Metabolic capability in host-restricted serovars of *Salmonella enterica*

Gemma C. Langridge

Darwin College, University of Cambridge

Wellcome Trust Sanger Institute

This dissertation is submitted for the degree of

Doctor of Philosophy

2010

Abstract

Metabolic capability in host-restricted serovars of Salmonella enterica

Gemma C. Langridge

The Gram negative bacterial species *Salmonella enterica* is comprised of over 2,500 serovars including *S. enterica* serovar Typhi (Typhi), the cause of typhoid, a disease solely affecting humans and *S. enterica* serovar Typhimurium (Typhimurium), capable of causing disease in a wide range of hosts. Exclusive infection of a host is seen in a number of *Salmonella* serovars and the accumulation of pseudogenes has been cited as a contributing factor. Such serovars also display a reduced ability to utilise multiple metabolic substrates. In this thesis, the influence of pseudogenes upon the metabolic and pathogenic capability of host-restricted serovars was investigated in comparison to non-adapted Typhimurium, using metabolic pathway analysis and transposon mutagenesis.

Metabolic pathway databases were generated for both Typhi and Typhimurium, based upon high quality genome sequence and annotation. This enabled pseudogenes to be identified in Typhi metabolism and compared with other *Salmonella* serovars. While few individual pseudogenes were shared between host-restricted *Salmonella*, both pathways and transporters were identified as commonly inactivated. A novel method, Transposon Directed Insertion-site Sequencing (TraDIS) was developed to enable one million transposon mutants to be simultaneously assayed using high-throughput Illumina sequencing. A Typhimurium mutant library was created and analysed in conjunction with a similar one in Typhi, to generate candidate essential gene lists for cellular survival. Only 75% of Typhi essential genes were shared with Typhimurium,

suggesting that while core metabolism is shared, there are differences in peripheral pathways that reflect different survival strategies. Additionally, the mutant libraries were screened in human macrophages to investigate the genes required for cell infection, revealing that Typhimurium utilises pathways inactivated by pseudogenes in Typhi.

In conclusion, metabolic phenotypes of host-restricted *Salmonella* serovars can be associated with pseudogenes and there is evidence to suggest that the activity of a host-generalist such as Typhimurium cannot necessarily be used to predict that of a host-restricted serovar like Typhi.

Acknowledgements

I would like to thank John Wain for allowing me to begin my PhD with him and, despite moving on from the Institute, supporting me throughout the project and always making time for a discussion of the data. I also want to thank Julian Parkhill for his supervision and support, particularly through the latter half of this project.

I would also like to thank Martin Welch and Anton Enright who formed part of my thesis committee, for their guidance and encouragement.

I would like to make a special mention of those who collaborated with and/or helped me during parts of my PhD: Keith 'The Master' Turner for his magical knowledge of transposon mutagenesis, Duy Phan for his support and insight as a fellow student, Satheesh Nair, for his support and encouragement, Jana Haase for her 'discussions' (not arguments) and commitment to lab work, Daniel Turner for his help and dedication in developing TraDIS, Sabine Eckert for carrying on the efforts, particularly against time deadlines, Leopold Parts for his wonderful statistical brain, Silvia Pinero and Fernanda Schreiber for their patience teaching me tissue culture, Nick Thomson for endless discussions, Theresa Feltwell for braving the Biolog and teaching me its ways and Lars Barquist for his help analysing the data it produced. My thanks also go to Craig Corton and Maria Fookes for their help unravelling the Typhisuis genome.

To all those others at the Institute and elsewhere who have helped, advised and supported me along the way, I cannot list you all but thank you for everything.

A huge thank-you goes to Team 100 with whom I started this journey, as a group you made science exciting and thoroughly enjoyable and I look forward to seeing you all again, wherever in the world that may be. An equally big thank-you also goes to Team 81, in particular to those that I have shared an office with and who have seen me through to the end.

I would like to dedicate this thesis to my parents, especially Mum, who always believed I could do this and gave me a push in the right direction when I needed it, and to my husband Ben, who kept me sane throughout and supplied the required chocolate and hugs I needed to finish writing up.

Declaration

I hereby declare that this dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and Acknowledgements.

This thesis is no longer than 60,000 words, as required by the School of Biological Sciences.

Gemma Langridge

August 2010

Table of Contents

Abstract i

Acknowledgements iii

Declaration v

List of Figures xvii

List of Tables xix

Abbreviations xxi

1 Introduction 1

1.1 *Salmonella* 1

1.1.1 Classification 1

1.1.1.1 The genus *Salmonella* 1

1.1.1.2 Serotyping and biochemical testing 3

1.1.1.3 Multi Locus Enzyme Electrophoresis and Sequence Typing 4

1.1.2 *S. bongori* 7

1.1.3 *S. enterica* subspecies *enterica* 8

1.1.3.1 *S. enterica* serovar Typhi 8

1.1.3.2 *S. enterica* serovar Paratyphi A 10

1.1.3.3 *S. enterica* serovar Paratyphi C 11

1.1.3.4 *S. enterica* serovar Choleraesuis 12

1.1.3.5 *S. enterica* serovar Typhisuis 13

1.1.3.6 *S. enterica* serovar Gallinarum 14

1.1.3.7 *S. enterica* serovar Typhimurium 15

1.2 Host adaptation and restriction 16

1.2.1 Definitions 16

1.2.2 Process of adaptation and restriction 16

1.2.2.1 IS elements 17

1.2.2.2 Genome rearrangement 18

1.2.2.3 Pseudogenes 18

1.3 Disease and infection of the host 21

1.3.1 Clinical features of enteric fever 21

1.3.2 Infection of the host 22

1.3.2.1 Typhi 22

1.3.2.2 Differences in Typhimurium 24

1.3.3 Molecular basis of infection 25

1.3.3.1 Invading the macrophage 25

1.3.3.2 The *Salmonella*-containing vacuole 27

1.4 Genetic diversity in *Salmonella* 30

1.4.1.1 Phage typing 30

1.4.1.2 Plasmid profiling 31

1.4.1.3 Signature tagged mutagenesis 32

1.4.1.4 Microarrays 33

1.4.1.5 High-throughput sequencing 35

1.5 Metabolism in *Salmonella* 37

1.5.1 Phenotypic analysis and biochemistry 37

1.5.2 Metabolic pathways 38

1.5.3 Pathway maps 39

1.5.3.1 Metabolic databases 39
1.5.3.2 Pathway prediction 41
1.5.4 Network modelling 42
1.6 Project focus 43
1.7 Aims of the project 45
2 Development of StyCyc and StmCyc and validation of known metabolic lesions 46
2.1 Introduction 46
2.2 Methods 49
2.2.1 Pathway Tools 49
2.2.1.1 Input file formatting 49
2.2.1.2 Pseudogene re-annotation 50
2.2.1.3 Assigning enzymes and multimers 50
2.2.1.4 Pathway prediction 51
2.2.1.5 Transport reactions 52
2.2.2 Pathway hole filling 52
2.2.2.1 Automated hole filling 52
2.2.2.2 Manual evaluation of pathway hole candidates 53
2.2.2.3 PGDB build process for Typhimurium 54
2.3 Results 55
2.3.1 Generation of StyCyc 1.0 55
2.3.1.1 Issues and resolutions 56
2.3.2 StyCyc 2.0 and onwards 58
2.3.2.1 Improvements from StyCyc 1.0 58

2.3.2.2 New pathways and transport reactions 58
2.3.2.3 Pathway Hole Filling 62
2.3.2.4 PGDB statistics 63
2.3.2.5 Metabolic map: Typhi 66
2.3.2.6 Pseudogenes in Typhi 69
2.3.3 StmCyc 71
2.3.3.1 Use of StyCyc to help assignments 71
2.3.3.2 Hole Filling 71
2.3.3.3 Additional development 72
2.3.3.4 PGDB statistics 74
2.3.3.5 Metabolic map: Typhimurium 74
2.3.4 Comparing StyCyc with StmCyc 76
2.4 Discussion 78
2.4.1 Rationale for building a pathway database 78
2.4.2 Rationale for using Pathway Tools 78
2.4.3 Presence of pseudogenes can explain metabolic capability 79
2.4.4 StyCyc versus StmCyc: a basis for comparison 81
2.4.5 StyCyc and StmCyc as resources 82
2.5 Conclusions 83
3 Pseudogenes in host restricted *Salmonella enterica* 84
3.1 Introduction 84
3.2 Methods 86
3.2.1 Strains 86

3.2.2 Sequencing of a Typhisuis reference strain 86
3.2.2.1 Pseudogene identification and validation 88
3.2.3 Whole genome comparison..... 89
3.2.4 Functional classification of pseudogenes..... 90
3.2.5 Metabolic pathway analysis..... 90
3.2.6 Biolog phenotyping..... 90
3.2.6.1 Preparation of strains to be phenotyped..... 91
3.2.6.2 Preparation of PM Inoculating Fluids..... 91
3.2.6.3 Preparation of cell suspension 92
3.2.6.4 Inoculation of plates PM-1 and PM-2A..... 92
3.2.7 Analysis of Biolog phenotypes 92
3.2.7.1 Phenotype calling..... 92
3.2.7.2 Differential phenotypes..... 93
3.3 Results..... 94
3.3.1 Individual pseudogenes..... 95
3.3.2 Relating metabolic phenotype to ‘pseudo’ genotype..... 96
3.3.2.1 Negative metabolic phenotypes 97
3.3.2.2 Differential metabolic phenotypes..... 97
3.3.2.3 Potential for improving annotation 102
3.3.2.4 Other metabolic phenotypes 103
3.3.3 Shared interruption of metabolic pathways 105
3.3.3.1 Pathways inactivated in all host restricted serovars..... 105
3.3.3.2 Pseudogenes inactivating pathways in human-restricted serovars 108

3.3.4 Loss of function in transport across the membrane 111
3.4 Discussion 114
3.5 Conclusions 118
4 Essential genes in Typhi and Typhimurium 119
4.1 Introduction 119
4.2 Methods 123
4.2.1 Strains 123
4.2.2 Transposome preparation 123
4.2.3 Cell transformation and transposon library creation 124
4.2.4 DNA manipulation and sequencing 125
4.2.4.1 Optimisation using Typhi 125
4.2.4.2 Typhimurium 127
4.2.5 Sequence analysis 127
4.2.6 Statistical analyses 129
4.2.6.1 High density mutagenesis 129
4.2.6.2 Essential genes per serovar 130
4.2.6.3 Comparison of essential genes between serovars 130
4.3 Results 131
4.3.1 Whole genome TraDIS 131
4.3.1.1 Optimising TraDIS 131
4.3.1.2 Insertion bias 133
4.3.1.3 Typhi 134
4.3.1.4 Typhimurium 136

4.3.2 Essential genes in Typhi 136
4.3.3 Essential genes in Typhimurium 143
4.3.4 Comparing Typhi and Typhimurium 146
4.3.4.1 Shared essential genes 146
4.3.4.2 Essential genes only present in one serovar 148
4.3.4.3 Genes essential in Typhi only 152
4.3.4.4 Genes essential in Typhimurium only 153
4.3.4.5 Functional classification of essential genes 156
4.4 Discussion 158
4.4.1 Improvement over microarrays 158
4.4.2 Assaying short regions 159
4.4.3 Essential prophage genes 159
4.4.4 A+T-rich islands protected in Typhimurium 160
4.4.5 Specific genes required by Typhi 161
4.5 Conclusions 163
5 Assaying large *Salmonella* transposon mutant libraries in human macrophages .. 164
5.1 Introduction 164
5.2 Methods 167
5.2.1 Strains and cell lines 167
5.2.2 Growth in RPMI 167
5.2.3 Preparation of THP-1 cells 168
5.2.4 Preparation of transposon libraries 169
5.2.5 Optimisation of gentamicin infection assay for TraDIS 169

5.2.5.1 Preliminary infection 169
5.2.5.2 Small scale 171
5.2.5.3 DNA manipulation (small scale) 172
5.2.6 Large scale infection assay 173
5.2.6.1 Single timepoint replicates 173
5.2.6.2 DNA manipulation (large scale) 173
5.2.7 TraDIS analysis 174
5.2.7.1 Sample processing and nucleotide sequencing 174
5.2.7.2 Sequence data processing 176
5.2.8 Statistical analysis of required genes 176
5.2.8.1 Large scale assays 176
5.2.8.2 Single timepoint replicates in Typhi 177
5.3 Results 178
5.3.1 Preliminary Typhimurium assay 178
5.3.2 Optimising the cell infection assay for TraDIS 179
5.3.2.1 Media 179
5.3.2.2 Infection assays 180
5.3.3 Large-scale infection 183
5.3.4 TraDIS analysis of large-scale infection assays 184
5.3.5 Typhimurium 185
5.3.5.1 Propanediol utilisation 190
5.3.5.2 Transport 191
5.3.6 Typhi 192

5.4 Discussion 195
5.5 Conclusions 200
6 Final discussion 201
6.1 Publications arising from this thesis 207
7 References 209
8 Appendices 237
8.1 Chapter 2 237
8.1.1 CD_001 237
8.2 Chapter 3 237
8.2.1 Biolog phenotypes for PM-1 237
8.2.2 Biolog phenotypes for PM-2A 239
8.3 Chapter 4 242
8.3.1 CD_002 242
8.3.2 CD_003 242
8.3.3 Typhi essential genes 242
8.3.4 CD_004 250
8.3.5 Typhimurium essential genes 251
8.3.6 Shared essential genes 258
8.3.7 Core gene functions in Typhimurium 264
8.3.8 Essential genes omitted from comparison 264
8.3.9 Putative essential genes 265
8.3.9.1 Typhi 265
8.3.9.2 Typhimurium 266

8.3.10 Attempted construction of a Typhi *recA* mutant..... 267

8.4 Chapter 5..... 268

8.4.1 Typhimurium genes essential for macrophage infection..... 268

8.4.2 Typhi genes required for macrophage infection (original assay) 274

List of Figures

Figure 1-1 Classification of the genus *Salmonella* 2

Figure 1-2 Relationships between the Enterobacteriaceae 3

Figure 1-3 MLSA tree for *Salmonella* 6

Figure 1-4 Pathogenesis of typhoid fever in humans..... 23

Figure 1-5 Intracellular pathway of *Salmonella* in epithelial cells 28

Figure 1-6 Metabolic capabilities of *Salmonella* serovars..... 43

Figure 2-1 Vi antigen biosynthesis 59

Figure 2-2 Superpathway of glycolysis, pyruvate dehydrogenase and TCA cycle 61

Figure 2-3 StyCyc 1.0: an automatically generated metabolic map of Typhi 67

Figure 2-4 StyCyc 7.0: a manually curated metabolic map of Typhi..... 68

Figure 2-5 Pseudogenes interrupting metabolic pathways and transport reactions..... 70

Figure 2-6 StmCyc 4.0: a metabolic map of Typhimurium..... 75

Figure 3-1 Functional classification of pseudogenes..... 95

Figure 3-2 Heatmap of carbon source utilisation..... 96

Figure 3-3 Galactitol (dulcitol) and tagatose degradation 101

Figure 3-4 Metabolic pathway of tetrathionate reduction..... 104

Figure 3-5 Early stages of vitamin (coenzyme) B12 biosynthesis 106

Figure 3-6 Metabolic pathway for 1,2-propanediol degradation 107

Figure 3-7 Metabolism of 4-hydroxyphenylacetate..... 110

Figure 4-1 Transposon insertion site bias in Typhi 134

Figure 4-2 Whole genome view of transposon insertion sites..... 135

Figure 4-3 Whole genome assay of Typhi 138

Figure 4-4 Peptidoglycan biosynthesis 140

Figure 4-5 Essential genes in Typhi 142

Figure 4-6 Insertions between genes 143

Figure 4-7 Whole genome assay of Typhimurium 144

Figure 4-8 Domains in *polA* are essential 146

Figure 4-9 Comparison of essential genes 147

Figure 4-10 Genes uniquely essential in Typhi 153

Figure 4-11 Gene uniquely essential in Typhimurium 154

Figure 4-12 SPI-2 genes uniquely essential in Typhimurium 155

Figure 4-13 Essential genes classified by function 157

Figure 5-1 Preliminary infection assay 179

Figure 5-2 Transposon library cell counts from initial THP-1 infection 180

Figure 5-3 Transposon library cell counts from large-scale THP-1 infection 183

Figure 5-4 Typhimurium genes essential for macrophage invasion and SCV generation 187

Figure 5-5 StmCyc view of Typhimurium candidate essential genes for macrophage invasion 189

Figure 5-6 Genes required in propanediol utilisation 190

List of Tables

Table 1-1 Host specificities and multi-locus sequence types for 6,7:c:1,5 *Salmonella*.... 12

Table 1-2 Metabolic properties that distinguish 6,7:c:1,5 serovars 14

Table 2-1 Statistics for StyCyc 1.0 56

Table 2-2 Classification of pseudogenes in Typhi CT18..... 57

Table 2-3 Predicted transporter proteins in Typhi CT18 60

Table 2-4 Incomplete pathways in StyCyc 62

Table 2-5 Summary of StyCyc pathway outcomes..... 62

Table 2-6 Statistics from StyCyc 7.0 63

Table 2-7 Pathway hierarchy in Pathway Tools 65

Table 2-8 Classification of Typhi pseudogenes among metabolic reactions..... 69

Table 2-9 Incomplete pathways in StmCyc 72

Table 2-10 Summary of StmCyc pathway outcomes 72

Table 2-11 Statistics from StmCyc 4.0 74

Table 2-12 Metabolic pathway differences..... 76

Table 2-13 Metabolic pathways interrupted in Typhi..... 77

Table 3-1 Strains 86

Table 3-2 Metabolic phenotyping 91

Table 3-3 Pseudogenes across host restricted *Salmonella* 94

Table 3-4 Differential Biolog phenotypes 99

Table 3-5 Metabolic functions of pseudogenes 112

Table 3-6 Pseudogenes in iron uptake systems..... 113

Table 4-1 Functional categories in genome annotation 128

Table 4-2 Validating TraDIS 132

Table 4-3 Known genes coding for fundamental biological processes in Typhi..... 141

Table 4-4 Genes uniquely essential in Typhimurium 149

Table 4-5 Genes uniquely essential in Typhi..... 151

Table 5-1 Small scale infection assays 181

Table 5-2 DNA yields from different extraction kits..... 182

Table 5-3 DNA yields from large-scale infection..... 184

Table 5-4 Sequencing output from large-scale invasion assays 185

Table 5-5 Candidate essential transporters 192

Table 5-6 Typhi genes required during macrophage assay 193

Abbreviations

ACT	Artemis Comparison Tool
EC	Enzyme Commission
IS	Insertion Sequence
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LPS	Lipopolysaccharide
Mbp	Mega base pairs
MLEE	Multi Locus Enzyme Electrophoresis
MLST	Multi Locus Sequence Typing
PFGE	Pulsed-Field Gel Electrophoresis
PGDB	Pathway/Genome Database
SCV	*Salmonella*-containing vacuole
SPI	*Salmonella* Pathogenicity Island
ST	Sequence Type
TraDIS	Transposon directed insertion-site sequencing
VPT	Variable pseudogene in Typhi

N.B. *Salmonella* nomenclature is complex when describing serovars within subspecies. For simplicity, the *Salmonella* serovars mentioned in this dissertation are referred to by their serovar names alone; the preceding *Salmonella enterica* subspecies *enterica* is implied.