)</td>
<td>4</td>
<td>0.36</td>
</tr>
<tr>
<td>GUC(V)</td>
<td>20</td>
<td>2.29</td>
<td>GCC(A)</td>
<td>19</td>
<td>3.17</td>
<td>GAC(D)</td>
<td>23</td>
<td>1.53</td>
<td>GCC(G)</td>
<td>16</td>
<td>1.45</td>
</tr>
<tr>
<td>GUA(V)</td>
<td>0</td>
<td>0</td>
<td>GCA(A)</td>
<td>1</td>
<td>0.17</td>
<td>GA(A)</td>
<td>23</td>
<td>1.35</td>
<td>GGA(G)</td>
<td>19</td>
<td>1.73</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>11</td>
<td>1.26</td>
<td>GCG(A)</td>
<td>4</td>
<td>0.67</td>
<td>GAG(E)</td>
<td>11</td>
<td>0.65</td>
<td>GGG(G)</td>
<td>5</td>
<td>0.45</td>
</tr>
</tbody>
</table>

(b) nifD

<table>
<thead>
<tr>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
<th>Codon</th>
<th>Count</th>
<th>RSCU</th>
</tr>
</thead>
<tbody>
<tr>
<td>UUU(F)</td>
<td>10</td>
<td>0.69</td>
<td>UCU(S)</td>
<td>4</td>
<td>0.83</td>
<td>UAU(Y)</td>
<td>5</td>
<td>0.53</td>
<td>UGU(C)</td>
<td>1</td>
<td>0.5</td>
</tr>
<tr>
<td>UUC(F)</td>
<td>19</td>
<td>1.31</td>
<td>UCC(S)</td>
<td>11</td>
<td>2.28</td>
<td>UAC(Y)</td>
<td>14</td>
<td>1.47</td>
<td>UGC(C)</td>
<td>3</td>
<td>1.5</td>
</tr>
<tr>
<td>UUA(L)</td>
<td>0</td>
<td>0</td>
<td>UCA(S)</td>
<td>2</td>
<td>0.41</td>
<td>UAG(*)</td>
<td>1</td>
<td>3</td>
<td>UGA(*)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>UUG(L)</td>
<td>1</td>
<td>0.16</td>
<td>UCG(S)</td>
<td>5</td>
<td>1.01</td>
<td>UAG(*)</td>
<td>0</td>
<td>0</td>
<td>UGG(W)</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>7</td>
<td>1.14</td>
<td>CCU(P)</td>
<td>3</td>
<td>0.44</td>
<td>CAU(H)</td>
<td>10</td>
<td>1.05</td>
<td>CGU(R)</td>
<td>3</td>
<td>0.86</td>
</tr>
<tr>
<td>CUC(L)</td>
<td>15</td>
<td>2.43</td>
<td>CCC(P)</td>
<td>8</td>
<td>1.19</td>
<td>CAC(H)</td>
<td>9</td>
<td>0.95</td>
<td>CGC(R)</td>
<td>3</td>
<td>0.86</td>
</tr>
<tr>
<td>CUA(L)</td>
<td>1</td>
<td>0.16</td>
<td>CCA(P)</td>
<td>4</td>
<td>0.59</td>
<td>CA(A)</td>
<td>3</td>
<td>0.75</td>
<td>CGA(R)</td>
<td>1</td>
<td>0.29</td>
</tr>
<tr>
<td>CUG(L)</td>
<td>13</td>
<td>2.11</td>
<td>CGG(P)</td>
<td>12</td>
<td>1.78</td>
<td>CAG(Q)</td>
<td>5</td>
<td>1.25</td>
<td>CGG(R)</td>
<td>6</td>
<td>1.71</td>
</tr>
<tr>
<td>AAU(I)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>0</td>
<td>0</td>
<td>AA(U)</td>
<td>10</td>
<td>0.8</td>
<td>AGU(S)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AUG(L)</td>
<td>25</td>
<td>2.88</td>
<td>ACC(T)</td>
<td>14</td>
<td>1.87</td>
<td>AAC(N)</td>
<td>15</td>
<td>1.2</td>
<td>AGG(S)</td>
<td>7</td>
<td>1.45</td>
</tr>
<tr>
<td>AAU(I)</td>
<td>0</td>
<td>0</td>
<td>ACA(T)</td>
<td>7</td>
<td>0.93</td>
<td>AAA(K)</td>
<td>7</td>
<td>0.44</td>
<td>AGA(R)</td>
<td>3</td>
<td>0.86</td>
</tr>
<tr>
<td>AUG(L)</td>
<td>19</td>
<td>1</td>
<td>ACC(T)</td>
<td>7</td>
<td>0.93</td>
<td>AAC(K)</td>
<td>25</td>
<td>1.56</td>
<td>AGG(R)</td>
<td>5</td>
<td>1.43</td>
</tr>
<tr>
<td>GUU(V)</td>
<td>5</td>
<td>0.83</td>
<td>GCA(A)</td>
<td>3</td>
<td>0.34</td>
<td>GAU(D)</td>
<td>6</td>
<td>0.43</td>
<td>GGU(G)</td>
<td>4</td>
<td>0.48</td>
</tr>
<tr>
<td>GUC(V)</td>
<td>13</td>
<td>2.17</td>
<td>GCC(A)</td>
<td>24</td>
<td>2.74</td>
<td>GAC(D)</td>
<td>22</td>
<td>1.57</td>
<td>GCC(G)</td>
<td>5</td>
<td>0.61</td>
</tr>
<tr>
<td>GUA(V)</td>
<td>0</td>
<td>0</td>
<td>GCA(A)</td>
<td>4</td>
<td>0.46</td>
<td>GA(A)</td>
<td>22</td>
<td>1.22</td>
<td>GGA(G)</td>
<td>17</td>
<td>2.06</td>
</tr>
<tr>
<td>GUG(V)</td>
<td>6</td>
<td>1</td>
<td>GCG(A)</td>
<td>4</td>
<td>0.46</td>
<td>GAG(E)</td>
<td>14</td>
<td>0.78</td>
<td>GGG(G)</td>
<td>1</td>
<td>0.85</td>
</tr>
</tbody>
</table>

(c) nifK
### A.2 Genbank accession numbers

**Table A.6** – Genbank Accession numbers of 'housekeeping' gene sequences used for phylogenetic tree construction.

<table>
<thead>
<tr>
<th>Species</th>
<th>16SrRNA</th>
<th>gyrB</th>
<th>rpoB</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Leptospirillum ferrodiazotrophum</em> C5</td>
<td>JN007036.1*</td>
<td>JN390684*</td>
<td>GG693867</td>
</tr>
<tr>
<td><em>Leptospirillum ferrodiazotrophum</em> UBA1</td>
<td>EF065178.1</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td><em>Leptospirillum ferriphilum</em> DSM 14647</td>
<td>AF356829</td>
<td>JN390679*</td>
<td>EAY55962.1</td>
</tr>
<tr>
<td><em>Leptospirillum ferriphilum</em> Fairview</td>
<td>AF356830</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td><em>Leptospirillum sp</em> BRG44</td>
<td>GU167989</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td><em>Leptospirillum group II 5 CG</em></td>
<td>DS995260</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td><em>Leptospirillum ferrooxidans</em> DSM 2705</td>
<td>X86776.1</td>
<td>JN390680*</td>
<td>JN390682*</td>
</tr>
<tr>
<td><em>Leptospirillum sp</em> NOen1</td>
<td>AF376016</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td><em>Leptospirillum sp</em> 3.2</td>
<td>EU372651</td>
<td>/</td>
<td>/</td>
</tr>
<tr>
<td><em>Thermodesulfovibrio yellowstonii</em> DSM 1147</td>
<td>CP00147</td>
<td>CP00147</td>
<td>CP00147</td>
</tr>
</tbody>
</table>

* Sequences generated from this study

**Table A.7** – Genbank Accession numbers of DNA sequences generated from this study (*) and used in phylogenetic analysis.

<table>
<thead>
<tr>
<th>Species</th>
<th>nifH</th>
<th>nifD</th>
<th>nifK</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Acidithiobacillus ferrooxidans</em> ATCC 53993</td>
<td>CP001132.1</td>
<td>CP001132.1</td>
<td>CP001132.1</td>
</tr>
<tr>
<td><em>Acidithiobacillus ferrovorans</em> SS3</td>
<td>CP002985.1</td>
<td>CP002985.1</td>
<td>CP002985.1</td>
</tr>
<tr>
<td><em>Azotobacter vinelandii</em> DJ</td>
<td>CP001157.1</td>
<td>CP001157.1</td>
<td>CP001157.1</td>
</tr>
<tr>
<td><em>Bradyrhizobium japonicum</em> USDA 110</td>
<td>BA000040.1</td>
<td>BA000040.1</td>
<td>BA000040.1</td>
</tr>
<tr>
<td><em>Burkholderia xenovorans</em> LB400</td>
<td>CP000271.1</td>
<td>CP000271.1</td>
<td>CP000271.1</td>
</tr>
<tr>
<td><em>Clostridium pasteurianum</em></td>
<td>AY603957.3</td>
<td>AY603957.3</td>
<td>AY603957.3</td>
</tr>
<tr>
<td><em>Leptospirillum ferrodiazotrophum</em> C5</td>
<td>JQ231108*</td>
<td>JQ231109*</td>
<td>JQ231110*</td>
</tr>
<tr>
<td><em>Leptospirillum ferriphilum</em> DSM 14647</td>
<td>JN390678*</td>
<td>JN390678*</td>
<td>JN390678*</td>
</tr>
<tr>
<td><em>Leptospirillum ferrooxidans</em> DSM 2705</td>
<td>JQ231105*</td>
<td>JQ231106*</td>
<td>JQ231107*</td>
</tr>
<tr>
<td><em>Methylacidiphilum fumaroliicum</em></td>
<td>GU299762.1</td>
<td>GU299762.1</td>
<td>GU299762.1</td>
</tr>
<tr>
<td><em>Methylocystis</em> sp ATCC 49242</td>
<td>NZ_AEVM01000015</td>
<td>NZ_AEVM01000015</td>
<td>NZ_AEVM01000015</td>
</tr>
</tbody>
</table>

* Sequences generated from this study
Table A.8 - Genbank Accession numbers of theoretical amino acid Nif sequences used for structural analysis and phylogenetic analysis.

<table>
<thead>
<tr>
<th>Species</th>
<th>NifH</th>
<th>NifD</th>
<th>NifK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acidithiobacillus ferrooxidans ATCC 53993</td>
<td>YP_002219685.1</td>
<td>YP_002219684</td>
<td>YP_002219683.1</td>
</tr>
<tr>
<td>Acidithiobacillus ferrovorans SS3</td>
<td>ZP_08488603.1</td>
<td>ZP_08488602.1</td>
<td>ZP_08488601.1</td>
</tr>
<tr>
<td>Azotobacter vinelandii DJ</td>
<td>YP_002797378.1</td>
<td>YP_002797379.1</td>
<td>YP_002797380.1</td>
</tr>
<tr>
<td>Bradyrhizobium japonicum USDA 110</td>
<td>NP_768409.1</td>
<td>NP_768383.1</td>
<td>NP_768384.1</td>
</tr>
<tr>
<td>Burkholderia xenovorans LB400</td>
<td>YP_553849.1</td>
<td>YP_553848.1</td>
<td>YP_553847.1</td>
</tr>
<tr>
<td>Clostridium pasteurianum</td>
<td>AAT37643.2</td>
<td>AAA83531.2</td>
<td>AAA23266.2</td>
</tr>
<tr>
<td>Leptospirillum ferrodiazotrophum</td>
<td>EE53484.1</td>
<td>EE53485</td>
<td>EE535486</td>
</tr>
<tr>
<td>Leptospirillum ferrophilum DSM 1647</td>
<td>AEB14646.1*</td>
<td>AAO38344.1*</td>
<td>AEN14648.1*</td>
</tr>
<tr>
<td>Leptospirillum ferrooxidans DSM 2705</td>
<td>AAO38343.1</td>
<td>AAO38344.1</td>
<td>Appendix A.7</td>
</tr>
<tr>
<td>Methylacidiphilum fumaroliicum</td>
<td>ADC97878.1</td>
<td>ADC97879</td>
<td>ADC97880.1</td>
</tr>
<tr>
<td>Methylocystis sp ATCC 49242</td>
<td>ZP_08073552.1</td>
<td>ZP_08073551.1</td>
<td>ZP_08073550.1</td>
</tr>
</tbody>
</table>

* Sequences generated from this study

MSQNPKHVLDFNLFREPEYVEMFENKKKNFENPHPEDEVSRRIIEWTKTEEKYKELNFNRSLSLTVPAPACQPLGAVFLALGFLPFMDPRAVAWAYYRSHLSRHFKEPTSCVSSSMT
EDAVFGLNNMDANAYGYMYKPKMIASSTCMAMEVIQQDLNAFIKTSKDGVSPEEY
DVPFAHTPAFVGSHITGYNLTKGIMEHFWAKKERQENINETINVMGDYSVGNIRELKRV
LKEMDINAILSDTSDVFTDTPDTFTRMYMGTPLEAREALNSKATISFQENFNTPSLEYA
QLSGQKTMAFHYPIRGATRWRWSISLKLTGKEIESIKLERGRLVDAVDSTSHHGKKFAL
YGDPDOQLGLSLFSLMGAEPHVHLATNGKDWEEKMNALFATSFGACGACHYGRDL
WHMRSLLFTEPFPFLIGNYKYLERDTGPTLHRRIFDRHHHRRYPTGVQGAMNVLV
KILDKIFDEIDKNVTNVPKTDYSDIIR

Figure A.4 - *L. ferrooxidans* DSM 2705 NifK protein sequence
<table>
<thead>
<tr>
<th>Species</th>
<th>Spot #</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrodiazotrophum</td>
<td>17, 326, 412, 484, 601, 665, 738, 751, 865</td>
</tr>
<tr>
<td>L. ferriphilum</td>
<td>29, 497, 583, 629, 784, 844, 924, 1042</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td>412, 437, 507</td>
</tr>
</tbody>
</table>

(a) Not isolated

<table>
<thead>
<tr>
<th>Species</th>
<th>Spot #</th>
</tr>
</thead>
<tbody>
<tr>
<td>L. ferrodiazotrophum</td>
<td>421, 1026, 1140, 1595, 2512, 2519, 2611, 2845, 3123,</td>
</tr>
<tr>
<td>L. ferriphilum</td>
<td>501, 660, 678, 687, 704, 713, 717, 759, 763, 909, 914, , 976, 977, 1080, 1098</td>
</tr>
<tr>
<td>L. ferrooxidans</td>
<td>658, 804</td>
</tr>
</tbody>
</table>

(b) No match

**Table A.9** – Protein spots identified from *Leptospirillum spp*, whose change in abundance was computationally identified as having altered >2 fold, but could not be isolated on gels for manual excision (a) or the identification of proteins via MASCOT was not statistically significant (b).
Appendix B

Supplementary Information - II

B.1 Computing

Table B.1 - Commands employed to perform Bayesian Inference using MrBayes 3.2 on DNA and AA sequences.

Mr Bayes 3.2 DNA block (GTR model)
>begin mrbayes;
>execute file.nex;
>lset nst=6 rates=invgamma;
>mcmc ngen=1000000 samplefreq=1000 printfreq=100;
>sump burnin=250;
>sumt burnin=250;
>end;

Mr Bayes 3.2 Protein block (WAG)
>begin mrbayes;
>execute file.nex;
>lset rates=gamma;
>prset aamodelpr=fixed(WAG);
>mcmc ngen=1000000 samplefreq=1000 printfreq=100;
>sump burnin=250;
>sumt burnin=250;
>end;
### Table B.2 – Computer programs used during this thesis.

<table>
<thead>
<tr>
<th>Program</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>4Peaks v1.7</td>
<td>Chromatogram editing</td>
</tr>
<tr>
<td>Adobe Photoshop CS3</td>
<td>Image editing</td>
</tr>
<tr>
<td>Bibdesk v1.5.4</td>
<td>Referencing</td>
</tr>
<tr>
<td>Bio-Rad iQ5 real time software</td>
<td>Analysis of mRNA transcription rates</td>
</tr>
<tr>
<td>BLAST (NCBI)</td>
<td>Sequence identification and homology</td>
</tr>
<tr>
<td>ClustalX</td>
<td>Sequence alignment</td>
</tr>
<tr>
<td>Cn3D v4.3</td>
<td>3D protein modelling</td>
</tr>
<tr>
<td>DataExplorer 4.8</td>
<td>Analysis of MS data files</td>
</tr>
<tr>
<td>Emboss 6.4.0</td>
<td>Relative codon usage and GC% content</td>
</tr>
<tr>
<td>Endnote X4</td>
<td>Reference importing and management</td>
</tr>
<tr>
<td>FigTree v1.3.1</td>
<td>Graphical rendering of phylogenetic analysis files</td>
</tr>
<tr>
<td>Geneious Pro 3.7.0</td>
<td>Genetic data manipulation and storage</td>
</tr>
<tr>
<td>MASCOT (MatrixScience)</td>
<td>PMF identification</td>
</tr>
<tr>
<td>MEGA 5.05</td>
<td>Maximum Likelihood phylogenetic analysis</td>
</tr>
<tr>
<td>Mfold</td>
<td>DNA folding and secondary structure predictions</td>
</tr>
<tr>
<td>Model Generator v0.85</td>
<td>Model selection for phylogenetic analysis</td>
</tr>
<tr>
<td>Mr Bayes 3.2</td>
<td>Bayesian Inference for phylogenetic analysis</td>
</tr>
<tr>
<td>Nanodrop 1000 v3.7</td>
<td>DNA/RNA concentration and purity</td>
</tr>
<tr>
<td>NetPrimer</td>
<td>Primer design</td>
</tr>
<tr>
<td>Perkin Elmer ProSCAN Imaging</td>
<td>Scanning of 2D-SDS page gels</td>
</tr>
<tr>
<td>Progenesis Same Spots v1.0</td>
<td>2D SDS-PAGE protein analysis</td>
</tr>
<tr>
<td>SPSS v19</td>
<td>Statistical evaluation of experimental data</td>
</tr>
<tr>
<td>TexMaker 3.0.2</td>
<td>Typesetting editing</td>
</tr>
<tr>
<td>TexShop 2.43</td>
<td>Typesetting editing</td>
</tr>
<tr>
<td>TextMate 1.5.10</td>
<td>Typesetting editing</td>
</tr>
<tr>
<td>XeLaTeX v3.14</td>
<td>Typesetting</td>
</tr>
</tbody>
</table>

### B.2 Typesetting

This PhD thesis was typeset using \texttt{XeLaTeX}, in combination with the fontspec package, in ‘Franklin Gothic Book’ font, size 11 point, and 1.5 line spacing. Editing was performed with TexMaker version 3.0.2, TexShop 2.43 and TextMate 1.5.10. References were managed by Endnote X4, converted into a BibTEX database and further edited with Bibdesk v1.5.4. The total word count (not including references) is 48,380.
References


References


References


Norris, P. R., and D. A. Clark. 1996. *Acidimicrobium ferrooxidans* gen. nov., sp. nov.: mixed-culture ferrous iron oxidation with *Sulfobacillus* species. *Microbiology* 142.


References


References


References


Watling, H. R. 2006. The bioleaching of sulphide minerals with emphasis on copper sulphides -- a review. Hydrometallurgy 84 (1-2).


Yun, A. C., and A. A. Szalay. 1984. Structural genes of dinitrogenase and dinitrogenase reductase are transcribed from two separate promoters in the broad host range cowpea *Rhizobium* strain IRc78. *Proc Natl Acad Sci USA* 81 (23): 7358–62.


"Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged."