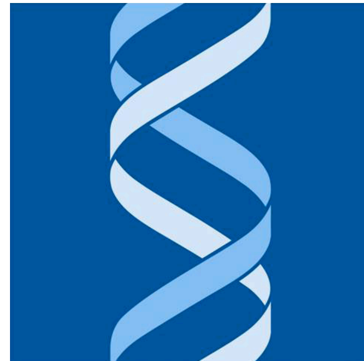
Metabolic capability in host-restricted serovars of *Salmonella enterica*



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.

1 Introduction

Some bacterial pathogens are defined by the host organism they infect. This thesis describes the infectious capacity of such organisms, the human-restricted *Salmonella enterica* serovars Typhi and Paratyphi A, and the chicken-restricted serovar Gallinarum. How this capacity can be related to loss of gene function and bacterial metabolism is demonstrated by comparison with the non-host-adapted serovar Typhimurium.

1.1 Salmonella

1.1.1 Classification

1.1.1.1 The genus Salmonella

The Gram negative genus *Salmonella* is currently divided into two species, *S. bongori* and *S. enterica*. A third species, *S. subterranea* was proposed in 2004 (Shelobolina et al. 2004), but this was later shown not to belong to the genus (Grimont and Weill 2007). *S. enterica* is further divided into six subspecies, *enterica*, *salamae*, *arizonae*, *diarizonae*, *houtenae* and *indica*, based upon DNA-DNA hybridisation, 16S RNA analysis and multi-locus enzyme electrophoresis (Crosa et al. 1973; Reeves et al. 1989). The vast majority of *S. enterica* serovars are found in subspecies *enterica* and account for >99.9% of known human and animal infection (Figure 1-1) (Selander et al. 1996).

A member of the Enterobacteriaceae, *Salmonella* is most closely related to *Shigella*, *Escherichia coli* and *Citrobacter* (Figure 1-2).

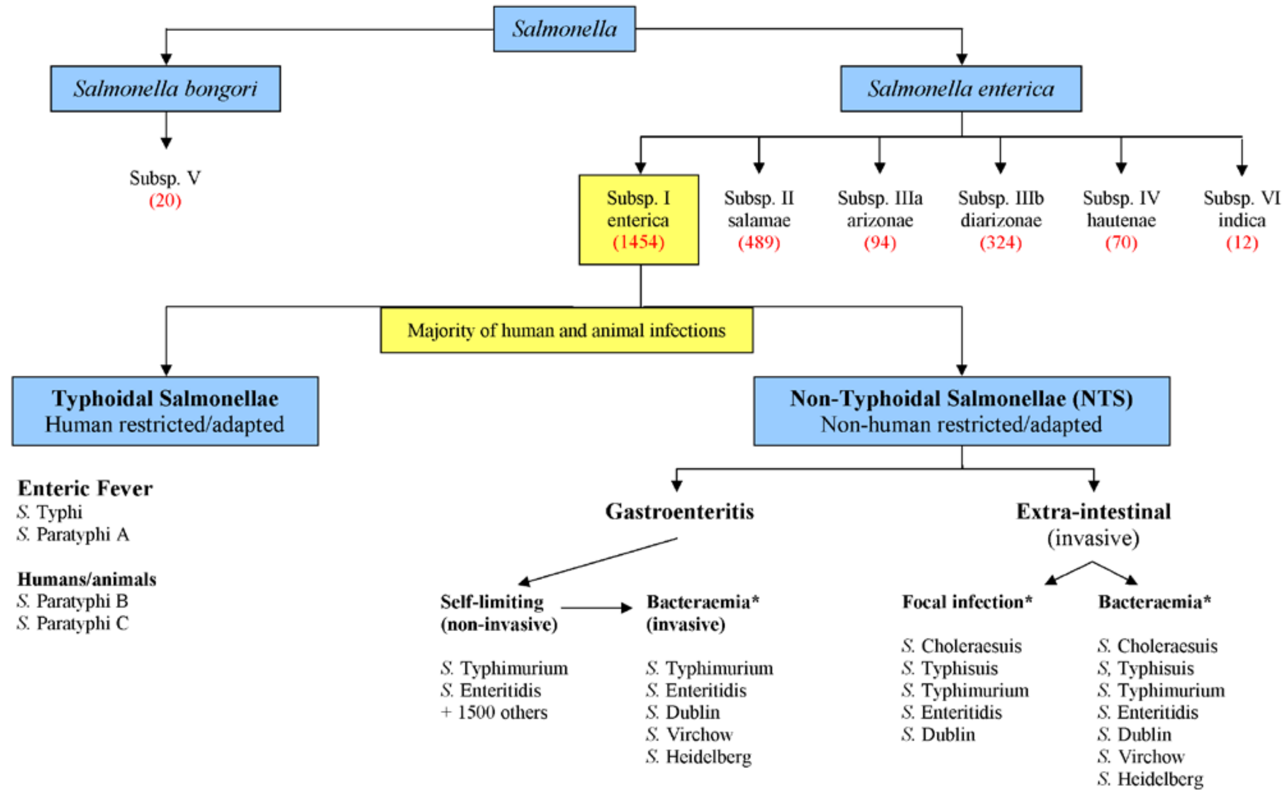


Figure 1-1 Classification of the genus *Salmonella*

Figure adapted from (Langridge et al. 2008). *Salmonella* subspecies have been defined by biotyping, DNA-DNA hybridisation, 16S RNA analysis and multi-locus enzyme electrophoresis. Serotyping is used for differentiation beyond the level of subspecies. Serovar numbers are given in red below each subspecies (Guibourdenche et al. 2009). * Common serotypes are listed but other serotypes may cause bacteraemia or focal infection; subsp., subspecies.

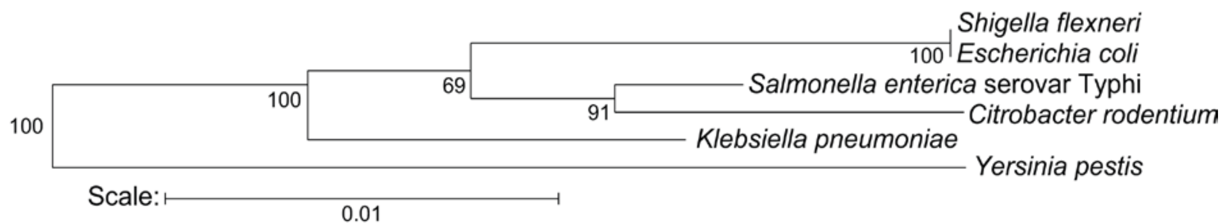


Figure 1-2 Relationships between the Enterobacteriaceae

Weighted neighbour joining tree based upon 16S rRNA sequences retrieved from the Ribosomal Database Project and built using Tree Builder (Cole et al. 2009). *S. enterica* serovar Typhi used as a representative of *S. enterica*. Numbers indicate bootstrap values for each branch and distance is based upon the Jukes-Cantor correction.

1.1.1.2 Serotyping and biochemical testing

As a genus, *Salmonella* is subdivided serologically into 2,610 serovars by the White-Kauffmann-Le Minor scheme (Grimont and Weill 2007; Guibourdenche et al. 2009). This scheme is based upon 'O' surface antigens and the expression of flagellar 'H' antigens. The polysaccharide O antigen is encoded by the *wba* (*rfb*) gene cluster and forms part of the lipopolysaccharide (LPS) found in the outer membrane of Gram negative bacteria. It consists of a variable number of oligosaccharide repeats, and sixty seven O antigen variants are currently known (Grimont and Weill 2007). The H antigens relate to flagellar phases 1 (motile) and 2 (non-motile) and are the respective products of *fliC* and *fljB*. Together, these antigens give rise to an antigenic formula that (in most cases) differentiates each serovar. This formula is written in the style O antigen(s): H antigen phase 1: H antigen phase 2: other, where other refers to rare 'R' or third H antigen phases (Grimont and Weill 2007). The Vi capsular antigen, found in few serovars, is given in square brackets after the O antigen designation. Some serovars have

identical antigenic formulas but are distinguished by biochemical properties, pathogenicity or environmental niche.

Salmonella strains are serotyped by slide or tube agglutination with antisera. One H antigen is expressed predominantly; strains can be grown in soft agar containing specific antibody against the expressed antigen to induce a switch to the other H antigen (if present). Agglutination results can be difficult to interpret, leading to difficulties in standardisation even between reference laboratories. In a clinical setting, prior to serotyping, suspected Enterobacteriaceae are often subjected to biochemical testing. Methods such as the API20E (BioMérieux, France) test the ability of an isolate to grow upon certain substrates, produce various metabolic products or alter the pH. Again, results of such tests are open to interpretation and require experience and a high level of technical skill to generate reproducible results.

1.1.1.3 Multi Locus Enzyme Electrophoresis and Sequence Typing

Molecular methods have also been used to distinguish *Salmonella* serovars. Multi locus enzyme electrophoresis (MLEE) distinguishes strains based upon the electrophoretic mobilities of various intracellular enzymes (Selander and Levin 1980; Whittam et al. 1983). Mutations in the genes encoding these enzymes may result in amino acid substitutions that affect the electrostatic charge of the enzyme and hence its electrophoretic mobility. However, the gel-based nature of this typing scheme means that this too is subject to interpretation and difficult to standardise across laboratories.

With the advent of fast, low-cost nucleotide sequencing, Maiden and colleagues published their multi locus sequence typing (MLST) scheme in *Neisseria meningitidis*, which directly captured nucleotide sequence variation in 6 housekeeping genes (i.e. genes under stabilising selection) (Maiden et al. 1998). This scheme was subsequently adapted to *Salmonella* using seven housekeeping genes (Kidgell et al. 2002). Defined fragments from each gene are PCR-amplified and sequenced to determine any nucleotide differences. Any base difference in the sequence is considered a new allele and is given a new number. The numbers from each of the seven alleles are put together to form a 'barcode' and each new barcode is designated a sequence type (ST). The primers used for both the PCR amplification and the sequencing are published as standards so results are reproducible between laboratories. A database for *Salmonella enterica* was set up in 2005 (http://pubmlst.org/databases.shtml) and currently has over 3,300 entries encompassing >800 STs (accessed 22 July 2010).

Even finer resolution can be gained by multi locus sequence analysis (MLSA). This approach discriminates between isolates based on the concatenated nucleotide sequence of each typing gene, thus making use of the number, position and type of sequence changes causing allelic variation (Lorenzon et al. 2003; Maidhof et al. 2002; Naser et al. 2005; Thompson et al. 2005). A phylogenetic tree based upon multi locus sequence analysis of multiple *Salmonella* isolates indicates the evolutionary distances between the *S. enterica* subspecies and *S. bongori* (Figure 1-3).

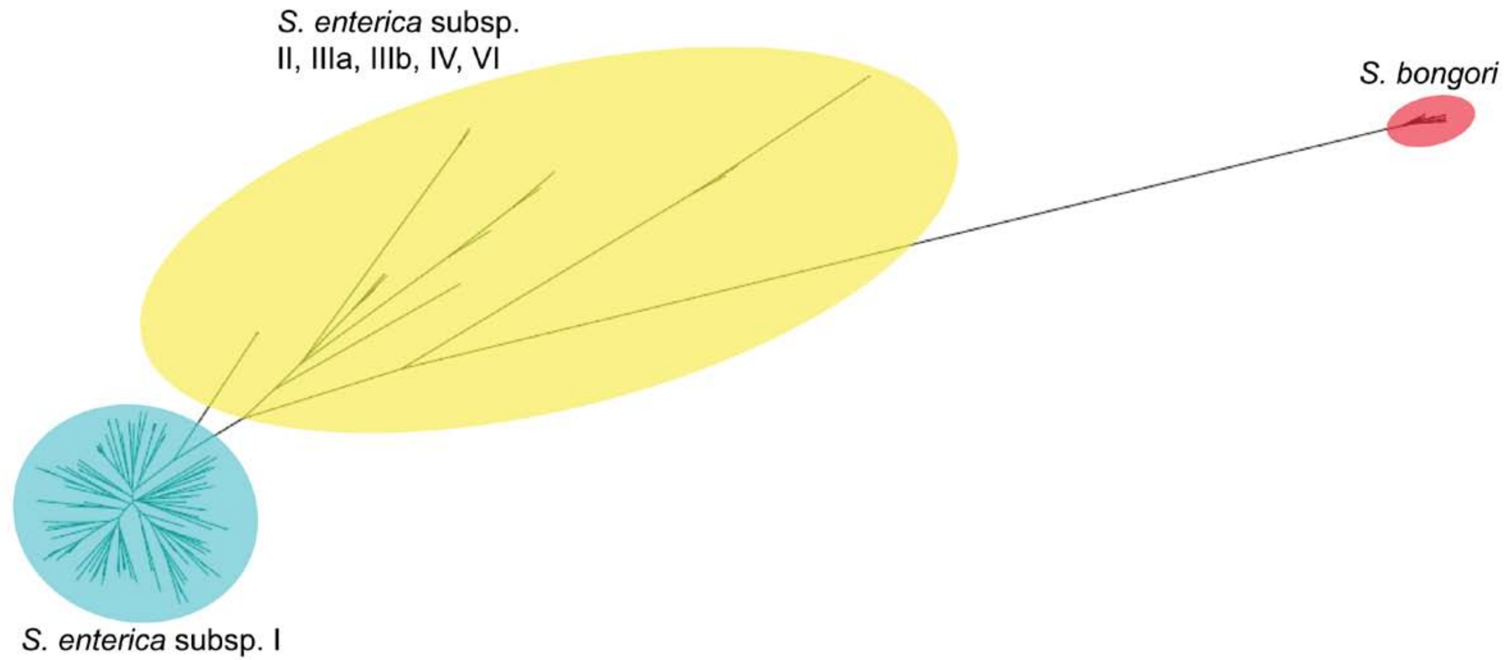


Figure 1-3 MLSA tree for *Salmonella*

Phylogenetic tree based upon concatenated sequence from seven housekeeping genes. Light blue ellipse, *S. enterica* subspecies I; yellow ellipse, *S. enterica* subspecies II, IIIa, IIIb, IV and VI; red ellipse, *S. bongori* (also known as subspecies V). Tree reproduced with permission from Maria Fookes.

1.1.2 *S. bongori*

In 1989, Reeves and colleagues demonstrated by MLEE that subgroup V (as it was then known) was so divergent from the remaining subgroups of *Salmonella* that they believed it should be elevated to separate species status (Reeves et al. 1989). *Salmonella bongori* is thus the second species of the genus *Salmonella*. Analyses using fluorescent amplified-fragment length polymorphism and MLST have both confirmed the earlier findings that *S. bongori* are the most divergent from the other subgroups and represent a binary split from a common ancestor with *S. enterica* (Falush et al. 2006; Scott et al. 2002).

S. bongori is composed of 23 serovars, which exhibit very little sequence variation (Guibourdenche et al. 2009) (Nick Thomson, personal communication). As a species, it is associated with infection mainly in cold-blooded vertebrates; human infections occur rarely and usually as a result of direct contact with reptiles (e.g. pets) (Woodward et al. 1997). However, one serovar, designated 48:z35;- has been documented in southern Italy as the cause of enteritis in young children (Pignato et al. 1998). A molecular analysis in the same region of 31 strains isolated over 15 years from humans, animals and the environment indicated that the isolates may be persisting in the environment and habitually causing human infection (Giammanco et al. 2002).

The whole genome sequencing of multiple *S. bongori* strains in progress at the Sanger Institute will provide an extremely useful resource for *Salmonella* genomics. While this species is not the ancestor of *S. enterica*, extensive genetic information regarding *S. bongori*, a pathogen of mainly cold-blooded animals, addresses the imbalance of

knowledge currently held about *Salmonella* that largely comes from the subspecies of *S. enterica* that cause disease in birds and mammals.

Based upon a core set of genes shared between 15 *Salmonella*, *E. coli* and *Shigella*, *S. bongori* occupies an intermediate taxonomic position between *S. enterica* subspecies I and *E. coli* (N. Thomson, personal communication). It also has the smallest genome of any *Salmonella* sequenced (< 4.5 Mbp) and the lowest average G+C content.

1.1.3 *S. enterica* subspecies *enterica*

Currently, 1,547 serovars make up subspecies I, by far the largest in *S. enterica* (Guibourdenche et al. 2009). These serovars differ widely in the range of hosts they can infect and the diseases caused. The serovars of interest to this project are introduced below.

1.1.3.1 ***S. enterica*** **serovar Typhi**

S. enterica serovar Typhi (Typhi) is the causative agent of typhoid fever, a febrile systemic illness that is a form of enteric fever. Typhi was first described by Eberth and Koch in 1880 and first cultured by Gaffki in 1884. It is identified by the antigenic formula 9[Vi]:d:-, and specific biochemical reactions, including an inability to ferment ornithine and rhamnose. It causes disease solely in humans and current estimates indicate that Typhi causes over 21 million cases of typhoid annually (Bhutta and Threlfall 2009; Crump et al. 2004; Kothari et al. 2008). Research efforts have been focused upon the pathogenicity of Typhi and its host restriction, with the aim of understanding more about

the host-pathogen interaction, and of finding and developing new drug targets and vaccines. Such studies typically concentrated upon the pathogenicity of one strain, with limited knowledge of population genetics.

Analysis by MLST addressed this issue and revealed that Typhi forms a monophyletic group that is approximately 50,000 years old (Kidgell et al. 2002). The 4.8 Mbp genome of Typhi CT18 is predicted to contain more than 4,500 protein coding sequences and 204 pseudogenes (Parkhill et al. 2001a). These pseudogenes are largely conserved (>90%) in Typhi Ty2 (Deng et al. 2003) and in the group of 19 Typhi isolates sequenced more recently (Holt et al. 2008), lending support to the concept that genome degradation has contributed to the host restriction of serovar Typhi. DNA acquisition has also played an important role in the evolution of salmonellae (Kingsley and Baumler 2000). In particular, Typhi contains *Salmonella* Pathogenicity Island (SPI)-7 (Liu and Sanderson 1995b; Wain et al. 2002) which encodes the *viaB* region: genes responsible for the synthesis and transport of the Vi capsular polysaccharide (Hashimoto et al. 1993; Virlogeux et al. 1995). SPI-7 also harbours the SPI-1 type III secretion system (TTSS) effector protein *sopE* (Hardt et al. 1998), and the *pil* genes which encode a type IVB pilus implicated in bacterial self-association (Morris et al. 2003; Tsui et al. 2003) and interaction with epithelial cells (Zhang et al. 2000). Culture collections of Typhi have been shown to contain SPI-7-negative strains (Nair et al. 2004), but the island is almost always present in fresh clinical isolates (Wain et al. 2005). This suggests that SPI-7 plays an important role during the infection process. Whilst all *S. enterica* serovar Paratyphi C and some *S. enterica* serovar Dublin isolates harbour very similar forms of SPI-7

(Pickard et al. 2003), the subtle differences may affect the properties encoded by specific island regions.

1.1.3.2 S. enterica serovar Paratyphi A

Infection with *S. enterica* serovar Paratyphi A (Paratyphi A) also causes enteric fever, being the causative agent of paratyphoid fever. It is identified by the antigenic formula 1,2,12:a:[1,5] which indicates that it is most commonly isolated as monophasic (H:a) but rare isolates are diphasic with phase 2 H:1,5 (Grimont and Weill 2007). Similarly to Typhi, Paratyphi A causes disease solely in humans. Whilst typhoid fever cases have long been known to outnumber paratyphoid, the prevalence of paratyphoid cases in several Asian countries is on the increase (Jin 2008; Ochiai et al. 2005; Palit et al. 2006; Woods et al. 2006). However, the risk factors for acquiring Paratyphi A may differ from those for Typhi (Vollaard et al. 2004), suggesting potential differences in transmission routes.

There are two full 4.6 Mbp genome sequences now available for Paratyphi A, which show the presence of ~ 4,200 protein coding sequences and over 170 pseudogenes per genome (Holt et al. 2009b; McClelland et al. 2004). The close genetic relationship believed to be shared between Paratyphi A and Typhi is the result of multiple recombination events occurring over a quarter of their genomes (Didelot et al. 2007). The remaining three-quarters are as diverse as any pair of *S. enterica* serovars. In addition, Paratyphi A does not harbour SPI-7, an intriguing observation that raises questions concerning how this serovar causes a disease clinically indistinguishable from typhoid and about the role of SPI-7 in typhoid fever (Maskey et al. 2006; Woods et al. 2006).

1.1.3.3 S. enterica serovar Paratyphi C

S. enterica serovar Paratyphi C (Paratyphi C) is capable of causing invasive disease in humans and is probably restricted to this host (MLST study in progress, Satheesh Nair, personal communication). Paratyphi C shares its antigenic formula (6,7:c:1,5) with four other serovars that are recognised as variants based on biochemical properties: Typhisuis, Choleraesuis, Choleraesuis var. Kunzendorf, and Choleraesuis var. Decatur (Grimont and Weill 2007). In most clinical settings, defining a *Salmonella* isolate by its antigenic formula is the furthest point to which typing is taken. Hence, distinguishing serovars beyond this has only been undertaken in a research environment or in reference labotatories. Besides biochemical testing, two other techniques, ribotyping and IS*200* fingerprinting have been used to separate these five serovars (Uzzau et al. 1999)(S. Nair, personal communication). More recently, a global collection of Paratyphi C was characterised by MLST, revealing that all strains formed a complex of three STs, separated only by single allele differences. The complex was dominated by ST146 (34/47 strains), with ST90 and ST114 occurring much more rarely (10/47 and 3/47 respectively, Table 1-1). The strains were also investigated for the presence of pathogenicity islands, revealing that all carried SPI-7, though not in identical form to the Typhi SPI-7 (S. Nair, personal communication). While the *viaB* locus was present, only one strain out of 47 was found to express the Vi antigen.

One of the rare Paratyphi C STs (ST 114) has been sequenced, revealing a 4.8 Mbp genome encoding ~ 4,600 genes and 152 pseudogenes (Liu et al. 2009). This gives insight into the evolution of a potentially human host-restricted serovar separate from Typhi and Paratyphi A, but may not be representative of Paratyphi C as a whole. Current

sequencing efforts are focused upon a strain from ST146, the commonest form currently in circulation (S. Nair, personal communication).

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

MLST complex*	ST	Serovar	Source		
			Human	Pig	Unknown
145	66	CK	28	25	3
	68	C	2	4	5
	133	CK	1	0	0
	139	C	0	0	4
	145	CK	6	7	20
	246	†	0	0	1
	363	CK	0	0	1
146	90	P	5	0	5
	114	P	0	0	3
	146	P	18	0	16
147	147	T	0	1	3

Adapted with permission from S. Nair and based upon a global collection of >150 strains. * Sequence Types (ST)s within a complex share 6 alleles; C, Choleraesuis; CK, Choleraesuis var. Kunzendorf; P, Paratyphi C; T, Typhisuis; † unknown serovar.

1.1.3.4 *S. enterica* serovar Choleraesuis

S. enterica serovar Choleraesuis (Choleraesuis) shares the antigenic formula of Paratyphi C. It is partly defined by its host range; Choleraesuis is adapted to pigs, where it causes swine paratyphoid fever, but it is also capable of causing disease in humans (Table 1-1). Human infection is often associated with underlying diseases in the patient (Chiu et al.

2004a; Chiu et al. 2004b; Wang et al. 2006) and is highly invasive. A measure of the invasiveness of a serovar can be obtained by dividing the number of isolates derived from blood by the total number of isolates to give an 'invasive index'. The higher the percentage, the more invasive the serovar. There are reports from Taiwan, the USA and England & Wales of invasive indexes ranging from 52% to 74%, indicating that infection with Choleraesuis often leads to systemic disease in humans (Chiu et al. 2006; Langridge et al. 2009a; Lauderdale et al. 2006; Threlfall et al. 1992; Vugia et al. 2004).

MLST analysis of Choleraesuis, Choleraesuis var. Kunzendorf and Choleraesuis var. Decatur showed that the first two form a complex of 7 STs, dominated by ST66 and ST145. The latter did not form part of this complex; it was separated into three different STs, each different by 4 or more alleles, a strong indicator of polyphyly. The genome of Choleraesuis SC-B67 (ST 66) is available and has been shown to contain 151 pseudogenes (Chiu et al. 2005), a level of genome degradation perhaps reflecting its strong adaptation to swine.

1.1.3.5 S. enterica **serovar Typhisuis**

S. enterica serovar Typhisuis (Typhisuis) is a host-restricted serovar that is not isolated from humans. Infection in the natural host, pigs, ranges from enterocolitis to chronic paratyphoid fever (Rodriguez-Buenfil et al. 2004; Uzzau et al. 2000). Typhisuis is a cystine auxotroph and forms a monophyletic group by MLEE and by MLST (from the few isolates tested) (Boyd et al. 1993; Selander et al. 1990; Uzzau et al. 2000; S. Nair, personal communication). Differentiation from Paratyphi C is based on the prototrophic

nature of the latter and that all Typhisuis strains typed by MLST thus far are ST147 (S. Nair, personal communication). The use of arabinose and trehalose fermentation tests has also been described, since Typhisuis utilises both but Paratyphi C only utilises arabinose (Table 1-2) (Uzzau et al. 1999). Differentiation from Choleraesuis is based upon tartrate utilisation, and for var. Kunzendorf, hydrogen sulphide (H_2S) production.

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

Serovar of *Salmonella*	Tartrate	Dulcitol	H_2S	Trehalose	Arabinose
Paratyphi C	+	+	+	-	+
Choleraesuis	+	-	-	-	-
Choleraesuis var. Kunzendorf	+	-	+	-	-
Typhisuis	-	-	-	+	+

+ indicates production of substrate (H_2S) or a positive growth phenotype upon substrate, - indicates the opposite. Taken from (Le Minor et al.) and (Uzzau et al. 1999).

1.1.3.6 S. enterica serovar Gallinarum

S. enterica serovar Gallinarum (Gallinarum) was described as the causative agent of an invasive typhoid-like disease in chickens nearly a century ago (Smith and TenBroeck 1915; St John-Brooks and Rhodes 1923). This serovar is non-motile and has the antigenic formula 1,9,12:-:-. Gallinarum is highly adapted to the chicken host and hence presents little public health threat (Shivaprasad 2000). However, in common with human-restricted Typhi, the genome sequence of this serovar has revealed a large amount of genome degradation (~7% of the genome is represented by pseudogenes) (Thomson et al. 2008). Gallinarum forms a related strain cluster with both host-adapted (*S. enterica*

serovars Dublin and Pullorum) and host-generalist (*S. enterica* serovar Enteritidis) serovars (Thomson et al. 2008).

1.1.3.7 S. enterica serovar Typhimurium

S. enterica serovar Typhimurium (Typhimurium) is one of the leading causes of foodborne gastroenteritis (Zhang et al. 2003) and is the focus of national surveillance systems across the globe. The antigenic formula for this serovar is 1,4,[5],12:i:1,2 but many subtypes have been described, mainly defined by phage typing (Anderson et al. 1977; Callow 1959) and less frequently by MLST. These are often referred to as 'variants' or 'pathovariants' of Typhimurium as they differ in both host range and level of host adaptation (Rabsch et al. 2002). For example, in both developed and less developed countries, Typhimurium is one of the most commonly reported causes of extra-intestinal non-typhoidal salmonellosis (Brown and Eykyn 2000; Kariuki et al. 2005). Such systemic infections are clinically distinct from those caused by Typhi and Paratyphi A (Gordon et al. 2002; Graham et al. 2000). Recently, MLST has been used to analyse a set of Typhimurium strains that caused invasive disease in humans in sub-Saharan Africa; ST313 was identified as the dominant type (Kingsley et al. 2009). At the other end of the spectrum, ST19 is a prototypical Typhimurium variant, having been recorded in the MLST database as being isolated from cattle, pigs, humans, poultry and horses. This ST, which includes the mouse-virulent strain SL1344, is often the host-generalist comparator used in studies comparing host-adapted and non-adapted strains of *Salmonella enterica.*

1.2 Host adaptation and restriction

1.2.1 Definitions

The host range of a bacterium is defined by which higher organisms (hosts) it is capable of naturally infecting. Bacteria which can infect multiple hosts are known as host-promiscuous or host generalists. When bacteria are highly associated with infection in one organism but remain capable of natural infection in others, they are referred to as host adapted. This is the case for a number of *Salmonella* serovars, including Dublin (adapted to cattle), Choleraesuis (to pigs), and Obortusovis (to sheep). However, when bacteria are only capable of infecting a single host, this is termed host restriction or specialism and may be due to both gene acquisition and loss of gene function. Often, a spectrum of host specificities exist within the same genus, e.g. *Bordetella*, *Escherichia* and *Mycobacterium*, as well as *Salmonella*.

1.2.2 Process of adaptation and restriction

The genomes of pathogenic bacteria are continually changing through various evolutionary processes. Acquisition of new virulence factors or even pathogenicity islands can occur via horizontal gene transfer and subsequently affect the host range and virulence of the recipient. In *Francisella tularensis*, two of the four subspecies are human pathogenic strains; the remaining two are rarely implicated in disease. Genome comparisons between 3 of the subspecies revealed 41 genes present in the pathogenic strains that were absent from the non-pathogen (Rohmer et al. 2007). Similarly in *Listeria*, the two pathogenic species, *L. monocytogenes* and *L. ivanovii* both contain the

'*Listeria* virulence locus', which encodes *hlyA*, a toxin required for virulence. While the locus is present in one of the non-pathogenic *L. seeligeri*, it is apparently inactivated by an insertion. The same locus is absent from all other non-pathogenic *Listeria* species (Vázquez-Boland et al. 2001).

Acquired traits only form part of the picture regarding host adaptation and restriction. Entirely host restricted organisms are often characterised by extensive genome decay, through insertion sequence (IS) element proliferation, genomic rearrangement and pseudogene formation.

1.2.2.1 IS elements

Increase in IS element copy number has been linked with the early stages of genome reduction (Moran and Plague 2004). *Shigella flexneri* and *Bordetella pertussis* are prime examples of genomes containing high densities of IS elements (Parkhill et al. 2003) (Wei et al. 2003). This is believed to be associated with relatively recent host adaptation (Moran and Plague 2004). The more IS elements that are present in a genome, the greater the likelihood of recombination between these repeats. Indeed, a genome comparison of *B. pertussis* and *B. bronchispetica* revealed that almost 90% of the 150 recombination events in *B. pertussis* were delimited by IS elements (Parkhill et al. 2003). Conversely, Typhi has relatively few IS elements; 26 copies of IS200F and 3 copies of other ISs make up the entire complement in strain Ty2 (Deng et al. 2003).

1.2.2.2 Genome rearrangement

While rearrangements due to recombination of IS elements are minimal, large rearrangements have occurred around the seven *rrn* operons present in the *Salmonella* genome (Liu and Sanderson 1995c). While the order of genomic fragments (delimited by the *rrn* operons) is stable and conserved in Typhimurium, 21 different arrangements were observed in a 127 strain collection of Typhi (Liu and Sanderson 1996). Gene order within these fragments is conserved, but the mechanism(s) governing the genomic balance between origin and terminus of replication and any gene dosage effects appear to have a reduced role in Typhi (Liu and Sanderson 1996). Out of 10 Paratyphi A strains tested, only one genomic arrangement was found, although this represents an inversion of half the genome with respect to Typhimurium (Liu and Sanderson 1995a). It is possible that more rearrangements would be found if a larger strain collection was tested.

1.2.2.3 Pseudogenes

Genes whose functions have been lost over evolutionary time (pseudogenes) were initially thought to be rare in bacterial genomes, due to their compact genome size and general lack of non-coding DNA (Lawrence et al. 2001). However, as more bacterial genome sequences became available, especially those of recently emerged pathogens, pseudogenes were recognised as particularly common among obligate symbionts (Cole et al. 2001; Parkhill et al. 2003; Wei et al. 2003). Identification of pseudogenes remains largely due to comparative analyses between closely related genomes, looking for stop

codons, small indels and truncations that may affect gene function (Dagan et al. 2006; Lerat and Ochman 2004; Lerat and Ochman 2005).

Pseudogene formation is a hallmark of many host restricted organisms. One extreme case is *Mycobacterium leprae*, where 41% of the potential protein-coding genes are non-functional (Cole et al. 2001). Smaller but nonetheless significant pseudogene complements have been discovered in *Shigella flexneri*, and in *Yersinia pestis*, which was extended upon comparison to the less pathogenic *Y. pseudotuberculosis* (Chain et al. 2004; Parkhill et al. 2001b; Wei et al. 2003). In all these cases, pseudogenes were predicted by comparison with intact orthologues. Hence, genes appear to be inactivated more quickly than they are removed by deletion (Cole et al. 2001). As expected of a host generalist, the sequence of Typhimurium strain LT2 (ST19) contains very few pseudogenes, making up only ~ 0.9% of its gene complement (McClelland et al. 2001). Host specialists Typhi and Paratyphi A however, have between 4.5 and 5% of their genes as pseudogenes (McClelland et al. 2004; Parkhill et al. 2001a). Comparisons of individual pseudogenes have shown that at least a third of those shared (estimates range from 28-66 depending upon genome annotation) between both serovars contain different inactivating mutations (Holt et al. 2009b; McClelland et al. 2004). Thus, these two pathogens, if adapting to the niche of the human host by loss of gene function, seem to be doing so along a path of convergent evolution. The degree of host specificity displayed by particular *Salmonella* serovars generally correlates positively with the amount of gene degradation revealed upon whole genome sequencing. Chicken-restricted Gallinarum remains the most extreme example, with over 300 pseudogenes (7.2%). Choleraesuis and Enteritidis, which display strong host associations with swine and chickens respectively,

have 151 (3.3%) and 113 (2.6%) pseudogenes each. As mentioned above, a host promiscuous variant of Typhimurium contains only 39 (0.9%) pseudogenes. One serovar that does not quite fit this pattern is Paratyphi C. MLST analysis suggests this serovar is human restricted, but the pseudogene complement of 149 (3.3%) places it with Choleraesuis, a serovar capable of infecting both pigs and humans (Liu et al. 2009). However, this estimate may be revised upwards when the sequence from the more representative ST146 isolate becomes available.

1.3 Disease and infection of the host

Typhoid and paratyphoid fever (collectively enteric fever) occur primarily in regions of the world where clean water supplies and sanitation are inadequate. Over 21 million cases occur annually (Bhutta and Threlfall 2009; Crump et al. 2004; Kothari et al. 2008).

1.3.1 Clinical features of enteric fever

With enteric fever, the fever itself is the only consistent symptom (Parry 2004; Parry et al. 2002), although abdominal pain or discomfort, muscle and/or joint pain, and headache are frequently observed. Recent studies in Nepal suggest that Typhi and Paratyphi A cause clinically indistinguishable syndromes (Maskey et al. 2006; Woods et al. 2006). Once an appropriate antibiotic course has been administered, these symptoms usually resolve quickly. However, persistence of several weeks can occur in an untreated patient (Parry 2004; Parry et al. 2002). Complications can develop from Typhi infection (Deng et al. 2003), the most severe of which is gastrointestinal haemorrhage and perforation (Chanh et al. 2004; Everest et al. 2001). This condition requires both surgical and antimicrobial intervention and carries a high risk of mortality (Butler et al. 1985). An estimated 5% of typhoid patients become chronically infected by retention of Typhi in the gall bladder and continue to excrete Typhi for many years (Parry et al. 2002). These typhoid carriers not only pose a significant health risk to others, but also have a higher risk of developing cancer of the gallbladder, pancreas and large bowel (Caygill et al. 1995; Dutta et al. 2000; Shukla et al. 2000). Carriers also provide a reservoir for Typhi, contributing significantly to the persistence of typhoid in endemic regions (Roumagnac et

al. 2006). Carriage has recently been demonstrated for Paratyphi A, but not to the same level as Typhi (Khatri et al. 2009). What effect Paratyphi A carriage has upon excretion rates and persistence in endemic areas has yet to be determined.

1.3.2 Infection of the host

1.3.2.1 Typhi

Modern publications on typhoid fever in humans are few, and much of the information on typhoid pathogenesis has been gained through studies on Typhimurium in mice. The infection of mice by Typhimurium is often cited as a model for typhoid fever, since this serovar causes a typhoid-like disease in the mouse, but translating these findings to typhoid fever in humans requires careful interpretation (Sabbagh et al. 2010). The current view of typhoid pathogenesis is shown in Figure 1-4. Typhi is transmitted via the faecal-oral route, with sufferers ingesting an infectious dose of between 10^5 and 10^9 bacterial cells in contaminated food or water (Wain et al. 2002). After passing through the stomach, Typhi invades the gut epithelium of the terminal ileum, possibly using the cystic fibrosis transmembrane conductance regulator (CFTR) for entry (Pier et al. 1998).

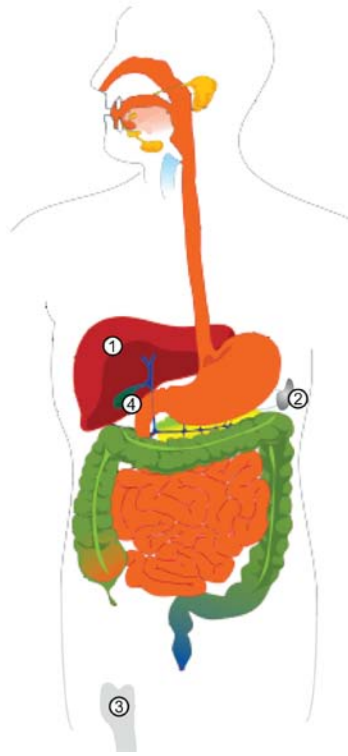


Figure 1-4 Pathogenesis of typhoid fever in humans

Adapted from (Wain et al. 2002). Transmission occurs via the faecal-oral route, leading to the stomach. After penetrating the intestinal epithelium, Typhi disseminates through the body in the bloodstream and seeds the liver (1), spleen (2), and bone marrow (3). Some patients make a full recovery, but a small percentage progress to an asymptomatic chronic infection (carriage) in the gall bladder (4). Human image from http://commons.wikimedia.org.

Synthesis of the Vi capsular polysaccharide (encoded on SPI-7) is down-regulated under the low osmotic conditions found at the intestinal epithelial barrier. These conditions also promote the secretion of effector proteins via the SPI-1 encoded type III secretion system and hence an adhesive and invasive phenotype (Sukhan 2000; Wehland and Bernhard 2000). This invasion triggers the secretion of interleukin(IL)-6 from host cells, and the bacteria are taken up by or invade macrophages. In addition, CD18+ host cells that have migrated into the gut lumen may also take up bacteria, which they transfer across the gut epithelium when they migrate back into gut tissue (Wain et al. 2002). High expression of

SPI-1 at this point results in caspase-1-mediated death of macrophages, release of more cytokines and Typhi is subsequently disseminated throughout the body via the bloodstream (House et al. 2001). Symptoms of typhoid fever are not yet apparent at this stage of the infection; it is a secondary bacteraemia resulting from bacterial replication within the liver, spleen and bone marrow that causes the onset of clinical symptoms, when Typhi can usually be cultured from the blood or bone marrow, albeit at very low levels (<1 CFU/mL) (Wain et al. 2002).

1.3.2.2 Differences in Typhimurium

Typhimurium is ingested in the same manner as Typhi, although gastroenteritis rather than systemic disease is the usual outcome. Typhimurium also invades the gut epithelium, but is associated with host cell release of IL-8 and recruitment of neutrophils to the site of infection, causing localised inflammation (House et al. 2001). Bacteria do not disseminate through the body, and the diarrhoeal symptoms caused by the infection are usually self-limiting.

1.3.3 Molecular basis of infection

1.3.3.1 Invading the macrophage

Wildtype Typhi is classified as a hazard group 3 organism, which is one reason why much of the literature on the interaction of *Salmonella* with macrophages is derived from studies performed with Typhimurium (hazard group 2) in murine cells.

The importance of infection and survival within the macrophage was noted when it was shown that Typhimurium mutants unable to survive inside these cells were avirulent in the mouse (Fields et al. 1986). A screen of almost 10,000 transposon mutants for survival in macrophages identified a number of auxotrophies that were also associated with decreased virulence in the murine model. These included requirements for purines, pyrimidines and histidine, which indicates the importance of bacterial metabolism in survival inside the macrophage (Fields et al. 1986). A proteomic study from the same group showed that over thirty Typhimurium proteins were synthesised during murine macrophage infection, with the heat-shock proteins DnaK and GroEL the most abundant (Buchmeier and Heffron 1990).

It has been postulated that Typhi is unable to infect other hosts because it lacks genes present in Typhimurium that allow it to colonise a broad host range (Morrow et al. 1999). Prior to the availability of full genome sequences, genomic subtractive hybridisation was used to isolate gene sequences in Typhimurium not present in Typhi. Using a technique called ‘selective capture of transcribed sequences’ (SCOTS), RNA from Typhimurium inside macrophages was labelled and hybridised against the subtracted sequences to identify those which were expressed inside the macrophage (Morrow et al. 1999). The

SCOTS technique identified a putative transcriptional regulator and a novel fimbrial operon expressed by Typhimurium inside macrophages, neither of which are present in the Typhi genome. However, when tested, mutations in the regulator did not affect virulence in the mouse or survival within the macrophage (Morrow et al. 1999). As such then, there may not be individual genes that can be linked with broad host avirulence in Typhi, but both the absence of genes present in broad-host range *Salmonella*, and the presence of pseudogenes in the Typhi genome likely contribute to the restriction of this serovar to its human host.

The SCOTS technique was also used to look (for the first time) at which Typhi genes are expressed inside human macrophages. In the study, over twenty Typhi genes were identified, including eight of unknown function (Daigle et al. 2001). Other predicted functions included Vi capsule biosynthesis and the stress response. In a follow-up study, effectively a complementation of the Typhimurium subtractive hybridisation experiment was performed to look for Typhi-specific genes expressed inside the human macrophage (Faucher et al. 2005). Thirty-six Typhi genes were found to be expressed intracellularly, of which twenty-five were encoded on pathogenicity islands including SPI-7, and prophage elements.

By combining SCOTS with microarray analysis, a more detailed picture began to emerge of the environmental conditions encountered by Typhi during macrophage infection. As a functional class, iron transport was repressed, as were motility and peroxide-induced functions, while antimicrobial peptide resistance was induced (Faucher et al. 2006). Typhi was therefore predicted to reside in an environment that is neither acidic nor oxidative.

Similar studies have been performed using Typhimurium, and have made other predictions about the intracellular environment. Microarray analysis showed that Typhimurium expression of amino acid, potassium and iron transport was not induced but that phosphate and magnesium may be limiting (Eriksson et al. 2003). Also, once inside murine macrophages, Typhimurium has been predicted to use gluconate and related carbohydrates for growth (Eriksson et al. 2003). However, Typhi does not encode the *dgo* genes required to break down these compounds, and there was no SCOTS evidence that the relevant transporters were being induced, suggesting that Typhi may use different carbon sources to Typhimurium for intracellular growth (Faucher et al. 2006).

1.3.3.2 The **Salmonella**-*containing vacuole*

In the early 1990s, microscopy techniques were used to investigate how *S. enterica* invade host epithelial cells and macrophages. It was initially discovered that *S. enterica* triggered membrane ruffling upon contact with the surface of host cells (Galan et al. 1992), and later shown that this was directly related to the formation of intracellular 'spacious phagosomes' containing *S. enterica* (Alpuche-Aranda et al. 1994; Garcia-del Portillo and Finlay 1994). A large body of work has been conducted regarding the generation and maintenance of what are now called '*Salmonella*-containing vacuoles' (SCVs), but this has, for the most part, been concentrated upon Typhimurium (see (Holden 2002) and (Knodler and Steele-Mortimer 2003) for reviews). However, one early study looked at both Typhi and Typhimurium and concluded that the mechanisms involved in invasion and intracellular trafficking are very similar for both serovars in epithelial cells, which are non-phagocytic (Mills and Finlay 1994).

Normally, eukaryotic cells, including epithelial cells, utilise the endocytic pathway for internalising, sorting, recycling and degrading molecules recovered from the cell surface or from extracellular space. Such molecules are endocytosed and progress through early endosomes, multivesicular bodies, late endosomes and finally lysosomes (Knodler and Steele-Mortimer 2003). Since invasion of epithelial cells requires expression of SPI-1 and SPI-1 related genes, the experimental conditions for achieving cell infection are relatively standardised and have produced a consensus view of SCV biogenesis and maturation in this cell type (Figure 1-5).

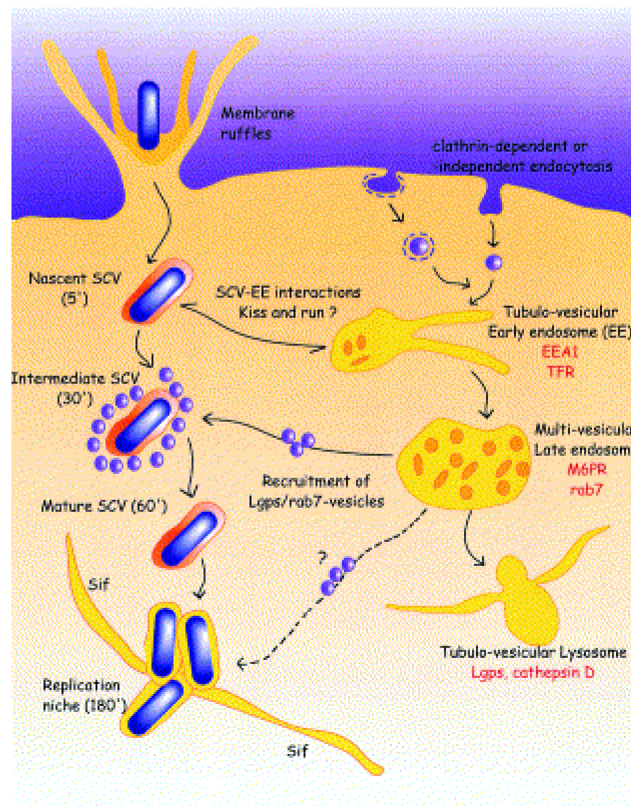


Figure 1-5 Intracellular pathway of *Salmonella* in epithelial cells

From (Gorvel and Méresse 2001). Membrane ruffles mark the beginning of invasion in non-phagocytic host cells. Once inside the host cell, *Salmonella* are found in nascent SCVs, which are modulated by *Salmonella* to achieve maturation and enable replication. The approximate time course is shown in minutes post infection.

In macrophages, molecules are phagocytosed and the subsequent phagosome interacts sequentially with the endocytic compartments and requires the activity of Rab GTPases. However, when *Salmonella* invade, while interactions do occur with the endocytic pathway, degradation is avoided by blocking the fusion of the SCV with terminal acidic lysosomes (Buchmeier and Heffron 1991; Ishibashi and Arai 1990). There has been greater debate over the process of SCV maturation in macrophages, for multiple reasons. Experimental conditions for infection of macrophages are much less consistent and since these cells are phagocytic, the mechanism of *Salmonella* entry varies. Conflicting evidence has also been presented regarding the interaction of SCVs with the late endocytic pathway, with some studies showing that fusion between the SCV and lysosomes is blocked (Buchmeier and Heffron 1991; Ishibashi and Arai 1990), but others demonstrating that this fusion does occur (Drecktrah et al. 2007; Oh et al. 1996). However, a recent study has documented an imbalance in the ratio of acidic lysosomes to SCVs caused by Typhimurium, such that the stock of lysosomes is exhausted before all the SCVs have been targeted (Eswarappa et al. 2010). This may provide a unifying explanation for the evidence for and against SCV/lysosome fusion, and serves to highlight the need for further investigation of the roles played by Typhi and Typhimurium inside macrophages.

1.4 *Genetic diversity in* Salmonella

The variety of diseases caused by different *S. enterica* serovars have helped to provide impetus for improving the methods by which serovars and strains can be distinguished. Initially, the main focus of such methods was for epidemiological and public health reasons, but as the resolution offered by new technologies has improved, such methods have become useful from evolutionary and population biology viewpoints as well.

1.4.1.1 Phage typing

In the late 1930s, early 1940s, a typing scheme for *Bacillus typhosus* (Typhi) was being used as an epidemiological tool to differentiate subtypes based upon resistance to Vi phage adapted to various *B. typhosus* strains (Anderson and Felix 1953; Felix 1943). A similar scheme was soon being applied to Typhimurium, which in the early stages used 29 phages to distinguish 34 types (Callow 1959). At this time, other schemes were also in use for Paratyphi A and Paratyphi B (Banker 1955; Felix and Callow 1943). By the late 1970s, phage typing was the method of choice for the epidemiological tracking of Typhimurium outbreaks, and the scheme had expanded to include 34 phages used to differentiate 207 types (Anderson et al. 1977). By the late 1980s, Enteritidis had become the top serovar causing human infection in the UK, so a phage typing scheme was developed that distinguished 27 types with 10 phages (Ward et al. 1987). Phage typing is still used today, particularly to differentiate the economically important subtypes of Enteritidis and Typhimurium such as PT4 and DT104 respectively. Enteritidis is a prominent cause of non-typhoidal salmonellosis, and phage type (PT) 4 caused 49% of

Enteritidis outbreaks in the USA in 1999 (Patrick et al. 2004). Similarly, Typhimurium definitive type (DT) 104 has caused smaller outbreaks of multidrug resistant salmonellosis since the early 1990s (Helms et al. 2005).

1.4.1.2 Plasmid profiling

Over 30 years ago, methods for extracting plasmid DNA were being optimised to aid in differentiating between bacterial isolates (Kado and Liu 1981; Schaberg et al. 1981). With a clean DNA sample, plasmids were run out on a gel by electrophoresis, resulting in a distinct 'fingerprint' pattern (Schaberg et al. 1981). The utility of this technique was shown even outside a clinical setting by Brunner and colleagues in Switzerland, where Typhimurium isolates were categorised in 6 plasmid patterns (PPs), with PP1 identified as the type responsible for a minor epidemic (Brunner et al. 1983). A later study looked at multiple outbreaks of Typhimurium in the USA and found that, in outbreaks where unrelated isolates were also analysed, the outbreak isolates were differentiated from the unrelated 8/9 times using plasmid profiling (Holmberg et al. 1984). Phage typing only differentiated 6/9 and antibiotic resistance screening 4/9. A Spanish study, looking mainly at Enteritidis, also compared plasmid profiling, phage typing and resistance screening to determine the relative merits of each (Borrego et al. 1992). They found plasmid profiling to be superior to phage typing and antibiotic resistance screening as an epidemiological marker although they also stated that none gave total discrimination. On a somewhat contradictory note, Threlfall and colleagues directly compared phage typing and plasmid profiling and found that a particular plasmid pattern was found in more than

one phage type and that multiple plasmid patterns could be found in a single phage type (Threlfall et al. 1989).

Plasmid profiles therefore can be associated with a specific strain of *Salmonella*, but these may also be transferred. Fine typing of plasmids and the strains in which they reside is shedding more light on this relationship (Minh-Duy Phan, personal communication).

1.4.1.3 Signature tagged mutagenesis

In the mid-1990s, transposon-based mutant screening emerged. For the first time, multiple mutants could be screened simultaneously for attenuation in an animal model (Hensel et al. 1995). In the first published signature-tagged mutagenesis experiment, unique, identifying DNA tags were ligated onto a Tn5-derived transposon that was used to conjugate a mouse-virulent variant of Typhimurium (Hensel et al. 1995). Over 1100 successfully transformed exoconjugants were separated into 96-well microtitre plates and pools of 96 mutants were screened through mice. Bacteria recovered from the spleen were subjected to a tag-specific PCR incorporating a radiolabelled nucleoside triphosphate (dCTP). A probed colony blot of ‘recovered’ bacterial PCR products was then compared with the equivalent blot from the inoculum. This typified a ‘negative’ selection screen, where transposon mutants that were present in the inoculum but absent from the recovered pool represented those attenuated in the infection model. A total of 39 attenuated mutants were identified and the transposon-chromosome junction sequenced for 28 of them. Five of these mutants were in genes related to type III secretion and led the authors to investigate further (Shea et al. 1996). An entire virulence locus was then

identified; yielding the first description of SPI-2, an island unique in *Salmonella* to the *enterica* species (Hensel et al. 1997a; Shea et al. 1996), that was later demonstrated to be required for the survival of *S. enterica* in macrophages (Cirillo et al. 1998; Hensel et al. 1998).

1.4.1.4 Microarrays

The advent of whole genome sequencing quickly led to the production of bacterial microarrays. These were initially the result of PCR products spotted onto glass slides that were probed with fluorescently labelled genomic DNA (Dziejman et al. 2002; McClelland et al. 2001; Smoot et al. 2002), but later oligonucleotide arrays were also produced. Microarrays were used as a tool for gene expression profiling and for phylogenetic typing (Lucchini et al. 2001). Using microarrays to type related bacteria meant that differences in gene content could be visualised and quantified without needing the full genome sequence of every strain tested. In 2003, Porwollik and colleagues produced a non-redundant microarray for Typhimurium and Typhi that contained probes for all the genes in Typhimurium and was supplemented with probes for any Typhi genes that were > 10% divergent from Typhimurium (Porwollik et al. 2003). This established the concept of using close homologues to characterise multiple serovars with the same underlying microarray (Porwollik et al. 2004). Characterisation of almost 80 strains showed that there may be hundreds of genes different between strains of the same serovar and that in some cases there was more intra-serovar variation than inter-serovar variation (Porwollik et al. 2004). DNA microarrays have also been used to investigate the role of prophage-like elements in generating diversity within *S. enterica* (Thomson et al. 2004).

Using a microarray based on Typhi, supplemented with novel Typhimurium probes, Thomson and colleagues were able to compare strains from 20 serovars against the phage complement present in Typhi. They found that the Typhi strains harboured a set of temperate bacteriophage unique from all the other serovars tested and were also able to detect more subtle intra-serovar variations (Thomson et al. 2004). Using the Sanger sequenced Enteritidis strain for microarray construction, a study found that phage type (PT)-8 strains harboured a particular set of phage genes believed to be the molecular basis of the distinction of PT8 from PT4, the two types responsible for the majority of infection caused by Enteritidis (Porwollik et al. 2005).

Transposon-based assays also took a step forward with the production of microarrays. Techniques such as TraSH (transposon site hybridisation) and TMDH (transposon-mediated differential hybridisation) relied upon the use of microarrays to determine the insertion sites of transposons recovered from mutants passed through a selective screen (Chaudhuri et al. 2009; Sassetti et al. 2001). In TMDH, RNA was generated by outwards *in vitro* transcription from the transposon into adjacent genomic DNA and hybridised to the microarray. Conditionally essential genes were identified when a strong hybridisation signal from the input sample was missing in the output. However, background levels of hybridisation made distinguishing the on/off signal difficult, and as with all microarrays, a key limitation was that information could only be gained about what was on the array – novel genes and intergenic regions could not be assessed.

1.4.1.5 High-throughput sequencing

As Typhi is genetically monomorphic, gaining deeper insight into global diversity and evolutionary history required a high resolution approach (Achtman 2008). Two hundred 500 bp DNA fragments were sequenced from a global collection of >100 strains to look for informative mutations. Nineteen mutations that marked the evolutionary history of Typhi were found, with a further 69 that helped to define 59 distinct haplotypes (Roumagnac et al. 2006). Unusually, the minimal spanning tree generated from this data was rooted in an extant haplotype, H45, from which several lineages have descended. What was also striking was that many haplotypes were found on multiple continents, including H45 which was isolated in Africa, Asia and North America. This sheds little light on the question of where Typhi evolved. An investigation of a further 161 Typhi isolates from Indonesia looked at 84 single nucleotide polymorphisms (SNPs) as markers of genome variation (Baker et al. 2008). These isolates were assigned to nine haplotypes, indicating that multiple haplotypes were in circulation and that such haplotypes persist, as some were isolated repeatedly over a 30 year period (Baker et al. 2008). SNP genotyping can be achieved with systems like the Illumina GoldenGate and Sequenom massARRAY. However, this does not allow for novel SNP detection. To achieve this, Holt and colleagues sequenced multiple strains of Typhi using both 454 and Illumina (formerly Solexa) sequencing (Holt et al. 2008) where the isolates were chosen to represent major nodes from the phylogenetic tree (Roumagnac et al. 2006). By comparing generated sequence reads to the reference sequence of Typhi CT18, almost 2000 SNPs were detected, 10-fold more than previously, which allowed an even higher-scale resolution of the tree with improved branch length estimates. While the cost of sequencing isolates

individually was prohibitive to genome-wide SNP detection, a method was developed to reduce this cost. Six individual strains of Paratyphi A and a pooled DNA sample, containing an equivalent amount from each strain, were sequenced using the short read high-throughput Illumina system (Holt et al. 2009a). Of 550 SNP loci checked in each strain, over 400 had sufficient high quality sequence coverage to estimate reliable frequencies across the six strains. With the pooled sample, genome coverage of 40x was achieved, and the sequences obtained were compared with the loci and allele frequencies from the individual strains. The sensitivity of SNP detection in the pool was 100% for SNPs that occurred in 3 or more strains, although this declined to 37% if a SNP occurred in just one strain. Overall, this represented a cheaper method of sampling a bacterial population with the aim of unbiased detection of genetic variation (Holt et al. 2009a). More recently however, the ability to 'tag' samples within a pool has become possible, allowing multiple strains to be sequenced simultaneously on the Illumina Genome Analyzer II platform. This has been used to characterise a set of 63 *Staphylococcus aureus* strains, with 23 x coverage achieved on average per strain relative to the reference and revealing over 4000 SNPs (Harris et al.). Making use of whole genome sequences yields the optimum resolution, especially between bacterial isolates sharing the same multi-locus sequence type.

1.5 Metabolism in Salmonella

1.5.1 Phenotypic analysis and biochemistry

By the 1940s, typhoid and paratyphoid epidemics were largely a thing of the past in the USA, as serotyping these 'salmonelloses of human origin' had been the focus of a concerted public health effort (Borman et al. 1943). Instead, attention was turning towards more precise definitions of those *Salmonella* isolates that did not fall into the typhoid group. At this point, various metabolic tests were in place only to identify isolates to the genus level. For example, *Salmonella* were known to be negative for both indole and the Voges-Proskauer test, and usually salicin fermentation. Once an organism had been identified as a *Salmonella*, the only metabolic test that would differentiate below the species level was utilisation of tartrates, as a negative result for this test would suggest *S. schottmuelleri* (Paratyphi B) and was 'seldom wrong' (Borman et al. 1943). Paratyphi B (which causes paratyphoid fever) and Java (gastroenteritis) share the same antigenic formula, 1,4,(5),12:Hb:1,2 and are still distinguished based on the fermentation of *d*-tartrate. This metabolic difference is believed to be due to a putative cation transporter, which is inactivated by a single nucleotide polymorphism in Paratyphi B (Han et al. 2006).

In the 1970s, efforts were being made to collect together multiple biochemical tests to allow rapid and relatively high throughput identification of clinical bacterial isolates (Lindberg et al. 1974). Numerous test kits of varying accuracy became available for the identification of the Enterobacteriaceae. The API system of twenty tests in a sterile plastic strip was 'found to be the most reliable', having a 99% correlation with standard

biochemical tests and a 94% identification rate (Nord et al. 1974). Over 60 bacterial species can currently be identified with the API20E, with identification extending to the serovar level for Typhi and Paratyphi A among others.

1.5.2 Metabolic pathways

Early interest in the metabolism of *Salmonella* and other bacteria was towards the identification of compounds that could act as sole carbon and nitrogen sources and the effects these had upon growth rate (Richmond and Maaloe 1962). A comprehensive screen of ~600 substrates identified 76 carbon sources and 26 nitrogen sources for Typhimurium LT2 (Gutnick et al. 1969). Interestingly, growth was only observed on some of these compounds for mutant derivatives of the parent strain. This information provided a baseline from which studies could be performed to elucidate the mechanisms by which these substrates were utilised. For example, the uptake and degradation of the pentose sugar xylose was investigated and found to mirror the transport system and two catabolic enzymes present in *E. coli* and *Aerobacter* (*Enterobacter*) *aerogenes* (Shamanna and Sanderson 1979). However, utilisation of ethanolamine as a sole carbon and nitrogen source was found to be complicated by the requirement for the cofactor vitamin B_{12} (Roof and Roth 1988). Mutants unable to grow in the presence of ethanolamine and vitamin B_{12} were located to a cluster of genes between *purC* and *cysA* on the chromosome, termed the *eut* (ethanolamine utilisation) region (Roof and Roth 1988).

The majority of experiments on bacterial metabolism have been performed on the genetically tractable model bacterium *E. coli* K12. As a close relative, pathways in

Salmonella have generally been studied in detail only where they differ from *E. coli*. A case in point is the *de novo* synthesis of coenzyme B_{12}, genes for which are present in *S. enterica* serovars but not in *E. coli* (Roth et al. 1996).

1.5.3 Pathway maps

During the late 1990's, numerous metabolic databases became available, either as a general resource or to depict the metabolism of a particular organism. By using computational symbolic theory, the aim of collating a vast quantity of metabolic data into a single database was to give scientists a way to analyse and understand the complexity of biochemical reactions and pathways (Karp 2001).

1.5.3.1 Metabolic databases

One of the best-known general purpose databases is the Kyoto Encyclopaedia of Genes and Genomes (KEGG) (Kanehisa and Goto 2000; Ogata et al. 1999). KEGG is based upon three linked databases, describing genes, higher order biological functions (e.g. metabolic pathways) and chemical compounds. In the higher order database, metabolism is represented as 'reference' pathways which are visualised as a network of enzyme names or Enzyme Commission (EC) numbers. Updated daily, the reference metabolic pathways are based upon gene information taken from completely sequenced genomes. Given suitable enzyme annotation, organism-specific pathways can be generated based upon the reference pathway. The success of this process depends upon whether the

pathway is conserved across multiple organisms. Species-specific variations are also difficult to determine using KEGG.

Other general purpose databases were created, such as the WIT (What Is There) and MPW (Metabolic Pathways Database) systems which formed part of another set of linked databases (Overbeek et al. 2000; Selkov et al. 1998). Similar to KEGG, but on a smaller scale, they comprised a collection of metabolic pathway reconstructions from sequenced genomes (Selkov et al. 1998). However, many of these are now inactive, including PUMA2 (Maltsev et al. 2006)(successor to WIT) and aMAZE (van Helden et al. 2001).

A tool that is still widely used, Cytoscape was developed to model biological interaction networks, and is especially powerful when there are high quality protein-protein, protein-DNA and genetic interaction data available for the organism of interest (Shannon et al. 2003). To this end, model organisms often comprise the best datasets.

EcoCyc is an example of a single organism metabolic database that uses experimental data to generate metabolic pathways (Karp et al. 1997; Keseler et al. 2009). It describes the metabolic capability of *E. coli* (K12) MG1655 and has been curated from the vast literature on this model organism. By concentrating on a single organism, EcoCyc has captured metabolic variations specific to MG1655 and provides links to the experimental evidence that generated these predictions. While EcoCyc remains the most highly curated, a number of other databases are now available from the BioCyc collection (http://www.biocyc.org). However, at the commencement of this thesis project, no serovar-specific *Salmonella* databases were publically available.

1.5.3.2 Pathway prediction

Software for pathway prediction can be broadly divided into two categories based upon the input data required. The first category requires only sequence data while the second requires both sequence data and genome annotation.

metaSHARK and the GEM system represent the first category (Arakawa et al. 2006; Pinney et al. 2005). metaSHARK is a fully automated system for detecting genes encoding enzymes and the subsequent visualisation of these within metabolic networks (Pinney et al. 2005). The novel bioinformatics in this software is almost entirely concentrated upon the gene prediction aspect. The metabolic networks are simply imported from KEGG, with some corrections, and used as a general framework for locating enzymatic reactions. The GEM (Genome-based modelling) system is slightly more oriented towards metabolic pathway prediction, with pathways predicted based upon enzyme function, and then checked against references in KEGG and BioCyc (Arakawa et al. 2006).

A member of the second category, Pathway Tools is the software developed by the bioinformaticians behind EcoCyc. This however requires genome annotation in order to predict the metabolic reactions and pathways encoded by the organism of interest. The reason stated by the authors is that manual efforts render a higher quality genome annotation than can be achieved by computation alone, and this should in turn produce a higher quality set of predicted pathways. Pathway Tools uses MetaCyc, “a database of non-redundant, experimentally elucidated metabolic pathways” (http://www.metacyc.org) as the reference for predicting pathways in the organism of interest. Like EcoCyc,

MetaCyc is curated from the literature, but includes pathways from over 1,800 different organisms.

1.5.4 Network modelling

Once a set of metabolic pathways have been predicted, the next level of analysis comes in assessing how these interact with each other, as they must do in nature, to form a dynamic metabolic network. Often, metabolic networks are built *in silico* and tested and refined by analysing the biomass composition of the organism grown in batch culture (Novak and Loubiere 2000; Reed and Palsson 2003). Without the benefit of *Salmonella*-specific metabolic pathways, the only network analysis published thus far is based upon a large scale comparison with *E. coli* (AbuOun et al. 2009). As a step up from batch culture, the authors were able to make use of a metabolic phenotyping microarray that allowed the model to be tested for its ability to predict growth on over 250 substrates. Once this level of detail can be validated with *Salmonella*-specific information, this will provide a powerful tool for analysing *Salmonella* under different growth conditions.

1.6 Project focus

Despite their close genetic relatedness, it has been known for almost 100 years that *Salmonella* serovars display different metabolic phenotypes, varying for example in their fermentative ability. Figure 1-6 is reproduced from a paper published in 1919 showing metabolic phenotypes associated with organisms across a range of bacteria (Winslow et al. 1919). Of the five *Salmonella* depicted (in red), only two show the same metabolic phenotype across all 13 'fermentative relationships', and Typhi appears to have a relatively restricted profile.

B. aerogenes, *B. cloacae*, *B. neapolitanus*, *B. communior*, *B. coli*, *B. acidi-lactici*, *B. schottmulleri*, *B. enteritidis*, *B. suipestifer*, *B. paratyphosus*, *B. tyhposus*, *B. dysenteriae*, *B. shigae*, *B.alcaligenes*

Hexoses
Mannitol
Maltose
Sorbitol
Gas production
Rhamnose
Arabinose
Xylose
Lactose
Salicin
Sucrose
Raffinose
Voges-Proskauer

Figure 1-6 Metabolic capabilities of *Salmonella* serovars

Figure adapted from (Winslow et al. 1919). Filled boxes represent bacterial ability to metabolise/ produce/react to substrate indicated. White boxes indicate absence of ability. *Salmonella* serovars are highlighted in red. Key: *B. schottmulleri*, Paratyphi B; *B. Enteritidis*, Enteritidis; *B. suipestifer*, Choleraesuis; *B. paratyphosus*, Paratyphi A; *B. typhosus*, Typhi.

While some metabolic phenotypes are absolute, others are more subtle. For example, hydrogen sulphide (H_2S) production is often used to identify particular salmonellae but variation in ability to reduce sulphate to sulphite between isolates of the same serovar and the test method used can alter the result. Thus, the tests used to differentiate salmonellae are not always stable and it is common to incorrectly identify a serovar. However, biochemical testing is still used globally and it performs well, to an extent. This suggests that there must be enough reproducible metabolic phenotypes to define some serovars. These phenotypes are of particular interest as they are likely to identify a large proportion of true serotype-specific metabolic capability.

The level of genome degradation in Typhi and Paratyphi A, coupled with evidence that these serovars have more restricted metabolic profiles than other serovars, has lead to the hypothesis of this thesis: that there is a causal link between host restriction and metabolic ability. However, at the beginning of this project there was little, if any, evidence in the literature. Whilst various complete *Salmonella* genome sequences were available, the lack of evidence was likely due in part to the paucity of information describing gene function and an inability to assign individual pseudogenes to metabolic pathways.

1.7 Aims of the project

Different *Salmonella* serovars adapted to different mammalian hosts contain unique but overlapping sets of pseudogenes. An understanding of the influence of these pseudogenes on the pathogenic and metabolic capability of the host-adapted serovars will help to explain the basic process of host-pathogen adaptation. This project explored the biology of host-restricted serovars of *Salmonella enterica* in comparison to the non-adapted Typhimurium, with the aims of:

- Creating metabolic pathway databases for Typhi and Typhimurium
- Assigning pseudogenes in host-restricted *Salmonella* to metabolic pathways
- Defining the essential gene lists of Typhi and Typhimurium
- Investigating the effect of macrophage invasion upon the metabolic networks of Typhi and Typhimurium

2 Development of StyCyc and StmCyc and validation of known metabolic lesions

2.1 Introduction

Highly curated metabolic pathway databases exist either as collections across multiple species or as more specific resources for particular organisms. Pathway Tools (SRI International, California, USA) is a freely available piece of software for the latter, which predicts a repertoire of metabolic pathways given a fully annotated genome sequence. The Pathway Tools group is responsible for the BioCyc collection of metabolic databases (http://www.biocyc.org), with a particular focus on EcoCyc, the "encyclopaedia of *Escherichia coli* K-12 genes and metabolism" (http://www.ecocyc.org) (Karp et al. 1997; Keseler et al. 2009; Ouzounis and Karp 2000).

A Pathway/Genome Database (PGDB) built using Pathway Tools is designed to represent data on genes, proteins and enzymatic reactions, through to the metabolic pathways they make up. This integrated approach has led to the development of numerous bioinformatic tools to extract the maximum amount of information from a well-annotated genome sequence (Karp et al. 2002). The hallmark of Pathway Tools is to make use of high quality annotation rather than rely entirely upon computational methods. Thus, the software often uses name-matching algorithms to deduce enzymatic function, for example the Transport Identification Parser is used to predict cellular transport reactions, identifying the substrates that an organism can successfully import or export from the cytoplasm (Lee et al. 2008). All reactions, whether in pathways or transporters, are then

visualised in an overview diagram which acts as a framework for any gene-, protein- or metabolite-based data to be 'painted' on (Paley and Karp 2006). The outstanding feature of Pathway Tools however, is the Pathway Hole Filler algorithm (Green and Karp 2004), which facilitates the identification of novel enzymatic functions within the genome sequence being analysed and hence provides updated annotation.

EcoCyc provides an information retrieval system for *E. coli* K12 that is publically available via the internet. This has allowed studies performed on this strain to be analysed in the context of metabolic pathways, from large scale systems biology (Hyduke et al. 2007) to gene expression (e.g.(Konig and Eils 2004; Schramm et al. 2007)) to metabolic engineering (Chassagnole et al. 2002). The level of curation that has been achieved from over 10 years of evidence-based literature searching makes EcoCyc one of the most comprehensive resources available on a single organism.

Salmonella and *E. coli* are closely related members of the Enterobacteriaceae, and as such, EcoCyc represents a wealth of information that serves as a useful resource for the creation of a *Salmonella* Pathway/Genome database (PGDB).

For *Salmonella*, only one pathway database had been published at the beginning of this work. However, this map of Typhimurium LT2 metabolism was edited from the *E. coli* version depicted in EcoCyc, rather than being built *de novo*. While this was a useful shorthand method that was used to describe the metabolism of four *Salmonella* serovars during typhoid or typhoid-like disease (Becker et al. 2006), it has not been made publically available. Also, using a single database to model different serovars makes it more difficult to ask serovar-specific questions. Thus, at the outset of this project, no curated pathway database existed for Typhi.

The level of genome degradation in Typhi and Paratyphi A, coupled with evidence that these serovars have more restricted metabolic profiles than other serovars, prompts the hypothesis that there is a causal link between host restriction and metabolic ability (Uzzau et al. 2000). However, there is currently very little evidence to support this in the literature. Whilst various complete *Salmonella* genome sequences are now available, this lack of evidence is likely due in part to the paucity of information describing gene function and an inability to assign individual pseudogenes to metabolic pathways. The aim therefore was to build one database to represent the metabolism of Typhi in order to visualise the metabolic pathways interrupted by pseudogenes and one to represent Typhimurium as a comparator.

2.2 Methods

2.2.1 Pathway Tools

Pathway Tools version 11.0 (SRI International, California, USA), was installed on a Debian Linux system under the academic license available at http://www.biocyc.org. The software contains the PathoLogic suite, which allows the user to build metabolic databases from scratch.

2.2.1.1 Input file formatting

The Sanger-sequenced Typhi CT18 genome sequence and annotation was used as the basis of the Pathway/Genome Database (PGDB), named StyCyc. A custom Perl script (Appendix 8.1.1 CD_001) was used to parse the Typhi GenBank annotation file to ensure that information contained in qualifiers would be assigned correctly within the PGDB framework. All STY unique gene IDs (/systematic_id) were given the qualifier '/label' and all '/note's were renamed '/product_comment'. A first-pass automated PGDB build indicated that genes annotated with the '/pseudo' qualifier would be removed by the software before pathway construction. In order to understand where potential enzymes encoded by these genes would be placed with metabolic pathways, this qualifier was removed. Instead, all pseudogenes had the string "(pseudogene)" added to the gene product name to enable identification in the database.

2.2.1.2 Pseudogene re-annotation

It was possible that the enzymes encoded by pseudogenes would be candidates for pathway reactions. To allow these enzymes to be considered, the annotated translation needed to 'correct' any stop codons or frameshifts present in the sequence. Typhi CT18 was one of the first bacterial sequences completed at the Sanger Institute, before standardised methods were employed for the annotation of pseudogenes. Thus, frameshifted pseudogenes were annotated only in the frame in which the start codon was encoded. Each of these was reannotated using the Artemis Comparison Tool (ACT (Carver et al. 2005)) to allow automated translation of the whole reading frame. Typhimurium strains LT2 and SL1344 were used as comparators to reconstruct the intact sequence. Pseudogenes caused by stop codons were not reannotated as the automated translation was unaffected.

2.2.1.3 Assigning enzymes and multimers

Using the Pathway Tools software, the initial round of pathway prediction was based upon a name-matching algorithm, relying on high quality gene annotation to determine the presence of particular enzymes and assign them to relevant reactions. The software generated a report on this process, giving details of potential enzymes which could not be matched. This list was manually processed for typographical errors and probable enzymes were assigned to reactions based upon literature searches and/or sequence comparison with Typhimurium and *E. coli*. Some enzymes had non-specific names e.g. alcohol dehydrogenase and therefore could not be assigned to specific reactions.

For over 150 metabolic reactions, more than one enzyme was predicted to perform the relevant function. These were assessed individually to determine whether the enzymes formed a multimeric complex or whether they represented functional isomers. Complexes were confirmed based upon genome position (i.e. within operons), literature searches and/or comparison with Typhimurium and *E. coli*; otherwise the enzymes were, by default, assumed to be isomers.

2.2.1.4 Pathway prediction

Based on the assignment of enzymes to reactions, the presence of certain metabolic pathways was inferred for the organism of interest from MetaCyc (http://www.metacyc.org), a reference database containing experimentally elucidated metabolic pathways from over 1,000 organisms (Caspi et al.). The software is described as over-predictive, as, for example, a pathway is often inferred when some of its constitutive enzymes were not name-matched in the organism of interest. Therefore, once these stages were complete, a manual check of each predicted pathway was performed to ensure only genuine pathways remained in the database. Each pathway (imported from MetaCyc) was marked with the known range of organisms in which that pathway had been found. This information was used to detect and remove false-positive predicted pathways by assessing whether particular enzymes had been placed in single or multiple pathways. However, once this checking was complete, some pathways still contained reactions without an assigned enzyme. These were known as 'pathway holes' and were initially processed by the inbuilt algorithm 'Pathway Hole Filler'.

2.2.1.5 Transport reactions

Another tool in the software, the Transport Identification Parser, was used to predict transport reactions within the PGDB. Based upon gene annotation, gene products were grouped by substrate and transport mechanism, e.g. ATP-driven, channel-type facilitator or secondary transport. The automated predictions were manually assessed and confirmed if the genome annotation was sufficient to support the assignment and/or experimental evidence was available in the literature and/or comparison with Typhimurium and *E. coli* indicated that amino acid sequence was conserved (> 70% identity) and syntenic. Predictions were rejected when the literature searches were unsuccessful and amino acid sequence was either not conserved (< 70% identity) or not present. Predictions from TransportDB (http://www.membranetransport.org) were also taken into account.

2.2.2 Pathway hole filling

2.2.2.1 Automated hole filling

One of the strongest motivations for using Pathway Tools was the inbuilt Pathway Hole Filler algorithm. This four-step method identified gene candidates from Typhi CT18 to fill enzymatic holes in metabolic pathways. To do this, the function of the missing enzyme was inferred from the complete version of the pathway present in MetaCyc. Stage one in the hole-filling process was the retrieval of enzyme sequences from other organisms which catalysed the appropriate reaction. These sequences were used in a BLAST-p search against the genome sequence of Typhi CT18 and the top hits were recorded. A data consolidation step was then performed, pulling together a summary of

the BLAST data to be used as evidence that a particular candidate had the function required to fill the pathway hole. All the evidence per candidate was then evaluated using a Bayes classifier. The evaluation included whether the candidate was part of an operon and/or adjacent to a gene coding for the enzyme catalysing an adjacent reaction in the pathway. Finally, the classifier generated a probability that the candidate had the function required to catalyse the missing reaction.

2.2.2.2 Manual evaluation of pathway hole candidates

Every pathway with at least one hole-filling candidate was manually assessed. Pathways which would become fully intact with the assignment of a hole-filling candidate were considered a priority for assessment. Where possible, the pathway was compared with the equivalent in EcoCyc and/or MetaCyc. ACT comparisons (Carver et al. 2005) were used to determine if the Typhi candidate was conserved (> 70% amino acid identity) and syntenic with the enzyme from *E. coli*. Possible amino acid sequence structure homology was detected using FUGUE (Shi et al. 2001). Literature searches were also performed to obtain experimental evidence for assignments. Pathways which had a taxonomic range outside of the Enterobacteriaceae and had multiple holes with no candidates were pruned from the database. Pathways were also deleted when no evidence for the pathway existed, either from unsuccessful literature searches or from lack of amino acid conservation (< 70% identity) of predicted enzymes with potential orthologues in *E. coli*. In addition, if any enzymes previously assigned (automatically) were not unique to the pathway, and no further evidence was found, the pathway was also deleted.

2.2.2.3 PGDB build process for Typhimurium

Subsequent to StyCyc, a new PGDB was built for Typhimurium using the SL1344 genome sequence and annotation. This strain contains 40 pseudogenes annotated in the correct frames and the annotation used standard qualifiers, so no additional input file formatting was required. The same PGDB build procedures of enzyme and multimer assignment, pathway prediction and hole filling were completed. All hole filling-candidates were assessed as for StyCyc, except that comparisons were made against both Typhi and *E. coli*.

2.3 Results

2.3.1 Generation of StyCyc 1.0

Sequence and annotation information from Typhi strain CT18 was used to generate a Pathway/Genome Database (PGDB), named StyCyc to follow the convention of EcoCyc and MetaCyc. An automated build procedure resulted in the importation (from MetaCyc) of all pathways containing a reaction catalysed by an enzyme in Typhi. Some pathways were subsequently deemed to have insufficient evidence (e.g. only one reaction with an enzyme assigned) and immediately removed from the PGDB. Following this, a number of manual refining steps were performed. This included scrutinising over 400 ‘probable enzymes’ which had been missed in the initial name-matching exercise for possible enzymatic assignments. Approximately 1/3 had names that were too generic to be assigned to reactions (e.g. putative hydrogenase). A minority of enzyme names contained typographical errors and were easily assigned while the remainder were searched in EcoCyc and MetaCyc to determine function.

Each Typhi protein coding gene was assigned a corresponding polypeptide, of which 1,033 were predicted to be enzymes and over 100 formed protein complexes. Together, these catalyse over 1,200 enzymatic reactions containing 925 compounds. A further 184 transporters were also predicted, manually confirmed and assigned to 145 transport reactions (Table 2-1).

A fully-automated run of the Pathway Hole Filler was completed for StyCyc 1.0, where top candidates (according to the inbuilt Bayes classifier - see Methods), for pathway holes were accepted without manual intervention. In total, 474 holes were identified in

predicted Typhi metabolic pathways and the Hole Filler identified candidates from the genome to fill 217 of these, resulting in 59 intact pathways out of a total of 312 (Table 2-1). This indicated that the hole-filling algorithm was capable of finding novel functions for Typhi gene products.

Table 2-1 Statistics for StyCyc 1.0

Pathways	312
Enzymatic reactions	1264
Transport reactions	145
Polypeptides	4404
Protein complexes	103
Enzymes	1033
Transporters	184
Compounds	925

Total numbers in database categories after an automated build process, manual curation of unassigned enzymes and a fully-automated run of Pathway Hole Filler.

2.3.1.1 Issues and resolutions

Upon initial examination of StyCyc 1.0, it became apparent that pseudogenes were not given associated gene products and hence were not assigned to any metabolic pathways. As part of the intended utility of the Typhi metabolic map was to determine the location of pseudogenes within pathways, this needed to be resolved. The '/pseudo' qualifier that identified pseudogenes to the software was removed from the input file. As an alternative identification mechanism, the string '(pseudogene)' was added to the name of the product of each pseudogene.

In addition, for pseudogenes to be considered as candidates by the Hole Filler, accurate protein sequences were required. The initial annotation of Typhi displayed all pseudogenes in the same frame as the start codon and therefore did not take into account the effects of frameshift mutations. Out of the 204 pseudogenes identified in CT18 (Parkhill et al. 2001a), 96 were caused by frameshifts, so these were re-annotated in the appropriate frames and used to update the input file (Table 2-2).

Table 2-2 Classification of pseudogenes in Typhi CT18

Genetic lesion	Number
Frameshift	96
In-frame stop codon	76
Insertion	4
Fragment/remnant	32
Truncation	9
Internal deletion	4
Fusion	1

In accordance with the annotation of Typhi CT18, all 204 pseudogenes were examined for genetic lesions in ACT (Carver et al. 2005). In a few cases, more than one genetic lesion was recorded. Only frameshift mutations required re-annotation.

A key requirement of the Pathway Tools software was for each gene to be given a unique identifier which would be taken from the annotation. Typhi genes were identified by their ‘STY’ numbers, but these were not propagated through to StyCyc 1.0 due to the software not recognising the qualifier ‘/systematic_id’ in the original annotation file. The qualifier was renamed to ensure the software correctly recognised the information it was being given.

Before rebuilding StyCyc, the enzyme assignment decisions made early in the creation of StyCyc 1.0 were transferred into the locally stored enzyme mapping file, to be used by all subsequent *Salmonella* PGDB builds.

2.3.2 StyCyc 2.0 and onwards

2.3.2.1 Improvements from StyCyc 1.0

As well as the improvements detailed above, new data on Typhi pseudogenes were included in the updated input file. A high-throughput sequencing study of 19 Typhi strains had generated a list of pseudogenes that were either core (i.e. present in all) or variable (absent in some) in the sequenced population (Holt et al. 2008). The variable pseudogenes were given the suffix '(VPT)' (for variable pseudogene in Typhi) in the gene product name. The updated Typhi annotation was then used to build StyCyc 2.0.

2.3.2.2 New pathways and transport reactions

Typhi expresses the Vi antigen, synthesised by genes encoded upon SPI-7. This pathogenicity island is found in only a few *Salmonella* serovars and in some *Citrobacter*. Vi antigen biosynthesis was therefore not present in MetaCyc and could not be imported into StyCyc. Using experimental evidence from the literature, this pathway and its transport reaction was reconstructed using the editing tools available in Pathway Tools (Virlogeux et al. 1995; Zhang et al. 2006) (Figure 2-1).

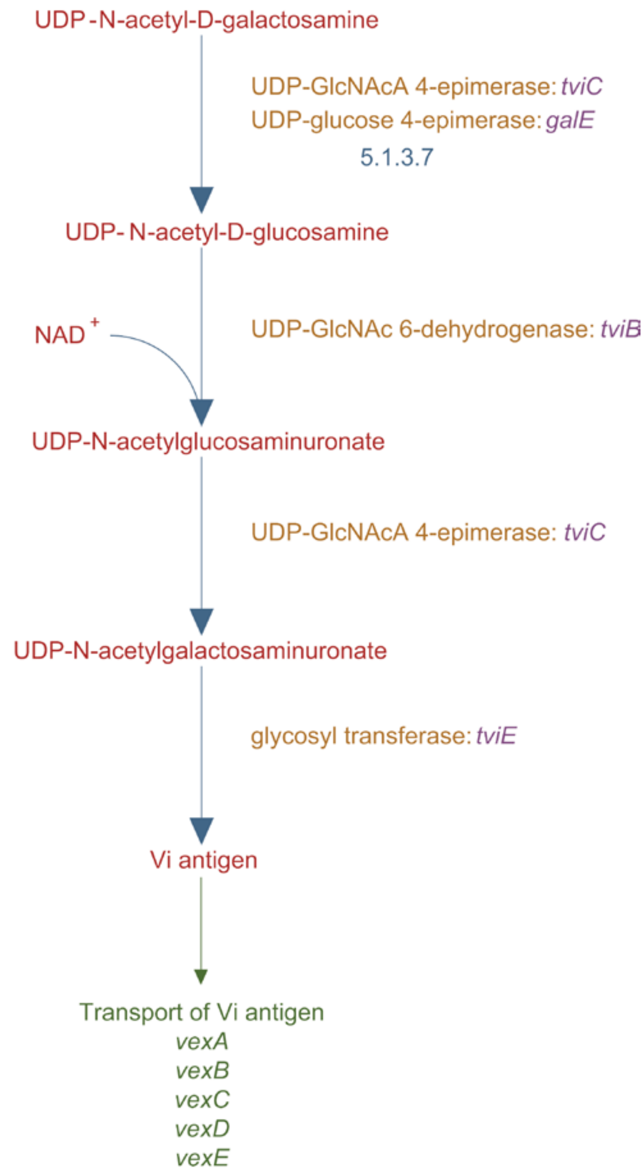


Figure 2-1 Vi antigen biosynthesis

Pathway Tools depiction of the Vi antigen biosynthetic pathway. Pathway direction is indicated by blue arrows. Gene names given in purple; enzyme names in gold; compound names in red. Blue number, enzyme commission (E.C.) number for the reaction. Transport reaction and related genes shown in green.

Using EcoCyc as a standard for presentation of central metabolic pathways, links between glycolysis, pyruvate dehydrogenase and the TCA cycle were created to form a 'superpathway' (Figure 2-2). In Pathway Tools, superpathways simply represent a 'bigger picture' in order to enable the user to better understand how highly related processes are connected. Similarly, a superpathway was created to visualise how the glyoxylate cycle relates to the TCA cycle.

Initially, predictions from the inbuilt transport identification module, based upon gene annotation, were used to assign transport reactions. Information from TransportDB was then used to identify other potential transport reactions based on protein sequence and seventeen transporters were added to the database (Table 2-3). Multidrug efflux systems which act in co-ordination with TolC were also added based upon evidence from the literature (Nishino et al. 2006).

Table 2-3 Predicted transporter proteins in Typhi CT18

Type	Number	Example substrates
Channel-type facilitators	11	nickel; calcium; glycerol; formate
Secondary transporters	47	glutamate; rhamnose; citrate; fucose
ATP-driven transporters	37	thiosulfate; glutathione; maltose; methionine
PEP-driven transporters	11	cellobiose; mannose; glucose; fructose
Unknown mechanism	27	serine; xanthine; gluconate; glucarate

PEP, phosphoenolpyruvate.

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

Representation of (1) glycolysis, (2) pyruvate dehydrogenase and (3) the TCA cycle as an interconnected 'superpathway'. Pathway direction is indicated by blue arrows. Gene names given in purple; enzyme names in gold; compound names in red. Blue numbers, enzyme commission (E.C.) numbers for reaction

2.3.2.3 Pathway Hole Filling

The first stage of the hole-filling process identified 85 metabolic pathways with hole-filling candidates. Between 1 and 500 protein sequences were retrieved from enzymes performing the missing functions in other organisms. The majority of the incomplete pathways contained only one hole; the remainder between 2 and 5 (Table 2-4). A summary of how many pathways were completely filled is given in Table 2-5.

Table 2-4 Incomplete pathways in StyCyc

Holes	Number of pathways	Examples
1	54	coenzyme A biosynthesis; ethylene glycol degradation
2	12	lipoate biosynthesis and incorporation; ECA biosynthesis
3	13	histidine biosynthesis; menaquinone biosynthesis
4	4	4-hydroxyphenylacetate degradation; thiamine biosynthesis
5	2	lipid A core biosynthesis; methylerythritol phosphate pathway

StyCyc pathways contained between 1 and 5 pathway holes before assignment of hole-filling candidates. ECA, enterobacterial common antigen.

Table 2-5 Summary of StyCyc pathway outcomes

	Pathways			
Holes	Deleted	Completely filled	Partly filled	Not filled
1	26	21	0	7
2	3	6	2	1
3	2	9	0	2
4	1	2	0	1
5	0	2	0	0

Pathways without sufficient evidence were deleted from the PGDB. Hole-filling candidates were manually assessed for remaining pathways and assigned only with evidence as described in the methods.

2.3.2.4 PGDB statistics

The current version of StyCyc is StyCyc 7.0 which has undergone both literature-based and direct curation. The direct curation came from the subsequent creation of a PGDB for Typhimurium and is described in greater detail in section 2.3.3 below. StyCyc 7.0 contains 200 predicted pathways and 133 transport reactions. A total of 1,024 enzymatic reactions, containing 821 compounds, are catalysed by 1,052 enzymes (Table 2-6). The changes from previous StyCyc versions are for the most part reductions in the number of pathways, reactions and compounds predicted to be present in Typhi. This is due to curation aimed at identifying and eliminating false pathways and their associated reactions and compounds. However, both the number of polypeptides and protein complexes has increased from StyCyc 1.0 to 7.0. The former is due to the recognition of 204 pseudogenes and 2 genes whose corrupted information has been repaired and the latter is an increase from 103 to 139, due to the identification of new protein complexes based upon literature searches.

Table 2-6 Statistics from StyCyc 7.0

Category	Number (change from StyCyc 1.0)	
Pathways	200	(112 fewer)
Enzymatic reactions	1011	(253 fewer)
Transport reactions	133	(12 fewer)
Polypeptides	4610	(206 more)
Protein complexes	139	(36 more)
Enzymes	1052	(19 more)
Transporters	150	(34 fewer)
Compounds	821	(104 fewer)

These statistics reflect the status of StyCyc 7.0 after an automated build process, manual curation and a manually supervised run of Pathway Hole Filler.

In the Pathway Tools software, pathways are classified in a hierarchy that at the highest level divides into broad areas such as biosynthesis, degradation and generation of precursor metabolites and energy. The breakdown of the highest two levels is given in Table 2-7. This hierarchy recognises classical pathways such as glycolysis and fermentation alongside some unique to this software. For example, aminoacyl-tRNA charging is not strictly a metabolic pathway but being classed as such allows all 20 tRNA-charging reactions to be grouped together.

Table 2-7 Pathway hierarchy in Pathway Tools

Class (Sty/Stm)	Sub-class	Sty	Stm
Biosynthesis (140/139)	Amines and polyamines	8	8
	Amino acids	40	40
	Aminoacyl-tRNA charging	1	1
	Aromatic compounds	2	2
	Carbohydrates	8	8
	Cell structures	11	10
	Cofactors, prosthetic groups, electron carriers	43	43
	Fatty acids and lipids	20	20
	Metabolic regulators	1	1
	Nucleosides and nucleotides	10	10
	Other	1	1
	Siderophore	1	1
Degradation/utilisation/ assimilation (76/81)	Alcohols	5	5
	Aldehyde	5	5
	Amines and polyamines	7	7
	Amino acids	15	16
	Aromatic compounds	1	1
	Carbohydrates	15	15
	Carboxylates	7	7
	Other	1	1
	Fatty acids and lipids	1	1
	Inorganic nutrients	9	9
	Nucleosides and nucleotides recycling	4	4
	Secondary metabolites	8	12
Detoxification (3/3)	Acid resistance	1	1
	Methylglyoxal	3	3
Generation of precursor metabolites and energy (27/27)	Chemoautotrophic energy	1	1
	Fermentation	2	2
	Glycolysis	2	2
	Pentose phosphate pathways	3	3
	Respiration	11	11
	TCA cycle	5	5
Superpathways		*43*	*43*

Only classes present in StyCyc 7.0 and StmCyc 4.0 are shown. Some pathways may be present in more than one sub-class category. Sty, Typhi; Stm, Typhimurium; numbers indicate how many pathways of this class/sub-class are present in each database. Superpathways represent overviews of connected pathways.

2.3.2.5 Metabolic map: Typhi

A 'cellular overview' of Typhi metabolism was generated for the automatically created StyCyc 1.0 (Figure 2-3) and, following extensive pathway curation and pathway hole filling, for StyCyc 7.0 (Figure 2-4). These depict a Gram negative cell with inner and outer membranes, across which transport reactions are shown. By EcoCyc convention, biosynthetic pathways are drawn on the left, then energy pathways, with degradation pathways on the right. Independent metabolic reactions that do not form known pathways are shown on the far right. Figure 2-4 therefore, represents the current metabolic map of Typhi CT18.

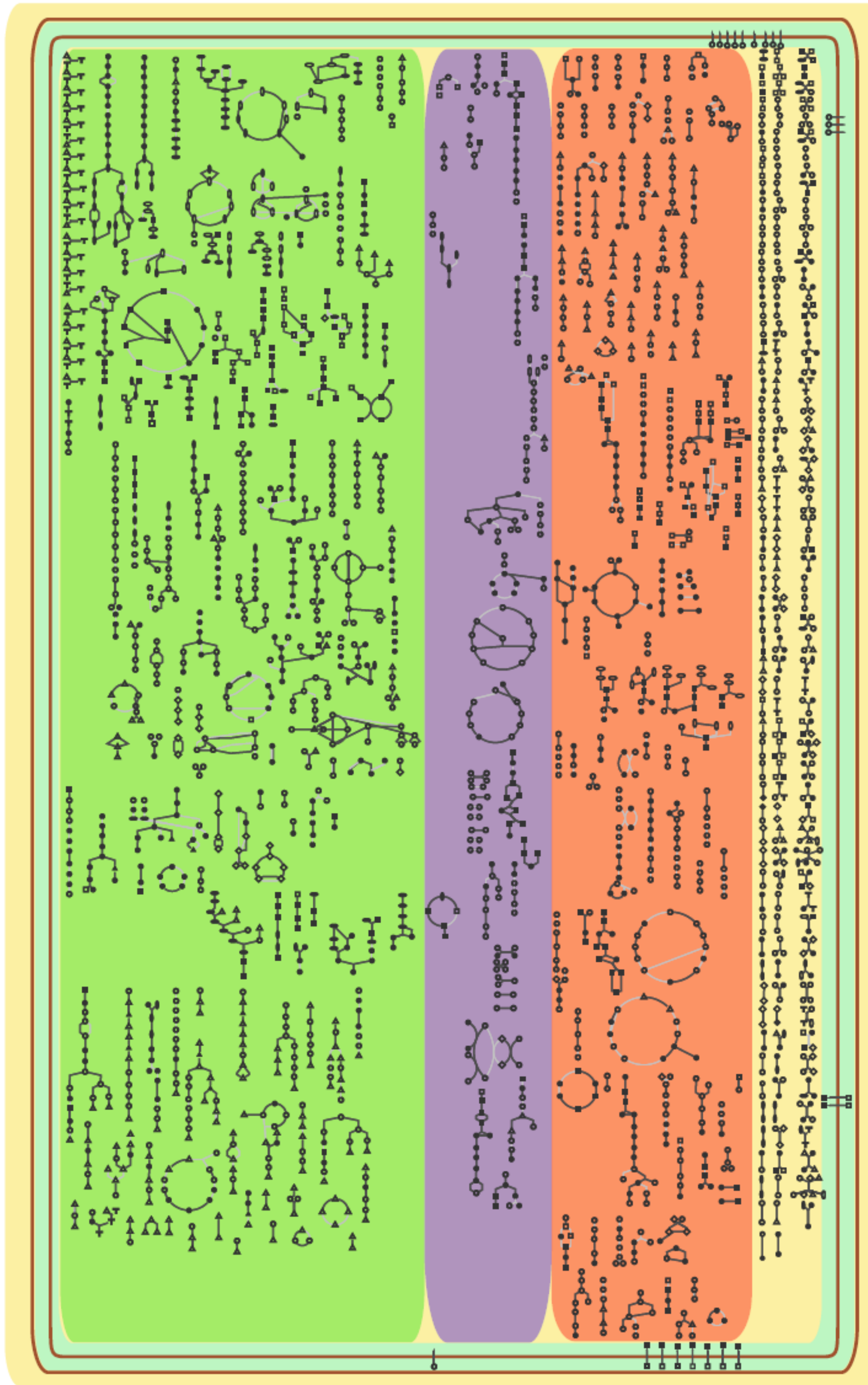


Figure 2-3 StyCyc 1.0: an automatically generated metabolic map of Typhi

Brown lines indicate bacterial cell membrane; black lines represent metabolic reactions. Green background, biosynthetic pathways; dark blue, energy pathways; red, degradation pathways; light blue, transport reactions. Symbols: upward-pointing triangle, amino acids; square, carbohydrates; diamond, proteins; vertical oblong, purines; horizontal oblong, pyrimidines; downward-pointing triangle, cofactors; T, tRNAs; open circle, other; enclosed circle, phosphorylated

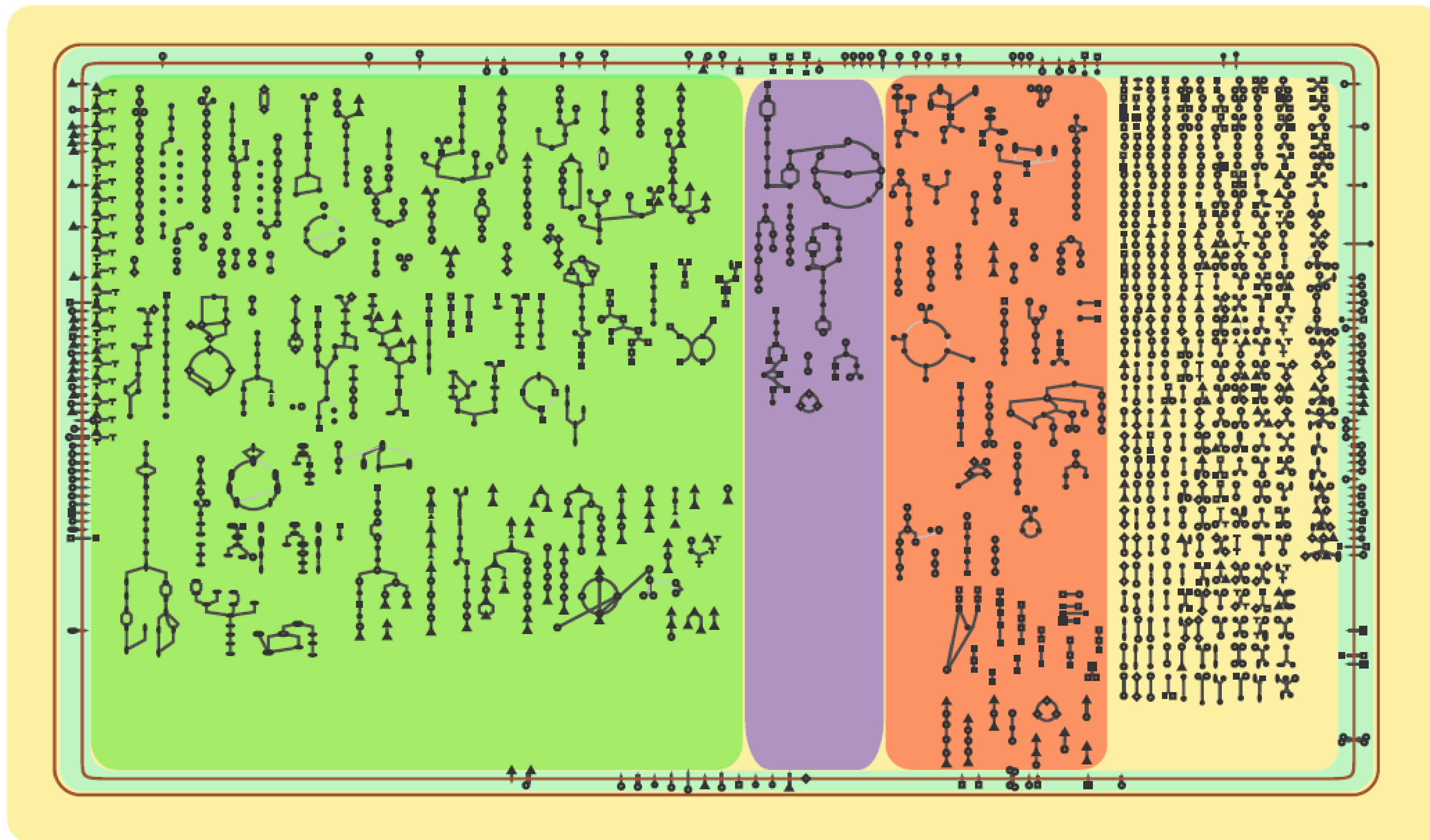


Figure 2-4 StyCyc 7.0: a manually curated metabolic map of Typhi

Brown lines indicate bacterial cell membrane; black lines represent metabolic reactions. Green background, biosynthetic pathways; dark blue, energy pathways; red, degradation pathways; light blue, transport reactions. Symbols: upward-pointing triangle, amino acids; square, carbohydrates; diamond, proteins; vertical oblong, purines; horizontal oblong, pyrimidines; downward-pointing triangle, cofactors; T, tRNAs; open circle, other; enclosed circle, phosphorylated.

2.3.2.6 Pseudogenes in Typhi

After comparison across multiple strains of Typhi, 211 pseudogenes were identified in Typhi CT18 alone, seven more than noted in the original annotation. However, the availability of 19 fully sequenced Typhi strains allowed pseudogenes from this serovar to be classified as 'core' (present in all) or 'variable' (present in one or more strains). In total, 274 pseudogenes were identified in all Typhi sequences. Using the metabolic map as a framework, the enzymatic reactions performed by these pseudogene 'products' were highlighted (Figure 2-5). Fifty seven appear in the overview, including 6 in biosynthetic pathways, 10 in degradation pathways and 20 in standalone reactions. Pseudogenes also account for 20 out of 133 transport reactions. The percentage of metabolic reactions attributed to pseudogenes is shown in Table 2-8.

Table 2-8 Classification of Typhi pseudogenes among metabolic reactions

Class	Total reactions	Pseudogene 'products'			Possible complementary enzyme(s)
		Core	Variable	% of total	
Biosynthesis	590	4	2	1	1
Energy	103	1	0	1	1
Degradation	211	5	5	4.7	2
Standalone reactions	282	11	9	7.1	7
Transporters	133	18	2	15	-
Regulators*	238	7	4	4.6	-

A reaction was considered to have complementary enzyme(s) when the pseudogene 'product' was not the only enzyme assigned to that reaction. -, not determined. * Regulators are not displayed in the metabolic overview; these were determined from genome annotation.

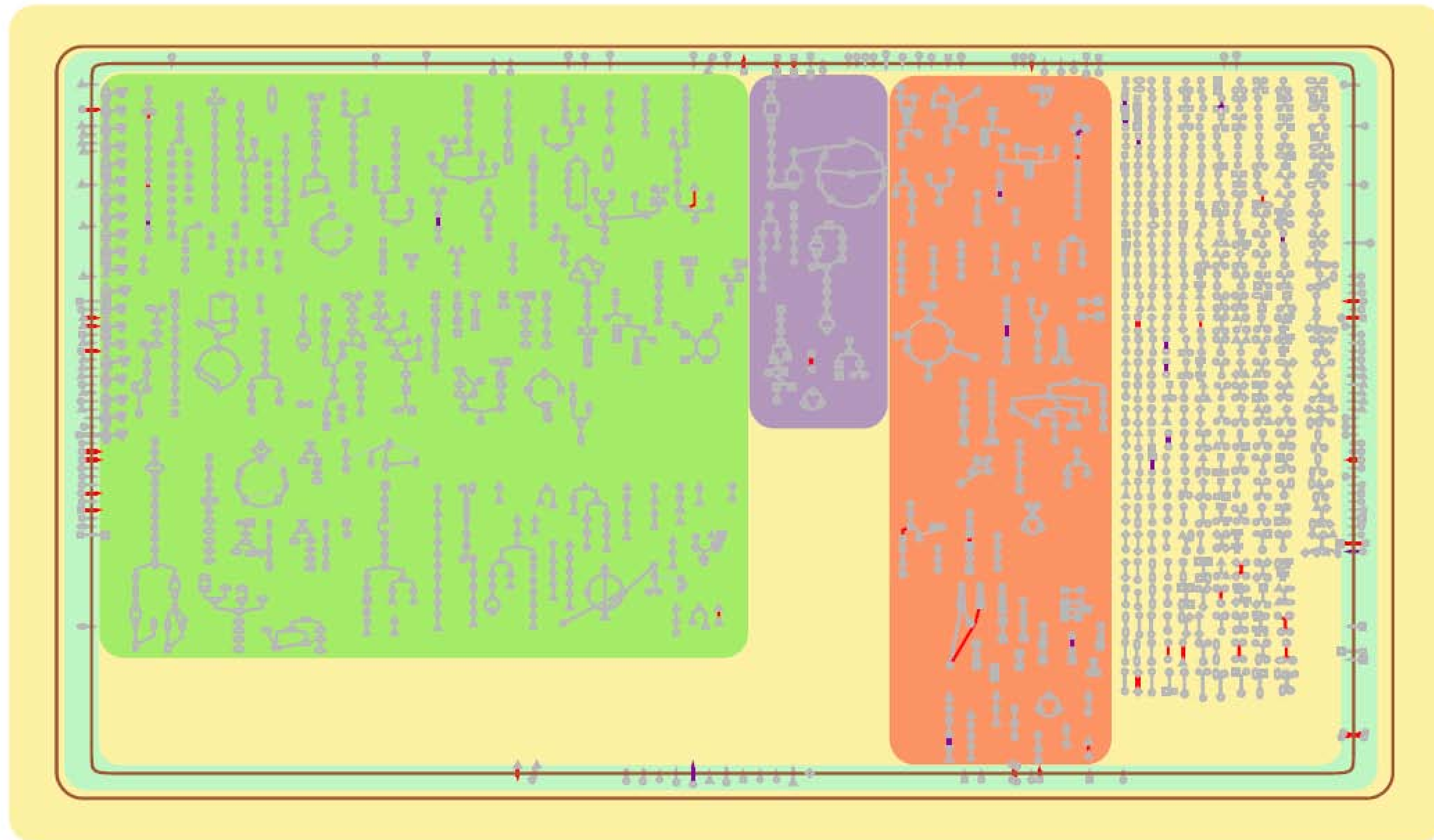


Figure 2-5 Pseudogenes interrupting metabolic pathways and transport reactions

Brown lines indicate bacterial cell membrane; red lines indicate core pseudogene ‘products’, purple lines are variable pseudogene ‘products’; greyed out lines represent metabolic reactions. Green background, biosynthetic pathways; dark blue, energy pathways; red, degradation pathways; light blue, transport reactions. Symbols: upward-pointing triangle, amino acids; square, carbohydrates; diamond, proteins; vertical oblong, purines; horizontal oblong, pyrimidines; downward-pointing triangle, cofactors; T, tRNAs; open circle, other; enclosed circle, phosphorylated.

2.3.3 StmCyc

2.3.3.1 Use of StyCyc to help assignments

The value of having already built StyCyc resulted in a much faster process for the initial stages of the Typhimurium database, StmCyc. The local enzyme mapping file contained extra *Salmonella*-specific information from StyCyc that would not have been found in EcoCyc or MetaCyc and hence the number of enzymes to manually assign was reduced from ~ 400 in StyCyc to ~280 in StmCyc.

This automated build took place approximately 11 months after StyCyc 2.0, during which time both EcoCyc and MetaCyc had been continuously updated with new metabolic pathways and reactions. Hence, more pathways were predicted to be present in StmCyc than had been in StyCyc 2.0. Further curation of StyCyc was performed to take this into account and is detailed in section 2.3.3.3.

2.3.3.2 Hole Filling

A semi-automated run of Pathway Hole Filler was performed, with hole-filling candidates predicted for 99 incomplete pathways. Sixty of these contained a single pathway hole, with two long pathways that had seven holes (Table 2-9). The ratio of deleted and filled pathways is shown in Table 2-10.

Table 2-9 Incomplete pathways in StmCyc

Holes	Number of pathways	Examples
1	60	arginine degradation; galactitol degradation
2	24	lyxose degradation; NAD biosynthesis
3	8	flavin biosynthesis; ketogluconate metabolism
4	5	ubiquinone biosynthesis; methylcitrate cycle
7	2	adenosylcobalamin biosynthesis; purine nucleotides biosynthesis

StmCyc pathways contained between 1 and 7 pathway holes before assignment of hole-filling candidates.

Table 2-10 Summary of StmCyc pathway outcomes

	Pathways			
Holes	Deleted	Completely filled	Partly filled	Not filled
1	25	26	0	7
2	10	13	0	1
3	2	6	0	2
4	0	4	1	1
7	0	2	0	0

Pathways without sufficient evidence were deleted from the PGDB. Hole-filling candidates were manually assessed for remaining pathways and assigned only with evidence as described in the methods.

2.3.3.3 Additional development

Many new pathways were predicted in StmCyc that were not present in the relevant databases at the time StyCyc was built. In order to have two *Salmonella* databases comparable in terms of the predicted pathways, these new pathways were assessed for their suitability for transfer into StyCyc. ACT comparisons and literature searches were used where necessary to inform this process. Nineteen new biosynthetic pathways and eight biosynthetic superpathways were transferred from StmCyc and assigned the

relevant enzymes in StyCyc. A further ten degradation pathways, including one superpathway, were added, alongside eleven reactions relating to electron transfer.

A global comparison of pathways present in each PGDB highlighted some pathways that had not been predicted in StmCyc that were present in StyCyc. Some of these had been manually curated in StyCyc, and twelve were subsequently added to StmCyc. The two databases are now curated to the same level and are fully comparable across pathways present in both.

Of the forty pseudogenes annotated in Typhimurium SL1344, four appear in the cellular overview: *caiB* (degradation: carnitine), *cusA* (transport: cation), *mdaA* (standalone: nitroreductase) and *appC* (standalone: cytochrome oxidase subunit). None of these are pseudogenes in the laboratory strain Typhimurium LT2, thus they are potential 'variable' pseudogenes in Typhimurium.

2.3.3.4 PGDB statistics

The current version is StmCyc 4.0 which contains 204 metabolic pathways and encompasses 1,133 enzymatic reactions. The summary statistics are shown in Table 2-11.

Table 2-11 Statistics from StmCyc 4.0

Pathways	204
Enzymatic reactions	1133
Transport reactions	134
Polypeptides	4536
Protein complexes	137
Enzymes	1119
Transporters	169
Compounds	870

2.3.3.5 Metabolic map: Typhimurium

Following the side-by-side curation with StyCyc, a metabolic map for Typhimurium was generated (Figure 2-6). The map follows the same visual conventions as StyCyc and EcoCyc.

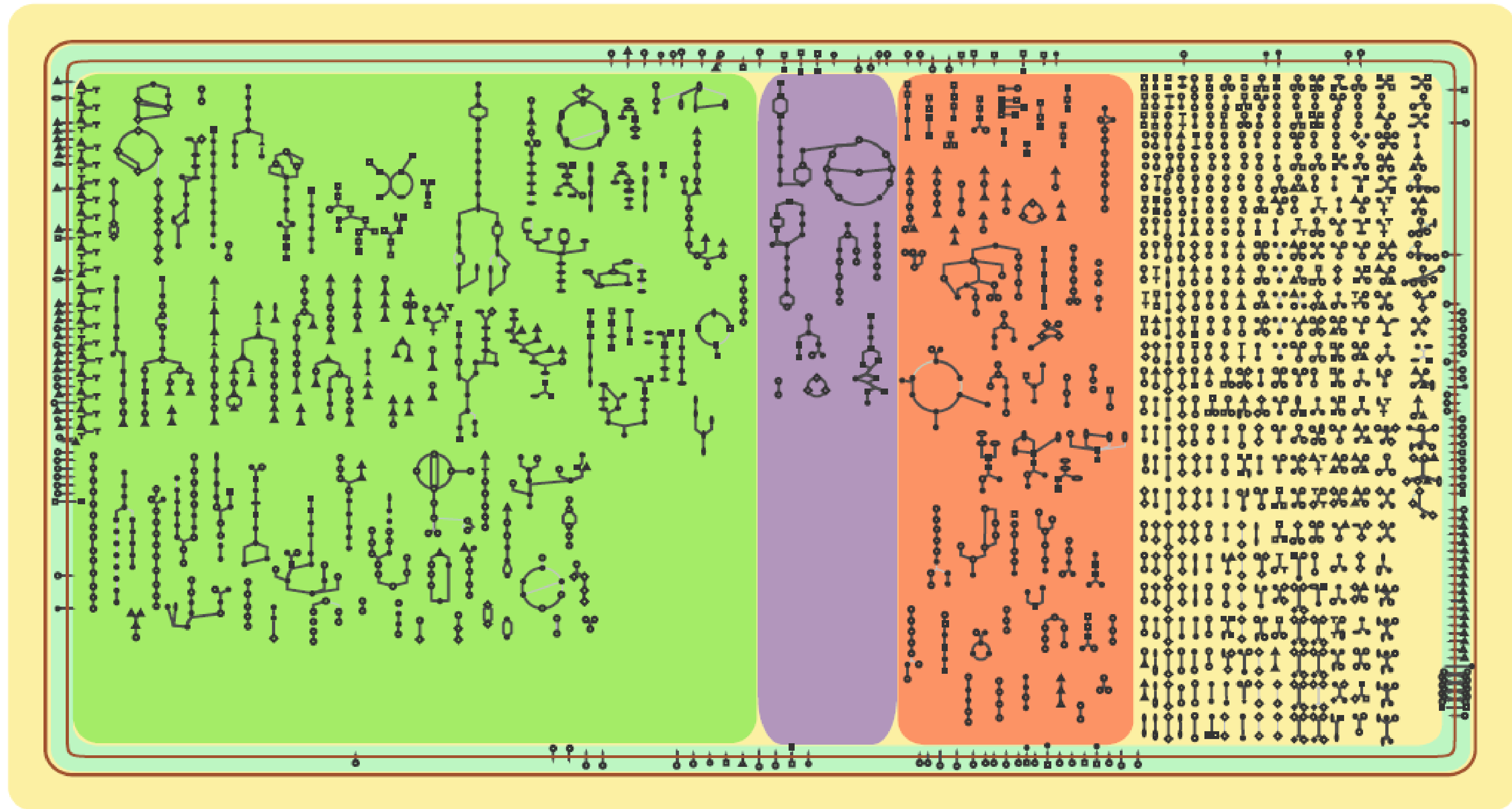


Figure 2-6 StmCyc 4.0: a metabolic map of Typhimurium

Brown lines indicate bacterial cell membrane; black lines represent metabolic reactions. Green background, biosynthetic pathways; dark blue, energy pathways; red, degradation pathways; light blue, transport reactions. Symbols: upward-pointing triangle, amino acids; square, carbohydrates; diamond, proteins; vertical oblong, purines; horizontal oblong, pyrimidines; downward-pointing triangle, cofactors; T, tRNAs; open circle, other; enclosed circle, phosphorylated.

2.3.4 Comparing StyCyc with StmCyc

Side-by-side curation of the two PGDBs facilitated the identification of pathway differences between them. Metabolically, they differ by the presence or absence of 5 pathways, one biosynthetic and four degradative (Table 2-12). They also differ in the number of intact pathways, with StmCyc maintaining 11 intact pathways that in StyCyc contain 'core' pseudogenes (Table 2-13).

Table 2-12 Metabolic pathway differences

Pathway	Present in StyCyc	Present in StmCyc	Genes
Vi antigen biosynthesis	✓	✗	*tvi*
L-idonate degradation	✗	✓	*idn*
Ketogluconate metabolism	✗	✓	*idn*
myo-inositol degradation	✗	✓	*dgo*
Glutamine degradation II	✗	✓	SL1455

Pathways present or absent in StyCyc and StmCyc as indicated by cross or tick.

Table 2-13 Metabolic pathways interrupted in Typhi

Pathway	Typhi pseudogene(s)
adenosylcobalamin biosynthesis (early cobalt insertion)	*cbiC*, *cbiJ cbiK*
allantoin degradation (anaerobic)	*allA*
arginine degradation / ethylene glycol degradation	STY1536
asparagine degradation	STY4203
hydrogen oxidation I (aerobic)	*hyaA*, *hyaB2*
4-hydroxyphenylacetate degradation	*hpaB*, *hpcC*
N-acetylneuraminate and *N*-acetylmannosamine degradation	*nanE*
putrescine biosynthesis	*speC*, *speF*
rhamnose degradation	*rhaD*
trehalose degradation	*treA*

Pathways intact in StmCyc but interrupted in StyCyc by the presence of single or multiple pseudogenes.

2.4 Discussion

2.4.1 Rationale for building a pathway database

Both human-restricted serovars Typhi and Paratyphi A display reduced metabolic profiles as well as significant genome degradation. While individual pseudogenes in these serovars have been previously identified, their global effect upon metabolism has not. Only by looking at pseudogenes in the context of metabolic pathways and transport reactions can wider comparisons be made and inactivating mutations of the same pathway be recognised. Building a metabolic pathway database provides the framework required for such a comparison, and having PGDBs customised for both Typhi and Typhimurium lends great scope for the comparison of the metabolic effect of pseudogenes, across host restricted *Salmonella*.

2.4.2 Rationale for using Pathway Tools

Prior to building a metabolic map for Typhi, the various tools for pathway visualisation and prediction were assessed. KEGG was suitable for visualisation purposes, as it was possible to view the organism-specific set of metabolic pathways. However, there were a number of factors that led to the decision to discount KEGG as a possibility: firstly, when organism-specific pathways were visualised, this was in the context of all the other pathways in the reference diagram, hence there was no way of looking solely at a Typhi-specific metabolic map; secondly, the presence of pseudogenes in pathways was not clearly denoted and there was no way to overlay this data on the framework; lastly, and

most importantly, KEGG generates static reference diagrams, and there were no editing tools available to tailor pathways in any way.

The second possibility was Cytoscape, which is a more abstract method for looking at metabolism as it is oriented towards interaction data. However, this is most powerful when various types of interaction data can be used and these weren't available at the time for Typhi.

The methods implemented by other software tools often use either KEGG or BioCyc as references for their own predictions. Pathway Tools software is part of the BioCyc suite and appeared well-suited to the task. Unlike some software, it does requires an annotation file, but since Typhi had been annotated to a very high level, this was not an issue. Another positive was that this would mean a pathway database built on the same framework as the extensively curated EcoCyc. Where there was overlap between *E. coli* and *Salmonella*, EcoCyc could therefore be used to speed up curation. The knowledge that the software was already installed and functional at the Institute also factored into the decision to use Pathway Tools to build pathway databases for Typhi and later Typhimurium.

2.4.3 Presence of pseudogenes can explain metabolic capability

For the first time, all Typhi pseudogenes affecting metabolic functions could be viewed simultaneously as part of the metabolic map. Data from 19 Typhi sequences were used to determine whether pseudogenes were core to all strains or had only occurred in some (variable). All strains sequenced came from patients with typhoid fever, suggesting that

any variable pseudogenes are giving information on genes and pathways whose functional status is not relevant during the infective process.

Of the core pseudogenes, there are a number that provide a genotypic basis for a known phenotype. In the API20E system, Typhi is negative for the metabolism of rhamnose. StyCyc shows that rhamnose degradation is carried out by the *rha* genes, which catalyse 3 reactions to generate L-lactaldehyde and dihydroxyacetone phosphate. The latter feeds into glycolysis, but the final enzymatic step required to produce this metabolite is catalysed by *rhaD*, which is a pseudogene in Typhi, due to an early frameshift. Hence Typhi cannot utilise rhamnose as a sole carbon source.

A weakly positive production of hydrogen sulphide during growth on thiosulphate is also used to identify Typhi in biochemical testing. StyCyc indicates that the genes *ttrABC* encoding tetrathionate reductase are all intact, so a search for regulators found that this is likely due to the pseudogene present in *ttrS*, the sensor element of the *ttrRS* 2-component regulator that controls *ttrABC*.

Production of vitamin B12 has been identified previously as inactive in both Typhi and Paratyphi A, and StyCyc confirms that multiple *cbi* pseudogenes interrupt the biosynthetic pathway. Thus, StyCyc has identified the genetic basis for known lesions in Typhi metabolism, validating the PGDB for use in determining the effects of other pseudogenes upon metabolic capability.

Perhaps the most striking effect seen in StyCyc is the number of Typhi pseudogenes that occur in transport enzymes and complexes. Fifteen percent of transport reactions are encoded by pseudogenes or complexes that contain pseudogenes. The majority of the affected transporters are predicted to promote the passage of sugars from the environment

into the cell. This has implications for the range of sugars Typhi is able to take up and utilise, with the possibility that only a small subset of sugars is available during human infection, hence other transport systems are no longer required. A knock-on effect if a substrate can no longer be imported is that flux through the metabolic pathway for utilisation of that substrate may be vastly reduced, leading to compensatory flux changes elsewhere.

2.4.4 StyCyc versus StmCyc: a basis for comparison

The eventual side-by-side curation of StyCyc and StmCyc allowed any pathway differences to be easily identified and explained. Currently in StyCyc 7.0, there is only one biosynthetic pathway present that is not found in Typhimurium; Vi antigen biosynthesis. The *tvi* genes required to produce the Vi antigen are encoded by the *viaB* locus of SPI-7, which has never been found in Typhimurium. In contrast, StmCyc 4.0 contains four additional pathways for the degradation of secondary metabolites, with respect to Typhi. Two of these are related: L-idonate degradation and ketogluconate metabolism, both of which involve *idn* genes no longer found in Typhi, due to phage insertion. The ability to degrade myo-inositol is conferred upon Typhimurium by a genomic island just over 22 kb in length, encoding the *iol* gene cluster (Kroger and Fuchs 2009). This island is only otherwise found in Paratyphi B and two other rare *Salmonella* serovars. Another region unique to Typhimurium encodes five *dgo* genes that account for the transport and degradation of D-galactonate, via a different pathway to that found in non-enteric bacteria. Finally, Typhimurium encodes SL1455, which is predicted to function as a glutaminase catalysing the conversion of glutamine to glutamate, as an

alternative to the glutamate synthase complex *gltBD* present in both Typhi and Typhimurium.

2.4.5 StyCyc and StmCyc as resources

Both PGDBs can be accessed via the Internet at http://pathways.genedb.org. At present, StmCyc is in use by external collaborators at the University of Birmingham to look at gene expression data in Typhimurium. StmCyc is also being used by internal collaborators to locate pseudogenes in metabolic pathways of host-adapted variants of Typhimurium, and to visualise gene expression.

In another internal collaboration, we have used the new genome sequence and annotation from *S. bongori* 12419 to generate a metabolic map modified from StyCyc 6.1. This provided a shorthand method to obtain a global view of metabolism in this *Salmonella* species.

2.5 Conclusions

The generation of metabolic maps for Typhi and Typhimurium has opened up new possibilities for the visualisation and interpretation of gene-based data in *Salmonella*. For the first time, the gene degradation apparent from the Typhi genome sequence was put into the context of metabolic pathways, allowing metabolic phenotypes to be associated with pseudogene genotypes. A side-by-side curation effort with both databases revealed which pathways were unique to each serovar, and showed that the major differences between the two were due partly to genes present in Typhimurium and not Typhi and partly to the interruption/inactivation of pathways and transport reactions in Typhi and not Typhimurium.

StyCyc and StmCyc provide a basis for metabolic comparisons not only between themselves, but also as a framework upon which data from multiple *Salmonella* serovars can be visualised and interrogated.

3 Pseudogenes in host restricted *Salmonella enterica*

3.1 Introduction

Within subspecies I, hugely variable levels of host specificity and disease outcomes can be found among the 1,547 serovars. Based upon serotyping, there are even groups of serovars that share the same antigenic formula, but still display the variation in host range and severity of disease seen across the subspecies as a whole, e.g. 6,7:c:1,5 (Paratyphi C, and Typhisuis and Choleraesuis) and 1,4,(5),12:Hb:1,2 (Paratyphi B and Java) (Grimont and Weill 2007). Paratyphi C and Typhisuis are restricted to humans and swine respectively (S. Nair, personal communication) while Choleraesuis is strongly adapted to swine, although it is capable of causing a highly invasive disease in humans. Isolates of both Paratyphi C and Choleraesuis have been sequenced, revealing that the formation of pseudogenes is a common feature (Chiu et al. 2005; Liu et al. 2009). Serovars Paratyphi B and Java are distinguished on the basis of a single metabolic test, *d*-tartrate fermentation, yet Paratyphi B is believed to cause paratyphoid fever in humans, and Java mild gastroenteritis (Han et al. 2006).

Another lineage displaying differential host specificity includes Gallinarum, which is restricted to chickens. By multi locus enzyme electrophoresis (MLEE), Gallinarum forms a related cluster with host-generalist Enteritidis and bovine-adapted Dublin (Boyd et al. 1993). The genome sequence of Gallinarum has also been completed, again revealing a significant level of degradation, with over 300 pseudogenes (Thomson et al. 2008).

In recent years, as more genome sequences have become available for *S. enterica*, genome degradation has emerged as a common theme for both host-adapted and

restricted serovars. This phenomenon is also seen in other bacteria, including *Mycobacterium*, *Shigella* and *Yersinia* (Chain et al. 2004; Cole et al. 1998; Wei et al. 2003).

Prior to the advent of genome sequencing, *Salmonella* serovars were differentiated by the metabolic reactions they could or could not perform, and there were hints of an association between reduced metabolic capacity and a narrow host range (Winslow et al. 1919). Until recently however, there have not been the tools available to investigate whether the effect of extensive pseudogene formation is responsible for the lack of metabolic capability displayed by host restricted *Salmonella*.

By using the metabolic pathway databases described in Chapter 2, pseudogenes from multiple host-restricted serovars can now be identified and their effect upon metabolic capability examined. Further advances in technology now allow multiple metabolic phenotypes to be assayed at once, which aid in describing the full metabolic potential of serovars regardless of host range. Comparing the loss-of-function phenotypes found in host-restricted serovars with the fuller metabolic potential observed in host-generalists has begun to uncover the pseudogenes responsible.

3.2 Methods

3.2.1 Strains

The serovars and related genome sequences used in the pseudogene comparison and for Biolog phenotyping are given in Table 3-1. The Typhimurium and Typhi strains are described in greater detail in the following chapter.

Table 3-1 Strains

Serovar	Strain (pseudogenes)	Strain (Biolog)	Host	Reference
Typhimurium	n/a	SL3261	Multiple	This dissertation
Typhi	multiple*	WT174†	Human	(Holt et al. 2008) (Langridge et al. 2009b)
Paratyphi A	ATCC 9150 AKU 12601	nd	Human Human	(McClelland et al. 2004) (Holt et al. 2009b)
Gallinarum	287/91	287/91	Chicken	(Thomson et al. 2008)
Typhisuis	61-6	61-6	Pig	This dissertation

* Pseudogenes analysed from 19 Typhi strains in total, including CT18 (Parkhill et al. 2001a); † This is an attenuated strain of Typhi classified as hazard group 2; nd, not done as an attenuated hazard group 2 Paratyphi A was not available (normally hazard group 3).

3.2.2 Sequencing of a Typhisuis reference strain

The Typhisuis 61-6 genome sequencing and initial assembly was performed by Craig Corton.

Typhisuis 61-6 (kindly donated by S. Nair) was sequenced as a reference strain, using two second-generation sequencing platforms: 454 Roche GS FLX Titanium and the Illumina Genome Analyzer II. The sequence data produced by 454 Roche consisted of a paired end library with a 3 kb insert, generating 232,241 reads. The Illumina platform used a 200-300 bp standard paired end library and was run in one lane on a flow cell

generating 18,525,268 37 bp reads. The theoretical sequence depth coverage for the Roche 454 and Illumina platform was 10x and 146x respectively.

The Illumina sequences were assembled using Velvet, a *de novo* short read assembly program (Zerbino and Birney 2008). The parameters used for the assembly were optimised for the Typhisuis dataset and produced an assembly in 421 contigs with an N50 of 20,868 bp, representing approximately 97% of the entire genome (based upon the genome size of Enteritidis P125109). The contigs generated by the Velvet assembly were then combined with the 454 Roche sequences using Roche's GS *de novo* assembler, Newbler. The final combined assembly statistics were 1,041 contigs with an N50 of 36,665 bp. The combined assembly statistics were skewed due to slight contamination in the Roche 454 library with *Yersinia enterocolitica*. However, the majority of this contamination assembled in contigs less than 2 kb in length, and the final assembly contained 210 contigs > 2 kb.

The combined assembly was converted into a Gap4 database (Bonfield et al. 1995) to allow improvement by a round of *in silico* finishing. The 454scaffolds.fna file produced as part of the Newbler assembly output was used to guide gap closure based on the 454 read pair information. ABACAS, a script for ordering and orientating fasta sequences against references (Assefa et al. 2009) was used to align the fragmented Typhisuis assembly against the complete genomes of Enteritidis P125109 and Choleraesuis SC-B67 to help scaffold the contigs where no read pair information was available. This allowed a large number of small repeat regions to be correctly assembled (the cause of the majority of gaps in the sequence), aiding the reduction in contig numbers. After improvement there were 36 contigs > 2kb, containing approximately 99% of the total genome, again

based upon genome length of Enteritidis P125109. These final contigs were ordered and oriented against Choleraesuis SC-B67 using ABACAS. iCORN (Otto et al.) was then used to correct the assembled sequence using the Illumina sequences, primarily to check all homopolymer base discrepancies (errors inherent with the Roche 454 technology) and to highlight any potential problematic regions within the assembly. The corrections made by iCORN were checked and confirmed using the assembly in the Gap4 database.

3.2.2.1 Pseudogene identification and validation

The corrected sequence aligned against Choleraesuis SC-B67 was used to mark up the position of putative pseudogenes. By comparing the Choleraesuis and Typhisuis genomes using the Artemis Comparison Tool (ACT) (Carver et al. 2005), each Choleraesuis pseudogene was checked in Typhisuis for the same, or different, inactivating mutation. Every Choleraesuis coding sequence was subsequently checked for a possible pseudogene in Typhisuis.

Craig Corton then performed sequence checks of putative pseudogenes.

These were checked against both the 454 and Illumina sequence to determine whether coverage was sufficient to call them pseudogenes.

Maria Fookes performed assemblies of multiple Typhisuis genomes sequenced using Illumina.

Seven strains of Typhisuis were sequenced using the Illumina short read platform. Reads relating to each individual strain were assembled using Velvet (Zerbino and Birney 2008), and contigs were ordered relative to the reference strain 61-6, using ABACAS

(Assefa et al. 2009). All contigs not assembled against the reference were concatenated to the draft genome, which was then used to run a protein BLAST against the reference .

Ambiguous pseudogenes in Typhisuis 61-6 were checked against these seven other Illumina-sequenced Typhisuis genomes in order to determine their validity. A pseudogene was deemed valid if sequence from at least 4 of the other genomes was continuous across the relevant region and consistent with the 61-6 sequence.

3.2.3 Whole genome comparison

A full list of pseudogenes was obtained from the genome annotation of Gallinarum (GenBank accession number AM933173). The pseudogene complement of Paratyphi A ATCC9150 was taken from the genome annotation (GenBank CP000026) and extended to include others found by comparison with Paratyphi A AKU 120601 (GenBank FM200053). For Typhi, the pseudogene list was based upon those annotated in CT18 (GenBank AL513382) and again extended based upon comparison with multiple sequenced isolates (Holt et al. 2009b). Pseudogenes in Typhisuis 61-6 were determined as described above.

Pseudogene orthologues between all the genomes were established either from the genome annotation or from reciprocal nucleotide BLAST searches. Genes were deemed to be orthologues if they were the best reciprocal BLAST hit for each other. Frameshifted pseudogenes were typically un-matched, so orthologues were determined manually by checking for conservation and synteny between genomes using ACT.

3.2.4 Functional classification of pseudogenes

The 'lost' functions of pseudogenes in all four serovars were assessed by comparison to intact orthologues annotated in Typhimurium SL1344 (in-house annotation) and Enteritidis P125109 (GenBank AM933172). Functional categories were taken from standard Sanger annotation.

3.2.5 Metabolic pathway analysis

Each serovar was matched to either StyCyc or StmCyc depending upon which contained the highest number of orthologous pseudogenes. The list of appropriate pseudogene orthologues was overlaid on the relevant metabolic map in order to visualise where pseudogenes were interrupting pathways and transport reactions. All pathways containing pseudogenes were recorded for each serovar. Where possible from the literature, regulators for pathways were identified and checked against pseudogene lists.

3.2.6 Biolog phenotyping

The Biolog Phenotype MicroArray (PM) assays (Technopath, Ballina, Ireland) for some of the strains were carried out by Theresa Feltwell, as indicated in Table 3-2.

Each strain was assayed in triplicate, with PM plates 1 and 2A, which contain 192 individual carbon sources. All strains were assayed at the body temperature of their natural host: 37 °C for mammalian hosts and 42 °C for avian (Gallinarum).

Table 3-2 Metabolic phenotyping

Serovar	Strain details	Additives	Temperature
Typhi	WT174	Aro mix	37 °C
Typhimurium	SL3261	none	37 °C
Gallinarum*	287/91	Nicotinic acid and thiamine	42 °C

*This strain and the third replicate of Typhi and Typhimurium were phenotyped by Theresa Feltwell; aro mix is comprised of 40µg/mL each of L-phenylalanine and L-tryptophan and 10µg/mL each of *p*-aminobenzoic acid and 2,3-dihydroxybenzoic acid, final concentration.

3.2.6.1 Preparation of strains to be phenotyped

Luria-Bertani (LB) agar plates were inoculated with frozen stocks (stored at -80°C) of each strain and incubated overnight at 37 °C. New LB agar plates were inoculated from the overnight growth and again incubated overnight at 37 °C. Colonies from these plates were used to inoculate the PM plates.

3.2.6.2 Preparation of PM Inoculating Fluids

125 mL 1.2x IF-0 (Technopath) was added to 25 mL distilled H_20, of which 16 mL was aliquoted into a sterile, capped test tube. A separate 'IF-0 + dye' mixture was prepared by adding 1.8 mL of dye mix (Technopath) and 23.2 mL distilled H_20 to 125 mL 1.2x IF-0. Where required, 'aro mix' was added at 1 in 100, and nicotinic acid and thiamine at 0.5 µM, final concentration (see Table 3-2 for serovar-specific supplements). 50 mL aliquots were pipetted into sterile glass bottles for plates PM-1 and PM-2A .

3.2.6.3 Preparation of cell suspension

Colonies were taken from LB agar plates using a sterile swab and transferred into the sterile capped tube containing 16 mL IF-0. Cells were added to this suspension until a 42%T (transmittance) had been achieved according to the Biolog Turbidimeter (Technopath). 10 mL of the 42%T cell suspension was added to the vial containing 50 mL 'IF-0 + dye' and gently mixed to a final cell density of 85%T.

3.2.6.4 Inoculation of plates PM-1 and PM-2A

22 mL of the 85%T cell suspension was transferred to a sterile reservoir and used to inoculate PM-1 and PM-2A at 100 µL / well. The PM plates were incubated for 48 hours at 37 ^{o}C or 42 ^{o}C, and readings were taken every 15 minutes. The readings measured the strength of the dye colour in each well, which increased as the substrate in that well was utilised.

3.2.7 Analysis of Biolog phenotypes

3.2.7.1 Phenotype calling

Raw data from all 3 replicates of the two PM plates for each strain were loaded into the OmniLog File Management software (version 12.0, Technopath). The value determining the area under the curve (AUC) was taken as a representative parameter for the 48 hour timepoint and was exported into comma-separated files. The AUC values were averaged across replicates and plotted on a frequency distribution to determine an arbitrary cutoff

below which the strain was deemed unable to utilise the relevant substrate. This value was placed at 100, which correlated well with the kinetic graphs produced by the OmniLog software. Thus, substrates where the AUC value exceeded 100 at 48 hours were deemed positive metabolic phenotypes.

3.2.7.2 Differential phenotypes

As above, the raw data from all 3 replicates were loaded into the OmniLog File Management Software (Technopath). In this instance, the AUC values were exported for every timepoint, which covered measurements taken every 15 minutes from 0 to 48 hours. A pair-wise analysis was used to compare both Gallinarum and Typhi with Typhimurium.

Lars Barquist performed this analysis in R, using `limma`.

The data were read into R using the `PMarray` package (unpublished, L. Barquist). Each well was modelled with a spline-fit curve using `grofit` (Kahm et al.). `limma` (Smyth 2004) was then used to assign Benjamani-Hochberg corrected *P*-values to the log fold differences in AUC values between strains. A cutoff of $P < 0.001$ was used, corresponding to a false discovery rate of 0.1%.

3.3 Results

A summary of the total number of pseudogenes determined for each genome or set of genomes analysed is given in Table 3-3. The genomes of Typhi, Paratyphi A and Typhisuis displayed similar levels of genome degradation of ~200 pseudogenes each, with Gallinarum showing the greatest loss of gene function with over 300 pseudogenes.

Table 3-3 Pseudogenes across host restricted *Salmonella*

Serovar	Strain	Pseudogenes	Genome size
Typhi	multiple	211	4.8 Mbp
Paratyphi A	multiple	187	4.6 Mbp
Gallinarum	287/91	306	4.65 Mbp
Typhisuis	61-6	190	4.65 Mbp

An analysis of the functions lost from each of these serovars revealed that the nature of the genome degradation seen was broadly similar across them all (Figure 3-1). Approximately one third of pseudogenes are found in membrane/surface structures, and alongside loss of function in regulation and pathogenicity/adaptation/chaperones account for at least half of the genome degradation. Pseudogenes that represent conserved hypothetical proteins and proteins of unknown function make up the next largest proportion, indicating that there may be functions relating to the top 3 categories that have yet to be elucidated.

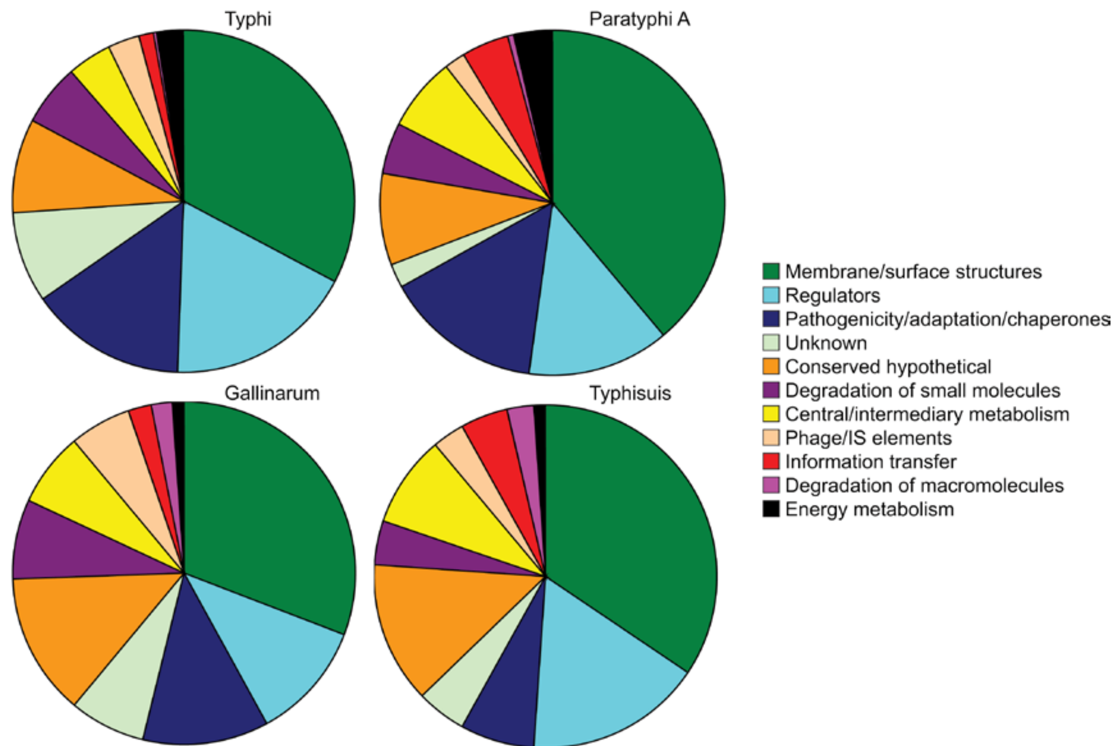


Figure 3-1 Functional classification of pseudogenes

Pseudogene function determined based upon comparison with intact orthologues in Typhimurium and Enteritidis. Categories and colours based upon genome annotation.

3.3.1 Individual pseudogenes

Across the four genomes of host restricted *Salmonella*, there are only 2 individual pseudogenes that are shared between them. Both are intact in the host generalist Typhimurium. The first is *sopA*, encoding a secreted effector protein known to be translocated into eukaryotic cells and to play a role in bovine enteritis in serovar Dublin (Wood et al. 2000). The second, *mglA*, encodes one of the inner membrane components of a galactose transport system and acts as an ATPase specifically stimulated by galactose (Richarme et al. 1993). The inactivating mutations for these genes are different in each serovar (with one exception), indicating a convergent loss of function. The exception is

sopA in Paratyphi A and Typhi which has the same inactivating mutations since the gene falls into one of the low divergence regions indicating recombination between the serovars (Didelot et al. 2007).

3.3.2 Relating metabolic phenotype to 'pseudo' genotype

A total of 192 carbon sources were tested for their ability to support the growth of Typhi, Typhimurium and Gallinarum (Appendix 8.2.1 and 8.2.2). The range of substrate utilisation is shown in Figure 3-2, indicating that Typhi and Gallinarum are capable of utilising less than two thirds of the number of substrates that support the growth of Typhimurium.

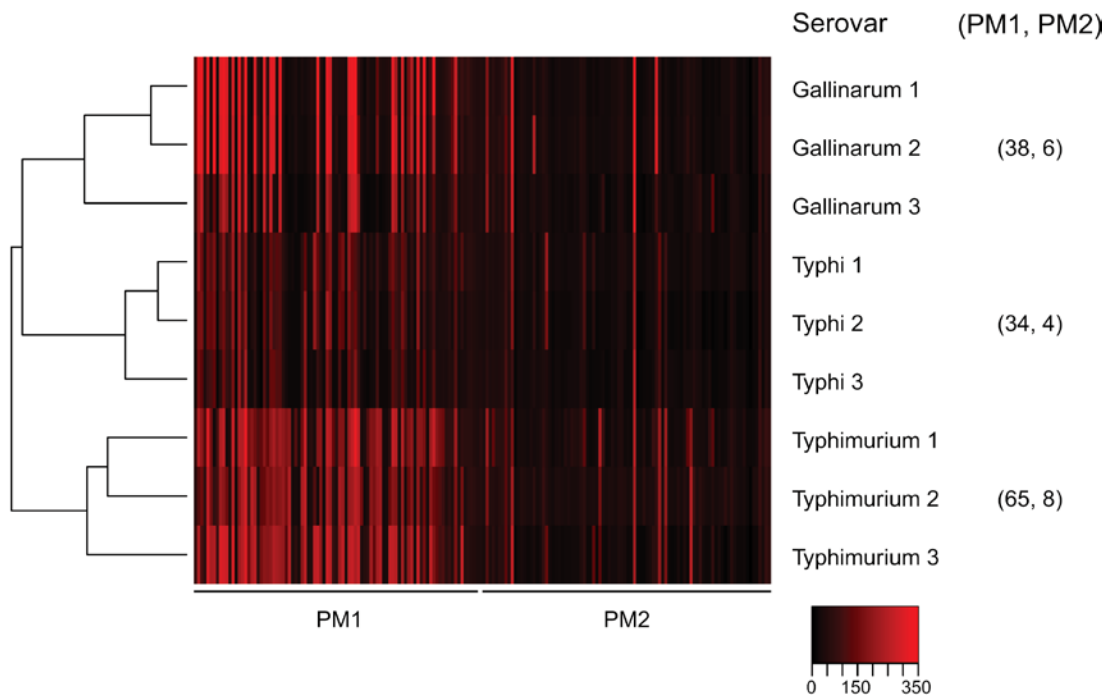


Figure 3-2 Heatmap of carbon source utilisation

Drawn using the R heatmap2 package (http://www.r-project.org). Values are a measure of the area under curve (AUC) for dye reduction of each substrate. Higher values (red) indicate capacity to utilise substrate. (PM1, PM2) numbers represent the number of positive metabolic phenotypes per Phenotype MicroArray.

3.3.2.1 Negative metabolic phenotypes

Both Typhi and Gallinarum yielded negative phenotypes for utilisation of L-ornithine (AUC < 100), a result consistent with the use of this substrate in the API20E test that distinguishes serovars of *Salmonella enterica*. According to StyCyc, L-ornithine is degraded into putrescine by either *speC* or *speF*. The former is constitutive and is inactivated in both Typhi and Gallinarum. The latter is inducible but appears not to compensate for the lack of *speC*, since it remains intact in Gallinarum. Alternatively, Gallinarum may harbour an additional inactivating mutation in the induction mechanism.

3.3.2.2 Differential metabolic phenotypes

When compared with Typhimurium, differential phenotypes were observed for 29 substrates in Typhi and 30 in Gallinarum ($P < 0.001$). In Typhi, all 29 displayed a negative log fold change (Log FC) but 2 of the phenotypes in Gallinarum showed a slight increase in utilisation relative to Typhimurium (Table 3-4). These were α-D-glucose and D-fructose (Log FCs 0.98 and 1.08 respectively), perhaps indicating the preference of Gallinarum for these carbon sources.

As mentioned in the previous chapter, Typhi contains a pseudogene in *rhaD* that affects its ability to ferment rhamnose. As a sole carbon source, rhamnose would provide energy by its degradation into dihydroxyacetone phosphate which is further utilised in central metabolism. This phenotype was borne out in the Biolog data for Typhi (Log FC -3.4), and for Gallinarum (Log FC -2.2). Gallinarum retains all the genes required for rhamnose

transport and degradation, but the likely cause of the loss of function is a pseudogene in *rhaS*, which encodes the transcriptional activator of the rhamnose operon.

Typhi displayed no growth on L-arabinose (LogFC -1.38, $P < 0.0005$), which would typically sustain growth as a single carbon source by its 3-step conversion to D-xylulose-5-phosphate, an intermediate of the pentose phosphate pathway. All the genes involved in this conversion are intact in Typhi. However, there are two uptake systems for L-arabinose in *E. coli*, a low affinity transporter *araE*, and a high affinity transport complex encoded by *araFGH*. While *araE* is intact in Typhi, only a truncated form of *araH* remains of the high-affinity transporter. Thus, it appears *araE* alone is not sufficient to support growth.

Table 3-4 Differential Biolog phenotypes

Ty LogFC	Ty *P*-value	Substrate	Gal LogFC	Gal *P*-value
-1.58	0.00017	acetic acid		
		alpha-D-glucose	0.98	0.00065
		alpha-hydroxy butyric acid	-2.14	1.91E-05
-2.29	8.22E-05	**alpha-keto-butyric acid**	-2.71	2.87E-05
		alpha-methyl-D-galactoside	-3.25	6.08E-06
-2.63	7.24E-06	**bromo succinic acid**	-2.52	1.05E-05
-2.99	6.03E-06	**D,L-malic acid**	-3.33	6.08E-06
-2.47	6.03E-06	**D-aspartic acid**	-2.62	6.16E-06
-1.53	0.000429	dextrin		
		D-fructose	1.08	0.00073
-2.06	1.47E-05	**D-galatonic acid-gamma-lactone**	-1.52	0.00012
-3.19	4.05E-06	**D-glucosaminic acid**	-2.86	6.34E-06
		D-lactic acid methyl ester	-1.77	4.09E-05
		D-melibiose	-2.14	4.04E-05
-1.57	0.000777	D-ribose		
-2.81	4.63E-06	**D-saccharic acid**	-2.46	9.93E-06
		D-sorbitol	-2.24	9.93E-06
-2.90	7.24E-06	**D-tagatose**	-2.48	1.95E-05
-2.81	4.05E-06	**D-tartric acid**	-2.51	6.08E-06
-2.67	5.99E-05	**fumaric acid**	-2.62	6.70E-05
-1.54	8.22E-05	glycyl-L-aspartic acid		
-2.32	4.16E-05	L-alanine		
-1.38	0.000343	L-arabinose		
-1.88	1.14E-05	**L-asparagine**	-1.89	1.08E-05
-2.94	4.63E-06	**L-aspartic acid**	-2.42	1.08E-05
-2.36	8.87E-06	**L-fucose**	-1.88	4.09E-05
-1.30	8.22E-05	**L-glutamic acid**	-1.06	0.00038
-1.63	6.11E-05	**L-glutamine**	-1.66	5.61E-05
-1.98	4.13E-05	**L-malic acid**	-1.25	0.00096
-3.41	1.44E-05	**L-rhamnose**	-2.19	0.00034
		melibionic acid	-2.62	9.93E-06
-1.77	5.90E-05	**m-hydroxy phenyl acetic acid**	-1.64	8.87E-05

Log fold change (FC) and adjusted *P*-values were calculated with respect to Typhimurium. Substrates (in alphabetical order) in bold are differential phenotypes shared by Typhi (Ty) and Gallinarum (Gal).

D-tagatose was poorly utilised by both Typhi and Gallinarum with respect to Typhimurium (LogFCs -2.9 and -2.5 respectively). Tagatose, a naturally occurring isomer of fructose, is transported into the cell by a phosphotransferase (PTS) system encoded by *tagTH*. Tagatose-1-phosphate is phosphorylated by *tagK* to form tagatose-1,6-bisphosphate which is degraded by the *gatYZ*-encoded aldolase in a reaction that also forms the final step in galactitol degradation (Figure 3-3) (Mayer and Boos 2005; Shakeri-Garakani et al. 2004). The final products in these pathways are D-glyceraldehyde-3-phosphate and dihydroxyacetone phosphate, both ubiquitous metabolites that feed into a variety of metabolic pathways to provide sufficient energy for growth. In Typhi, *gatZ* is inactivated by a premature stop codon, and this also explains why Typhi produced a negative phenotype for galactitol (named dulcitol in the Biolog system). Gallinarum also displayed a negative phenotype for galactitol, but both PTS systems for tagatose and galactitol are intact, as are all the genes depicted in Figure 3-3. The mechanism behind these phenotypes in Gallinarum therefore remains unclear.

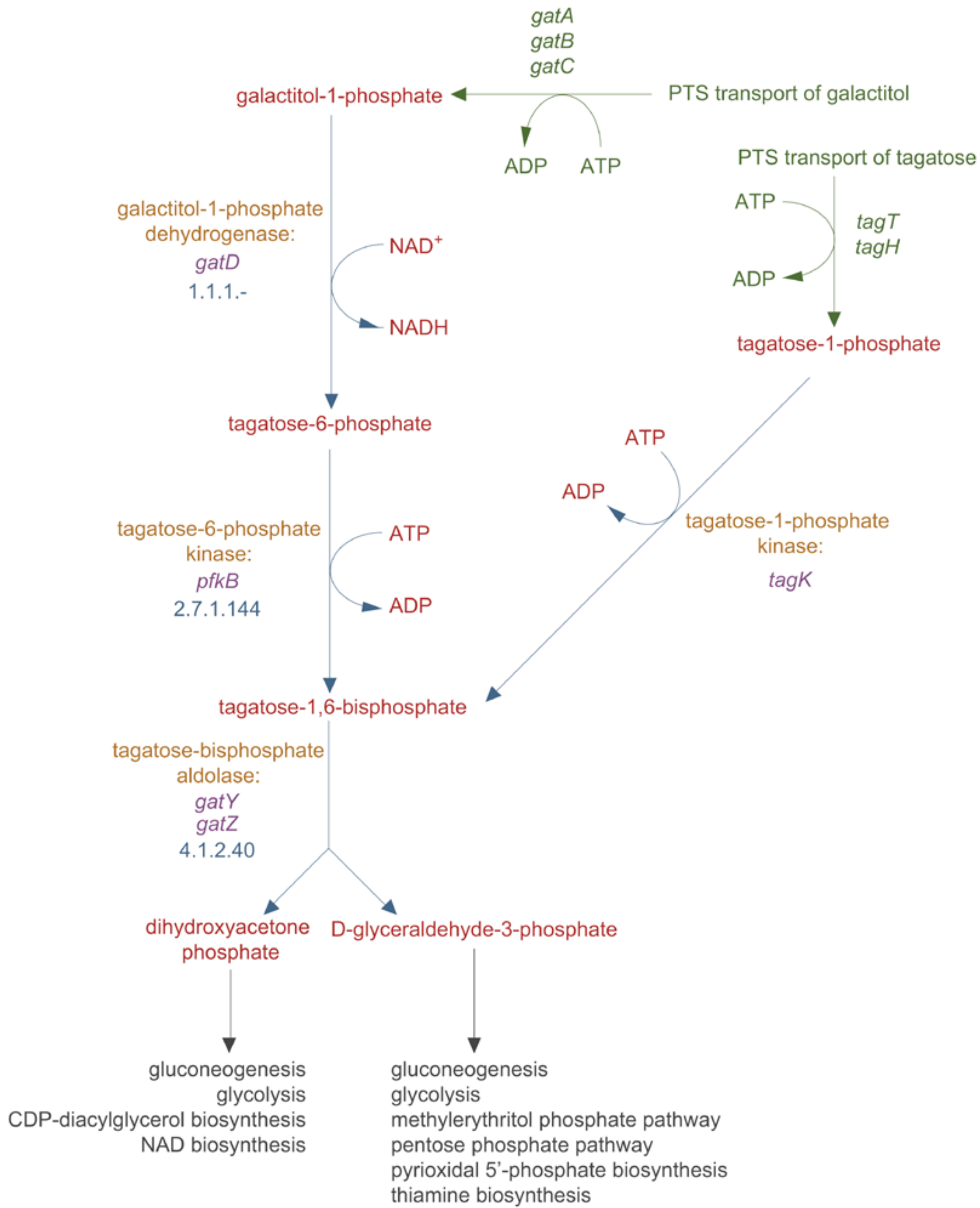


Figure 3-3 Galactitol (dulcitol) and tagatose degradation

Metabolic pathway diagram from StyCyc 7.0 demonstrating that the degradation of galactitol and tagatose share the final enzyme, which in Typhi is inactive. Blue arrows indicate enzymatic reactions, substrates shown in red, enzymes in gold, genes in purple and enzyme commission numbers in blue. Green arrows indicate transport reactions, substrates and genes also shown in green. Grey arrows and names represent pathways into which dihydroxyacetone phosphate and D-glyceraldehyde-3-phosphate subsequently feed into.

Gallinarum alone displayed a differential phenotype for sorbitol (glucitol), with a Log FC of -2.2. Sorbitol is a sugar alcohol transported into the cell via a phosphotransferase (PTS) permease encoded by *srlAEB* and degraded into D-fructose-6-phosphate by *srlD*. Again, all of these genes are intact, but given that this reaction may be used to differentiate Gallinarum from its close relative Pullorum, this suggests there is some other interaction affecting its ability to utilise this substrate.

One deficiency shared between Typhi and Gallinarum is lack of growth on L-glutamine (Log FCs both -1.6). In the Typhimurium strain tested, the utilisation of L-glutamine was relatively slow, with a lag of ~15 hours before reaching the AUC threshold of 100. This phenotype is supported by evidence from *E. coli*, where L-glutamine can support growth as a single carbon source, but growth is slow (McFall and Newman 1996). One multimeric complex has been demonstrated to transport L-glutamine, encoded by *glnQPH*. Gallinarum has a pseudogene in *glnH*, the periplasmic binding protein of the transport complex. Since *E. coli* mutants with enhanced transport are known to grow well, such a defect in this transporter is a highly probable cause of this phenotype. However, *glnQPH* appears intact in Typhi, raising the possibility that it is not a pseudogene causing this phenotype, but that transcription of the intact required genes is not at a high enough level to support growth.

3.3.2.3 Potential for improving annotation

Another shared phenotype was observed for mucic acid, also known as D-galactarate. Degradation of this substrate is by the products of four *gar* genes and *glxK*, and results in the production of pyruvate and 2-phosphoglycerate. A transporter for mucic acid has been

proven to be present in *E. coli*, but no gene has yet been shown to encode it (Hubbard et al. 1998). Given the shared phenotype, there is a possibility that this transporter is encoded in *Salmonella* by one of the pseudogenes in common between Typhi and Gallinarum. One candidate is STY2501/SG2301 which is a putative transmembrane protein that has a Pfam hit to the Major Facilitator superfamily (Finn et al. 2007). A site-specific mutation/knockout of this gene in Typhimurium could answer this hypothesis, and indicates the potential of high throughput metabolic phenotyping for improving genome annotation.

3.3.2.4 Other metabolic phenotypes

One metabolic reaction not part of the Biolog system but commonly used in *Salmonella* differentiation is the focus of a test for tetrathionate reduction, and the production of hydrogen sulphide (H_2S). Both Gallinarum and Paratyphi A test negative for H_2S production, with Typhi only weakly positive. The H_2S phenotypes can all be explained by pseudogenes present in the tetrathionate reduction pathway (Figure 3-4). Both Paratyphi A and Gallinarum have pseudogenes in tetrathionate reductase, encoded by *ttrABC*, located within SPI-2. In Typhi however, an inactivating mutation occurs in *ttrS*, the sensory element of a two-component regulator with *ttrR* that positively regulates the activity of *ttrABC* (Hensel et al. 1999). It is the activity of the tetrathionate reductase that is usually measured, but further degradation of the pathway has occurred in both Paratyphi A and Gallinarum.

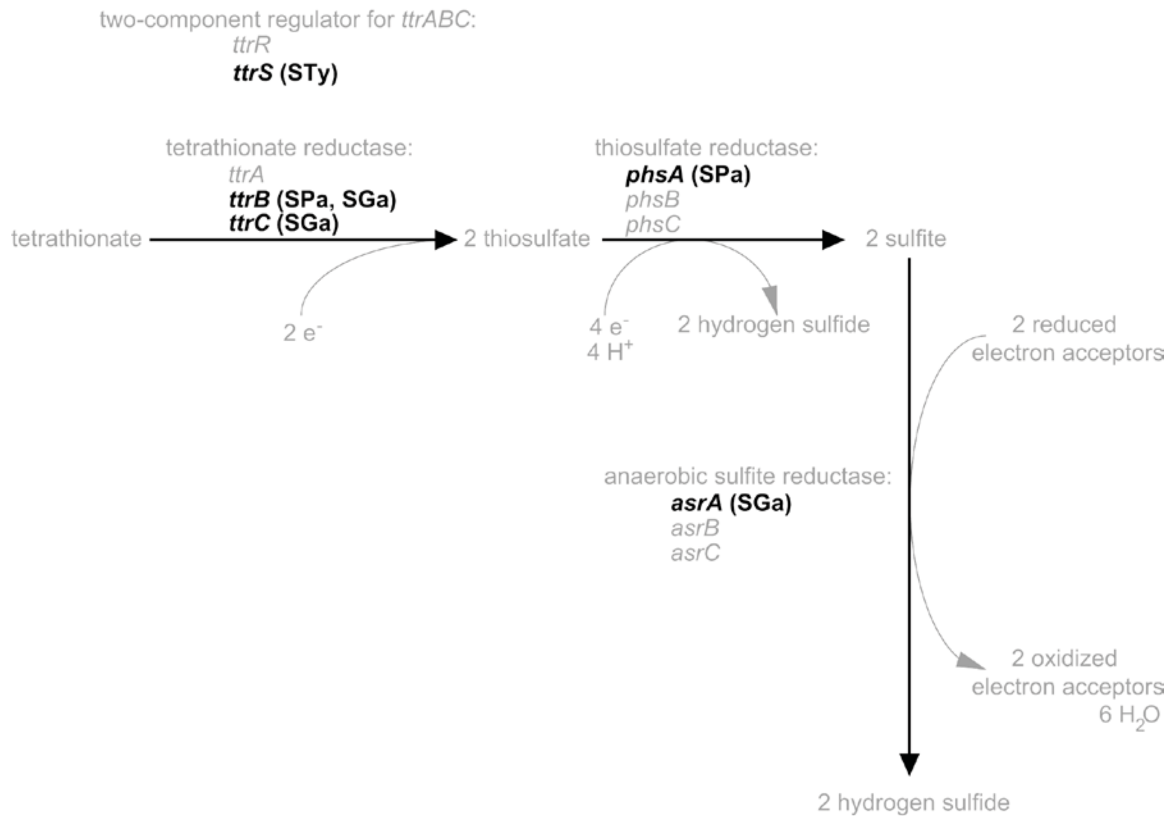


Figure 3-4 Metabolic pathway of tetrathionate reduction

Reaction lines and gene names in bold represent pseudogenes in host-restricted *Salmonella*. STy, Typhi; SGa, Gallinarum; Spa, Paratyphi A.

While Paratyphi A could not be tested with the Biolog system (as a hazard group 3 organism), the main reaction used to distinguish this serovar is a negative phenotype for L-lysine. From StyCyc, the genes involved in lysine degradation are *ldcC* and *cadA*, both of which are intact in Paratyphi A. The primary transporter for lysine is encoded by *cadB*, which is also intact, but the most likely cause of this phenotype is the pseudogene present in *cadC*, the transcriptional activator of *cadAB*.

3.3.3 Shared interruption of metabolic pathways

While few individual pseudogenes are shared between all serovars, similar losses of metabolic capability are found and can be explained by different pseudogenes interrupting the same metabolic pathway.

3.3.3.1 Pathways inactivated in all host restricted serovars

The anaerobic biosynthesis of vitamin B12 has been noted previously as inactive in both Typhi and Paratyphi A via mutations in different *cbi* genes. Both Gallinarum and Typhisuis also have pseudogenes in this pathway, shown in Figure 3-5.

Typhi has three and Gallinarum contains two pseudogenes that halt the addition of the eight methyl groups required to form the corrin ring in cobyrinate, while it is the enzyme required for the subsequent amidation of cobyrinate to cobyrinate *a,c*-diamide that is inactivated in Paratyphi A. Cobalt ions are required during biosynthesis and are provided from an extracellular location via a transport complex of CbiMNOQ. Typhisuis, Gallinarum and Typhi have different pseudogenes in this complex and given the requirement for these cobalt ions, it is likely that the loss of transport alone inactivates the pathway. It is therefore possible that the other pseudogenes in Gallinarum and Typhi were inactivated after this loss of function, when any selective pressure to maintain the remainder of the pathway would also have been lost.

The latter stages of vitamin B12 biosynthesis involve the *cob* genes, in which a further pseudogene occurs in Gallinarum (*cobD*).

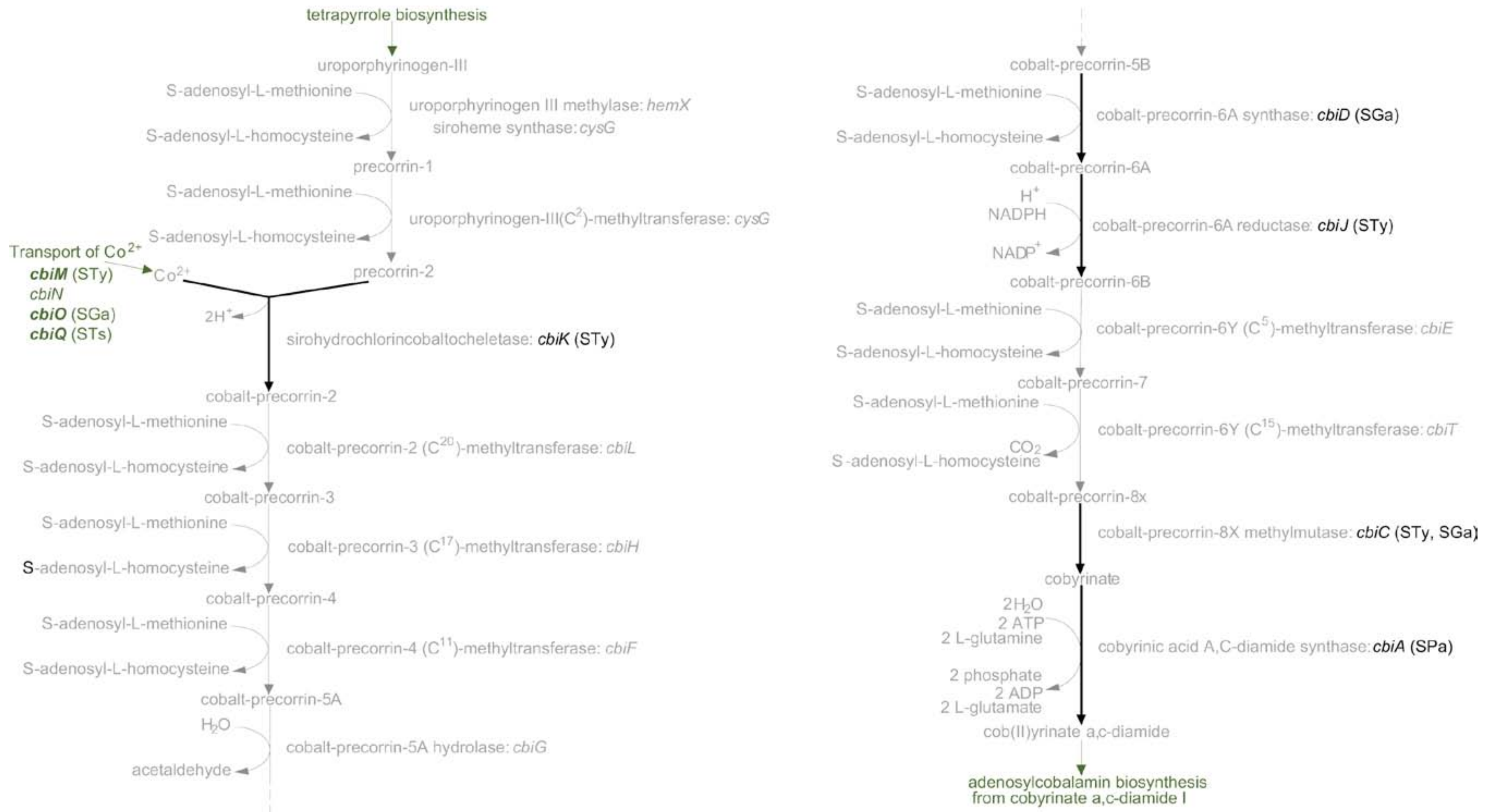


Figure 3-5 Early stages of vitamin (coenzyme) B12 biosynthesis

Reaction lines and gene names in bold represent pseudogenes in host-restricted *Salmonella*. STy, Typhi; SGa, Gallinarum; STs, Typhisuis; Spa, Paratyphi A. Green lines and names represent links to/from other metabolic pathways and transporters.

A related process is the degradation of L-1,2-propanediol (PDL), which is dependent upon the presence of vitamin B12. The metabolic pathway is shown in Figure 3-6 and each of the genomes analysed contains pseudogenes that affect PDL degradation.

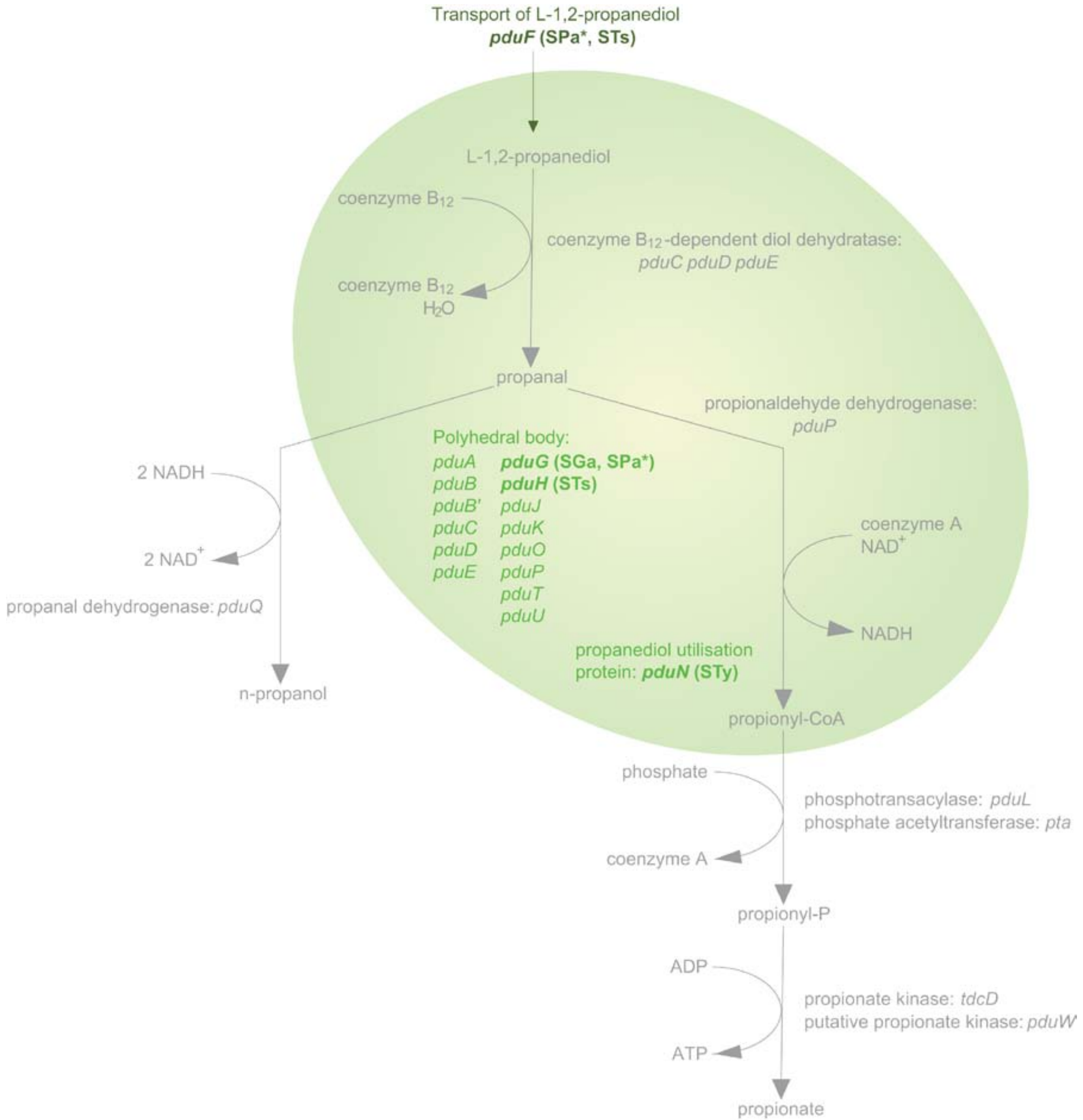


Figure 3-6 Metabolic pathway for 1,2-propanediol degradation

Gene names in bold represent pseudogenes in host-restricted *Salmonella*. STy, Typhi; SGa, Gallinarum; STs, Typhisuis; Spa, Paratyphi A. *, in the two sequenced Paratyphi A strains, one harbours a pseudogene in *pduF*, the other in *pduG*. Reactions inside the green ellipse occur inside the polyhedral body; gene names in bright green are those involved in its structure and formation.

Both Paratyphi A and Typhisuis contain mutations in *pduF*, which encodes the transporter protein for PDL. The two initial steps of the pathway are believed to occur inside the proteinaceous PDL degradation polyhedral body, which protects the cell from aldehyde toxicity (Havemann and Bobik 2003). Paratyphi A, Gallinarum and Typhisuis pseudogenes occur in *pduG* and *pduH*, which together perform the reactivation of diol dehydratase in the polyhedral body. In addition, PduG may also be involved in the adenosylation of vitamin B12 (Bobik et al. 1999). In Typhi, *pduN* is inactivated, which encodes a close relative of the CcmL-CchB family of proteins required for the proper assembly and function of carboxysomes (Bobik et al. 1999). Hence, *pduN* likely functions to aid the formation of the polyhedral bodies.

3.3.3.2 Pseudogenes inactivating pathways in human-restricted serovars

Typhi and Paratyphi A have in common two inactivated pathways that show no pseudogenes in the other host-restricted serovars. The first is the degradation of 4-hydroxyphenyl-acetate (4-HPA), which was identified in *Klebsiella* as being degraded into the TCA cycle intermediates pyruvate and succinate, thus providing energy as a carbon source (Martin et al. 1991)(Figure 3-7). 4-HPA is a metabolite produced during tyrosine degradation in some bacterial species, although this process has not been demonstrated to occur in *Salmonella* or *E. coli*.

The two pseudogenes in Typhi disrupt the pathway at the beginning and the one in Paratyphi A in the middle. Interestingly, while all the genes in this pathway, and in the associated transporter (encoded by *hpaX*) are intact in Gallinarum, it displayed a

differential phenotype (LogFC -1.64) equivalent to Typhi (Log FC -1.77, suggesting that 4-HPA utilisation is also inactivated in this serovar.

Aromatic compounds like tyrosine are abundant in soil and water, so the likelihood that these will be utilised as sole carbon sources is quite high, but only for bacteria that have a substantial life cycle in the environment (e.g. *E. coli*) (Diaz et al. 2001). It has been postulated that similar compounds can also be found in the animal gut, as the product of various degradative reactions by the intestinal flora, but it is possible that the systemic infection caused by host-restricted *Salmonella* means that the organism does not remain long enough in the gut to utilise any 4-HPA that may be available and therefore no longer retains the use of this pathway.

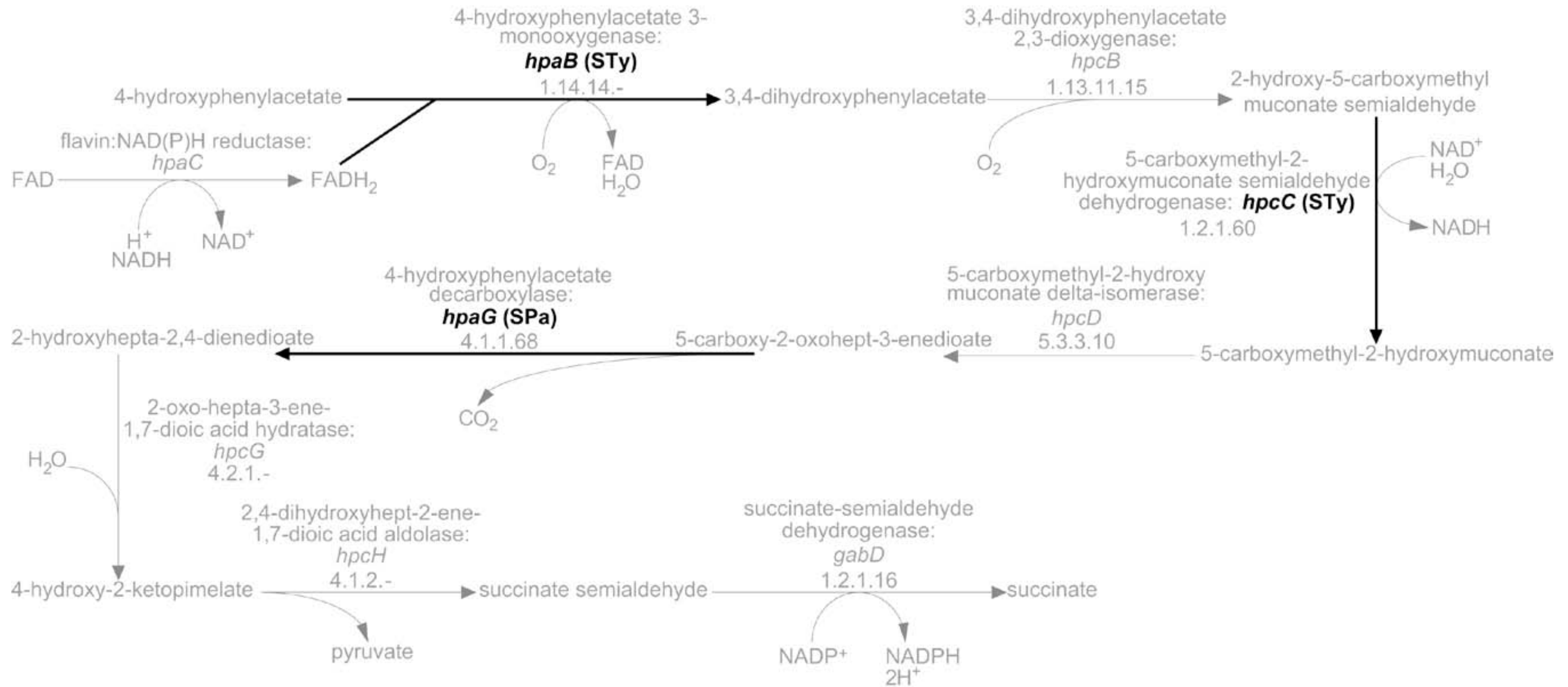


Figure 3-7 Metabolism of 4-hydroxyphenylacetate

Gene names and reactions in bold represent pseudogenes in Typhi (STy) and Paratyphi A (SPa).

One pathway that is not tested by the Biolog system and was therefore only analysed by pseudogene comparison is colanic acid biosynthesis, encoded by the *wca* gene cluster. Typhi contains pseudogenes in *wcaA*, *wcaD*, and *wcaK* while Paratyphi A harbours inactivating mutations in *wcaJ* and *wcaK*. The particular effect of this loss of function may be understood from the work that has been done on this pathway in *E. coli*, where colanic acid is made primarily at low temperatures and therefore believed to be important for survival in the environment rather than in the animal host (Ryu and Beuchat 2004; Whitfield and Keenleyside 1995). Host-restricted serovars have little theoretical need for a capsule that promotes long-term survival outside the host.

3.3.4 Loss of function in transport across the membrane

By overlaying pseudogenes onto StyCyc and StmCyc, it was possible to determine which broad functional areas (as defined by Pathway Tools) contained the highest levels of genome degradation (Table 3-5). Displaying the data in this manner indicated that the dominance of pseudogenes in membrane/surface structures (shown earlier in Figure 3-1), could be largely explained by those encoding transporter proteins as in each serovar, between 10 and 20% of pseudogenes occur in transporters. A large number of these occur in sugar transport systems, perhaps reflecting the reduced availability of many carbon sources in the restricted host environment.

Table 3-5 Metabolic functions of pseudogenes

	Typhi		Paratyphi A		Gallinarum		Typhisuis	
Function	Reactions with a pseudogene (possible redundancy[a])	% of total	Reactions with a pseudogene (possible redundancy[a])	% of total	Reactions with a pseudogene (possible redundancy[a])	% of total	Reactions with a pseudogene (possible redundancy[a])	% of total
Biosynthesis	4(1)	1%	23(12)	12%	14(8)	4.5%	10(7)	6%
Energy	1(1)	1%	1(1)	0.5%	5(5)	1.5%	4(4)	2%
Degradation	5(2)	5%	12(3)	6%	13(5)	4%	11(2)	6.5%
Transport	18(-)	15%	19(-)	10%	34(-)	11%	33(-)	20%
Standalone	11(7)		10(4)		20(6)		10(3)	
Regulator[b]	37(-)		23(-)		34(-)		31(-)	
Total pseudogenes	211		187		306		167	
Metabolic[c]	35		39		51		43	

[a], pseudogenes for reactions which retain an intact enzyme were recorded as possibly redundant. Enzyme redundancy was determined from StyCyc and StmCyc. [b], pseudogenes in regulators were identified from genome annotation. [c], represents the actual number of pseudogenes present in the relevant metabolic map from StyCyc and StmCyc. This number may vary from the number of reactions with a pseudogene as one enzyme may catalyse multiple reactions. Typhi and Paratyphi A were mapped to StyCyc, the remainder to StmCyc.

Shared across all serovars are pseudogenes in iron uptake systems (Table 3-6). *Salmonella* contain multiple complexes for obtaining iron, which is required for growth (Earhart 1996). Inactivation of *fhuE* was the most common, occurring in three of the four serovars. FhuE is a receptor protein for both ferric-coprogen and ferric-rhodotorulic acid (Hantke 1983), which are both produced by various fungal species, hence it is unlikely that a host-restricted serovar comes across these potential iron sources. Inactivation of the *fhu* complex would therefore not be associated with a biological cost.

Table 3-6 Pseudogenes in iron uptake systems

Serovar	Iron uptake system pseudogene(s)
Typhi	*fepE*, *fhuA*, *fhuE*
Paratyphi A	*fhuA*, *fhuE*
Gallinarum	*iroD*
Typhisuis	*iroD*, *fhuE*

Both Gallinarum and Typhisuis contain an inactivated *iroD*, one of two hydrolases present in the siderophore salmochelin uptake system. IroD acts to cleave salmochelin into several substrates, allowing the bound iron to be reduced and removed (Zhu et al. 2005). Typhi also contains a pseudogene in *fepE*, but this does not appear to affect the activity of the *fepBCDG* complex for uptake of another siderophore (enterobactin) that was found to be essential under laboratory conditions (see Chapter 4 (Langridge et al. 2009b)).

3.4 Discussion

Comparison of pseudogenes across host-restricted *Salmonella* has been concentrated in the past upon individual pseudogenes commonly shared. While these are likely to be important, expanding this comparison to look at commonly inactivated metabolic pathways provides a wider context in which to interpret the presumed loss of function. Examining these pathways in relation to the role they play when intact in host-generalists also gives an insight into how the loss of function may affect these serovars.

In the Biolog experiment, each serovar was grown at the temperature associated with their natural host; 37 ^{o}C for mammalian Typhimurium and Typhi, and 42 ^{o}C for avian Gallinarum. Previous work in our laboratory (unpublished) had shown that another host-associated serovar tested at varying temperatures was the most metabolically active at the temperature associated with the host. The temperatures chosen for the serovars in this study were therefore expected to be those the serovars were best adapted to, and yield the most informative results. This approach was validated by the number of metabolic phenotypes that were in accordance with those currently in use for the identification of Enterobacteriaceae, and *Salmonella* serovars in particular.

Other phenotypes were also identified, with Gallinarum showing a negative reaction for growth on L-glutamine, which could be traced to a pseudogene in the periplasmic binding protein of the transport system. While Typhimurium may not be the optimum serovar to use as a control for pseudogene formation, it did provide a basis for determining relatively straightforward phenotypes. In this study, it was shown that Typhimurium was capable of utilising over twenty substrates that neither Typhi nor Gallinarum could.

A number of the differential phenotypes displayed by Typhi and Gallinarum, with respect to Typhimurium, could be linked to ‘pseudo’ genotypes. Unsurprisingly, given the large proportion of pseudogenes that occur in genes encoding transporter proteins, mutations that affect the cells ability to take up substrates were often the cause of growth deficiency. However, the rhamnose example in Gallinarum showed the importance of being able to identify which genes are involved in regulation of metabolic pathways or transport reactions. Here, a pseudogene in the transcriptional activator of the rhamnose operon was the most likely cause of Gallinarum’s inability to utilise rhamnose. Inactivation of transport and regulatory genes also circumvent issues of pathway intermediates accumulating inside the cell, to possible harmful effects.

The Biolog Phenotype MicroArrays represent a high-throughput system for identifying metabolic phenotypes, and analysed in conjunction with the metabolic pathway databases of StyCyc and StmCyc allowed the metabolic effect of pseudogenes to be better understood.

However, there were some phenotypes that could not be directly explained with a relevant pseudogene. In order to gain further understanding of the functional relevance of these losses of function, and perhaps uncover the genotypic cause, these results should be examined in the context of the relevant evolutionary lineage. For example, the effect of pseudogenes in Gallinarum would ideally be assessed in the context of Enteritidis, the closest non-adapted relative. Obtaining Biolog data for Enteritidis would be the first step in such an analysis and is currently under way.

Not all metabolic functions can be or are tested by the Phenotype MicroArrays. Therefore, comparing pseudogenes from host-restricted serovars across metabolic

pathways complements and extends the high throughput approach. The most striking pathway inactivation seen in all the host restricted serovars analysed in this study, is that involving vitamin B12 biosynthesis and 1,2-propanediol (PDL) degradation.

Originally, *Salmonella* was documented to grow on PDL as a sole carbon source aerobically but not under anaerobic conditions. However, its degradation requires adenosylcobalamin (vitamin B12) in the first enzymatic step, and this is only produced anaerobically. This paradox was solved when it was discovered that using tetrathionate as an alternative electron acceptor allows anaerobic growth on PDL, making vitamin B12 biosynthesis essential for this process, and revealing that *Salmonella* therefore require 40-50 genes (including the *cbi*, *cob*, and *pdu* genes) for PDL degradation (Bobik et al. 1999; Price-Carter et al. 2001). Both the *pdu* and *cob* genes are induced by PDL, implying that PDL utilisation is the primary reason for maintaining vitamin B12 biosynthesis. Other evidence to support this comes from the evolutionary history of this metabolic interaction. It has been proposed that the ancestor of most enteric bacteria could synthesise vitamin B12 and hence degrade PDL, but that these abilities were lost in the lineage leading to *E. coli* and *Salmonella* (Roth et al. 1996). *Salmonella* however, subsequently acquired the *pdu* and *cob* genes as a single chromosomal fragment and regained the ability to synthesis vitamin B12 and degrade PDL (Roth et al. 1996). Hence, if the ability to degrade PDL provides the selective pressure for maintaining such a large number of genes, then this pressure is apparently no longer exerted upon the genomes of host-restricted serovars. Unravelling the reasons behind this may shed more light on the host-pathogen interaction.

A pathway inactivated only in the human-restricted serovars was colanic acid biosynthesis, carried out by proteins encoded in the *wca* cluster. In Typhimurium, a *wcaE* mutant was shown to be attenuated for intestinal colonisation in calves and chicks, although the authors suggested such a mutant may also be more susceptible to acid stress (Morgan et al. 2004). Biofilm formation has also been connected with colanic acid production, where evidence has been presented indicating the *wca* genes are required for biofilm formation on HEp-2 cells, mammalian tissue culture cells and in the chicken intestinal epithelium (Ledeboer and Jones 2005). It is possible then that the inability to produce colanic acid is of twofold consequence: one, Typhi and Paratyphi A have a reduced need for a capsule providing long-term protection in the environment since they are host-restricted, and two, a reduced ability to colonise the intestine would correlate with evidence that these bacteria act as 'stealth' pathogens that cross the intestinal epithelium without causing inflammation and diarrhoea.

A pathway level analysis of the effects of pseudogenes is an important step towards understanding these effects upon the interconnected network of bacterial metabolism. This study has shown the importance of transport and of regulation in determining how a metabolic phenotype may be caused by a 'pseudo' genotype. A systems-level description of *Salmonella* metabolism would be the optimum scale at which to examine the metabolic activity of host restricted and host generalist serovars. But before this can be put in place, further curation is required to associate regulators and transporters with metabolic pathways, and experimental validation of any new pathways predicted from growth upon particular substrates in the Biolog system.

3.5 Conclusions

This is the first *Salmonella*-specific pathway analysis of the effect pseudogenes have upon metabolism. By putting pseudogenes from host-restricted serovars into context, it has been shown that there are more shared characteristics between host-restricted *Salmonella* than found by comparing individual genes. Use of a high throughput system like Biolog is advantageous for large-scale metabolic phenotyping, and there are possibilities that in the future, substrate analysis could be directly linked to the relevant pathway. The importance of both regulation and transport is key and curation efforts should focus upon improving StyCyc and StmCyc in that regard. This will help to untangle the mechanisms behind the negative metabolic phenotypes that cannot be explained directly by pseudogenes related to the substrate.

4 Essential genes in Typhi and Typhimurium

4.1 Introduction

Both Typhi and Typhimurium are important human pathogens and multiple strains of both serovars have been sequenced, revealing approximately 99% similarity between orthologous coding sequences at the amino acid level (Deng et al. 2003; Holt et al. 2008; McClelland et al. 2001; Parkhill et al. 2001a). Despite this similarity, Typhi is the causative agent of typhoid fever, exclusively affecting humans while Typhimurium is a leading cause of foodborne gastroenteritis and infects a wide range of mammals and birds. Key areas of investigation have been to find explanations for how these organisms are capable of causing such different disease phenotypes in humans, and why Typhi is human-restricted.

The question of which genes are required for cellular viability is of fundamental importance to biology. Recent developments have seen estimates of the minimal gene set required for bacterial cell viability, via single gene deletion (Baba et al. 2006; de Berardinis et al. 2008), random mutagenesis, resulting in vast libraries where a single gene is represented by multiple mutants (French et al. 2008; Gallagher et al. 2007) or ordered mutagenesis to produce one mutant per gene (Liberati et al. 2006). Such studies performed in the laboratory, by their nature, identify those bacterial genes required for viability under specific laboratory growth conditions. Signature-tagged mutagenesis (STM) (Hensel et al. 1995) and transposon-site hybridisation (TraSH) (Sassetti et al. 2001) are methods that make use of hybridisation to identify genes disrupted by transposon insertions. They are also negative selection methods, whereby transposon

mutants lost under selection through a functional screen represent the genes required for that function, and have been used to identify "niche-specific" virulence genes in bacterial pathogens (reviewed in (Andrews-Polymenis et al. 2009)). Most recent bacterial transposon mutant library screens have used a few thousand transposon mutants, which represent, on average, several insertions per gene (Gallagher et al. 2007; Glass et al. 2006; Laia et al. 2009; Liberati et al. 2006; Salama et al. 2004). A more recent method, transposon mediated differential hybridisation (TMDH) (Charles and Maskell 2001), has been used to analyse approximately one million mutants to identify essential genes in *Staphylococcus aureus* (Chaudhuri et al. 2009). However, these approaches are all sub-optimal, due to inaccuracy in the estimation of the transposon insertion site from microarrays, and because some genes, especially those which are smaller, will be missed by chance.

Current knowledge regarding how Typhi and Typhimurium differ is, for the most part, concentrated upon virulence and pathogenicity factors. Prior to genome sequencing, the presence of plasmids was shown to affect *Salmonella* virulence. A 100kb plasmid, encoding the *spv* (*Salmonella* plasmid virulence) genes, is found in some Typhimurium strains and contributes significantly towards systemic infection in animal models (Gulig and Curtiss 3rd 1987; Gulig et al. 1993). Plasmids encoding similar functions have been found in many serovars of *Salmonella*, but never in Typhi, which historically was rarely found to harbour plasmids. However, since the early 1970s, plasmids of incompatibility group (Inc) HI1 have been isolated in Typhi, although these do not harbour the *spv* genes. Initially characterised based on genes encoding resistance to the first generation antibiotics used to treat typhoid (review: (Phan and Wain 2008)), there is also evidence

that Typhi strains carrying an IncHI1 plasmid present a higher bacterial load in the blood during human infection (Wain et al. 1998).

The horizontal acquisition of pathogenicity islands during the evolution of the salmonellae is also believed to have impacted upon their disease potential. *Salmonella* pathogenicity islands (SPI)-1 and -2 are common to both serovars, and are required for invasion of epithelial cells (review in (Darwin and Miller 1999)) and survival inside macrophages respectively (Ochman et al. 1996; Shea et al. 1996). SPI-7 and SPI-10, however, are unique to Typhi (with respect to Typhimurium). SPI-7 harbours putative virulence genes encoding the Vi capsular polysaccharide, a SopE effector protein and a type IVB pilus system (Pickard et al. 2003; Seth-Smith 2008). The *sef* chaperone/usher fimbrial operon is encoded on SPI-10 and has been implicated in virulence and host adaptation (Townsend et al. 2001).

However, the acquisition of virulence determinants is not the sole explanation for the differing disease phenotypes displayed in humans by Typhimurium and Typhi. Indeed, genome degradation is a feature of the Typhi genome, in common with other host-restricted serovars such as Paratyphi A (humans) and Gallinarum (chickens). In each of these serovars, pseudogenes account for 4-7% of the genome (Holt et al. 2009b; Thomson et al. 2008). Loss of gene function has occurred in *shdA*, *ratB* and *sivH* of Typhi, genes which have been shown to encode intestinal colonisation and persistence determinants in Typhimurium (Kingsley et al. 2003). Numerous sugar transport and degradation pathways have also been interrupted (Parkhill et al. 2001a), but remain intact in Typhimurium, potentially underlying the restricted host niche occupied by Typhi.

Given the close phylogenetic relationship between Typhi and Typhimurium, a whole genome approach was required to ask whether these organisms are different not only in clinical phenotype but also the base gene set required for survival. Using Illumina-based transposon directed insertion-site sequencing (TraDIS) with large mutant libraries of both Typhimurium and Typhi, we investigated whether these *Salmonella* share the same essential gene set, or whether there are differences which reflect intrinsic differences in the pathogenic niches these bacteria inhabit.

4.2 Methods

4.2.1 Strains

The Typhimurium strain used was SL3261, which contains a deletion relative to the parent strain, SL1344. The 2166bp deletion stretches from 153bp within *aroA* (normally 1284bp) to the last 42bp of *cmk*, hence forming two pseudogenes and deleting the intervening gene SL0916 completely.

The Typhi strain used was WT26 pHCM1, a derivative of the attenuated Ty2-derived strain CVD908-*htrA* which has stable deletion mutations in *aroC*, *aroD* and *htrA* (Tacket et al. 1997). WT26 (Turner et al. 2006) has a point mutation in *gyrA* conferring reduced susceptibility to fluoroquinolone antibiotics and the multiple antibiotic resistance plasmid, pHCM1, has been introduced. These additions were intended to allow the transposon mutant library to be used for fluoroquinolone resistance and plasmid studies.

4.2.2 Transposome preparation

The TraDIS transposon is a derivative of EZ-Tn5 <R6Kγori/KAN-2> (Epicentre Biotechnologies, Wisconsin, USA) with outward oriented T7 and SP6 promoters at each end, and with R6Kγori deleted. The transposon was PCR amplified using Pfu Ultra Fusion II (Stratagene, California, USA) and the following oligonucleotides:

5’-CTGTCTCTTATACACATCTCCCT

5’-CTGTCTCTTATACACATCTCTTC

The resulting amplicon was phosphorylated using polynucleotide kinase (New England Biolabs, Hitchin, UK). 400 ng of this DNA were incubated with EZ-Tn5™ transposase (Epicenter Biotechnologies) at 37°C for 1h then stored at -20 °C.

4.2.3 Cell transformation and transposon library creation

In the following section, the Typhi mutant library was generated by Keith Turner and Duy Phan, the Typhimurium library by myself.

Bacterial cells (Typhi or Typhimurium) for electrotransformation were grown in 2 x TY broth to an OD_{600} of 0.3 – 0.5, then cells were harvested and washed three times in ½ x vol 10% glycerol. Cells were finally resuspended in 1/1000 x vol 10% glycerol and stored at -80 °C. 60 µl cells were mixed with 0.2 µl transposomes and electrotransformed in a 2 mm electrode gap cuvette using a BioRad GenePulser II set to 1.4 kV, 25 µF and 200 Ω. Cells were resuspended in 1 mL SOC medium (Invitrogen, Paisley, UK) and incubated at 37 °C for 2h then spread on L-agar supplemented with kanamycin at 7.5 µg/mL and 20 µg/mL for Typhi and Typhimurium respectively. For Typhi, the L-agar was also supplemented with “aro mix” (40 µg/ml each of L-phenylalanine and L-tryptophan, and 10 µg/ml each of *p*-aminobenzoic acid and 2,3-dihydroxybenzoic acid final concentration). After incubation overnight at 37 °C, the number of colonies on several plates was estimated by counting a proportion of them, and from this the total number of colonies on all plates was estimated conservatively. Kanamycin resistant colonies were scraped off plates and resuspended in sterilised deionised water using a bacteriological spreader.

Typically, ten or more electrotransformations were performed to generate one batch of mutants. The number of mutants in each batch ranged from 42,000 to 148,000. From the estimated total number of mutants and using the OD_{600} to estimate the cell concentration in each batch, volumes containing approximately similar numbers of mutants from 13 batches were pooled to create the Typhi mutant library mixture, estimated to include 1.1 million mutants. Ten batches were similarly pooled to create the Typhimurium mutant library, estimated to include 930,000 mutants.

4.2.4 DNA manipulation and sequencing

Daniel Turner developed and helped to optimise TraDIS and performed the nucleotide sequencing for Typhi and Typhimurium. Duy Phan and I performed the sequence analysis required to optimise TraDIS.

4.2.4.1 Optimisation using Typhi

Five µg of Typhi genomic DNA from the mutant pool was fragmented to an average size of either 200 or 300 bp by Covaris AFA (Quail et al. 2008) and the Illumina DNA fragment library preparation was performed following the manufacturer's instructions, but using 1.5x the recommended reagent volumes in each step. DNA fragments from the library, ligated to Illumina adapters, were run in a 12 cm 2 % agarose gel in 1 x TBE buffer, at 6 V cm^{-1} without the preceding column clean up step. After 45 minutes, fragments corresponding to an insert size of 250-350 bp were excised, and DNA was extracted from the gel slice without heating (Quail et al. 2008). The DNA was quantified

on an Agilent DNA1000 chip, following the manufacturer's instructions. Template DNA of either 0.1 ng, 1 ng, 25 ng, 65 ng or 100 ng was used in a PCR of 32 or 22 cycles to amplify insertion sites at the 3' end or the 5' end of the transposon, using the transposon-specific forward primer 5F (Sigma-Aldrich, Dorset, UK, HPLC purified):

5'-AATGATACGGCGACCACCGAGATCTACACCTGAATTACCCTGTTATCCCTATTTAGGTGAC

or 3F (Sigma-Aldrich, HPLC purified):

5'- AATGATACGGCGACCACCGAGATCTACACCTGACCTCTAGAGTCGACTGGCAAAC

and a custom Illumina reverse primer V3.3 (Sigma-Aldrich, HPLC purified):

5'-AAGCAGAAGACGGCATACGAGATCGGTACACTCTTTCCCTACACGACGCTCTTCCGATCT

The resultant libraries were sequenced on a paired or single end Illumina flowcell using an Illumina GAII sequencer for 36 or 54 cycles of sequencing, using 2x Hybridization Buffer and the custom sequencing primer 5TMDH2seq (Sigma-Aldrich, HPLC purified):

5'- ATCCCTATTTAGGTGACACTATAGAAGAGATGTGTA

or 3TMDH1seq (Sigma-Aldrich, HPLC purified):

5' - TTATGGGTAATACGACTCACTATAGGGAGATGTGTA

The custom sequencing primers were designed such that the first 10 bp of each read was transposon sequence.

4.2.4.2 Typhimurium

Five µg of Typhimurium genomic DNA from the mutant pool was fragmented to an average size of 300 bp by Covaris AFA (Quail et al. 2008) and the Illumina DNA fragment library preparation was performed as for the Typhi libraries above. To amplify the transposon insertion sites, 22 cycles of PCR were performed using 100 ng of DNA fragment library, and the transposon-specific forward primer 5F and custom Illumina reverse primer detailed above.

The amplified library was cleaned up with a QiaQuick PCR product purification column following the manufacturer's instructions, eluted in 30 µl EB, and then quantified by qPCR (Quail et al. 2008). The amplified DNA fragment library was sequenced on a single end Illumina flowcell using an Illumina GAII sequencer, for 36 cycles of sequencing, using 2x Hybridization Buffer and the custom sequencing primer 5TMDH2 detailed above.

4.2.5 Sequence analysis

Using a custom Perl script co-written with Duy Phan (Appendix 8.3.1), the Illumina FASTQ sequence files were parsed for 100% identity at the 5' end to the last 10bp of the transposon (TAAGAGACAG). Sequence reads which matched were stripped of the transposon tag and subsequently mapped to the appropriate reference chromosome using Maq version maq-0.6.8 (Li et al. 2008). Ty2 was used as the reference for Typhi and SL1344 for Typhimurium. Precise insertion sites were determined using the output from the Maq mapview command, which gives the first nucleotide position to which each read

mapped. The number and frequency of insertions mapping to each nucleotide in the reference genome was then determined. Gene boundaries were defined from the Sanger in-house annotation of the Typhimurium SL1344 sequence, and the publically available Typhi Ty2 sequence (Accession number: AE014613) allowing the number and frequency of transposon insertions to be established for every gene. Genes were grouped into functional classes based on genome annotation (Table 4-1), and the average number of insertions per functional class was calculated by dividing the total number of insertions recovered for the class by the summed total of all gene lengths within the class.

Table 4-1 Functional categories in genome annotation

Function	Colour in genome annotation
Pathogenicity/adaptation	Dark blue*
Energy metabolism	Dark grey/black
Information transfer	Red
Membrane/surface structures	Green
Degradation of small molecules	Purple
Degradation of macromolecules	Cyan
Central/intermediary metabolism	Yellow
Unknown function	Pale green
Regulators	Light blue
Conserved/hypothetical	Orange
Pseudogenes	Brown
Phage/IS elements	Pink

Categories as annotated in the Typhi CT18 genome. *Pathogenicity/adaptation genes are usually coloured white, but for display purposes (e.g. Figure 4-2) have been coloured dark blue.

4.2.6 Statistical analyses

The nature of the statistical analysis was discussed with Leopold Parts who advised on the best way to proceed and subsequently wrote scripts in R to perform the following analyses.

4.2.6.1 High density mutagenesis

We calculated the *P*-value for the distances between insertion sites using F = G/N where G is the number of bases in the genome (4,791,961 / 4,878,012) and N is the number of unique insert sites (371,775 / 549,086) for Typhi and Typhimurium respectively. Across the whole genome, the *P*-value for at least X consecutive bases without an insert site is $e^{(-X/F)}$, giving a 5% cut-off at 39 bp and a 1% cut-off at 60 bp for Typhi. For Typhimurium, there was a 5% cut-off at 27 bp and a 1% cut-off at 41 bp.

To investigate the G+C content insertion site bias, the G+C content and number of insertion sites was calculated for a sliding window of 1 kb (with a 500 bp skip) along the Typhi genome. The average number of insertion sites for a given integer G+C content was determined, and used to calculate the average number of insertion sites in genomic regions with G+C content above and below the genome average (52%). The *P*-value for the distances between insertion sites was again calculated using F = G/N where G is the number of bases in the genomic regions with above/below average G+C content (4,357,000 / 5,225,000) and N is the average number of insert sites (615,547 / 273,576). The *P*-value for at least X consecutive bases without an insert site is again $e^{(-X/F)}$, giving a 5% cut-off at 58 bp and a 1% cut-off at 88 bp.

4.2.6.2 Essential genes per serovar

The number of insertion sites for any gene is dependent upon its length, so the values were made comparable by dividing the number of insertion sites by the gene length, giving an "insertion index" for each gene. The distribution of insertion indices was bimodal, corresponding to the essential (mode at 0) and non-essential models. We fitted gamma distributions for the two modes using the R MASS library (http://www.r-project.org). Log_2-likelihood ratios (LLRs) were calculated between the essential and non-essential models and we called a gene essential if it had an LLR of less than -2, indicating it was at least 4 times more likely according to the essential model than the non-essential model. 'Non-essential' genes were assigned for an LLR of greater than 2.

4.2.6.3 Comparison of essential genes between serovars

For every gene *g* present in both serovars, with $n_{g,A}$ reads observed in Typhi and $n_{g,B}$ reads observed in Typhimurium, we calculated the log_2 fold change ratio $S_{g,A,B} = log_2$ *\frac*$(n_{g,A}+100)(n_{g,B}+100)$. The correction of 100 reads smoothes out the high scores for genes with very low numbers of observed reads. We fitted a normal model to the mode +/- 2 sample standard deviations of the distribution of $S_{A,B}$, and calculated p-values for each gene according to the fit. We considered genes to be uniquely essential to one serovar with a *P*-value of 0.05, according to the normal model.

4.3 Results

4.3.1 Whole genome TraDIS

4.3.1.1 Optimising TraDIS

To determine the optimum sequencing parameters, initially, 32 cycles of PCR were performed on 0.1ng and 1ng of fragmented Typhi DNA, in order to guarantee sufficient template for sequencing. One PCR primer was designed to anneal to the Illumina adapter sequence, and the other was transposon-specific, but tailed to allow cluster amplification later on in the protocol. The very high yield of PCR product (>100nM,), demonstrated that fewer cycles of amplification could be performed in subsequent PCRs. This would reduce amplification bias and potential dropout of rare transposon insertion sites from the resulting sequencing library.

Separate DNA fragment libraries were prepared for sequencing off each end of the transposon, to determine the degree of concurrence, and hence evaluate the completeness of the dataset. Initially these libraries were sequenced with 36 bp reads as paired end, with one end corresponding to the transposon / genomic DNA junction, and the other located 200-300bp away, in the Typhi genomic sequence.

Analysis of the resulting sequences established that read mapping for these sequences was equally robust with single or paired end reads, therefore the extra cost of a paired end run was not justifiable and all subsequent sequencing runs were performed as single end. The analysis also confirmed that more complete coverage was obtained when 1ng template DNA was used in the PCRs, compared to 0.1ng (Table 4-2). With the aim of minimising amplification biases as far as possible, a mixture of both quantitative PCR

and regular PCR was used to establish the minimum number of PCR cycles and maximum realistic mass of template DNA that would give a workable yield of amplicon DNA for sequencing. Virtually no difference in amplification yield was observed for 65ng template DNA or above, and 22 cycles were found to be the fewest that gave a reasonable yield of amplified DNA.

Table 4-2 Validating TraDIS

	Sample	Total number of reads	Number of tagged reads (%)	Number of reads mapped to Ty2 (%)	Number of unique insert sites	Average distance between inserts (bp)
a	200bp	9108747	6978334 (77)	4248840 (61)	39103	122.5
	300bp	8678131	6139885 (71)	3686804 (60)	61250	78.2
b	200bp	6188899	4920008 (80)	2049222 (42)	168452	28.4
	300bp	6345624	5357249 (84)	2510720 (47)	225423	21.3
c	100ng	5170743	4019030 (78)	1391070 (35)	268284	17.9
	65ng	5526380	4091846 (74)	1384263 (34)	256431	18.7
	25ng	6368930	4835018 (76)	1837189 (38)	247522	19.4
d	2 lanes	9522310	7219913 (76)	2424118 (34)	312089	15.4
e	4 lanes	21417620	16146777 (75)	5645570 (35)	371775	12.9

A representative lane for each stage of refinement is shown. **(a)** 0.1ng genomic DNA samples, fragmented into 200/300bp lengths, underwent 32 PCR cycles prior to 36 cycles on sequencing. **(b)** 1ng genomic DNA samples, fragmented into 200/300bp lengths, underwent 32 PCR cycles prior to 36 cycles on sequencing. **(c)** Genomic DNA samples underwent 22 PCR cycles and 25-100ng PCR product underwent 54 cycles of sequencing. **(d)** Combined data from 2 sequencing lanes with same conditions as the 100ng sample from (c). **(e)** Combined data from 4 sequencing lanes; the samples from (d) and the 65ng and 25ng samples from (c). All samples were sequenced from the 5' end of the transposon, with the exception of samples from (a), which were sequenced from the 3' end.

To verify these results, DNA fragment libraries were prepared from both ends of the transposon, using 22 cycles of PCR and 25, 65 and 100ng of template DNA (fragmented into ~200 bp pieces) in the reaction. These libraries were sequenced in a single end run with a 54bp read length, in order to assess whether the longer reads enhanced read

mapping. The greatest number of unique insertion sites were recovered from the 100ng sample (Table 4-2), although the longer sequencing reads did not improve read mapping. Overall, the least biased sequence dataset was obtained from the use of 100ng template DNA and 22 cycles of amplification in the PCR reaction. DNA fragment libraries prepared from each end of the transposon gave highly overlapping results but sequencing two lanes of a library from one end significantly increased the number of unique mutants recovered (Table 4-2). We therefore used 100ng DNA and 22 PCR cycles in all further DNA fragment library preparations from the 5' end of the transposon, and sequenced two lanes per library.

4.3.1.2 Insertion bias

Transposon insertion site bias is often cited as a limitation of transposon mutagenesis techniques. We detected a bias towards insertion in A+T rich regions in both mutant libraries (Typhi shown in Figure 4-1), but the frequency of insertion achieved by the number of transposon mutants ensured that even G+C-rich regions contained numerous insertions. In Typhi, the library with slightly lower coverage, the average insertion rate in G+C regions above 52% (genome average for both these *Salmonella*) was 1 every 19bp, meaning that a region of >88 bp without insertion had a less than 1% probability of occurring by chance. The equivalent values in G+C regions below 52% were 1 insert every 7 bp, with a less than 1% probability that a region >33 bp without insertion occurred by chance. Given that transposon insertion sites successfully delimited essential genes, we have no reason to believe that transposon insertion bias has had any bearing upon our conclusions.

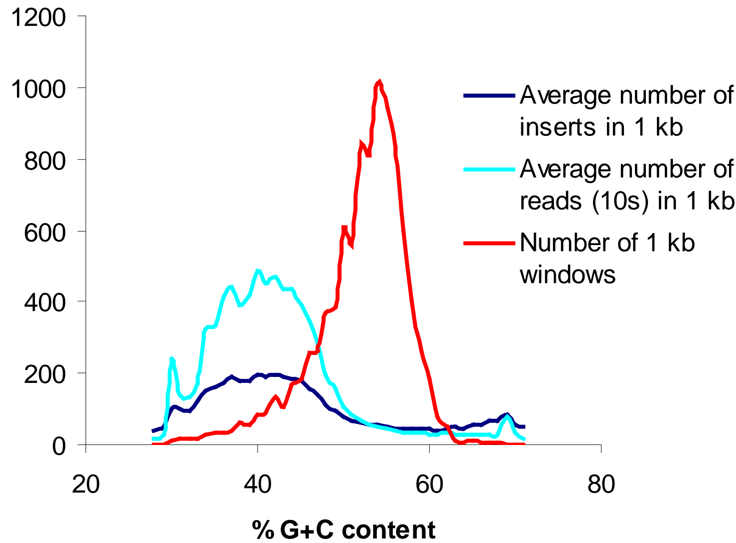


Figure 4-1 Transposon insertion site bias in Typhi

The %G+C content and number of insertions and sequence reads (in tens) was determined for every 1 kb window (with a 500 bp skip between windows) across the genome. Red line: the number of 1 kb windows with a particular %G+C content; dark blue line: the average number of transposon insertions found in 1 kb windows of a particular %G+C content; light blue line: the average number of mapped reads found in windows of a particular %G+C content, in tens. The average G+C content in *S.* Typhi is 52% but the highest number of insertion sites and mapped reads occur in windows with 40% G+C content.

4.3.1.3 Typhi

Genomic DNA was extracted from the ~1 million mutant pool for nucleotide sequencing from the transposon into the adjacent sequences of the insertion sites. Since Typhi had been used for optimisation of TraDIS, all four sequencing lanes using the optimised parameters were combined. These four lanes from the Illumina sequencing flow cell generated over 21 million nucleotide sequence reads, of which 75% included an identical match to the 10 base transposon nucleotide sequence tag (Table 4-2). Of these tagged sequence reads, 5.6 million were mapped to the Typhi Ty2 chromosome sequence. This allowed the identification of 371,775 individual transposon insertion sites; an average of 1 insertion site for every 13 bp. This represents an average of more than 80 inserts per

gene, which is far in excess of the number of insertions achieved previously for bacterial transposon mutant libraries which have reported an average of 5 to 17 inserts per gene (Gallagher et al. 2007; Laia et al. 2009; Salama et al. 2004; Sassetti et al. 2001), and makes possible the assay of every gene in the genome. The distribution of mapped sequence reads across the whole genome is shown in Figure 4-2A.

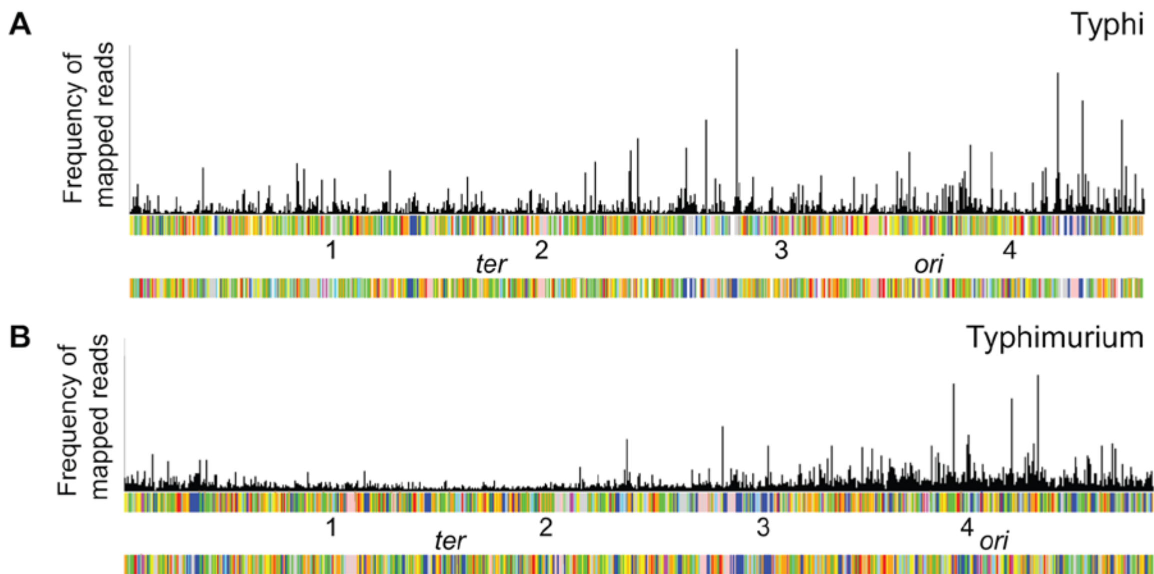


Figure 4-2 Whole genome view of transposon insertion sites

A) Frequency and distribution of transposon directed insertion-site sequence reads across A) the entire Typhi Ty2 genome and B) the entire Typhimurium SL1344 genome, scaled to heights of 2500 and 800, respectively. Numbers represent megabases; the y-axes show the number of mapped sequence reads within a window size of 3; *ori* and *ter* indicate the approximate positions of the replication origin and terminus respectively. Genes are colour-coded according to function: dark blue, pathogenicity/adaptation; black, energy metabolism; red, information transfer; dark green, membranes/surface structures; cyan, degradation of macromolecules; purple, degradation of small molecules; yellow, central/intermediary metabolism; light blue, regulators; pink, phage/IS elements; orange, conserved hypothetical; pale green, unknown function; brown, pseudogenes.

4.3.1.4 Typhimurium

Approximately 12 million sequence reads were generated from the Typhimurium sequencing run which used two lanes on the Illumina flowcell. Almost 90% of the reads contained a 100% identical match to the transposon tag and 49% of these could be mapped to the SL1344 reference sequence (N.B. 35% mapped to three plasmids present in this strain and a further 16% contained Illumina adapter sequence). Combining the two sequencing lanes meant that, in total, 549,086 unique insertion sites were recovered from the ~1 million mutant library, an average of one insertion every 9bp, or over 100 unique inserts per gene (Figure 4-2B). There is an apparent bias in the frequency of transposon insertion towards the origin of replication. This likely occurred as the bacteria were in exponential growth phase immediately prior to transformation with the transposon. In this phase of growth, multiple replication forks would have been initiated, meaning genes closer to the origin were in greater copy number and hence more likely to be a target for insertion. This is not evident in the Typhi library, but we believe this is disguised partly by a few over-represented mutants (high peaks in Figure 4-2A) and partly because Typhi has a slower growth rate than Typhimurium.

4.3.2 Essential genes in Typhi

Three hundred and fifty-six genes had an LLR of less than −2; thus these genes were considered essential for growth under standard laboratory conditions, based on the criteria given in the methods. Conversely, 4162 genes had an LLR greater than 2 and were thus non-essential. Nineteen genes had LLRs between the two cut-off values and so it was not possible to assign these as essential or non-essential with the same degree of

confidence. In addition, the density of insertions across the genome was such that a 60bp region without insertion had only a 1% chance of occurring randomly. We could not therefore make conclusions with confidence for very short genes (less than 60bp) with no insertions which may have been missed by chance. However, there were only 2 annotated genes less than 60bp long that had no mapped insertion sites. Thus, we effectively assayed all but 2 very short annotated genes, and were able to draw conclusions with statistical confidence for 4518 of 4537 (99.6%) annotated genes in the genome (Figure 4-3, Appendix 8.3.2).

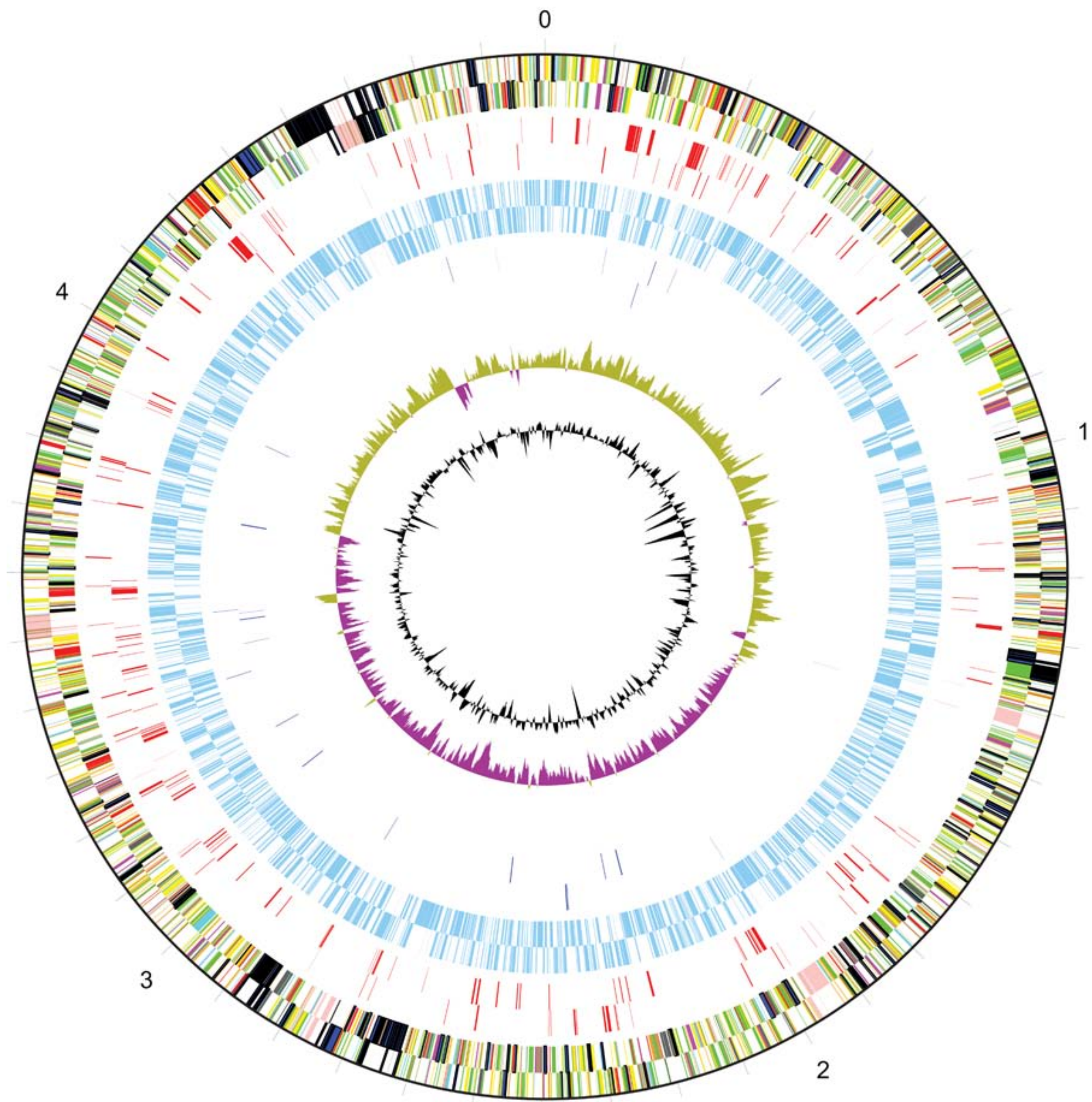


Figure 4-3 Whole genome assay of Typhi

The outer scale is marked in megabases. Circles range from 1 (outer circle) to 6 (inner circle) and represent genes on both forward and reverse strands. Circle 1, all genes; circle 2, Typhi essential genes (red); circle 3, Typhi non-essential genes (light blue); circle 4, 26 genes essential in Typhi only (dark blue); circle 5, GC bias ((G-C)/(G+C)), khaki indicates values >1; purple <1; circle 6 %G+C content. Genes in outer circle are colour-coded according to function: dark blue, pathogenicity/adaptation; black, energy metabolism; red, information transfer; dark green, membranes/surface structures; cyan, degradation of macromolecules; purple, degradation of small molecules; yellow, central/intermediary metabolism; light blue, regulators; pink, phage/IS elements; orange, conserved hypothetical; pale green, unknown function; brown, pseudogenes.

Many of the 356 essential genes are required for fundamental biological processes, including cell division, DNA replication, transcription and translation (Table 4-3). The full list is available in Appendix 8.3.3. A few are worthy of note, including DNA polymerase III, a multimeric enzyme encoded by eight subunit genes, six of which are identified as essential. The remaining two genes are *holE*, (LLR = 4.83), and *holC* (LLR = 5.36) which are unlikely to be essential. All the aminoacyl-tRNA synthetase genes were identified as candidate essential genes except for *trpS* (t4024) and *trpS* (t4557) which are both tryptophanyl-tRNA synthetases and therefore mutually redundant. Similarly, of the 11 genes that are involved in peptidoglycan biosynthesis, 9 were assigned as essential, while *ddlA* and *ddlB* were assigned as non-essential; both these genes perform the same function (Figure 4-4). Of the 356 Typhi candidate essential genes identified by TraDIS, 256 (~70%) are also essential in *Escherichia coli* (Baba et al. 2006), including 110 of the genes in Table 4-3. Of the 100 genes essential in Typhi but not *E. coli*, almost half are involved in energy metabolism or regulation of gene expression.

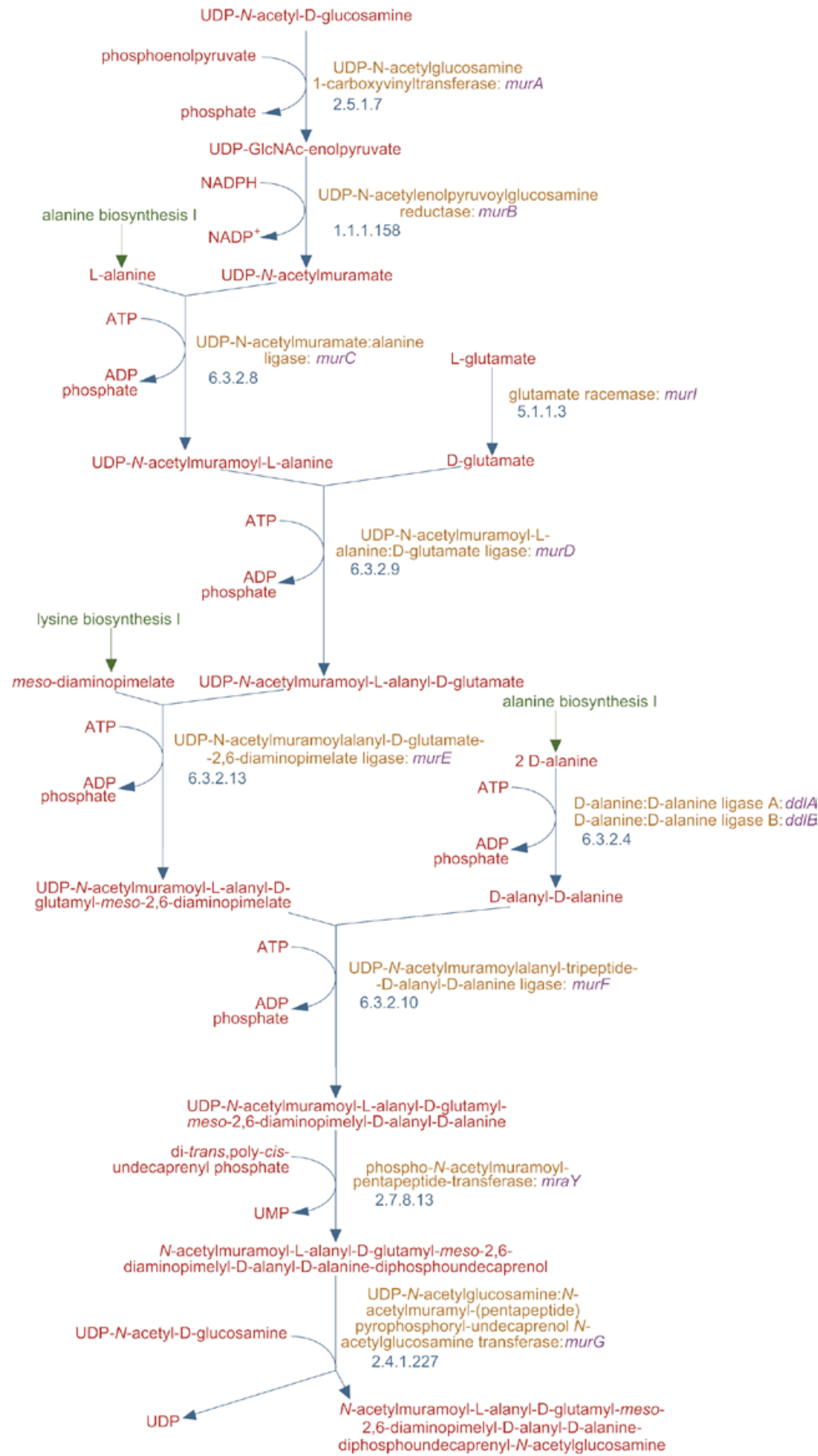


Figure 4-4 Peptidoglycan biosynthesis

All genes are essential in this pathway, except for *ddlA* and *ddlB*, which perform the same enzymatic function. Reaction direction indicated by arrows, blue numbers represent enzyme commission (E.C.) numbers, green words and arrows represent incoming substrates from other metabolic pathways.

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

Biological process	Sub-process (total number of genes)		Essential genes	Non-essential genes
Cell division		20	***ftsAH***_J_***LQWXYZ***, ***mukB***, **t0429**	***ftsNK***, *minC****DE***, *sdiA, cedA, sulA,* t3932
DNA replication	DNA Polymerase I	1	*polA*	
	DNA Polymerase II	1		*polB*
	DNA Polymerase III	8	***dnaEN***_Q_***X***, ***holAB***_D_	*holC*
	Supercoiling	4	***gyrAB***, ***par***_C_***E***	
	Primosome-associated	10	***dnaBC***_GT_, *priAB, rep,* ***ssb*****(t4161)**	*priC, ssb*(t4237)
Transcription	RNA polymerase	3	***rpoABC***	
	Sigma, elongation, anti- and termination factors	9	***rpoDE***_H_, ***nusA***_B_***G***, *rho*	*rpoNS*
Translation	tRNA-synthetases	23	***glyQ***_S_, ***hisS*** , *lysS*, ***metG***, ***pheST***, ***proS***, ***serS***, ***thrS***, ***tyrS***, ***aspS***, ***asnS***, *alaS*, ***valS***, ***leuS***, *ileS*, ***gltX***, ***glnS***, ***cysS***, ***argS***	***trpS*****(t4024)**, *trpS*(t4557)
	Ribosome components	56	***rplBCDEFJ***_K_***LMNOP QRSTUVWX***_Y_, ***rpmABCDH***_IJ_(t4086), *r****psABCDE***_F_***GHIJKLM N***_O_***PQRS***_U_	*rplAI,* *rpmE*(t3522), *rpmE*(t2391), *rpmFGJ*(t2390), *rpsT*
	Initiation, elongation and peptide chain release factors	13	***fusA, infABC, prfA***_B_, ***tsf***	*efp, prfCH, selB, tufAB*

Gene names in bold are also essential in *E. coli* (Baba et al. 2006).

The high density of insertions across the Typhi genome allows a clear demarcation between many candidate essential and non-essential genes. As an example, topoisomerase IV, an essential enzyme for maintaining DNA supercoiling, is encoded by *parC* and *parE* and almost no insertion sites were identified for these genes, or for *plsC*, a lipid biosynthesis gene (Figure 4-5).

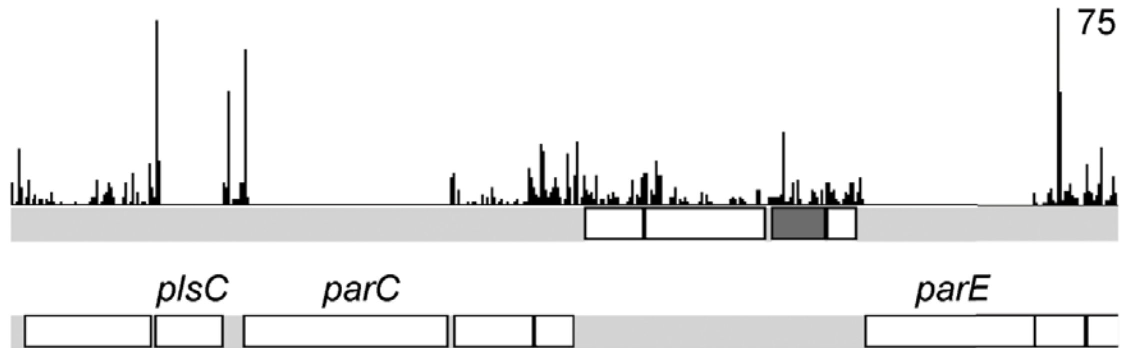


Figure 4-5 Essential genes in Typhi

Detailed plot generated using Artemis (Rutherford et al. 2000). The essential *plsC* gene and topoisomerase IV genes, *parC* and *parE*, showing the absence of transposon insertions. The maximum number of sequence reads within this plot is 75; white boxes represent genes, and grey boxes pseudogenes.

Inportantly, the genome coverage of the Typhi million mutant library is so great that insertions into small intergenic regions between essential genes such as *pyrH*, *frr* and *dxr* can also be seen clearly (Figure 4-6). This demonstrates that the insertion of this transposon is unlikely to have polar effects within operons. Elsewhere, the intergenic region between essential genes *leuS* and *rlpB* is only 14 bp but we observed 6 sequence reads mapping to 1 insertion site here without any insertions into the adjacent coding sequence.

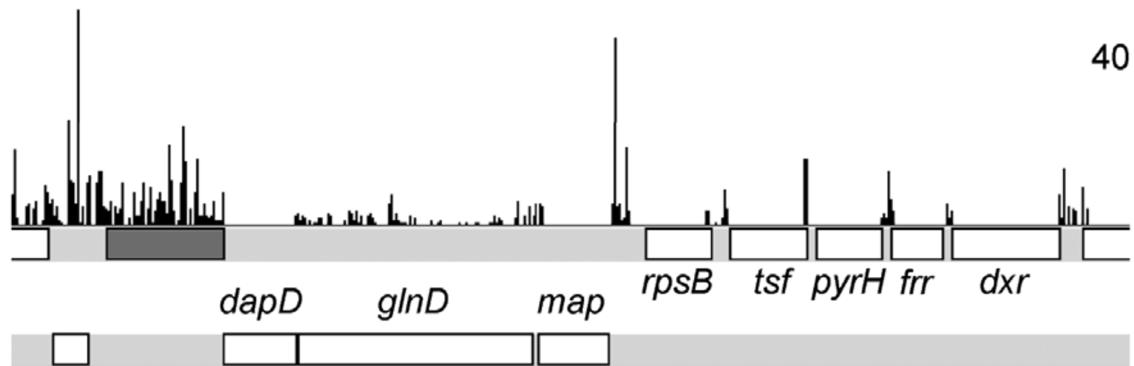


Figure 4-6 Insertions between genes

Detailed plot generated using Artemis (Rutherford et al. 2000). Sequence reads mapping to regions between essential genes. The maximum number of sequence reads within this plot is 40; white boxes represent genes, and grey boxes pseudogenes.

4.3.3 Essential genes in Typhimurium

The TraDIS analysis of the Typhimurium mutant library allowed us to identify 318 essential genes, and 4,135 non essential genes (Appendix 8.3.4 and 8.3.5). We were unable to assign 39 genes which had LLRs between -2 and 2, but all other genes contained enough insertions or were of sufficient length (>41bp) to generate credible LLR scores. Thus, every gene was assayed and we were able to draw conclusions with statistical confidence for 99.1% (4453/4492 genes) of the coding genome in a single sequencing run (Figure 4-7).

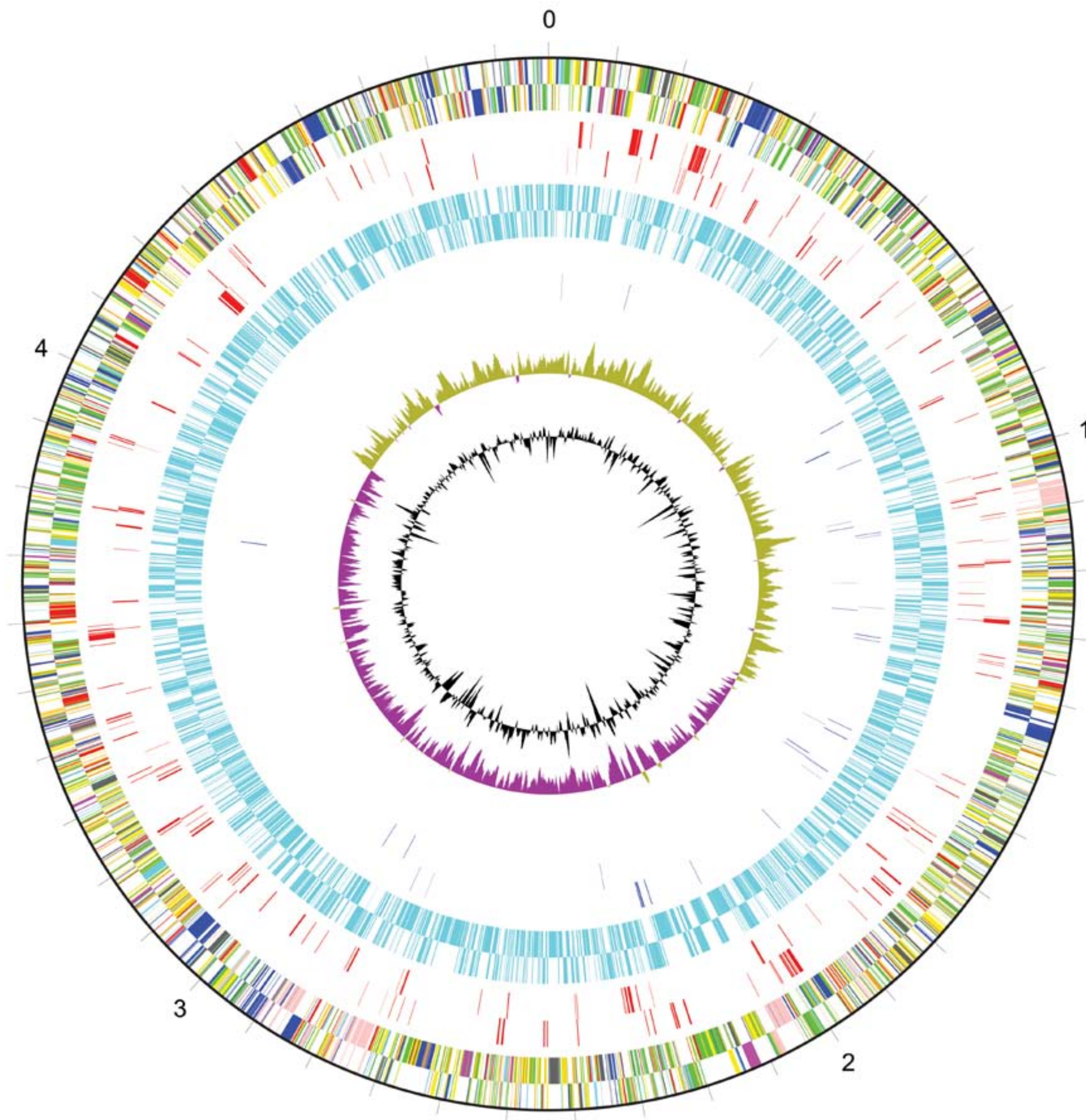


Figure 4-7 Whole genome assay of Typhimurium

The outer scale is marked in megabases. Circles range from 1 (outer circle) to 6 (inner circle) and represent genes on both forward and reverse strands. Circle 1, all genes; circle 2, Typhimurium essential genes (red); circle 3, Typhimurium non-essential genes (light blue) circle 4, 37 genes essential in Typhimurium only (dark blue); circle 5, GC bias ((G-C)/(G+C)), khaki indicates values >1; purple <1; circle 6 %G+C content. Genes in outer circle are colour-coded according to function: dark blue, pathogenicity/adaptation; black, energy metabolism; red, information transfer; dark green, membranes/surface structures; cyan, degradation of macromolecules; purple, degradation of small molecules; yellow, central/intermediary metabolism; light blue, regulators; pink, phage/IS elements; orange, conserved hypothetical; pale green, unknown function; brown, pseudogenes.

In some cases (for longer genes) we were also able to assay below the gene level to identify parts of genes that were essential. One example is *polA* where transposon insertions occurred across the majority of the 3' part of the gene but the 5' end contained no insertions (Figure 4-8). This 'essential' region corresponds to two protein domains (PF02739 and PF01367) which are involved in 5' to 3' exonuclease activity and have a role in DNA binding. Another example is *ams* (*rne*) which again only contained insertions through the 3' part of the gene. A Pfam search (Finn et al. 2007) based on the amino acid sequence of this gene predicts the presence of two domains in the preserved 5' end that are responsible for RNA-binding and cleavage. Our data suggest that it is the activity of these domains and not the whole gene that is essential for growth under laboratory conditions. It should be noted that genes with essential domains at the 3' end are unlikely to contain transposon insertions along the whole length of the gene, as insertions prior to the essential domain would interrupt translation of the required region.

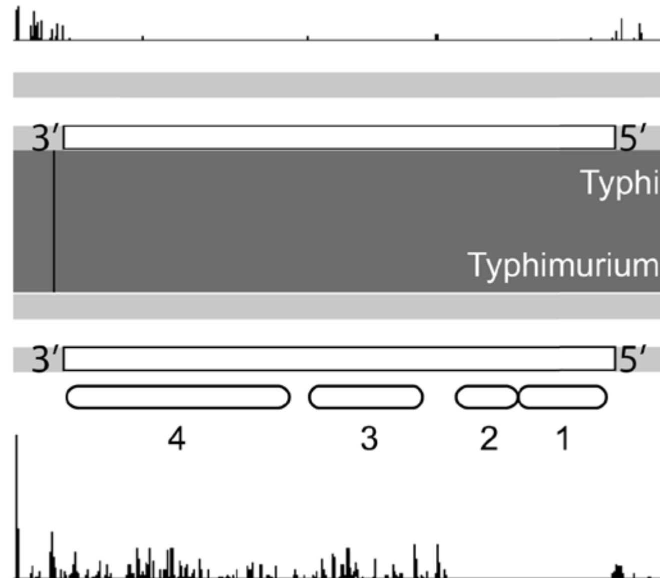


Figure 4-8 Domains in *polA* are essential

Image generated using Artemis Comparison Tool (ACT (Carver et al. 2005)) for mapped insertion sites across *polA*, scaled to a height of 50 with a window size of 3. Dark grey indicates sequence similarity. Direction of transcription is given and Pfam domains are shown in rounded boxes marked 1-4: 1, PF02739; 2, PF01367; 3, PF01612; 4, PF00476. Domains 1 and 2 do not tolerate insertion in either Typhi or Typhimurium.

4.3.4 Comparing Typhi and Typhimurium

4.3.4.1 Shared essential genes

The essential gene set of Typhi (Langridge et al. 2009b), was compared with the list generated from Typhimurium. We considered that the two serovars shared an essential gene if the gene had an LLR of < -2 in both organisms. Using this criterion, a total of 267 essential genes were shared, which represents 75% of the Typhi set (Figure 4-9A, Appendix 8.3.6). These sets were also compared with one generated in *E. coli* by systematic gene knockouts (Baba et al. 2006), which revealed a total of 226 essential

genes shared between all three, validating TraDIS as an approach for the identification of candidate essential genes (Figure 4-9B). As expected, the *Salmonella* serovars share slightly more essential genes with each other (267) than either does with *E. coli* (~258).

The majority of shared essential genes between all three bacteria are responsible for fundamental cell processes, including cell division, transcription and translation. A number of key metabolic pathways are also represented, such as fatty acid and peptidoglycan biosynthesis (Appendix 8.3.7). Interestingly, 16 genes annotated as conserved hypothetical are essential in both *Salmonella* serovars (11 shared with *E. coli*), indicating the presence of important genes whose functions in cell survival have yet to be elucidated.

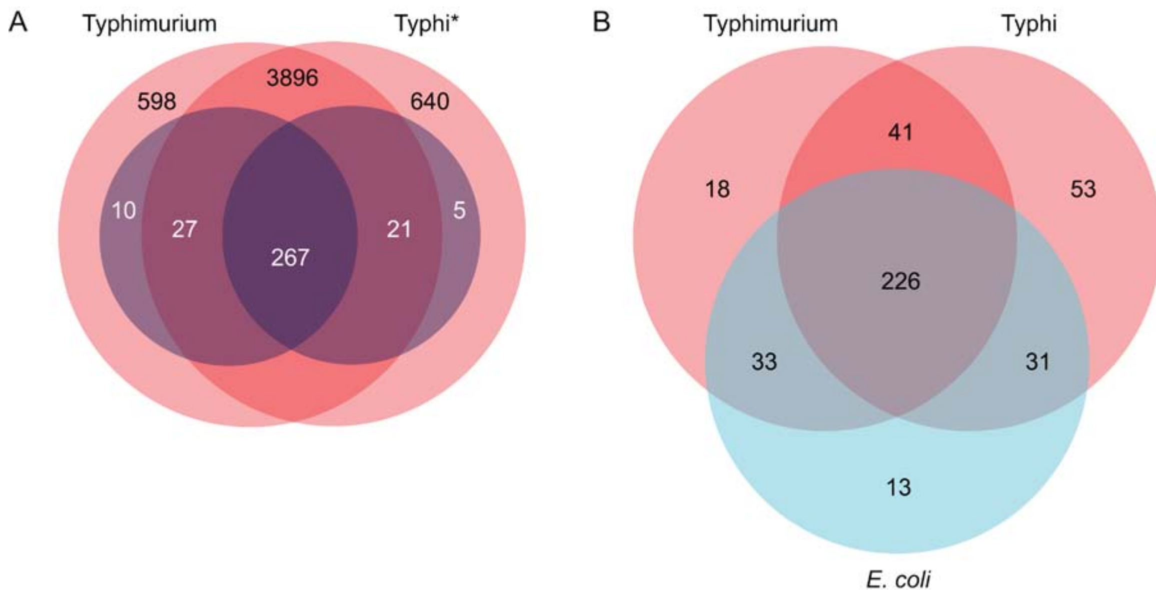


Figure 4-9 Comparison of essential genes

Venn diagrams showing (A) the overlap of all genes (red) and essential genes (blue) between Typhimurium and Typhi. Black numbers refer to all genes, white numbers to essential genes. *, details of the five Typhi essential genes omitted from the comparison are given in Appendix 8.3.8. (B) the overlap of essential genes between Typhimurium, Typhi and *E. coli*.

4.3.4.2 Essential genes only present in one serovar

Ten genes essential in Typhimurium were absent from Typhi (Table 4-4). Five of these were encoded on phage, of which three are repressors. One of the remaining phage essential genes encodes a PhoP/PhoQ regulated protein and the other is involved in natural bacterial transformation. These warrant further investigation as they are genes that have been acquired and then become essential for survival in rich media. The non-phage related essential genes included one encoding the antitoxin element of a chromosomally encoded toxin/antitoxin system. Others encoded a lipoprotein, a cation transporter and an electron transfer flavoprotein and are likely to have been lost from the Typhi genome, since they are found in many other *Salmonella* serovars.

Table 4-4 Genes uniquely essential in Typhimurium

	Ty inserts	Ty reads	SL inserts	SL reads	SL ID	SL gene length	Ty ID	Ty gene length	Name	Function
No orthologue in Typhi	-	-	18	123	SL0742	1250	-	-	-	putative cation transporter
	-	-	4	21	SL0831	836	-	-	-	putative electron transfer flavoprotein (beta subunit)
	-	-	0	0	SL0950	323	-	-	-	putative prophage protein
	-	-	11	75	SL1179	770	-	-	envF	lipoprotein
	-	-	2	4	SL1480	230	-	-	-	putative cytoplasmic protein
	-	-	1	3	SL1560	698	-	-	-	putative membrane protein
	-	-	3	27	SL1967	677	-	-	-	putative prophage protein
	-	-	3	34	SL2549	209	-	-	-	endodeoxyribonuclease
	-	-	10	146	SL2633	977	-	-	-	putative repressor protein
	-	-	0	0	SL2695	959	-	-	-	putative competence protein
Present in Typhi but essential only in Typhimurium†	22	156	16	174	SL1561	1208	t1534‡	122	sseJ	*Salmonella* translocated effector protein (SseJ)
	-	-	4	149	SL2593	449	STY2066*	-	-	putative DNA-binding protein
	33	463	5	26	SL0032	422	t0033	287	-	putative transcriptional regulator
	68	325	9	36	SL0623	623	t2232	557	lipB	lipoate-protein ligase B
	147	3451	10	64	SL0702	878	t2156	875	-	putative glycosyl transferase
	188	2959	9	61	SL0703	1115	t2155	1115	-	galactosyltransferase
	230	3478	15	67	SL0706	1760	t2152	1761	-	putative glycosyltransferase. From NCBI STM gene
	84	1041	2	4	SL0707	815	t2151	815	-	conserved hypothetical protein
	46	361	13	69	SL0722	1550	t2136	1550	cydA	cytochrome d ubiquinol oxidase subunit I
	73	1604	5	22	SL1069	674	t1789	674	-	putative secreted protein
	17	182	1	1	SL1203	131	t1146	137	-	hypothetical protein
	18	286	1	5	SL1264	296	t1209	296	-	putative membrane protein
	35	305	2	5	SL1341	209	t1275	209	ssaH	putative pathogenicity island protein
	44	387	1	3	SL1342	230	t1276	230	ssaI	putative pathogenicity island protein
	142	3178	5	14	SL1343	731	t1277	731	ssaJ	putative pathogenicity island lipoprotein
	70	747	4	44	SL1355	761	t1289	761	ssaT	putative type III secretion protein
	81	708	6	35	SL1532	932	t1511	932	sifB	putative virulence effector protein
	118	1635	10	44	SL1563	743	t1536	743	-	putative periplasmic amino acid-binding protein
	107	2440	5	44	SL1564	629	t1537	629	-	putative ABC amino acid transporter permease
	181	1562	19	92	SL1628	1355	t1612	1364	-	hypothetical protein
	23	177	1	5	SL1659	164	t1640	164	-	conserved hypothetical protein
	35	269	3	9	SL1785	377	t1022	377	-	conserved hypothetical protein
	164	2808	9	27	SL1793	896	t1016	896	pagO	inner membrane protein
	23	155	1	4	SL1823	953	t0988	953	msbB	lipid A acyltransferase
	55	338	10	57	SL2064	983	t0786	983	rfbV	putative glycosyl transferase
	82	483	6	58	SL2065	1274	t0785	1280	rfbX	putative O-antigen transporter
	40	195	4	11	SL3828	1811	t3658	1811	glmS	glucosamine-fructose-6-phosphate aminotransferase

SL, Typhimurium; Ty, Typhi; †*P*-value (associated with log2 read ratio) < 0.05; ‡ *sseJ* is a pseudogene in Typhi; *This gene is not present in Typhi Ty2 but is in CT18 so the STY identifier is given.

Five essential genes were only present in Typhi (Table 4-5) of which four were phage-related, including two phage repressors. The other two phage-related genes encode a glycosyl transferase and a putative DNA repair protein. One essential Typhimurium glycosyl transferase is orthologous to a pseudogene in Typhi, suggesting that the Typhi phage glycosyl transferase is acting as a functional replacement. Typhi also contains a pseudogene for *priC*, whose gene product normally interacts with RecA. The essential phage DNA repair protein is predicted to interact with RecA also, again suggesting some overlap of function. The remaining essential gene present only in Typhi is predicted to encode a secreted protein and is of interest as genomic comparisons with other *Salmonella* serovars indicate that only Paratyphi A, another human-restricted serovar, contains this gene.

Table 4-5 Genes uniquely essential in Typhi

	SL inserts	SL reads	Ty inserts	Ty reads	Ty ID	Ty gene length	SL ID	SL gene length	Name	Function
No orthologue in Tm	-	-	0	0	t1378	212	-	-	-	hypothetical protein
	-	-	1	2	t1920	386	-	-	-	putative DNA-binding protein
	-	-	3	70	t3402	551	-	-	cI	repressor protein
	-	-	3	45	t3415	722	-	-	-	hypothetical protein
	-	-	1	6	t4531	131	-	-	-	hypothetical secreted protein
Present in Typhimurium but essential only in Typhi†	43	493	3	22	t0123	440	SL0119	440	yabB	conserved hypothetical protein
	117	571	11	32	t0203	1262	SL0203	1262	hemL	glutamate-1-semialdehyde 2,1-aminomutase
	122	965	1	1	t0224	1334	SL0224	1334	yaeL	putative membrane protein
	65	446	1	12	t0270	557	SL2604	557	rpoE	RNA polymerase sigma-E factor
	139	757	0	0	t0587	2267	SL2246	2267	nrdA	ribonucleoside-diphosphate reductase 1 alpha chain
	113	641	14	38	t2140	2783	SL0718	2783	sucA	2-oxoglutarate dehydrogenase E1 component
	112	711	10	16	t2177	1622	SL0680	1622	pgm	phosphoglucomutase
	75	694	10	36	t2274	938	SL0582	938	fepB	ferrienterobactin-binding periplasmic protein precursor
	80	542	8	13	t2276	989	SL0580	989	fepD	ferric enterobactin transport protein FepD
	93	591	2	2	t2277	971	SL0579	971	fepG	ferric enterobactin transport protein FepG
	64	508	4	4	t2278	776	SL0578	776	fepC	ferric enterobactin transport ATP-binding protein FepC
	198	1116	12	116	t2410	2336	SL0444	2336	lon	Lon protease
	94	504	7	16	t2730	1043	SL2809	1043	recA	recA protein
	131	699	13	29	t2996	1973	SL3052	1928	tktA	transketolase
	76	358	3	9	t3120	1415	SL3173	1415	rfaE	ADP-heptose synthase
	211	1928	6	50	t3265	1052	SL3321	1052	degS	serine protease
	41	405	3	10	t3326	587	SL3925	587	yigP	conserved hypothetical protein
	121	557	16	34	t3384	2006	SL3872	2006	rep	ATP-dependent DNA helicase
	172	1194	5	18	t3621	2768	SL3947	2768	polA	DNA polymerase I
	116	775	9	13	t3808	1028	SL3677	1028	waaF	ADP-heptose-LPS heptosyltransferase II
	138	1082	8	29	t4411	932	SL4294	932	miaA	tRNA delta-2-isopentenylpyrophosphate transferase

SL/ Tm, Typhimurium; Ty, Typhi; †*P*-value (associated with $\log_2$ read ratio) < 0.05.

Forty one essential genes in Typhimurium were present as orthologues in both serovars; this number was 79 for Typhi. The cutoff for essentiality was placed at an LLR < -2, meaning there was the possibility that a gene with an LLR of < -2 in one serovar had an LLR just above this threshold in the other. Hence, we also calculated $\log_2$ read ratios for each essential gene to establish whether the number of mapped reads (frequency of transposon insertions) per gene was significantly different between serovars. The number of mapped reads acts as a proxy for the frequency of transposon insertion in a gene, so the $\log_2$ read ratio indicated whether a gene had greater or fewer insertions in Typhimurium or Typhi.

Using these ratios, we found genes that appeared essential in one serovar (i.e. LLR < -2) and as unassigned/non-essential in the other (LLR > -2) but did not have a significantly different frequencies of transposon insertion, according to our cutoff ($P < 0.05$). We termed these genes "putative" essential genes, of which there are 14 and 58 for Typhimurium and Typhi respectively (Appendix 8.3.9).

4.3.4.3 Genes essential in Typhi only

Twenty-one essential Typhi genes had a significantly lower frequency of transposon insertion compared to orthologues in Typhimurium ($P < 0.05$), including two encoding conserved hypothetical proteins. The *fepBDGC* operon (Figure 4-10) was essential only in Typhi, indicating an apparent requirement for ferric (Fe(III)) rather than ferrous (Fe(II)) iron. These genes function to recover ferric enterobactin from the periplasm by acting with two other proteins known to aid the passage of this siderophore through the outer membrane. FepA is the outer membrane receptor for ferric enterobactin and

provides a gated pore which is activated in the presence of TonB. While neither *fepA* (LLR = 27.7) nor *tonB* (LLR = 7) were found to be essential using TraDIS, it is probable that when Typhi enters the bloodstream, these two genes are then required for uptake of ferric enterobactin.

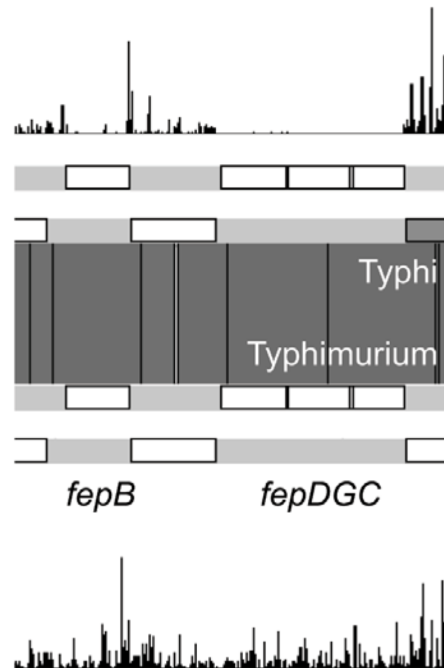


Figure 4-10 Genes uniquely essential in Typhi

ACT view of frequency and distribution of mapped insertion sites across *fepBDCG*. Dark grey blocks indicate sequence similarity. Scaled to a height of 50 with a window size of 3. White boxes indicate genes, grey boxes pseudogenes.

4.3.4.4 Genes essential in Typhimurium only

Twenty seven essential Typhimurium genes had a significantly lower frequency of transposon insertion compared to the equivalent genes in Typhi ($P < 0.05$), including five encoding hypothetical proteins (Table 4-4). This indicates that these gene products play a vital role in Typhimurium but not in Typhi when grown under laboratory conditions. One

protein of note is encoded by SL0706, which is a pseudogene in Typhi (Ty2 unique ID: t2152) due to a 1bp deletion at codon 62 that causes a frameshift (Figure 4-11). SL0706 is predicted to encode a glycosyl transferase and in total, three glycosyl transferases (SL0702, *wbaV*) and one galactosyl transferase (SL0703) are essential in Typhimurium. However, only one enzyme of this type is essential in Typhi, suggesting that surface structure biogenesis is of greater importance in Typhimurium. This is possibly because Typhi expresses the cell-surface Vi antigen and so selection of surface structures may be less intense.

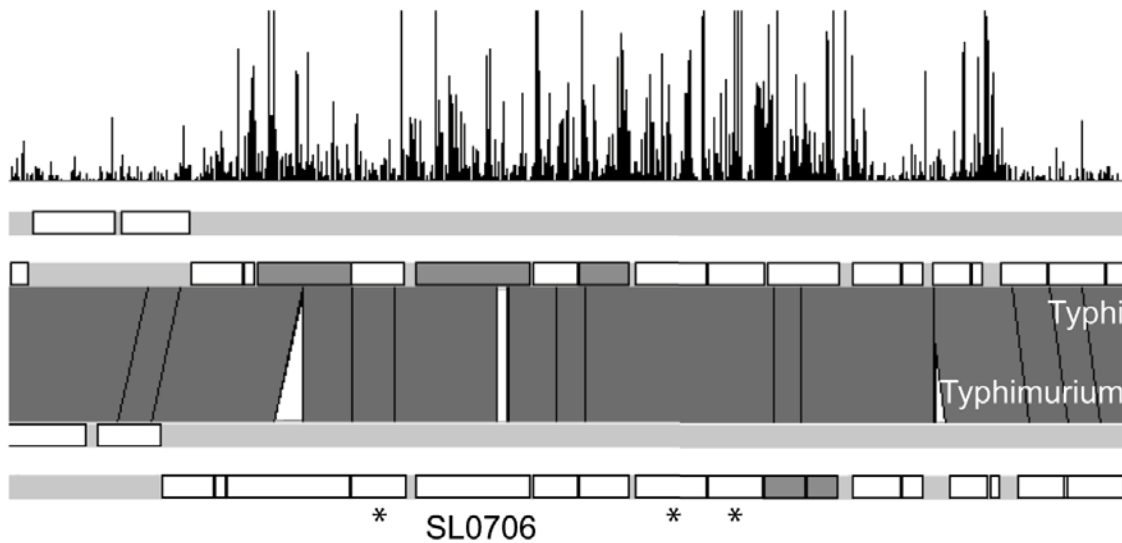


Figure 4-11 Gene uniquely essential in Typhimurium

ACT view of frequency and distribution of mapped insertion sites across the genomic region surrounding SL0706. Scaled to a height of 50 with a window size of 3. White boxes indicate genes, grey boxes pseudogenes. * These genes are also uniquely essential in Typhimurium.

We also identified four genes from SPI-2 that appear uniquely essential in Typhimurium under laboratory conditions (Figure 4-12). These genes (*ssaHIJT*) are thought to encode structural components of the SPI-2 type III secretion system apparatus (T3SS) (Kuhle and Hensel 2004). In addition, the effector genes *sseJ* and *sifB*, whose products are secreted through the SPI-2 type 3 secretion system (T3SS) (Freeman et al. 2003; Miao and Miller 2000), were also found to be uniquely essential in Typhimurium.

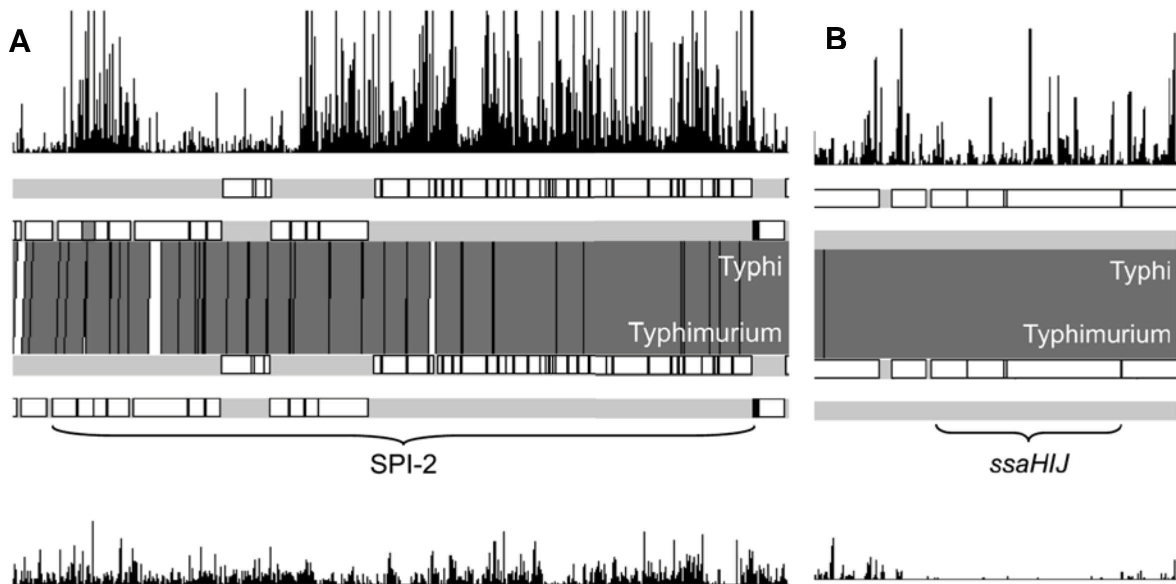


Figure 4-12 SPI-2 genes uniquely essential in Typhimurium

ACT view of frequency and distribution of mapped insertion sites across A) SPI-2 and B) *ssaHIJ*. Scaled to a height of 50 with a window size of 3. White boxes indicate genes, grey boxes pseudogenes.

All of these genes display high A+T nucleotide sequence and have been previously shown (in Typhimurium) to be strongly bound by the nucleoid associated protein HN-S, encoded by *hns* (Lucchini et al. 2006; Navarre et al. 2006). Therefore, rather than being essential, it is instead possible that access for the transposon was sufficiently restricted

that very few insertions occurred at these sites. Indeed, the generation of null Typhimurium mutants in *sseJ* and *sifB*, as well as many others generated at the SPI-2 locus suggest that these genes are not truly essential in this serovar (Freeman et al. 2003; Hensel et al. 1997b; Hensel et al. 1998; Ohlson et al. 2005). While this is a reminder that the interpretation of why a gene is essential needs to made with care, the effect of HN-S upon transposon insertion is not genome-wide. If this were the case, there would be an under-representation of transposon mutants in high A+T regions (known for HN-S binding), which is not what was observed. In total, only 15 candidate essential genes fall into the '*hns*-repressed' category described by Navarre and colleagues (Navarre et al. 2006), the remainder (almost 400) contained sufficient transposon insertions to conclude they were non-essential. In addition, we noted that all SPI-1 genes which encode another Type III secretion system and are of high A+T content were found to be non-essential.

4.3.4.5 Functional classification of essential genes

A functional breakdown of the shared and the serovar-specific essential genes is shown in Figure 4-13. The pattern of function essential in both serovars is heavily skewed towards the fundamental biological processes, described above, of energy metabolism, regulation and synthesis of membrane/surface structures. Nonetheless, the distribution of function required individually by Typhi and Typhimurium, while similar in some respects, also reveals some stark differences. Proportionally, regulatory genes and again those for synthesis of membrane/surface structures are equally represented. However, Typhi requires relatively more energy metabolism genes, and in fact has representatives in every functional category. This may partly be a reflection of the loss of overlapping functions in

the Typhi genome due to the presence of over 200 pseudogenes. For example, if two genes originally shared an essential function and one of them became a pseudogene, the other would then become essential. Conversely, almost 20% of the genes essential only in Typhimurium are of 'unknown' function indicating the presence of extremely important genes whose precise role in cellular viability remains to be elucidated.

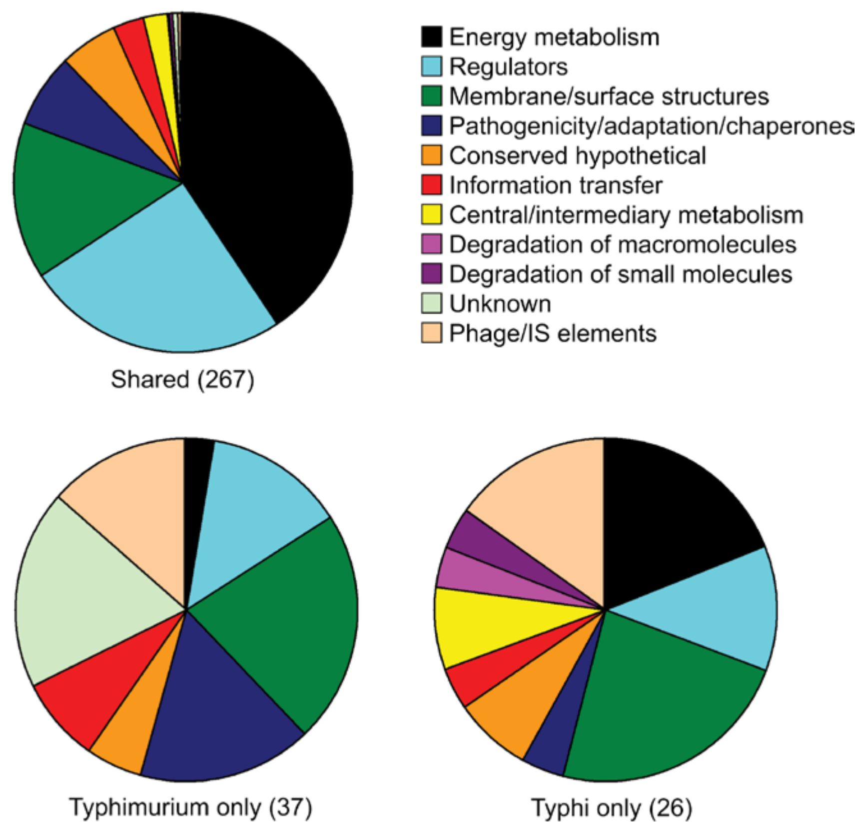


Figure 4-13 Essential genes classified by function

Functional classification for shared (267), Typhimurium only (37) and Typhi only (26) essential genes.

4.4 Discussion

4.4.1 Improvement over microarrays

A variety of previous methods has identified a number of essential and niche-specific genes, but to do this effectively on a genome-wide scale has required the use of microarrays to indirectly assay the sites of transposon insertion. Microarrays have their drawbacks: resolution is limited, and distinguishing a positive from a negative signal for some microarray features can be difficult. With sequencing, the signal is of a "digital" nature; any sequence read that has the 10bp transposon tag with adjacent genomic sequence is almost certainly an indication of the exact position of a transposon insertion site.

The combination of extremely large transposon mutant pools and high-throughput Illumina sequencing from the transposon insertion sites has brought an unparalleled degree of resolution to a transposon mutagenesis screen. Indeed, the number of insertions in the Typhimurium library was sufficiently great that gaps between insertion sites of 27bp had a less than 5% probability of occurring by chance, indicating the resolution available from this approach. This has allowed us to distinguish between essential and non-essential genomic regions to within a few base pairs and to confidently assign over 99% of the genes in both *Salmonella* genomes as essential or non-essential. In addition, there are sufficient insertions to allow the assay of nearly every gene in the genome for a particular growth condition; only small genes, with few or no transposon insertions, cannot be assayed. Thus, TraDIS can be used for the accurate estimation of minimal gene sets, and as a very effective negative selection method.

4.4.2 Assaying short regions

The ability to assay over 99% of the coding genome has implications beyond determining the minimal gene set. When a 60 bp region of the genome that does not contain insertions is statistically significant, it becomes possible to assay some functional domains encoded within genes for their contribution to cell survival. Here we have demonstrated an example of two protein domains in Typhimurium *polA* that produce no viable insertions, while insertions are found in two other domains.

The average level of transposon insertion in the Typhi and Typhimurium libraries of 1 every 10-20 bp also has implications for small RNAs. Initially, a non-coding 60 bp region without insertion is significant, both statistically and because it may identify an essential small RNA. In addition, given that there is a set of known small RNAs in *Salmonella* (Perkins et al. 2009), when these transposon libraries are used in biological screens, many of these will have been assayed for their response to that screen, which may contribute towards understanding of small RNA function. While this is not within the scope of this work, this kind of analysis is part of future efforts aimed at making full use of the transposon libraries as scientific resources.

4.4.3 Essential prophage genes

Many of the essential genes present in only one serovar encoded phage repressors. Repressors maintain the lysogenic state of the prophage, preventing transcription of early lytic genes (Echols and Green 1971). Transposon insertions into these genes will relieve

this repression and trigger the lytic cycle, resulting in cell death, and consequently mutants are not represented in the sequenced library. This again questions the definition of 'essential' genes; such repressors may not be required for cellular viability in the traditional sense, but once present in these genomes, their maintenance is required for continued survival.

4.4.4 A+T-rich islands protected in Typhimurium

In *Salmonella*, high A+T content is a hallmark of horizontal acquisition, and Tn5 inserts preferentially into such DNA. However, there are sites in Typhimurium where horizontally acquired A+T rich regions showed no increase in insertion frequency. The SPI-2 pathogenicity island and the region surrounding SL0702-3 and SL0706-7 both have average A+T contents of ~53% (compared to a genome average of 48%), and showed an increased frequency of transposon insertion in Typhi. In contrast, the average insertion frequency for these sites was similar to that of the surrounding chromosome in Typhimurium. This indicates a potential Typhimurium-specific mechanism that partially protects some A+T-rich regions from frequent transposon insertion.

Assuming a single sequencing read per transposon mutant, the average frequency of insertion in a single gene residing in the 15kb regions surrounding SPI-2 was 250-350 in both serovars. For the 44 genes in SPI-2, the average was increased over 5-fold in Typhi to ~1900 per gene, as expected in an A+T-rich region. However, this average remained at ~250 per gene in Typhimurium. The nucleoid-associated protein HN-S has been implicated in binding A+T-rich DNA in Gram negative bacteria, and virulence loci in particular have been demonstrated to be repressed by HN-S in Typhimurium (Navarre et

al. 2006). It is possible that the presence of HN-S affects the ability of the transposon to integrate into chromosomal DNA, and that *ssaHIJT* represents an extreme case as the number of insertion sites mapped to all other Typhimurium SPI-2 genes was sufficient to assign them as non-essential.

4.4.5 Specific genes required by Typhi

Our data indicate that in Typhi, *recA* is a candidate essential gene (LLR =11.5). Mutants of *recA* exist in *E. coli*, suggesting that it is not an essential gene in this bacterium (Baba et al. 2006). However, in support of the TraDIS data, multiple attempts in our laboratory to generate a *recA* mutant in Typhi, using the suicide vector allelic-exchange method (Turner et al. 2006), have failed (Appendix 8.3.10). During bacterial growth, RecA is involved in DNA replication and the re-activation of stalled replication forks. This occurs via the ‘restart’ primosome, a multimeric enzyme complex made up of 7 proteins encoded by *dnaTBCG* and *priABC* (Sandler and Marians 2000). In *E. coli*, *priC* mutants have little phenotypic effect on growth (Sandler et al. 1999) and in Typhi *priC* is a pseudogene (Parkhill et al. 2001a). However, without *priC*, there is only a *priA*-dependent pathway for replication fork restart and our results suggest that in this background, a *recA* mutant is not viable.

The *fepBDGC* operon, responsible for Fe(III) uptake, is also essential only in Typhi. Fe(III) is present in the mammalian bloodstream, where Typhi can be found during systemic human infection. Both Typhi and Typhimurium encode four transport systems for the uptake of Fe(III), chelated to different siderophores; the Fep system uses a self-encoded siderophore, enterobactin, synthesised by the *entAF* operon (Earhart 1996;

Hantke et al. 2003; Zhou et al. 1999). Fe(III) is transported into the bacterial cell in the form of ferric enterobactin by the TonB-dependent Fep system. During host adaptation, Typhi has accumulated pseudogenes in other iron chelating systems, presumably because they are not necessary for survival in the niche Typhi occupies in the human host.

In contrast, Typhimurium causes intestinal rather than systemic infection, suggesting that a mechanism for obtaining Fe(III), the only form of iron present in the blood, is not a requirement. Instead, FeoAB is more advantageous for Typhimurium (transposon insertion frequency across the *feoAB* operon is much reduced compared to the flanking genes), which encodes a high affinity system for the uptake of Fe(II), a soluble form of iron present under anaerobic conditions such as those found in the intestine (Tsolis et al. 1996).

4.5 Conclusions

Essential genes have been recognised previously using methods screening up to 17 mutants per gene (Laia et al. 2009; Sassetti et al. 2001). The extremely high resolution of TraDIS has allowed us to assay every gene for essentiality in two very closely related salmonellae with different host ranges. High density transposon mutagenesis screens such as ours produce gene lists that must be interpreted in the context under which they were performed. We found, under laboratory conditions, that 48 genes present in both serovars were essential in only one, suggesting that identical gene products do not necessarily have the same phenotypic effects in the two different serovar backgrounds. Predicting the phenotype associated with a gene must therefore be serovar-specific.

The generation of two large-scale transposon mutant libraries has created a valuable biological resource. These libraries are suited to use in high-throughput functional studies and are currently being used in a number of biological screens, including antibiotic resistance, quorum sensing and serum killing. The ability to assay over 99% of each genome also lends itself well to the possibility of screening these libraries through eukaryotic cell infection, a vital stage in the infective process of both serovars.

5 Assaying large *Salmonella* transposon mutant libraries in human macrophages

5.1 Introduction

During human infection, Typhi traverses the gut epithelium and then invades, or is taken up by macrophages. Entry and survival inside both epithelial cells and macrophages are therefore key stages in the successful establishment of infection by Typhi. However, Typhimurium infection rarely progresses systemically in humans and so does not typically encounter macrophages, but this may be becoming a more common occurrence with some strains, particularly in Africa (Kingsley et al. 2009).

The growth conditions encountered by *Salmonella* in the macrophage have been the subject of studies performed using both serovars. Typhimurium has been demonstrated to generate and maintain *Salmonella*-containing vacuoles (SCVs) in this cell type, but data on Typhi is limited and largely extrapolated from Typhimurium (Drecktrah et al. 2007; Faucher et al. 2006; Oh et al. 1996). The mechanism of cell entry is also less than clear, with active invasion and uptake by the host both possible in this phagocytic cell type.

Successful Typhimurium infection of the murine host requires the ability to synthesise aromatic amino acids and nucleotides (Hautefort et al. 2008). Amino acids and purines appear to be limiting because the relevant mutant strains are impaired for intracellular survival and virulence in mice (Muñoz-Elías and McKinney 2006). Typhimurium SCVs of both macrophages and epithelial cells have been reported as limiting for aromatic amino acids, purines and pyrimidines (Hautefort et al. 2008). Proteomics from Typhimurium at various stages post infection of murine macrophages showed that the

physiological functions of recovered proteins were related to housekeeping roles, involving biosynthesis and metabolism of amino acids, carbohydrates, lipids, proteins and nucleotides, energy production and cellular processes (Shi et al. 2006). Genes encoding the transport of gluconate, glucoronate/galacturonate and galactonate have also been shown to be upregulated, alongside genes involved in the interconversion of these sugars to pyruvate and glyceraldehyde-3-phosphate (Eriksson et al. 2003). The Entner-Doudoroff pathway may also be acting as a rich source of NADPH for use in biosynthetic pathways and redox cycling (Eriksson et al. 2003). An overall picture has been uncovered, suggesting that the Typhimurium environment inside the macrophage is aerobic, low in phosphate, high in potassium, rich in amino acids and that gluconate and related carbohydrates may be the principal source of carbon for growth (Thompson et al. 2006).

Interestingly, results from studies on Typhi and human macrophages describe a slightly different story. One microarray study suggests iron is not limiting in the Typhi SCV and it may not be limiting for manganese or phosphate either, conclusions that were inferred from the non-induction of *sitABCD*, *phoN* and *pstAB* respectively (Faucher et al. 2006). Also, Typhi, but not Typhimurium, upregulates *aceA*, isocitrate lyase, which is involved in the metabolism of fatty acids and acetate as carbon sources via the glyoxylate shunt (Muñoz-Elías and McKinney 2006). This is probably because Typhi lacks the *dgo* operon involved in gluconate utilisation, and the associated transport proteins GntT and GntU were not up-regulated during intracellular growth. Other sugar transport systems (including fructose and hexose) were also repressed – suggesting that Typhi uses different carbon sources inside macrophages to Typhimurium (Faucher et al. 2006).

The environmental differences observed thus far in Typhi and Typhimurium macrophage conditions may be due in part to the different cell lines used. Hence, to ensure observed differences are genuine, both serovars need to be tested in the same cell line.

In this chapter, an assay of the genes required by Typhi and Typhimurium for infection of human macrophages was performed in a high throughput manner by making use of the one million mutant transposon libraries and Illumina sequencing. This picture was interpreted further by placing the relevant genes into a metabolic context, using StyCyc and StmCyc as references. Human macrophages represent a biologically relevant culture condition for Typhi, and by comparing the results with Typhimurium, gives further insight into the different strategies employed by these pathogens for adherence to and infection of this host cell type.

5.2 *Methods*

All reagents were obtained from Sigma-Aldrich, Dorset, UK unless otherwise stated.

5.2.1 Strains and cell lines

Assays were carried out using Typhi WT174 and Typhimurium SL3261, and the transposon mutant libraries created in these strains, previously described in Chapter 4. Human monocytic cell line THP-1 was used for cell infections.

5.2.2 Growth in RPMI

Typhi WT174 and Typhimurium SL3261 were tested for growth in RPMI-1640 supplemented with 0.3 g/L L-glutamine only, and in RPMI supplemented with 0.3 g/L L-glutamine and buffered with 0.1 M MOPS (final concentration). Cultures of 4.5 mL were grown for 16 h shaking at 37 ^{o}C before the OD_{600} was determined on a spectrophotometer, using non-inoculated RPMI as a control.

Three independent growth curves per strain were also determined using RPMI supplemented with L-glutamine and buffered with 0.1M MOPS, in a culture volume of 100 mL. The OD_{600} was measured at 18-20 timepoints over an 8 h period and at a final timepoint of 24 h.

5.2.3 Preparation of THP-1 cells

THP-1 cells were grown up from frozen stocks in RPMI-1640 supplemented with 10% heat-inactivated foetal bovine serum and 2 mM L-glutamine, and incubated without shaking in vented flasks (VWR, Lutterworth, UK) at 37 ^{o}C in the presence of 5% CO_2. To facilitate growth, culture volumes were split and given fresh media every 3-4 days until the required cell density was reached. Typically, a starter culture of 5 mL was expanded to 30 mL over a week, then split into 3 flasks with each expanded to 60 mL over four days. These were then split into 5 flasks and each finally expanded to 100 mL. Cell density was measured using a 10 µL aliquot of culture mixed with 10 µL Trypan Blue stain pipetted into a chamber on a C-Chip haemocytometer (Labtech International, Ringmer, UK). Visualised under a light microscope, live cells (which exclude Trypan Blue) were manually counted and the density within each flask was estimated. Between one and five 6-well plates were used for each timepoint of the infection assay. The appropriate volume of culture for the entire assay, corresponding to 2 x 10^6 cells per well, was transferred into 50 mL tubes for centrifugation at 1200 rpm for 8 minutes. The supernatant was poured off and cells were resuspended in 2 mL x number of wells required of warmed, supplemented RPMI. PMA (phorbol myristate acetate) was used to differentiate the THP-1 monocytes; frozen 10 µL aliquots of 50 ng/mL PMA were thawed and mixed with 990 µL warmed, supplemented RPMI. Five hundred µL aliquots were then added to 49.5 mL aliquots of warmed, supplemented RPMI. A volume equivalent to 2 mL x number of wells required was mixed with the cell culture (i.e. a 1 in 2 dilution) and 4 mL cells were seeded into each well required for the infection assay.

The 6-well plates were incubated for six days at 37 ^{o}C in 5% CO_2. Upon the day of infection, the PMA-containing media was removed, cells were washed with dPBS and fresh warmed, supplemented RPMI was added to maintain the cells whilst the bacterial inoculum was prepared.

5.2.4 Preparation of transposon libraries

Frozen stocks of the Typhi library were measured by OD_{600} to be at half the concentration of the Typhimurium library. To ensure the overnight cultures started at similar concentrations, a 1 in 5000 dilution of the Typhi library and a 1 in 10,000 dilution of the Typhimurium library was used to inoculate the growth medium. Cultures for each transposon library were grown in 10 mL or 100 mL of RPMI-1640 supplemented with 0.3 g/L L-glutamine and buffered with 1 mL or 10 mL 1 M MOPS, to give a final concentration of 0.1 M, on a shaker at 37 ^{o}C for 16 h. These cultures were sub-cultured at 1 in 20 into fresh RPMI supplemented and buffered as before, and grown for between 3 and 4 hours to mid-log phase (OD_{600} of 2.4).

5.2.5 Optimisation of gentamicin infection assay for TraDIS

5.2.5.1 Preliminary infection

Silvia Pinero prepared the THP-1 cells for infection.

THP-1 cells were grown up and differentiated as described above, except that the cells were seeded at a density of ~1 x 10^5 in 1 mL, per well of a 24-well plate. Upon the day of

infection, cells were washed in dPBS and 1 mL fresh supplemented RPMI was added per well.

A culture of Typhimurium strain SL3261 was grown in LB broth (Oxoid, Basingstoke, UK), and incubated shaking at 37 ^{o}C for 16 h. A control Typhi strain (BRD948) was also cultured. 150 μL of these cultures were then used to inoculate 3 mL fresh LB and incubated shaking at 37 ^{o}C for 3 h to an OD_{600} of 0.2 for late exponential growth. Per strain, three experimental wells were inoculated for each timepoint (30 minutes, 2 hours and 4 hours), using 5 μL of the bacterial culture per well. The 24-well plate was centrifuged for 5 minutes at 600 x *g* and incubated at 37 ^{o}C in 5% CO_2 for 30 minutes. After 30 minutes, media from all wells was removed and replaced with either 1 mL dPBS + 100 μg/mL gentamicin (30 minute wells) or 1 mL supplemented RPMI + 100 μg/mL gentamicin (2 h and 4 h wells). The 30 minute wells were washed twice in dPBS, and the plate was incubated for a further 90 minutes. At this point, the 2 h wells were washed twice with dPBS, and the plate incubated for a further 2 h. After the full 4 h, the 4 h wells and the blank control were washed twice with dPBS and then 100 μL of 1% Triton-X-100 (VWR) was added to every well and the plate incubated at 37 ^{o}C for 2 minutes. This 100 μL was serially diluted to 10^{-5} and 3 x 10 μL drops were plated out on LB plates (supplemented with 'aro' mix for Typhi) for each dilution. The LB plates were incubated for 16 h at 37 ^{o}C.

5.2.5.2 Small scale

Initially, the assay was performed for each transposon library using one 6-well plate per timepoint: 30 minutes, 2 hours and 4 hours. Three wells of the 6-well plate were used as experimental replicates, one as a blank control and the remaining two for plating out viable counts.

At the start of the assay, media was removed from all wells except for the blank control, and a 3 mL bacterial inoculum was added to each experimental well. The plates were centrifuged for 5 minutes at 600 x *g* and incubated at 37 ^{o}C in 5% CO_2 for 30 minutes. A 2-3 mL aliquot of the inoculum was processed for genomic DNA as the input sample for TraDIS. After the 30 minutes, media was removed from all wells, and fresh RPMI additionally supplemented with 100 µg/mL gentamicin was added to wells for the 2 h and 4 h timepoints. The 30 min timepoint wells were washed twice in dPBS supplemented with 100 µg/mL gentamicin and once in plain dPBS. After 2 h and 4 h, the relevant wells were washed 3 times in plain dPBS. Following washing, 500 µL of 1% Triton-X-100 was added to each well to lyse the eukaryotic cells, mixed well by pipetting, and incubated at 37 ^{o}C in 5% CO_2 for 2 minutes. Cell suspensions from the three experimental wells at the same timepoint were pooled for bacterial DNA extraction. Three 10 µL droplets of serial dilutions (from 10^{1} to 10^{-6}) from the inoculum, two wells from each timepoint and from the blank controls were plated out onto L-agar (supplemented with ‘aro’ mix for Typhi) and incubated for 16 h at 37 ^{o}C to establish viable counts.

It was unknown whether sufficient DNA could be extracted directly after the infection, or if the recovered bacteria should be grown up overnight. To determine this, in the first infection assay, 200 µL of the pooled cell suspension from each timepoint was used to

inoculate 10 mL RPMI (supplemented with 0.3 g/L L-glutamine and buffered to a final concentration of 0.1 M MOPS) and incubated shaking at 37 ^{o}C for 16 h before DNA extraction.

5.2.5.3 DNA manipulation (small scale)

Genomic DNA was extracted from the inocula and pooled cell suspensions from each timepoint using either the Qiagen Generation Capture Columns or the DNeasy Blood and Tissue Kit (Qiagen, Crawley, UK), according to the manufacturer's protocol for Gram negative bacteria, with the optional RNAse step. DNA samples were quantified using the Qubit fluorescence-based dsDNA HS assay (Invitrogen, Paisley, UK).

Quantitative PCR was performed on one set of samples, using genomic DNA from the Typhi transposon mutant pool as DNA standards serially diluted from 30 ng/µL to 1.875 ng/µL. The PCR was run on an ABI cycler (Applied Biosystems, Warrington, UK) using the recommended 25 µL maximum reaction volume. For each sample, 12.5 µL of the QuantiTect SYBR Green PCR mix (Qiagen) was added to 0.125 µL (5µM) of a forward and reverse primer, 11.25 µL distilled H_20 and 1 µL of the DNA sample. Each sample was tested using the following primers for two genes, *aroC* and *phsA*:

aroC-05: 5'- GTGATCCATCAGTACGATCG and *aroC*-06: 5'- GACAACTCTTTCGCGTAACC

phsA-01: 5' – GATTTGAACAGTTAGCACAG and *phsA*-02: 5'- ACAGGTGACGGTCTGGAAAG

5.2.6 Large scale infection assay

The assay was carried out as before, except five 6-well plates were utilised for every timepoint assayed. In total, 29 wells were infected with the bacterial inoculum and one served as a blank control for eukaryotic cell contamination. A 4-6 mL aliquot of the inoculum was processed for genomic DNA as the input sample for TraDIS. The cell suspensions from all experimental wells at the same timepoint were pooled for bacterial DNA extraction. A 100 µL aliquot of the pooled cell suspension from each timepoint was serially diluted to 10^{-6} and plated onto L-agar (supplemented with 'aro' mix for Typhi) to establish viable counts; three 10 µL droplets of serial dilutions were also plated out from the inoculum and the control from each timepoint, and incubated as before.

5.2.6.1 Single timepoint replicates

The assay was again carried out as before, except that three independent 100 mL cultures were grown up overnight, used to inoculate 3 independent 270 mL starter cultures and then to infect 3 independent sets of five 6-well plates for 2 hours. For each replicate, a total of 29 wells were infected with the bacterial inoculum and one served as a blank control for eukaryotic cell contamination. All other aspects were performed as before.

5.2.6.2 DNA manipulation (large scale)

The pooled cell suspension from each timepoint was pelleted and washed twice in RPMI buffered with 0.1 M MOPS (large-scale), or in dPBS (single timepoint replicates). For the large-scale assays, the sample was split into 3 at this point to determine if this increased

DNA yield. For the multiple timepoint replicate assay, the sample was processed as one. Pellets were resuspended in 85 µL dH_20, 5 µL DNase I (New England Biolabs, Hitchin, UK) and 10 µL 10 x Buffer (New England Biolabs), incubated at room temperature for 15 minutes and then at 65 ^oC for 10 minutes. Samples were centrifuged for 10 minutes at 7500 rpm and resuspended in 180 µL Buffer ATL (Qiagen). Genomic DNA was extracted from the inoculum and from the Buffer ATL samples using the Qiagen DNeasy Blood and Tissue kit, according to the manufacturer's protocol for Gram negative bacteria. DNA samples were quantified using the Qubit fluorescence-based dsDNA HS assay (Invitrogen).

5.2.7 TraDIS analysis

5.2.7.1 Sample processing and nucleotide sequencing

Sabine Eckert and Daniel Turner performed the preparation and nucleotide sequencing for the large-scale and single timepoint replicate samples.

All sample clean-ups were performed using Qiagen MinElute columns.

Two µg of DNA per sample was sheared by Covaris AFA (Quail et al. 2008) on an S2 instrument (at settings 5, 20%, 200 bursts per cycle and 90 seconds) to an average length of 230 ± 100 nucleotide bases. The sheared DNA was cleaned up and end-repaired using the NEBNext Sanger Sequencing Sample Preparation kit (New England Biolabs). Samples were cleaned up again and A nucleotides added using the same NEB kit. Following another clean-up, samples were quantified with the Agilent Bioanalyzer 2100

and ligated to a 10-fold excess of annealed adapters (nucleotide sequences below) with the NEB kit.

Adapter 1: PE_AD_T ACACTCTTTCCCTACACGACGCTCTTCCGATC*T

Adapter 2: PE_AD_B GATCGGAAGAGCGGTTCAGCAGGAATGCCGAG

A final clean-up preceded a qPCR to again quantify sample concentration. This was performed using the following primers, against standards of known concentration:

Ad_T_qPCR1: 5' - CTTTCCCTACACGACGCTCTTC

Ad_B_qPCR2: 5' - ATTCCTGCTGAACCGCTCTTC

200 ng of DNA from the bacterial input samples, and 400 ng from samples taken post-infection were PCR-amplified using Jumpstart Taq DNA polymerase (Sigma) with the following primers:

5TMDHp5F2:

5' - CAAGCAGAAGACGGCATACGAGATCGCTGAATTACCCTGTTATCCCTATTTAGGTGAC

PE PCRv3.3:

5' - CAAGCAGAAGACGGCATACGAGATCGGTACACTCTTTCCCTACACGACGCTCTTCCGATCT

at 94 °C for 2 minutes, followed by 22 cycles of 94 °C for 30 seconds, 65 °C for 20 seconds and 72 °C for 30 seconds, then 72 °C for 10 seconds. The PCR products were quantified by qPCR using the primers below with standards of known concentration.

Syb_FP5: 5' - ATGATACGGCGACCACCGAG

Syb_RP7: 5' - CAAGCAGAAGACGGCATACGAG

Samples were then separated using a 2% TBE agarose gel and fragments in the range of 350 – 500 nucleotide bases were isolated using the Qiagen Gel-Extraction kit, according

to the manufacturer's instructions. These were quantified by qPCR, using primers Syb_FP5 and Syb_RP7 as before. Prior to flowcell preparation, samples were denatured in 0.2 M NaOH for 5 minutes, neutralised with 0.1 M HCl and diluted to a final concentration of 10 pM in hybridization buffer. Flowcells were prepared according to the Illumina protocols and samples were sequenced using a Genome Analyzer II with the custom primer 5TMDH2seq (see Chapter 4).

5.2.7.2 Sequence data processing

This was undertaken as described in Chapter 4, section 4.2.5.

5.2.8 Statistical analysis of required genes

The nature of this analysis was discussed with Leopold Parts who wrote scripts in R to perform the following analyses.

5.2.8.1 Large scale assays

The processed sequence data from all timepoints post infection were normalised for total sequencing yield and combined. For every gene g present in the input sample and the combined output, with $n_{g,A}$ reads observed in the input and $n_{g,B}$ reads observed in the output, we calculated the log_2 fold change ratio $S_{g,A,B} = log_2 \backslash frac(n_{g,A}+100)(n_{g,B}+100)$. The correction of 100 reads smoothed out the high scores for genes with very low numbers of observed reads. We fitted a normal model to the mode +/- 2 sample standard

deviations of the distribution of $S_{A,B}$, and calculated *P*-values for each gene according to the fit.

For Typhimurium, a cutoff was chosen at a $\log_2$ fold change ratio of at least 4 that gave a false discovery rate (FDR) of 4.5%, corresponding to a *P*-value of 0.001.

For Typhi, the same calculations were applied, but using the number of insertion sites rather than mapped reads since these provided a clearer cutoff, which was chosen at a $\log_2$ fold change ratio of at least 2 that gave an FDR of 3.8%, corresponding to a 10^{-4} *P*-value.

5.2.8.2 Single timepoint replicates in Typhi

The total number of insertion sites per gene for the three input samples and the three 2 h 'output' samples were combined and normalised to the number of insertions found in the first input replicate. In the same manner as above, $\log_2$ fold change ratios in insertion sites were calculated per gene, with corresponding *P*-values. A cutoff was chosen at a $\log_2$ fold change ratio of at least 1, that gave an FDR of 3.6%, corresponding to a 10^{-6} *P*-value.

5.3 Results

5.3.1 Preliminary Typhimurium assay

Prior to using the Typhimurium SL3261 transposon library itself with the human cell line THP-1, a preliminary infection assay was carried out using the Typhimurium strain in which the library was constructed. This was to ensure that this strain was capable of infecting this human cell line. Differentiated THP-1 cells were infected with the Typhimurium strain (approximately 10 bacteria per macrophage, a multiplicity of infection (MOI) of 10) and with Typhi BRD948 (MOI 60) as a control, and viable counts were performed at 30 minutes, 2 hours and 4 hours post-infection. Figure 5-1 shows the results from each timepoint and demonstrates that Typhimurium is capable of infecting THP-1 cells. In addition, uptake was approximately 20-fold higher for Typhimurium than for the control (Typhi) strain. This was likely due to the expression of the Vi antigen by the latter, which is known to negatively affect uptake (Arricau et al. 1998).

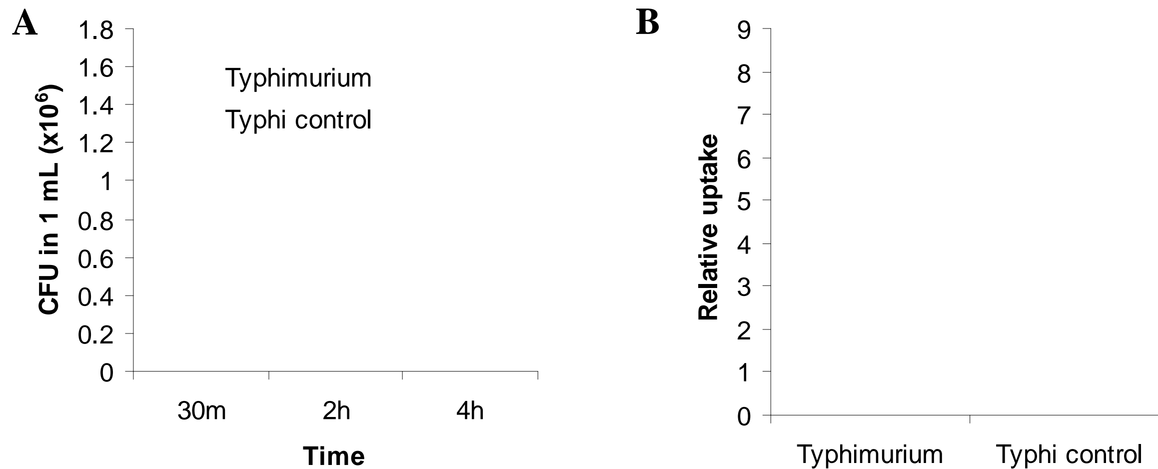


Figure 5-1 Preliminary infection assay

A) Total bacteria recovered from THP-1 gentamicin infection assay at three timepoints. CFU, colony forming units given in millions per mL. B). Relative uptake of bacteria, calculated by dividing the CFU at 30 minutes by the inoculum.

5.3.2 Optimising the cell infection assay for TraDIS

5.3.2.1 Media

THP-1 cells are cultured in RPMI media for optimum growth. If the bacterial inoculum was grown in LB and then added to the eukaryotic cells in RPMI, there was the possibility that mutants would be negatively affected by the culture media, rather than their infection potential. Hence, both bacterial strains used to construct the transposon libraries were tested for their ability to grow in RPMI. The ability of the Typhi transposon library strain to grow without the addition of ‘aro’ mix was also tested. Multiple growth curves of each transposon library strain in buffered RPMI, without ‘aro’ mix indicated that both strains were capable of growth in this culture media, allowing the first infection assay to be carried out using buffered RPMI to grow up the bacterial inocula.

5.3.2.2 Infection assays

Viable counts from the first small scale infection assay using the transposon libraries indicated that the assay had been successful: bacteria were recovered from the eukaryotic cells at all timepoints tested (Figure 5-2). Bacterial numbers did not increase over the timepoints as has been observed in the preliminary infection, but the total number of bacteria recovered at 4 hours with the transposon libraries (~ 4 x 10^7) was actually in excess of that recovered for the same timepoint in the preliminary Typhimurium infection (1.5 x 10^6). This is likely due to the high OD_{600} used for infection, which resulted in the multiplicity of infection in this assay being higher than any other. High MOIs were however used in all the infections to ensure every transposon mutant was well represented in the assay (Table 5-1).

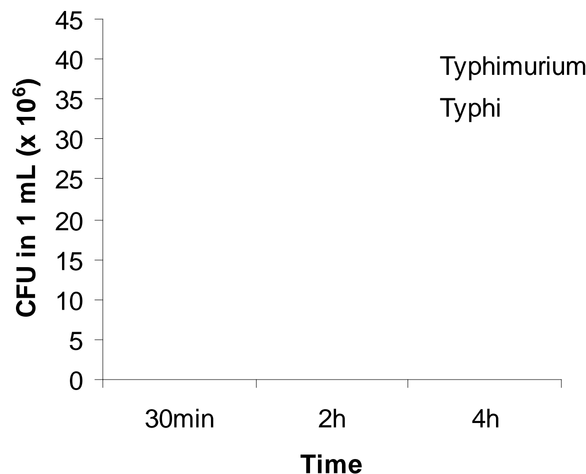


Figure 5-2 Transposon library cell counts from initial THP-1 infection

Total number of bacteria recovered after timepoints indicated. CFU, colony forming units given in millions per mL.

Table 5-1 Small scale infection assays

Assay	OD_{600} at infection	MOI	DNA extraction	Comments
1 SL	0.41	750	direct and overnight	Low DNA yield from both extraction times
1 WT	0.53	960	direct and overnight	Low DNA yield from both extraction times
2 SL	0.2	-	direct only	Eukaryotic cells contaminated
2 WT	0.2	-	direct only	Eukaryotic cells contaminated
3 SL	0.2	405	direct only	-
3 WT	0.2	270	direct only	Checked by qPCR
4 SL	0.2	315	direct only	-
4 WT	0.2	345	direct only	-
5 SL	0.2	225	direct only	-
5 WT	0.2	285	direct only	-

SL, Typhimurium; WT, Typhi; MOI, multiplicity of infection. MOI could not be determined for the second set of infection assays due to contamination on the viable count plates.

The question of when to extract genomic DNA from invading bacteria was addressed in the first infection assay using the transposon libraries. DNA was extracted from transposon mutants harvested directly after each timepoint in the infection, and from aliquots of the same mutants that had been grown overnight in RPMI. Quantification of these samples indicated that yields were extremely low in both cases, although the overnight samples were slightly improved (Table 5-2). However, the small increase in DNA yield was not sufficient to justify the possible bias generated by the transposon mutants from overnight growth.

Further infection assays were performed using a different DNA extraction kit to determine whether increased DNA yields could be obtained directly from the harvested mutants. In total, three independent assays were completed without contamination (Table 5-1), each resulting in greater DNA yields than the original assay (Table 5-2).

Table 5-2 DNA yields from different extraction kits

Time of DNA extraction	Sample	Typhimurium DNA yield (µg)	Typhi DNA yield (µg)	DNA extraction kit
Directly following infection	inoculum	0.034	0.25	Capture columns
	30 minutes	0.33	0.07	
	2 h	0.13	0.03	
	4 h	0.11	0.024	
After overnight growth	inoculum	0.77	1.2	Capture columns
	30 minutes	0.36	1.7	
	2 h	0.9	2.3	
	4 h	0.99	1	
Directly following infection*	inoculum	2.4	4.6	DNeasy Blood and Tissue
	30 minutes	15	7.6	
	2 h	12	7.7	
	4 h	11.1	6.9	

DNA extraction kits used were all from Qiagen. DNA yield quantified using the Qubit Quant-IT system (Invitrogen). *This infection is one representative assay of three performed under the improved protocol (see text).

With the chance that some eukaryotic DNA might have been carried over in the bacterial genomic DNA extraction, quantitative PCR was performed on the set of Typhi samples from the third infection to determine how much bacterial DNA was present. Two sets of PCR primers known to be reliable were chosen, for the genes *aroC* and *phsA*. The Typhi samples were run with a standard set of dilutions and revealed that, while the inoculum contained the most DNA, very little was present in the samples obtained post infection suggesting that most of the DNA measured by the Qubit system was in fact eukaryotic carry-over.

For this reason, two changes were introduced to the infection assay protocol. Firstly, the infections would be scaled up to improve bacterial DNA yield and enable TraDIS analysis of the results. Secondly, more rigorous wash steps and a DNase I treatment would be used in processing the samples obtained post infection, in order to reduce the amount of eukaryotic DNA being carried through and again improve bacterial DNA yield.

5.3.3 Large-scale infection

In the large-scale infection assays, each transposon library was assayed on a separate occasion. The assays themselves were scaled up almost 10-fold, from 3 experimental wells to 29, at each timepoint. The CFU/mL was slightly lower in these assays, but the 10-fold greater total volume generated a much improved yield of DNA.

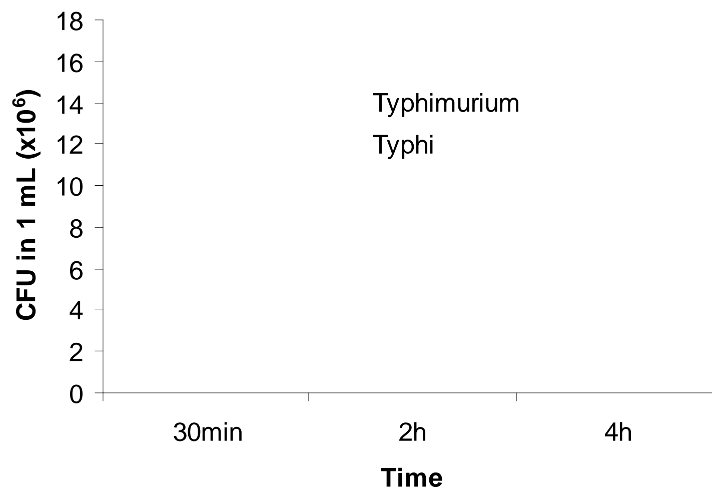


Figure 5-3 Transposon library cell counts from large-scale THP-1 infection

Total number of bacteria recovered after timepoints indicated. CFU, colony forming units given in millions per mL.

Despite the wash steps incorporated into the DNA extraction procedure, the amount of DNA recovered from the timepoints post infection suggested significant levels of eukaryotic contamination was still occurring (Table 5-3). Nonetheless, the bacterial yield proved sufficient for TraDIS analysis.

Table 5-3 DNA yields from large-scale infection

Timepoint	Typhi (µg)	Typhimurium (µg)
Input	7	2.2
30 minutes*	124	105
2 hours*	110	45.5
4 hours*	78	68.5

* yield estimated by combining results from one sample split into 3, extracted down 3 columns and quantified as 3 samples (see Methods).

5.3.4 TraDIS analysis of large-scale infection assays

The eight samples from the two large-scale assays were each run on a single lane of an Illumina flowcell, which generated between 1.7 and 14.2 million reads (Table 5-4). The lowest sequencing read yields were observed from the Typhi samples obtained post infection. While the sequence reads were not at the same level as the Typhi input sample, there was sufficient information in 1.7 – 3.5 million reads to enable identification of 64,000 – 87,000 transposon mutants. No such issues were observed with the Typhimurium sequencing results.

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

	Reads	Tagged (%)	Mapped (%)	Inserts
Typhi				
Input	14,182,028	10,557,817 (74)	7,990,449 (76)	233,497
30 minutes*	3,557,883	3,218,768 (90)	2,771,563 (86)	87,634
2 hours*	2,030,200	1,854,321 (91)	1,657,018 (90)	97,176
4 hours*	1,752,198	1,610,693 (92)	1,459,621 (91)	64,708
Typhimurium				
Input	13,565,707	12,457,266 (92)	7,312,946 (59)	179,702
30 minutes*	14,294,219	13,340,388 (93)	10,537,514 (79)	46,352
2 hours*	13,012,186	12,124,834 (93)	9,633,788 (72)	43,110
4 hours*	10,350,391	9,440,397 (91)	6,411,911 (68)	25,922

Tagged reads are those containing a 100% identical match to the last 10bp of the transposon, indicating a genuine transposon insertion site. The Typhimurium samples were sequenced twice, with very similar sets of results; only one is shown here.

5.3.5 Typhimurium

Initial analysis of the TraDIS data comparing the input individually with 30 minutes, 2 hours and 4 hours post infection showed that, in order to quantify changes between timepoints in a rigorous statistical manner, replicates at each timepoint were required. In the absence of these replicates, statistical power was instead increased by treating each of the 3 timepoint samples as independent replicates and combining the data. In this way, we were essentially assaying the genes required for adherence to and invasion of the macrophage, as this was the process each sample had been subjected to. The finer detail of changes across timepoints was lost, but a stronger signal for mutants selected against during infection was observed.

Accordingly, a significant loss of transposon mutants during the infection assay was observed for a total of 316 genes (log_2 FC < -4, *P*-value < 0.001, FDR 6%). The full list can be found in Appendix 8.4.1. These represent candidate genes essential for adherence to and invasion of THP-1 cells.

The candidate on the Typhimurium gene list with the most significant loss of mutants was *surA* ($log_2FC = -7.67$, $P << 0.001$), which encodes a peptidylprolyl-cis,trans-isomerase involved in energy metabolism. This gene has been previously characterised as a potential vaccine candidate, since it was found to be highly attenuated for virulence in the murine model, and for adherence to and invasion of epithelial cells (Sydenham et al. 2000). Its appearance in this study suggests that *surA* mutants are also defective for adherence/invasion of macrophage-like THP-1 cells. The ligand-binding protein encoded by *sapA* was also identified, as well as the associated ATP-binding protein encoded by *sapD*. The operon *sapABCDF* encodes a peptide transport complex that has been shown to have a role in both antimicrobial peptide resistance and in virulence (Parra-Lopez et al. 1993). Other genes previously associated with a reduced capacity to survive inside macrophages and decreased virulence in the murine include *prc* and *purD* (Baumler et al. 1994), both of which were identified in this study. Together, these provide validation for high-throughput screening as a method to assay genes important for cell adherence, invasion and even survival.

A functional breakdown of the 316 genes is shown in Figure 5-4. When these were compared to the functional breakdown of the whole genome, genes in the categories of pathogenicity/adaptation/chaperones, conserved hypothetical and to a lesser extent information transfer were found to be over-represented (by 3.5%, 3.5% and 1.9%

respectively). Almost one fifth were categorised as conserved hypothetical, indicating that many functions key to cell infection have yet to be elucidated.

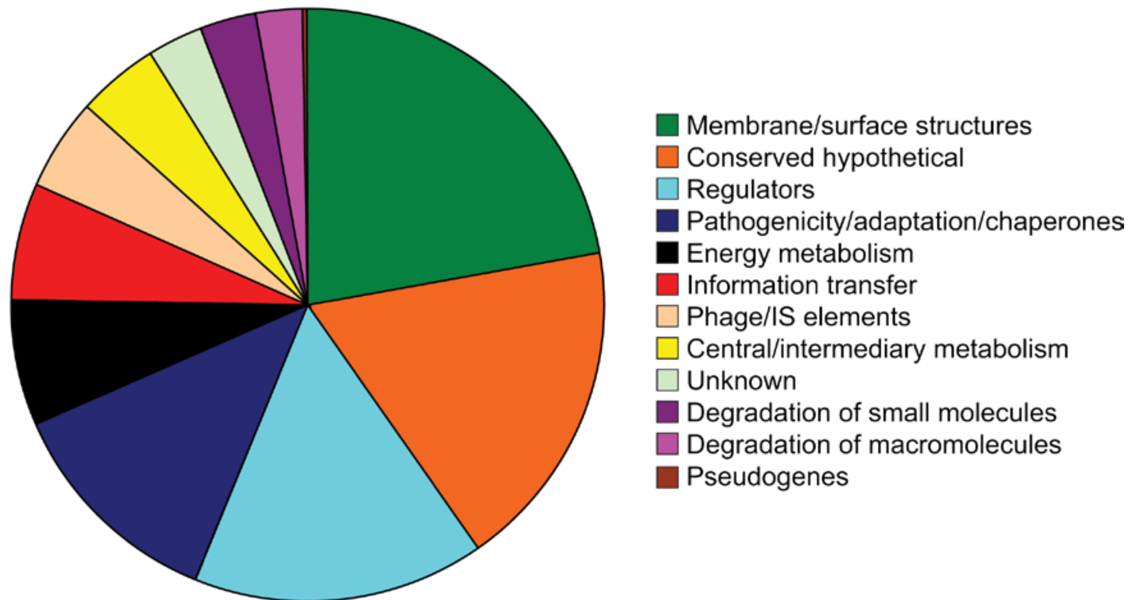


Figure 5-4 Typhimurium genes essential for macrophage invasion and SCV generation

Functional categories and colours assigned based on genome annotation.

While the membrane/surface structure genes required for host cell infection were under-represented compared to the whole genome (by 4.3%), fourteen were involved in flagellar regulation, biosynthesis and assembly, including *fliA*, the alternative sigma factor specific for flagellar operons. There were representatives of both SPI-1 and SPI-2 in the pathogenicity/adaptation/chaperone candidate essential genes. From SPI-1, both *iagB* and *spaM*, encoding a cell infection protein and a secreted effector respectively were found to be essential, as well as SL2883, which is part of SPI-1 but has not yet been characterised. Five SPI-2 *sse*/*ssa* genes were also found to be essential, in accordance with this island being known to be required for invasion of macrophages *in vivo*.

In order to determine the metabolic effects these candidate essential genes may have, they were mapped onto StmCyc 4.0 (Figure 5-5). Of note were the multiple genes involved in propanediol utilisation and vitamin B12 biosynthesis and transport, and the number of transporters whose function appeared essential for successful invasion of the macrophage.

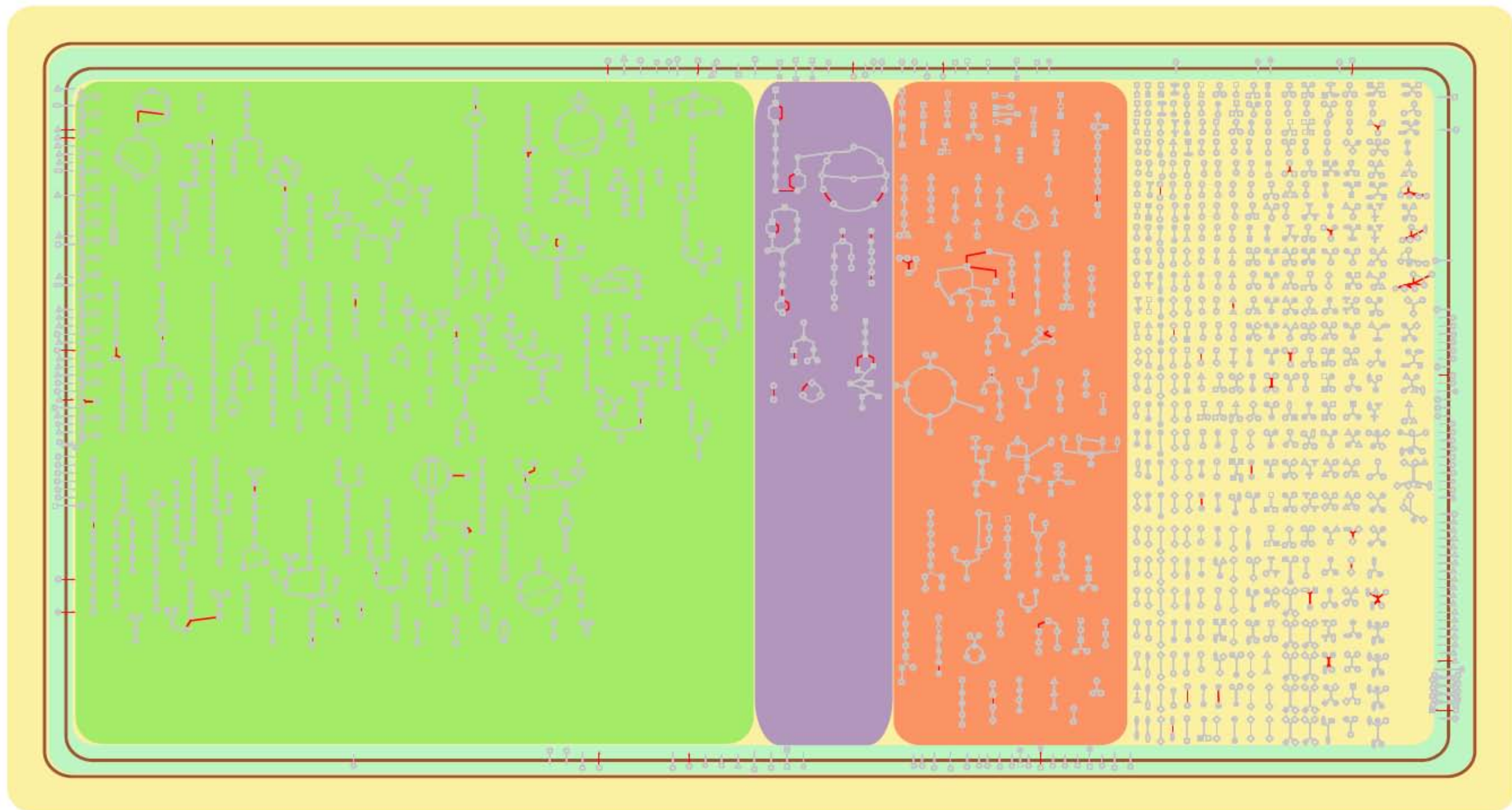


Figure 5-5 StmCyc view of Typhimurium candidate essential genes for macrophage invasion

Brown lines indicate bacterial cell membrane; red lines indicate candidate essential genes; greyed out lines represent metabolic reactions. Green background, biosynthetic pathways; dark blue, energy pathways; red, degradation pathways; light blue, transport reactions. Symbols: upward-pointing triangle, amino acids; square, carbohydrates; diamond, proteins; vertical oblong, purines; horizontal oblong, pyrimidines; downward-pointing triangle, cofactors; T, tRNAs; open circle, other; enclosed circle, phosphorylated.

5.3.5.1 Propanediol utilisation

As mentioned in Chapter 3, the utilisation of 1,2-propanediol (PDL), is an anaerobic process dependent upon the synthesis or exogenous provision of vitamin B12. The degradation of PDL was apparently required during the infection assay, since four *pdu* genes were identified on the candidate essential gene list (Figure 5-6).

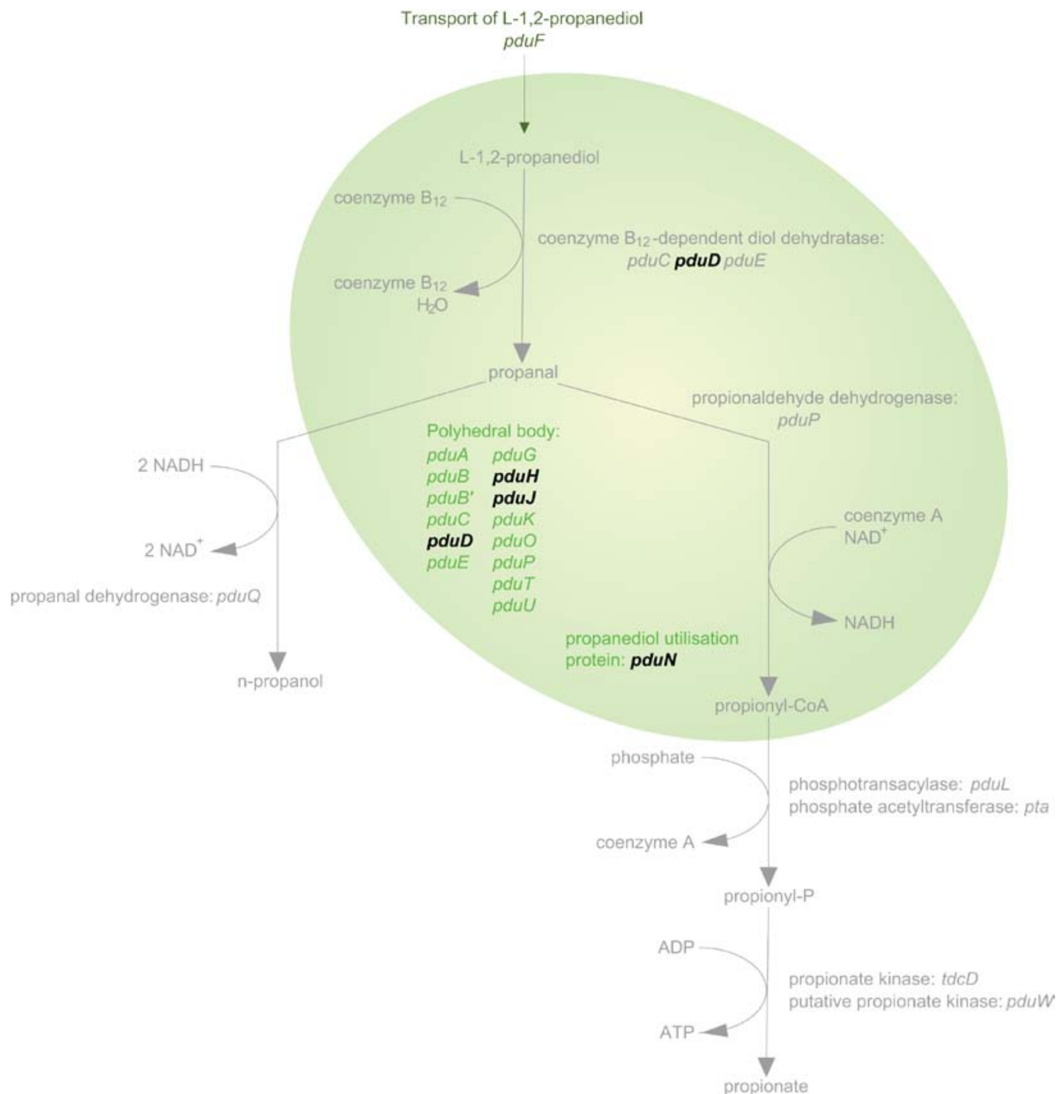


Figure 5-6 Genes required in propanediol utilisation

Gene names in black bold type are those identified as candidate essential genes in the Typhimurium large-scale infection assay.

The transport complex for vitamin B12 is encoded by *btuBCDEF*, of which both *btuD* and *btuE* were identified as candidate essential genes for Typhimurium in this assay. In *E. coli*, *btuD* mutants exhibit a defective transport phenotype for vitamin B12, but *btuE* mutants do not (Rioux and Kadner 1989). BtuE is a putative periplasmic protein, and the results of this assay suggest that its role in Typhimurium is required for macrophage invasion, whether this be as part of vitamin B12 transport or some as yet unknown function.

Since vitamin B12 is apparently being taken up from the host cell environment, the requirement for *de novo* biosynthesis could conceivably be reduced. However, the identification of *cbiF*, *cbiO* and *cobS* as candidate essential genes suggests that the metabolic pathways involved in biosynthesis are also required.

5.3.5.2 Transport

Numerous genes involved in transport reactions were identified as required for successful Typhimurium invasion of THP-1 macrophages. The specific substrates which are transported by these gene products therefore give some insight into the availability of nutrients during and immediately following the infection process. A summary of the genes and substrates transported are given in Table 5-5.

Table 5-5 Candidate essential transporters

Substrate	Type	Genes	Function/family of essential gene
glutamine	ABC	*gln**P**QH*	permease
lysine	Secondary	***lysP***	APC
proline / glycine betaine	ABC	*proVW**X***	substrate-binding periplasmic protein
an amino acid	ABC	SL1563-**6**	unknown
spermidine/putrescine	ABC	*pot**A**BCD*	ATP-binding protein
PTS carbohydrates	PTS	*pts**H**I*; multiple	HPr protein
iron	ABC	*sitAB**C**D*	inner membrane component
cobalt	ABC	*cbiMNQ**O***	ATP-binding protein
zinc	ABC	*znuA**C**B*	ATP-binding protein
vitamin B12	ABC	*btuBC**DE**F*	unknown

Candidate essential genes shown in bold type. ABC, ATP-binding cassette; PTS, phosphotransferase system; APC, amino acid-polyamine-organocation family.

The phosphotransferase system (PTS) histidine protein (HPr), encoded by *ptsH* was found to be essential during the assay. HPr and enzyme I (EI) are both soluble proteins that typically form part of all PTS transporters. The substrate of each PTS is determined by the carbohydrate-specific enzyme II component (EII). Since no EII components were found to be essential, it appears that Typhimurium preferentially utilises carbohydrates during infection, although no particular one is absolutely required.

5.3.6 Typhi

The Typhi data was sequenced and analysed after the Typhimurium dataset. Therefore the data from the three output samples were combined immediately for analysis. However, the analysis revealed that only 33 genes were specifically required (Log FC < -2, FDR 3.6%, $P < 0.001$) (Appendix 8.4.2). The assay had been carried out in exactly the

same manner as Typhimurium, so no experimental variation was expected. To be confident of the result, the assay was repeated with 3 independent sample sets of input and the 2 h timepoint. This timepoint was taken to best represent the time after which *Salmonella* should be inside the host cell and have potentially generated an SCV.

All three replicates confirmed that Typhi requires far fewer genes for macrophage infection than Typhimurium, as mutants in only 8 genes were consistently selected against (Table 5-6).

Table 5-6 Typhi genes required during macrophage assay

Gene	Unique ID	Inserts (input)	Inserts (2 h)	Log_2 FC	Combined *P*-value	Function
gidA	t3465	108	21	-1.93	2.7×10^{-18}	glucose-inhibited division protein
waaP	t3797	43	9	-1.47	2.4×10^{-11}	phosphorylation of core heptose
thdF	t3677	99	30	-1.46	3.9×10^{-11}	thiophene/furan oxidation protein
waaG	t3796	41	11	-1.26	9.4×10^{-9}	LPS glucosyltransferase I
wecB	t3377	49	15	-1.23	2.1×10^{-8}	UDP-*N*-acetylglucosamine 2-epimerase
dsbD	t4374	105	43	-1.14	2.2×10^{-7}	thiol:disulphide interchange protein
fkpA	t4052	59	21	-1.14	2.1×10^{-7}	peptidyl-prolyl isomerase
fimD	t2317	116	49	-1.09	6.8×10^{-7}	outer membrane usher protein D
rfe	t3379	71	29	-1.05	1.6×10^{-6}	undecaprenyl-phosphate α-*N*-acetylglucosaminyltransferase

Total number of insert sites per gene from input and 2 h timepoint were combined from the three replicates and normalised to the values from the first input replicate.

Many of these were in genes relating to outer membrane components; the *waa* genes are both involved in LPS core biosynthesis, and two early steps in enterobacterial common antigen (ECA) biosynthesis are encoded by *wecB* and *rfe*. The role of DsbD is to help fold proteins at the membrane, and correctly folded membrane proteins are needed in

order to adhere to host cells. This is supported by *fkpA*, which encodes a peptidyl-prolyl isomerase (PPIase) that assists protein folding, and is of the same family as *surA* (identified in Typhimurium). Together, these suggest that surface antigens and attachment of Typhi to host cells is important, possibly in triggering phagocytosis rather than invasion.

In *E. coli*, *gidA* and *thdF* have been renamed *mnmF* and *mnmE* respectively. These genes encode proteins that work in concert to maintain the correct reading frame during transcription, by reducing +2 frameshift errors (Bregeon et al. 2001). The authors suggest that *gidA* has further activities, which may be important in Typhi, but since both *gidA* and *thdF* were selected for, it is more likely that the function they perform together is the one required inside the macrophage, for reasons that remain unclear.

This extremely small set of genes apparently required by Typhi for macrophage infection clearly demonstrates the dissimilar manner in which Typhi and Typhimurium enter the cells used in this experiment, while the candidate genes required by Typhimurium suggests utilisation of pathways no longer active in the Typhi genome.

5.4 Discussion

The results documented here represent the first genome-wide assay of genes required for infection of human macrophages in serovars Typhimurium and Typhi. Previous studies have concentrated upon understanding which genes are highly expressed or down-regulated during the infection process, which presents its own technical challenges (Daigle et al. 2001; Faucher et al. 2005; Faucher et al. 2006). In this assay, multiple infections were carried out to establish a set of conditions under which each organism in the million mutant pools had an equivalent chance of infecting a macrophage, and enough DNA could be extracted to perform the TraDIS analysis. The nature of an infection assay requires two growth stages of the bacterial inoculum, so to discount the selective effects these might have exerted upon the mutant pools, samples taken post-infection were compared to an 'input' sample taken just prior to infection. The bacteria were also grown in the same cell culture media as the eukaryotic cells, thus reducing the possibility that mutants that were not recovered post infection were unable to respire in the culture media.

The results observed for Typhimurium and Typhi were starkly different, suggesting that these serovars enter THP-1 cells by dissimilar mechanisms. Though it was already known that Typhi is phagocytosed by this eukaryotic cell type (Daigle et al. 2001; Faucher et al. 2006), it appears that Typhi relies entirely upon this mechanism for cell entry. This is unlikely due to the strain used since mutations in the three deleted genes (*aroC*, *aroD* and *htrA*) have been shown not to affect the ability to infect human macrophages (Lowe et al. 1999). Only eight genes were consistently identified as important for cell infection; sufficient mutants in all other genes survived the uptake process and for two hours post-

infection. Of the eight genes, only one has previously been implicated in *Salmonella* virulence; *fkpA* mutants in some Typhimurium strains have shown a reduced ability to survive inside murine macrophages (Horne et al. 1997; Humphreys et al. 2003). The majority of the remaining genes were associated with the bacterial membrane and surface antigens, which suggests that their role lies in adherence to host cells and subsequent phagocytosis. However, the identification of 316 genes important for Typhimurium infection implies that Typhimurium actively invades this cell type.

The genes identified in the Typhimurium assay were from a combination of three timepoints post infection: 30 minutes, 2 hours and 4 hours. Some of these genes had been previously identified as important for survival inside macrophages, rather than invasion. It is possible that either the gene product was required earlier than previously described, or that the signal from some of these genes (e.g. *prc* and *purD*) was in fact strongest from the 2 and 4 hour timepoints. An analysis of single timepoint replicates from 2 hours post infection with Typhimurium is currently underway and will provide further insight.

There is evidence from murine macrophages that with Typhimurium, regardless of the method of entry (invasion/phagocytosis), an SCV is generated that diverges from the typical eukaryotic endocytic degradation pathway (Rathman et al. 1997). It is therefore possible that Typhi, while apparently not invading THP-1s, still generates an SCV after phagocytic uptake. However, the lack of genes identified as required for such a process suggests that the generation of an SCV may occur very slowly, longer than the 2 hour timepoint tested, or is not dependent upon bacterial products. A much later timepoint would need to be tested to confirm this. Whether the genes that are necessary for biogenesis of a Typhi SCV in macrophages would be similar to those in Typhimurium is

unclear, as there is evidence from Typhimurium infection of epithelial cells that SCVs containing non-invasive mutants do not have the same markers as those containing wildtype strains (Gorvel and Méresse 2001). Thus, while the vacuole generated by phagocytosed bacteria may perform the same general function of protection against the host cell environment, structurally it may be somewhat different.

The lack of candidate essential genes for infection of macrophages by Typhi may reflect the level of genome degradation in this serovar. While active invasion is required for infection of epithelial cells, the passive uptake by macrophages may allow certain pseudogenes to be tolerated. This is particularly intriguing when considered in the context of genes required by Typhimurium for macrophage invasion. Two pathways inactivated by pseudogenes in Typhi include vitamin B12 biosynthesis and 1,2-propanediol degradation. As discussed in Chapter 3, these are linked by the absolute requirement for vitamin B12 in propanediol utilisation. However, candidate essential genes involved in both pathways were identified in the infection assay of Typhimurium, an observation supported by experiments performed in murine macrophages (Klumpp and Fuchs 2007). Similarly, genes encoding transporters for proline, cobalt and multiple carbohydrates were required by Typhimurium, but these are also inactivated by pseudogenes in Typhi.

Why all the genes involved in these pathways and transport reactions were not identified as essential in Typhimurium is probably due to a combination of factors. Firstly, there may be redundancy in enzymatic function, which would not be considered essential using this type of assay. Secondly, substrates from elsewhere in the metabolic network may feed into pathways at certain points, alleviating the need for particular enzymes, although their presence may be extremely advantageous. Finally, and most likely, this may be

because the Typhimurium analysis was completed using three datasets taken from different timepoints post-infection. This may have diluted the signal coming from some genes important at the 2 hour and 4 hour stages, such that they did not meet the cutoff criteria. As previously stated, a repeat experiment with multiple Typhimurium replicates taken at 2 hours post-infection should provide greater detail.

It is important to note that a transposon-based assay is one designed to look not for the genes used during a process, but those that are essential for it. There is likely a redundancy in function encoded in the *Salmonella* genome that may account for particular genes found to be important by other methods. Additionally, if mutations in particular genes mean that utilisation of a particular substrate is blocked, other substrates may be used instead. For example, previous studies have indicated that Typhimurium preferentially uses gluconate as a carbon source inside macrophages (Eriksson et al. 2003). None of the necessary *dgo* genes were identified in this assay, but this serves to show that gluconate is not the only substrate upon which Typhimurium can survive inside this cell type. However, the requirement for *ptsH*, a common component of all PTS carbohydrate transporters suggests that the inability to transport at least one of these carbohydrates is detrimental to bacterial survival.

The use of a phagocytic cell line has revealed a distinct difference in the mechanism by which Typhi and Typhimurium infect such cells and suggests that few genes in Typhi are essential to this process. During human infection, Typhi must also infect the epithelial cells of the gut wall, in order to reach the bloodstream. Since these cells are non-phagocytic, Typhi presumably actively invades, likely resulting in a different gene requirement to that found in macrophages. A high-throughput screen of the transposon

library through an epithelial cell line is therefore the next logical step in determining how the metabolic capacity of Typhi affects its ability to infect and survive inside host cells.

5.5 Conclusions

Large transposon libraries that comprise over 80 independent mutants per gene provide a useful tool for assaying genes in a high-throughput manner. By making use of the two *Salmonella* libraries, the genes required for infection of a human macrophage-like cell line have been determined. An assay of this nature is difficult to optimise given the variability of the host cells, but a protocol has been established that allows each mutant an equivalent chance of infection, and obtains sufficient DNA post-infection to allow a TraDIS analysis.

The TraDIS analyses from Typhi and Typhimurium infection of THP-1 cells revealed that Typhimurium requires over 300 candidate genes for successful invasion and early survival. A subset of these genes were mapped to metabolic pathways and transporters, and indicates that Typhimurium utilises metabolic capabilities no longer present in Typhi due to inactivation by pseudogenes. Conversely, Typhi requires less than 10 genes for infection, apparently relying on phagocytosis for cell entry, with almost no genes clearly required for early survival inside the host cell.

6 Final discussion

In this thesis, I demonstrate that host restriction in *Salmonella* is linked to a reduction in metabolic capability, and that the inactivation of genes within metabolic pathways, regulators and transporters explains this link. I have also shown that the essential gene lists for a host-generalist and a host-specialist *Salmonella* have key differences, and by screening for genes important in the macrophage, I have shown the implications for host infection.

From the time that *Salmonella* bacteria were first isolated until the advent of nucleotide sequencing, the main methods for distinguishing between serovars were phenotypic in nature. It is from such studies that a correlation between reduced metabolic capability of a serovar and restricted host range can be identified. Indeed, publications as far back as 1919 show that human-restricted Typhi was able to utilise far fewer substrates than the host generalist Enteritidis. However, it wasn't until whole genome sequencing of *Salmonella* strains was published in the 2000s that the genome degradation common to host-restricted serovars was revealed. However, there was little evidence in the literature to link the effect of pseudogenes with metabolism, which had the potential to provide explanations for the reduced metabolic capacity observed for many years in host-restricted *Salmonella*.

Prior to this project, pseudogenes had generally been considered on an individual basis. To better understand the global effect of pseudogenes upon metabolism, metabolic pathway databases were generated specifically for human-restricted Typhi and host-generalist Typhimurium. These were curated by comparison with EcoCyc, a

comprehensive single-organism metabolic database for *E. coli*, a closely related member of the Enterobacteriaceae. In total, each *Salmonella* database describes 200 or more pathways and their associated genes and enzymes, as well as over 130 transport reactions. A direct comparison of the two serovars indicated that they only differ by the presence or absence of 5 pathways. However, for the first time, the position of all Typhi pseudogenes within metabolic pathways could be established, and a further 11 pathways were found to be inactivated in Typhi that remain intact in Typhimurium. These indicate the metabolic capacity no longer available to Typhi, and shed light upon how Typhi exists in the human host.

Having *Salmonella*-specific pathway databases also provided great value to the assessment of host-restricted *Salmonella* using high-throughput metabolic phenotyping. The phenotyping alone demonstrated that, over a wide range of substrates, host-restricted serovars displayed a metabolic capacity approximately 2/3 of that of a host-generalist, which supports the evidence from the older literature. Using the databases, pathways involved in the breakdown of substrates utilised by the host-generalist alone could be readily found and pseudogenes often identified that provided the cause of the loss of function in the host-specialists. Multiple cases of different pseudogenes occurring in the same metabolic pathway were observed, suggesting that these serovars have convergently lost specific metabolic functions. Particularly striking was the inactivation of transport reactions and regulation, which may reduce the likelihood of potentially toxic pathway intermediates building up inside the cell. Where pathways remained intact yet substrates were clearly not utilised by the host-restricted serovars, further investigation should be

directed to look for associated regulators and transporters whose function may have been inactivated.

In their own right, the pathway databases represent useful resources to the *Salmonella* research community, and as such they are publically available on the internet. They can also be used as a basis for generating other *Salmonella*-specific databases, as has already been undertaken for *S. bongori*.

In conjunction with the work on reconstructing metabolic pathways, high density transposon mutagenesis was used to investigate the essential gene complement of Typhi and Typhimurium, and to look for the genes required for infection of human macrophages. Originally, the transposon libraries were assayed using tiling microarrays, specifically designed to cover the entire genome. The Typhimurium library generated was of an equivalent size to the one previously made in Typhi, of approximately one million mutants. Such a large number was targeted to ensure as few cold spots for insertion as possible, and to increase the likelihood of multiple mutants per gene. However, the traditional method of generating *in vitro* transcribed RNA to hybridise to the microarray did not provide sufficient resolution to identify the transposon insertion sites in so many mutants. High background levels meant that the on/off signals overlapped by too much to determine a statistically robust cutoff. Hence, a novel technique that made use of the more 'digital' nature of next-generation sequencing was developed to deal with the transposon libraries, and simultaneously identify the position of every transposon insertion. This was termed transposon directed insertion-site sequencing (TraDIS) which makes use of the Illumina sequencing platform and is based upon generating short sequence reads oriented outward from the transposon into the

adjacent genomic DNA. Each read mapped to the bacterial genome contains a 10 bp 'tag' of transposon sequence, followed by genomic DNA and hence is almost certainly an indication of the exact position of an insertion site. With this approach, up to half a million unique transposon insertions were identified in the Typhimurium and Typhi libraries. These insertions demarcated 318 candidate essential genes in Typhimurium and 356 in Typhi, where the increased number suggested loss of redundant function through pseudogene formation. While this was partly the case, only 267 candidate essential genes were shared between the two serovars, and each contained 5-10 that were not present at all in the other serovar. In addition, almost 50 candidate essential genes were identified as essential only in one serovar, indicating that the same gene products have a different phenotypic effect in the different serovar backgrounds. By achieving such a high density of insertions in the transposon libraries, the genome-wide analysis of small RNAs also becomes possible. Such an analysis is part of future analysis planned for the two libraries described here, but should be feasible in any transposon-mutant pool large enough to ensure multiple insertions in regions of less than 300 bp. Indeed, TraDIS is applicable to any transposon-based mutagenesis, although it may be limited by the size of the mutant pool achievable with a particular organism and transposon.

The *Salmonella* transposon mutant libraries are valuable biological resources themselves, and have uses beyond those described in this thesis. Theoretically capable of identifying genes required in any selective screen, one or both libraries are being used to look at antibiotic resistance, serum killing and quorum sensing. However, the utility of having mutant libraries large enough to allow the assay of almost every gene was tested here by screening them through human macrophages, a condition relevant to typhoid infection.

Over 300 candidate genes were identified in Typhimurium, but only 8 in Typhi, suggesting that the method of host cell infection is different in these serovars; Typhimurium actively invades while it is more likely that Typhi is taken up solely by phagocytosis. Close examination of the genes required by Typhimurium indicate that metabolism of vitamin B12 and propanediol play an important role – pathways that are inactive in Typhi. Since propanediol provides energy as a carbon source, it is most likely utilised to support Typhimurium growth (i.e. replication) inside the macrophage. The lack of genes identified in Typhi suggests that this serovar does not actively invade, and does not replicate once inside the host cell, probably because to do so requires metabolic pathways that are no longer active. The question then becomes whether such inactivating mutations are selected for, or occur by genetic drift. The fact that all host-restricted serovars studied thus far are deficient in vitamin B12 and propanediol utilisation points towards the former, and examination of more host-restricted serovars will provide further insight.

In conclusion, the link between host restriction and reduced metabolic capability in *Salmonella enterica* serovars can be largely explained by the presence of pseudogenes in metabolic pathways, transporters and associated regulators. The effect of these pseudogenes upon *Salmonella* infection of human macrophages has been explored using large transposon mutant libraries which allow every gene in the bacterial genome to be assayed simultaneously. These libraries have also been used to generate candidate essential gene lists for Typhi and Typhimurium, revealing key differences that suggest the behaviour of a host-generalist cannot always be used to predict the behaviour of a host-specialist.

The metabolic pathway databases and TraDIS technique presented here provide resources and methods that better enable the investigation of the biology behind host-restricted *Salmonella* serovars. The generation of more transposon libraries in host-adapted and host-generalist serovars will also shed greater light on this, alongside further selective screens through biologically relevant conditions.

6.1 Publications arising from this thesis

Langridge GC, Wain J, and Nair S. 2008. Invasive Salmonellosis in Humans. Chapter 8.6.2.2 (revised version) In *EcoSal - Escherichia coli and Salmonella: cellular and molecular biology* (eds. A. Böck, R. Curtiss III, J.B. Kaper, F.C. Neidhardt, T. Nyström, K.E. Rudd, and C.L. Squires). posted 18 August. ASM Press, Washington DC.

Holt KE, Thomson NR, Wain J, Langridge GC, Hasan R, Bhutta ZA, Quail MA, Norbertczak H, Walker D, Simmonds M, White B, Bason N, Mungall K, Dougan G & Parkhill J. 2009. Pseudogene accumulation in the evolutionary histories of *Salmonella enterica* serovars Paratyphi A and Typhi. *BMC Genomics* **10:**36: doi:10.1186/1471-2164-1110-1136.

Langridge GC, Nair S, and Wain J. 2009. Nontyphoidal *Salmonella* Serovars Cause Different Degrees of Invasive Disease Globally. *The Journal of Infectious Diseases* **199:**602-603.

Langridge GC*, Phan MD*, Turner DJ*, Perkins TT, Parts L, Haase J, Charles I, Maskell DJ, Peters SE, Dougan G, Wain J & Parkhill J. 2009. Simultaneous assay of every *Salmonella* Typhi gene using one million transposon mutants. *Genome Research* **19:**2308-2316.

* equal first authors

Langridge GC, Turner DJ, Phan MD, Turner AK, Parkhill J & Wain J. 2010. An Assay of Every Gene Shows That *Salmonella* Serovars Typhi and Typhimurium Have Unique but Overlapping Sets of Essential Genes. *Submitted*.

Chaudhuri RR, Morgan E, Peters SE, Pleasance SJ, Hudson DL, Davies HM, Wang J, van Diemen PM, Buckley AM, Bowen AJ, Turner DJ, Langridge GC, Turner AK, Parkhill J, Charles IG, Maskell DJ & Stevens MP. 2010. Genes required for colonization of food animals by *Salmonella* Typhimurium. *Submitted.*

7 References

AbuOun M, Suthers PF, Jones GI, Carter BR, Saunders MP et al. 2009. Genome Scale Reconstruction of a Salmonella Metabolic Model. *Journal of Biological Chemistry* **284:**29480-29488.

Achtman M. 2008. Evolution, Population Structure, and Phylogeography of Genetically Monomorphic Bacterial Pathogens. *Annual Review of Microbiology* **62:**53-70.

Alpuche-Aranda CM, Racoosin EL, Swanson JA, and Miller SI. 1994. Salmonella stimulate macrophage macropinocytosis and persist within spacious phagosomes. *The Journal of Experimental Medicine* **179:**601-608.

Anderson ES and Felix A. 1953. The Vi Type-determining Phages carried by Salmonella typhi. *Journal of General Microbiology* **9:**65-88.

Anderson ES, Ward LR, Saxe MJ, and De Sa JD. 1977. Bacteriophage-typing designations of Salmonella typhimurium. *The Journal of Hygiene* **78:**297-300.

Andrews-Polymenis HL, Santiviago CA, and McClelland M. 2009. Novel genetic tools for studying food-borne Salmonella. *Current Opinion in Biotechnology* 149-157

Arakawa K, Yamada Y, Shinoda K, Nakayama Y, and Tomita M. 2006. GEM System: automatic prototyping of cell-wide metabolic pathway models from genomes. *BMC Bioinformatics* **7:**168, doi:110.1186/1471-2105-1187-1168.

Arricau N, Hermant D, Waxin H, Ecobichon C, Duffey PS et al. 1998. The RcsB–RcsC regulatory system of Salmonella typhi differentially modulates the expression of invasion proteins, flagellin and Vi antigen in response to osmolarity. *Molecular Microbiology* **29:**835-850.

Assefa S, Keane TM, Otto TD, Newbold C, and Berriman M. 2009. ABACAS: algorithm-based automatic contiguation of assembled sequences. *Bioinformatics* **25:**1968-1969.

Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y et al. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2:**2006.0008, doi:2010.1038/msb4100050.

Baker S, Holt K, van de Vosse E, Roumagnac P, Whitehead S et al. 2008. High-throughput genotyping of Salmonella enterica serovar Typhi allowing

geographical assignment of haplotypes and pathotypes within an urban district of Jakarta, Indonesia. *Journal of Clinical Microbiology* **46:**1741-1746.

Banker DD. 1955. Paratyphoid-A phage typing. *Nature* **175:**309-310.

Baumler AJ, Kusters JG, Stojiljkovic I, and Heffron F. 1994. Salmonella typhimurium loci involved in survival within macrophages. *Infection and Immunity* **62:**1623-1630.

Becker D, Selbach M, Rollenhagen C, Ballmaier M, Meyer TF et al. 2006. Robust Salmonella metabolism limits possibilities for new antimicrobials. *Nature* **440:**303-307.

Bhutta ZA and Threlfall J. 2009. Addressing the global disease burden of typhoid fever. *Journal of the American Medical Association* **302:**898-899.

Bobik TA, Havemann GD, Busch RJ, Williams DS, and Aldrich HC. 1999. The Propanediol Utilization (pdu) Operon of Salmonella enterica Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1,2-Propanediol Degradation. *Journal of Bacteriology* **181:**5967-5975.

Bonfield JK, Smith KF, and Staden R. 1995. A new DNA sequence assembly program. *Nucleic Acids Research* **23:**4992-4999.

Borman EK, Wheeler KM, West DE, and Mickle FL. 1943. Salmonella Typing in a Public Health Laboratory. *American Journal of Public Health* **33:**127-134.

Borrego JJ, Castro D, Jimenez-Notario M, Luque A, Martinez-Manzanares E et al. 1992. Comparison of epidemiological markers of Salmonella strains isolated from different sources in Spain. *Journal of Clinical Microbiology* **30:**3058-3064.

Boyd EF, Wang FS, Beltran P, Plock SA, Nelson K et al. 1993. Salmonella reference collection B (SARB): strains of 37 serovars of subspecies I. *Journal of General Microbiology* **139 (Pt 6):**1125-1132.

Bregeon D, Colot V, Radman M, and Taddei F. 2001. Translational misreading: a tRNA modification counteracts a +2 ribosomal frameshift. *Genes & Development* **15:**2295-2306.

Brown M and Eykyn SJ. 2000. Non-typhoidal Salmonella bacteraemia without gastroenteritis: a marker of underlying immunosuppression. Review Of cases at St. Thomas' Hospital 1970-1999. *Journal of Infection* **41:**256-259.

Brunner F, Margadant A, Peduzzi R, and Piffaretti JC. 1983. The plasmid pattern as an epidemiologic tool for Salmonella typhimurium epidemics: comparison with the lysotype. *Journal of Infectious Diseases* **148:**7-11.

Buchmeier NA and Heffron F. 1990. Induction of Salmonella Stress Proteins upon Infection of Macrophages. *Science* **248:**730-732.

Buchmeier NA and Heffron F. 1991. Inhibition of macrophage phagosome-lysosome fusion by Salmonella typhimurium. *Infection and Immunity* **59:**2232-2238.

Butler T, Knight J, Nath SK, Speelman P, Roy SK et al. 1985. Typhoid fever complicated by intestinal perforation: a persisting fatal disease requiring surgical management. *Reviews of Infectious Diseases* **7:**244-256.

Callow BR. 1959. A new phage-typing scheme for Salmonella typhimurium. *The Journal of Hygiene* **57:**346-359.

Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG et al. 2005. ACT: the Artemis comparison tool. *Bioinformatics* **21:**3422-3423.

Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA et al. 2008. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Research* **38:**D473-479.

Caygill CP, Braddick M, Hill MJ, Knowles RL, and Sharp JC. 1995. The association between typhoid carriage, typhoid infection and subsequent cancer at a number of sites. *European Journal of Cancer Prevention* **4:**187-193.

Chain PSG, Carniel E, Larimer FW, Lamerdin J, Stoutland PO et al. 2004. Insights into the evolution of Yersinia pestis through whole-genome comparison with Yersinia pseudotuberculosis. *Proceedings of the National Academy of Sciences USA* **101:**13826-13831

Chanh NQ, Everest P, Tan Khoa T, House D, Murch S et al. 2004. A Clinical, Microbiological, and Pathological Study of Intestinal Perforation Associated with Typhoid Fever. *Clinical Infectious Diseases* **39:**61-67.

Charles IG and Maskell DJ. Transposon mediated differential hybridisation. Patent No. WO2001/007651 (2001).

Chassagnole C, Noisommit-Rizzi N, Schmid JW, Mauch K, and Reuss M. 2002. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnology and Bioengineering* **79:**53-73.

Chaudhuri R, Allen A, Owen PJ, Shalom G, Stone K et al. 2009. Comprehensive identification of essential *Staphylococcus* aureus genes using Transposon-Mediated Differential Hybridisation (TMDH). *BMC Genomics* **10:**291, doi:210.1186/1471-2164-1110-1291.

Chiu CH, Chuang CH, Chiu S, Su LH, and Lin TY. 2006. Salmonella enterica serotype Choleraesuis infections in pediatric patients. *Pediatrics* **117:**e1193-1196.

Chiu CH, Su LH, and Chu C. 2004a. Salmonella enterica serotype Choleraesuis: epidemiology, pathogenesis, clinical disease, and treatment. *Clinical Microbiology Reviews* **17:**311-322.

Chiu CH, Tang P, Chu C, Hu S, Bao Q et al. 2005. The genome sequence of Salmonella enterica serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic Acids Research* **33:**1690-1698.

Chiu S, Chiu CH, and Lin TY. 2004b. Salmonella enterica serotype Choleraesuis infection in a medical center in northern Taiwan. *Journal of Microbiology, Immunology and Infection* **37:**99-102.

Cirillo DM, Valdivia RH, Monack DM, and Falkow S. 1998. Macrophage-dependent induction of the *Salmonella* pathogenicity island 2 type III secretion system and its role in intracellular survival. *Molecular Microbiology* **30:**175-188.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B et al. 2009. The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* **37:**D141-145.

Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C et al. 1998. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature* **393:**537-544.

Cole ST, Eiglmeier K, Parkhill J, James KD, Thomson NR et al. 2001. Massive gene decay in the leprosy bacillus. *Nature* **409:**1007-1011.

Crosa JH, Brenner DJ, Ewing WH, and Falkow S. 1973. Molecular relationships among the Salmonelleae. *Journal of Bacteriology* **115:**307-315.

Crump JA, Luby SP, and Mintz ED. 2004. The global burden of typhoid fever. *Bulletin of the World Health Organization* **82:**346-353.

Dagan T, Blekhman R, and Graur D. 2006. The" domino theory" of gene death: gradual and mass gene extinction events in three lineages of obligate symbiotic bacterial pathogens. *Molecular Biology and Evolution* **23:**310-316.

Daigle F, Graham JE, and Curtiss R. 2001. Identification of *Salmonella typhi* genes expressed within macrophages by selective capture of transcribed sequences (SCOTS). *Molecular Microbiology* **41:**1211-1222.

Darwin KH and Miller VL. 1999. Molecular basis of the interaction of Salmonella with the intestinal mucosa. *Clinical Microbiology Reviews* **12:**405-428.

de Berardinis V, Vallenet D, Castelli V, Besnard M, Pinet A et al. 2008. A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1. *Molecular Systems Biology* **4:**174, doi:110.1038/msb.2008.1010.

Deng W, Liou S-R, Plunkett Iii G, Mayhew GF, Rose DJ et al. 2003. Comparative Genomics of Salmonella enterica Serovar Typhi Strains Ty2 and CT18. *Journal of Bacteriology* **185:**2330-2337.

Diaz E, Ferrandez A, Prieto MA, and Garcia JL. 2001. Biodegradation of Aromatic Compounds by Escherichia coli. *Microbiology and Molecular Biology Reviews* **65:**523-569.

Didelot X, Achtman M, Parkhill J, Thomson NR, and Falush D. 2007. A bimodal pattern of relatedness between the Salmonella Paratyphi A and Typhi genomes: Convergence or divergence by homologous recombination? *Genome Research* **17:**61-68.

Drecktrah D, Knodler LA, Howe D, and Steele-Mortimer O. 2007. Salmonella trafficking is defined by continuous dynamic interactions with the endolysosomal system. *Traffic* **8:**212-225.

Dutta U, Garg PK, Kumar R, and Tandon RK. 2000. Typhoid carriers among patients with gallstones are at increased risk for carcinoma of the gallbladder. *American Journal of Gastroenterology* **95:**784-787.

Dziejman M, Balon E, Boyd D, Fraser CM, Heidelberg JF et al. 2002. Comparative genomic analysis of Vibrio cholerae: Genes that correlate with cholera endemic and pandemic disease. *Proceedings of the National Academy of Sciences USA* **99:**1556-1561.

Earhart CF. 1996. Uptake and metabolism of iron and molybdenum, Chapter 71. In *Escherichia coli and Salmonella: Cellular and Molecular Biology* (eds. F.C. Neidhardt R. Curtiss III J.L. Ingraham E.C.C. Lin K.B. Low B. Magasanik W.S. Reznikoff M. Riley E. Schaechter, and H.E. Umbarger), pp. 1075-1090. ASM Press, Washington, D.C.

Echols H and Green L. 1971. Establishment and maintenance of repression by bacteriophage lambda. The role of cI, clI and cIII proteins. *Proceedings of the National Academy of Science USA* **68:**2190-2194.

Eriksson S, Lucchini S, Thompson A, Rhen M, and Hinton JCD. 2003. Unravelling the biology of macrophage infection by gene expression profiling of intracellular Salmonella enterica. *Molecular Microbiology* **47:**103-118.

Eswarappa SM, Negi VD, Chakraborty S, Chandrasekhar Sagar BK, and Chakravortty D. 2010. Division of the Salmonella-Containing Vacuole and Depletion of Acidic Lysosomes in Salmonella-Infected Host Cells Are Novel Strategies of Salmonella enterica To Avoid Lysosomes. *Infection and Immunity* **78:**68-79.

Everest P, Wain J, Roberts M, Rook G, and Dougan G. 2001. The molecular mechanisms of severe typhoid fever. *Trends in Microbiology* **9:**316-230.

Falush D, Torpdahl M, Didelot X, Conrad DF, Wilson DJ et al. 2006. Mismatch induced speciation in Salmonella: model and data. *Philosophical Transactions of the Royal Society of London B Biological Sciences* **361:**2045-2053.

Faucher SP, Curtiss III R, and Daigle F. 2005. Selective capture of *Salmonella enterica* serovar Typhi genes expressed in macrophages that are absent from the *Salmonella enterica* serovar Typhimurium genome. *Infection and Immunity* **73:**5217.

Faucher SP, Porwollik S, Dozois CM, McClelland M, and Daigle F. 2006. Transcriptome of Salmonella enterica serovar Typhi within macrophages revealed through the

selective capture of transcribed sequences. *Proceedings of the National Academy of Sciences USA* **103:**1906-1911.

Felix A. 1943. Typing of Typhoid Bacilli by Vi Bacteriophage. *British Medical Journal* **1:**435-438.

Felix A and Callow BR. 1943. Typing of Paratyphoid B Bacilli by Vi Bacteriophage. *British Medical Journal* **2:**127.

Fields PI, Swanson RV, Haidaris CG, and Heffron F. 1986. Mutants of Salmonella typhimurium that cannot survive within the macrophage are avirulent. *Proceedings of the National Academy of Sciences USA* **83:**5189-5193.

Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ et al. 2007. The Pfam protein families database. *Nucleic Acids Research* **30:**276-280.

Freeman JA, Ohl ME, and Miller SI. 2003. The Salmonella enterica serovar Typhimurium translocated effectors SseJ and SifB are targeted to the Salmonella-containing vacuole. *Infection and Immunity* **71:**418.

French CT, Lao P, Loraine AE, Matthews BT, Yu H et al. 2008. Large-scale transposon mutagenesis of Mycoplasma pulmonis. *Molecular Microbiology* **69:**67-76.

Galan JE, Ginocchio C, and Costeas P. 1992. Molecular and functional characterization of the Salmonella invasion gene invA: homology of InvA to members of a new protein family. *Journal of Bacteriology* **174:**4338-4349.

Gallagher LA, Ramage E, Jacobs MA, Kaul R, Brittnacher M et al. 2007. A comprehensive transposon mutant library of Francisella novicida, a bioweapon surrogate. *Proceedings of the National Academy of Science USA* **104:**1009-1014.

Garcia-del Portillo F and Finlay BB. 1994. Salmonella invasion of nonphagocytic cells induces formation of macropinosomes in the host cell. *Infection and Immunity* **62:**4641-4645.

Giammanco GM, Pignato S, Mammina C, Grimont F, Grimont PAD et al. 2002. Persistent Endemicity of Salmonella bongori 48: z35:- in Southern Italy: Molecular Characterization of Human, Animal, and Environmental Isolates. *Journal of Clinical Microbiology* **40:**3502-3505.

Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR et al. 2006. Essential genes of a minimal bacterium. *Proceedings of the National Academy of Science USA* **103:**425-430.

Gordon MA, Banda HT, Gondwe M, Gordon SB, Boeree MJ et al. 2002. Non-typhoidal salmonella bacteraemia among HIV-infected Malawian adults: high mortality and frequent recrudescence. *Aids* **16:**1633-1641.

Gorvel J-P and Méresse S. 2001. Maturation steps of the Salmonella-containing vacuole. *Microbes and Infection* **3:**1299-1303.

Graham SM, Walsh AL, Molyneux EM, Phiri AJ, and Molyneux ME. 2000. Clinical presentation of non-typhoidal Salmonella bacteraemia in Malawian children. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **94:**310-314.

Green M and Karp P. 2004. A Bayesian method for identifying missing enzymes in predicted metabolic pathway databases. *BMC Bioinformatics* **5:**76 doi:10.1186/1471-2105-1185-1176.

Grimont PAD and Weill FX. 2007. Antigenic formulae of the Salmonella serovars. *WHO Collaborating Centre for Reference and Research on Salmonella, Institut Pasteur* **156**.

Guibourdenche M, Roggentin P, Mikoleit M, Fields PI, Bockemühl J et al. 2009. Supplement 2003-2007 (No. 47) to the White-Kauffmann-Le Minor scheme. *Research in Microbiology* **161:**26-29.

Gulig PA and Curtiss 3rd R. 1987. Plasmid-associated virulence of Salmonella typhimurium. *Infection and Immunity* **55:**2891-2901.

Gulig PA, Danbara H, Guiney DG, Lax AJ, Norel F et al. 1993. Molecular analysis of spv virulence genes of the Salmonella virulence plasmids. *Molecular Microbiology* **7:**825-830.

Gutnick D, Calvo JM, Klopotowski T, and Ames BN. 1969. Compounds which serve as the sole source of carbon or nitrogen for Salmonella typhimurium LT-2. *Journal of Bacteriology* **100:**215-219.

Han KH, Choi SY, Lee JH, Lee H, Shin EH et al. 2006. Isolation of Salmonella enterica subspecies enterica serovar Paratyphi B dT+, or Salmonella Java, from Indonesia

and alteration of the d-tartrate fermentation phenotype by disrupting the ORF STM 3356. *Journal of Medical Microbiology* **55:**1661-1665.

Hantke K. 1983. Identification of an iron uptake system specific for coprogen and rhodotorulic acid in Escherichia coli K12. *Molecular and General Genetics* **191:**301-306.

Hantke K, Nicholson G, Rabsch W, and Winkelmann G. 2003. Salmochelins, siderophores of Salmonella enterica and uropathogenic Escherichia coli strains, are recognized by the outer membrane receptor IroN. *Proceedings of the National Academy of Science USA* **100:**3677-3682

Hardt WD, Chen LM, Schuebel KE, Bustelo XR, and Galan JE. 1998. S. typhimurium encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* **93:**815-826.

Harris SR, Feil EJ, Holden MTG, Quail MA, Nickerson EK et al. 2010. Evolution of MRSA During Hospital Transmission and Intercontinental Spread. *Science* **327:**469-474.

Hashimoto Y, Li N, Yokoyama H, and Ezaki T. 1993. Complete nucleotide sequence and molecular characterization of ViaB region encoding Vi antigen in Salmonella typhi. *Journal of Bacteriology* **175:**4456-4465.

Hautefort I, Thompson A, Eriksson-Ygberg S, Parker ML, Lucchini S et al. 2008. During infection of epithelial cells Salmonella enterica serovar Typhimurium undergoes a time-dependent transcriptional adaptation that results in simultaneous expression of three type 3 secretion systems. *Cellular Microbiology* **10:**958-984.

Havemann GD and Bobik TA. 2003. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1, 2-propanediol in Salmonella enterica serovar Typhimurium LT2. *Journal of Bacteriology* **185:**5086-5095.

Helms M, Ethelberg S, and Molbak K. 2005. International Salmonella typhimurium DT104 infections, 1992–2001. *Emerging Infectious Diseases* **11:**859-867.

Hensel M, Hinsley AP, Nikolaus T, Sawers G, and Berks BC. 1999. The genetic basis of tetrathionate respiration in Salmonella typhimurium. *Molecular Microbiology* **32:**275-287.

Hensel M, Shea JE, Baumler AJ, Gleeson C, Blattner F et al. 1997a. Analysis of the boundaries of Salmonella pathogenicity island 2 and the corresponding chromosomal region of Escherichia coli K-12. *Journal of Bacteriology* **179:**1105-1111.

Hensel M, Shea JE, Gleeson C, Jones MD, Dalton E et al. 1995. Simultaneous identification of bacterial virulence genes by negative selection. *Science* **269:**400-403

Hensel M, Shea JE, Raupach B, Monack D, Falkow S et al. 1997b. Functional analysis of ssaJ and the ssaK/U operon, 13 genes encoding components of the type III secretion apparatus of Salmonella pathogenicity island 2. *Mol Microbiol* **24:**155-167.

Hensel M, Shea JE, Waterman SR, Mundy R, Nikolaus T et al. 1998. Genes encoding putative effector proteins of the type III secretion system of *Salmonella* pathogenicity island 2 are required for bacterial virulence and proliferation in macrophages. *Molecular Microbiology* **30:**163-174.

Holden DW. 2002. Trafficking of the *Salmonella* Vacuole in Macrophages. *Traffic* **3:**161-169.

Holmberg SD, Wachsmuth IK, Hickman-Brenner FW, and Cohen ML. 1984. Comparison of plasmid profile analysis, phage typing, and antimicrobial susceptibility testing in characterizing Salmonella typhimurium isolates from outbreaks. *Journal of Clinical Microbiology* **19:**100-104.

Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX et al. 2008. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. *Nature Genetics* **40:**987-993.

Holt KE, Teo YY, Li H, Nair S, Dougan G et al. 2009a. Detecting SNPs and estimating allele frequencies in clonal bacterial populations by sequencing pooled DNA. *Bioinformatics* **25:**2074-2075

Holt KE, Thomson NR, Wain J, Langridge GC, Hasan R et al. 2009b. Pseudogene accumulation in the evolutionary histories of Salmonella enterica serovars Paratyphi A and Typhi. *BMC Genomics* **10:**36: doi:10.1186/1471-2164-1110-1136.

Horne SM, Kottom TJ, Nolan LK, and Young KD. 1997. Decreased intracellular survival of an fkpA mutant of Salmonella typhimurium Copenhagen. *Infection and Immunity* **65:**806-810.

House D, Bishop A, Parry C, Dougan G, and Wain J. 2001. Typhoid fever: pathogenesis and disease. *Current Opinion in Infectious Diseases* **14:**573-578.

Hubbard BK, Koch M, Palmer DRJ, Babbitt PC, and Gerlt JA. 1998. Evolution of Enzymatic Activities in the Enolase Superfamily: Characterization of the (D)-Glucarate/Galactarate Catabolic Pathway in Escherichia coli. *Biochemistry* **37:**14369-14375.

Humphreys S, Rowley G, Stevenson A, Kenyon WJ, Spector MP et al. 2003. Role of Periplasmic Peptidylprolyl Isomerases in Salmonella enterica Serovar Typhimurium Virulence. *Infection and Immunity* **71:**5386-5388.

Hyduke DR, Jarboe LR, Tran LM, Chou KJY, and Liao JC. 2007. Integrated network analysis identifies nitric oxide response networks and dihydroxyacid dehydratase as a crucial target in Escherichia coli. *Proceedings of the National Academy of Sciences USA* **104:**8484-8489.

Ishibashi Y and Arai T. 1990. Specific inhibition of phagosome-lysosome fusion in murine macrophages mediated by Salmonella typhimurium infection. *FEMS Microbiology Letters* **64:**35-43.

Jin Y. 2008. Enteric Fever in South China: Guangxi Province. *Journal of Infection in Developing Countries* **2:**283-288.

Kado CI and Liu ST. 1981. Rapid procedure for detection and isolation of large and small plasmids. *Journal of Bacteriology* **145:**1365-1373.

Kahm M, Hasenbrink G, Lichtenberg-Fraté H, Ludwig J, and Kschischo M. 2010. grofit: Fitting Biological Growth Curves with R. *Journal of Statistical Software* **33:**1-21.

Kanehisa M and Goto S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28:**27-30.

Kariuki S, Revathi G, Kariuki N, Muyodi J, Mwituria J et al. 2005. Increasing prevalence of multidrug-resistant non-typhoidal salmonellae, Kenya, 1994-2003. *International Journal of Antimicrobial Agents* **25:**38-43.

Karp PD. 2001. Pathway databases: a case study in computational symbolic theories. *Science* **293:**2040-2044.

Karp PD, Paley S, and Romero P. 2002. The pathway tools software. *Bioinformatics* **18:**S225-S232.

Karp PD, Riley M, Paley SM, Pellegrini-Toole A, and Krummenacker M. 1997. EcoCyc: Enyclopedia of Escherichia coli Genes and Metabolism. *Nucleic Acids Research* **25:**43-51.

Keseler IM, Bonavides-Martinez C, Collado-Vides J, Gama-Castro S, Gunsalus RP et al. 2009. EcoCyc: A comprehensive view of Escherichia coli biology. *Nucleic Acids Research* **37:**D464-470.

Khatri NS, Maskey P, Poudel S, Jaiswal VK, Karkey A et al. 2009. Gallbladder Carriage of Salmonella paratyphi A May Be an Important Factor in the Increasing Incidence of This Infection in South Asia. *Annals of Internal Medicine* **150:**567-568.

Kidgell C, Reichard U, Wain J, Linz B, Torpdahl M et al. 2002. Salmonella typhi, the causative agent of typhoid fever, is approximately 50,000 years old. *Infection, Genetics and Evolution* **2:**39-45.

Kingsley RA and Baumler AJ. 2000. Host adaptation and the emergence of infectious disease: the Salmonella paradigm. *Molecular Microbiology* **36:**1006-1014.

Kingsley RA, Humphries AD, Weening EH, De Zoete MR, Winter S et al. 2003. Molecular and phenotypic analysis of the CS54 island of Salmonella enterica serotype Typhimurium: identification of intestinal colonization and persistence determinants. *Infect. Immun.* **71:**629-640.

Kingsley RA, Msefula CL, Thomson NR, Kariuki S, Holt KE et al. 2009. Epidemic multiple drug resistant Salmonella Typhimurium causing invasive disease in sub-Saharan Africa have a distinct genotype. *Genome Research* **19:**2279-2287.

Klumpp J and Fuchs TM. 2007. Identification of novel genes in genomic islands that contribute to Salmonella typhimurium replication in macrophages. *Microbiology* **153:**1207-1220.

Knodler L and Steele-Mortimer O. 2003. Taking Possession: Biogenesis of the *Salmonella*-Containing Vacuole. *Traffic* **4:**587-599.

Konig R and Eils R. 2004. Gene expression analysis on biochemical networks using the Potts spin model. *Bioinformatics* **20:**1500-1505.

Kothari A, Pruthi A, and Chugh TD. 2008. The Burden of Enteric Fever. *Journal of Infection in Developing Countries* **2:**253-259.

Kroger C and Fuchs TM. 2009. Characterization of the myo-Inositol Utilization Island of Salmonella enterica serovar Typhimurium. *Journal of Bacteriology* **191:**545-554.

Kuhle V and Hensel M. 2004. Cellular microbiology of intracellular Salmonella enterica: functions of the type III secretion system encoded by Salmonella pathogenicity island 2. *Cellular and Molecular Life Sciences* **61:**2812-2826.

Laia ML, Moreira LM, Dezajacomo J, Brigati JB, Ferreira CB et al. 2009. New genes of Xanthomonas citri subsp. citri involved in pathogenesis and adaptation revealed by a transposon-based mutant library. *BMC Microbiology* **9:**12, doi:10.1186/1471-2180-1189-1112.

Langridge GC, Nair S, and Wain J. 2009a. Nontyphoidal Salmonella Serovars Cause Different Degrees of Invasive Disease Globally. *The Journal of Infectious Diseases* **199:**602-603.

Langridge GC, Phan MD, Turner DJ, Perkins TT, Parts L et al. 2009b. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research* **19:**2308-2316.

Langridge GC, Wain J, and Nair S. 2008. Invasive Salmonellosis in Humans. Chapter 8.6.2.2 (revised version) In *EcoSal - Escherichia coli and Salmonella: cellular and molecular biology* (eds. A. Böck R. Curtiss III J.B. Kaper F.C. Neidhardt T. Nyström K.E. Rudd, and C.L. Squires). posted 18 August. ASM Press, Washington, D.C.

Lauderdale TL, Aarestrup FM, Chen PC, Lai JF, Wang HY et al. 2006. Multidrug resistance among different serotypes of clinical Salmonella isolates in Taiwan. *Diagnostic Microbiology & Infectious Disease* **55:**149-155.

Lawrence JG, Hendrix RW, and Casjens S. 2001. Where are the pseudogenes in bacterial genomes? *Trends in Microbiology* **9:**535-540.

Le Minor L, Beaud R, Laurent B, and Monteil V. Salmonella possessing the 6, 7: c: 1, 5 antigenic factors. *Annales de l'Institut Pasteur / Microbiologie* **136:**225-234.

Ledeboer NA and Jones BD. 2005. Exopolysaccharide sugars contribute to biofilm formation by Salmonella enterica serovar Typhimurium on HEp-2 cells and chicken intestinal epithelium. *Journal of Bacteriology* **187:**3214-3226.

Lee TJ, Paulsen I, and Karp P. 2008. Annotation-based inference of transporter function. *Bioinformatics* **24:**i259-267.

Lerat E and Ochman H. 2004. Psi-Phi: exploring the outer limits of bacterial pseudogenes. *Genome Research* **14:**2273-2278.

Lerat E and Ochman H. 2005. Recognizing the pseudogenes in bacterial genomes. *Nucleic Acids Research* **33:**3125-3132.

Li H, Ruan J, and Durbin R. 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research* **18:**1851-1858.

Liberati NT, Urbach JM, Miyata S, Lee DG, Drenkard E et al. 2006. An ordered, nonredundant library of Pseudomonas aeruginosa strain PA14 transposon insertion mutants. *Proceedings of the National Academy of Science USA* **103:**2833-2838.

Lindberg AA, Nord CE, Hellgren M, and Sjöberg L. 1974. Identification of gram-negative aerobic fermentors in a clinical bacteriological laboratory. *Medical Microbiology and Immunology* **159:**201-210.

Liu SL and Sanderson KE. 1995a. The chromosome of Salmonella paratyphi A is inverted by recombination between rrnH and rrnG. *Journal of Bacteriology* **177:**6585-6592.

Liu SL and Sanderson KE. 1995b. Genomic cleavage map of Salmonella typhi Ty2. *Journal of Bacteriology* **177:**5099-5107.

Liu SL and Sanderson KE. 1995c. Rearrangements in the genome of the bacterium Salmonella typhi. *Proceedings of the National Academy of Sciences USA* **92:**1018-1022.

Liu SL and Sanderson KE. 1996. Highly plastic chromosomal organization in Salmonella typhi. *Proceedings of the National Academy of Sciences USA* **93:**10303-10308.

Liu WQ, Feng Y, Wang Y, Zou QH, Chen F et al. 2009. Salmonella paratyphi C: Genetic Divergence from Salmonella choleraesuis and Pathogenic Convergence with Salmonella typhi. *PLoS ONE* **4:**e4510: doi:4510.1371/journal.pone.0004510.

Lorenzon S, Arzul I, Peyraud A, Hendrikx P, and Thiaucourt F. 2003. Molecular epidemiology of contagious bovine pleuropneumonia by multilocus sequence analysis of Mycoplasma mycoides subspecies mycoides biotype SC strains. *Veterinary Microbiology* **93:**319-333.

Lowe DC, Savidge TC, Pickard D, Eckmann L, Kagnoff MF et al. 1999. Characterization of Candidate Live Oral Salmonella typhi Vaccine Strains Harboring Defined Mutations in aroA, aroC, and htrA. *Infection and Immunity* **67:**700-707.

Lucchini S, Rowley G, Goldberg MD, Hurd D, Harrison M et al. 2006. H-NS mediates the silencing of laterally acquired genes in bacteria. *PLoS Pathog.* **2:**e81.

Lucchini S, Thompson A, and Hinton JCD. 2001. Microarrays for microbiologists. *Microbiology* **147:**1403-1414.

Maiden MCJ, Bygraves JA, Feil E, Morelli G, Russell JE et al. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences USA* **95:**3140-3145

Maidhof H, Guerra B, Abbas S, Elsheikha HM, Whittam TS et al. 2002. A Multiresistant Clone of Shiga Toxin-Producing Escherichia coli O118:[H16] Is Spread in Cattle and Humans over Different European Countries. *Applied and Environmental Microbiology* **68:**5834-5842.

Maltsev N, Glass E, Sulakhe D, Rodriguez A, Syed MH et al. 2006. PUMA2--grid-based high-throughput analysis of genomes and metabolic pathways. *Nucleic Acids Research* **34:**D369-372.

Martin M, Gibello A, Fernandez J, Ferrer E, and Garrido-Pertierra A. 1991. Catabolism of 3- and 4-hydroxyphenylacetic acid by Klebsiella pneumoniae. *Journal of General Microbiology* **137:**621-628.

Maskey AP, Day JN, Tuan PQ, Thwaites GE, Campbell JI et al. 2006. Salmonella enterica serovar Paratyphi A and S. enterica serovar Typhi cause indistinguishable clinical syndromes in Kathmandu, Nepal. *Clinical Infectious Diseases* **42:**1247-1253.

Mayer C and Boos W. 2005. Hexose/Pentose and Hexitol/Pentitol Metabolism. Chapter 3.4.1. In *EcoSal - Escherichia coli and Salmonella: cellular and molecular*

biology (eds. A. Böck R. Curtiss III J.B. Kaper F.C. Neidhardt T. Nyström K.E. Rudd, and C.L. Squires). posted 29 March. ASM Press, Washington, D.C.

McClelland M, Sanderson KE, Clifton SW, Latreille P, Porwollik S et al. 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of Salmonella enterica that cause typhoid. *Nature Genetics* **36:**1268-1274.

McClelland M, Sanderson KE, Spieth J, Clifton SW, Latreille P et al. 2001. Complete genome sequence of Salmonella enterica serovar Typhimurium LT2. *Nature* **413:**852-856.

McFall E and Newman EB. 1996. Amino acids as Carbon Sources. Chapter 22. In *EcoSal - Escherichia coli and Salmonella: cellular and molecular biology* (eds. A. Böck R. Curtiss III J.B. Kaper F.C. Neidhardt T. Nyström K.E. Rudd, and C.L. Squires). ASM Press, Washington, D.C.

Miao EA and Miller SI. 2000. A conserved amino acid sequence directing intracellular type III secretion by Salmonella typhimurium. *Proceedings of the National Academy of Science USA* **97:**7539-7544

Mills SD and Finlay BB. 1994. Comparison of Salmonella typhi and Salmonella typhimurium invasion, intracellular growth and localization in cultured human epithelial cells. *Microbial Pathogenesis* **17:**409-423.

Moran NA and Plague GR. 2004. Genomic changes following host restriction in bacteria. *Current Opinion in Genetics and Development* **14:**627-633.

Morgan E, Campbell JD, Rowe SC, Bispham J, Stevens MP et al. 2004. Identification of host-specific colonization factors of Salmonella enterica serovar Typhimurium. *Molecular Microbiology* **54:**994-1010.

Morris C, Yip CM, Tsui IS, Wong DK, and Hackett J. 2003. The shufflon of Salmonella enterica serovar Typhi regulates type IVB pilus-mediated bacterial self-association. *Infection and Immunity* **71:**1141-1146.

Morrow BJ, Graham JE, and Curtiss R, III. 1999. Genomic Subtractive Hybridization and Selective Capture of Transcribed Sequences Identify a Novel Salmonella typhimurium Fimbrial Operon and Putative Transcriptional Regulator That Are

Absent from the Salmonella typhi Genome. *Infection and Immunity* **67:**5106-5116.

Muñoz-Elías EJ and McKinney JD. 2006. Carbon metabolism of intracellular bacteria. *Cellular Microbiology* **8:**10-22.

Nair S, Alokam S, Kothapalli S, Porwollik S, Proctor E et al. 2004. Salmonella enterica Serovar Typhi Strains from Which SPI7, a 134-Kilobase Island with Genes for Vi Exopolysaccharide and Other Functions, Has Been Deleted. *Journal of Bacteriology* **186:**3214-3223.

Naser SM, Thompson FL, Hoste B, Gevers D, Dawyndt P et al. 2005. Application of multilocus sequence analysis (MLSA) for rapid identification of Enterococcus species based on rpoA and pheS genes. *Microbiology* **151:**2141-2150.

Navarre WW, Porwollik S, Wang Y, McClelland M, Rosen H et al. 2006. Selective Silencing of Foreign DNA with Low GC Content by the H-NS Protein in Salmonella. *Science* **313:**236-238.

Nishino K, Latifi T, and Groisman EA. 2006. Virulence and drug resistance roles of multidrug efflux systems of Salmonella enterica serovar Typhimurium. *Molecular Microbiology* **59:**126-141.

Nord CE, Lindberg AA, and DahlbÄck A. 1974. Evaluation of five test-kits—API, AuxoTab, Enterotube, PathoTec and R/B—for identification ofEnterobacteriaceae. *Medical Microbiology and Immunology* **159:**211-220.

Novak L and Loubiere P. 2000. The Metabolic Network of Lactococcus lactis: Distribution of 14C-Labeled Substrates between Catabolic and Anabolic Pathways. *Journal of Bacteriology* **182:**1136-1143.

Ochiai RL, Wang X, von Seidlein L, Yang J, Bhutta ZA et al. 2005. Salmonella paratyphi A rates, Asia. *Emerging Infectious Diseases* **11:**1764.

Ochman H, Soncini FC, Solomon F, and Groisman EA. 1996. Identification of a pathogenicity island required for Salmonella survival in host cells. *Proceedings of the National Academy of Science USA* **93:**7800-7804.

Ogata H, Goto S, Sato K, Fujibuchi W, Bono H et al. 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **27:**29-34.

Oh YK, Alpuche-Aranda C, Berthiaume E, Jinks T, Miller SI et al. 1996. Rapid and complete fusion of macrophage lysosomes with phagosomes containing Salmonella typhimurium. *Infection and Immunity* **64:**3877-3883.

Ohlson MB, Fluhr K, Birmingham CL, Brumell JH, and Miller SI. 2005. SseJ deacylase activity by Salmonella enterica serovar Typhimurium promotes virulence in mice. *Infect. Immun.* **73:**6249.

Otto TD, Sanders M, Berriman M, and Newbold C. 2010. Iterative Correction of Reference Nucleotides (iCORN) using second generation sequencing technology. *Bioinformatics* **26:**1704-1707.

Ouzounis CA and Karp PD. 2000. Global properties of the metabolic map of Escherichia coli. *Genome Research* **10:**568-576.

Overbeek R, Larsen N, Pusch GD, D'Souza M, Jr ES et al. 2000. WIT: integrated system for high-throughput genome sequence analysis and metabolic reconstruction. *Nucleic Acids Research* **28:**123-125.

Paley SM and Karp PD. 2006. The Pathway Tools cellular overview diagram and Omics Viewer. *Nucleic Acids Research* **34:**3771-3778.

Palit A, Ghosh S, Dutta S, Sur D, Bhattacharya MK et al. 2006. Increasing prevalence of Salmonella enterica serotype Paratyphi-A in patients with enteric fever in a periurban slum setting of Kolkata, India. *International Journal of Environmental Health Research* **16:**455-459.

Parkhill J, Dougan G, James KD, Thomson NR, Pickard D et al. 2001a. Complete genome sequence of a multiple drug resistant Salmonella enterica serovar Typhi CT18. *Nature* **413:**848-852.

Parkhill J, Sebaihia M, Preston A, Murphy LD, Thomson N et al. 2003. Comparative analysis of the genome sequences of Bordetella pertussis, Bordetella parapertussis and Bordetella bronchiseptica. *Nature Genetics* **35:**32-40.

Parkhill J, Wren BW, Thomson NR, Titball RW, Holden MTG et al. 2001b. Genome sequence of Yersinia pestis, the causative agent of plague. *Nature* **413:**523-527.

Parra-Lopez C, Baer MT, and Groisman EA. 1993. Molecular genetic analysis of a locus required for resistance to antimicrobial peptides in Salmonella typhimurium. *The EMBO Journal* **12:**4053-4062.

Parry CM. 2004. Typhoid Fever. *Current Infectious Disease Reports* **6:**27-33.

Parry CM, Hien TT, Dougan G, White NJ, and Farrar JJ. 2002. Typhoid fever. *New England Journal of Medicine* **347:**1770-1782.

Patrick ME, Adcock PM, Gomez TM, Altekruse SF, Holland BH et al. 2004. Salmonella Enteritidis infections, United States, 1985-1999. *Emerging Infectious Diseases* **10:**1-7.

Perkins TT, Kingsley RA, Fookes MC, Gardner PP, James KD et al. 2009. A Strand-Specific RNA–Seq Analysis of the Transcriptome of the Typhoid Bacillus Salmonella Typhi. *PLoS Genetics* **5:**e1000569. doi:1000510.1001371/journal.pgen.1000569

Phan MD and Wain J. 2008. IncHI plasmids, a dynamic link between resistance and pathogenicity. *Journal of Infection in Developing Countries* **2:**272-278.

Pickard D, Wain J, Baker S, Line A, Chohan S et al. 2003. Composition, Acquisition, and Distribution of the Vi Exopolysaccharide-Encoding Salmonella enterica Pathogenicity Island SPI-7. *Journal of Bacteriology* **185:**5055-5065.

Pier GB, Grout M, Zaidi T, Meluleni G, Mueschenborn SS et al. 1998. Salmonella typhi uses CFTR to enter intestinal epithelial cells. *Nature* **393:**79-82.

Pignato S, Giammanco G, Santangelo C, and Giammanco GM. 1998. Endemic presence of Salmonella bongori 48: z35:-causing enteritis in children in Sicily. *Research in Microbiology* **149:**429-431.

Pinney JW, Shirley MW, McConkey GA, and Westhead DR. 2005. metaSHARK: software for automated metabolic network prediction from DNA sequence and its application to the genomes of Plasmodium falciparum and Eimeria tenella. *Nucleic Acids Research* **33:**1399-1409.

Porwollik S, Boyd EF, Choy C, Cheng P, Florea L et al. 2004. Characterization of Salmonella enterica subspecies I genovars by use of microarrays. *Journal of Bacteriology* **186:**5883-5898.

Porwollik S, Frye J, Florea LD, Blackmer F, and McClelland M. 2003. A non-redundant microarray of genes for two related bacteria. *Nucleic Acids Research* **31:**1869-1876.

Porwollik S, Santiviago CA, Cheng P, Florea L, and McClelland M. 2005. Differences in gene content between Salmonella enterica serovar Enteritidis isolates and comparison to closely related serovars Gallinarum and Dublin. *Journal of Bacteriology* **187:**6545-6555.

Price-Carter M, Tingey J, Bobik TA, and Roth JR. 2001. The alternative electron acceptor tetrathionate supports B12-dependent anaerobic growth of Salmonella enterica serovar Typhimurium on ethanolamine or 1, 2-propanediol. *Journal of Bacteriology* **183:**2463-2475.

Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ et al. 2008. A large genome center's improvements to the Illumina sequencing system. *Nature Methods* **5:**1005-1010

Rabsch W, Andrews HL, Kingsley RA, Prager R, Tschape H et al. 2002. Salmonella enterica Serotype Typhimurium and Its Host-Adapted Variants. *Infection and Immunity* **70:**2249-2255.

Rathman M, Barker LP, and Falkow S. 1997. The unique trafficking pattern of Salmonella typhimurium-containing phagosomes in murine macrophages is independent of the mechanism of bacterial entry. *Infection and Immunity* **65:**1475-1485.

Reed JL and Palsson BO. 2003. Thirteen Years of Building Constraint-Based In Silico Models of Escherichia coli. *Journal of Bacteriology* **185:**2692-2699.

Reeves MW, Evins GM, Heiba AA, Plikaytis BD, and Farmer 3rd JJ. 1989. Clonal nature of Salmonella typhi and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of Salmonella bongori comb. nov. *Journal of Clinical Microbiology* **27:**313-320.

Richarme G, el Yaagoubi A, and Kohiyama M. 1993. The MglA component of the binding protein-dependent galactose transport system of Salmonella typhimurium is a galactose-stimulated ATPase. *Journal of Biological Chemistry* **268:**9473-9477.

Richmond MH and Maaloe O. 1962. The Rate of Growth of Salmonella typhimurium with Individual Carbon Sources Related to Glucose Metabolism or to the Krebs Cycle. *Journal of General Microbiology* **27:**285-297.

Rioux CR and Kadner RJ. 1989. Vitamin B 12 transport in Escherichia coli K12 does not require the btuE gene of the btuCED operon. *Molecular and General Genetics MGG* **217:**301-308.

Rodriguez-Buenfil JC, Alvarez-Fleites M, Villarreal-Morales ZY, and Segura-Correa JC. 2004. Incidence and identification of Salmonella species in pigs on two farm systems in Mexico. *The Veterinary Record* **154:**150-152.

Rohmer L, Fong C, Abmayr S, Wasnick M, Freeman TL et al. 2007. Comparison of Francisella tularensis genomes reveals evolutionary events associated with the emergence of human pathogenic strains. *Genome Biology* **8:**R102: doi:110.1186/gb-2007-1188-1186-r1102.

Roof DM and Roth JR. 1988. Ethanolamine utilization in Salmonella typhimurium. *Journal of Bacteriology* **170:**3855-3863.

Roth JR, Lawrence JG, and Bobik TA. 1996. Cobalamin (coenzyme B12): synthesis and biological significance. *Annual Reviews in Microbiology* **50:**137-181.

Roumagnac P, Weill FX, Dolecek C, Baker S, Brisse S et al. 2006. Evolutionary history of Salmonella typhi. *Science* **314:**1301-1304.

Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P et al. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16:**944-945

Ryu JH and Beuchat LR. 2004. Factors affecting production of extracellular carbohydrate complexes by Escherichia coli O157: H7. *International Journal of Food Microbiology* **95:**189-204.

Sabbagh SC, Forest CG, Lepage C, Leclerc JM, and Daigle F. 2010. So similar, yet so different: uncovering distinctive features in the genomes of Salmonella enterica serovars Typhimurium and Typhi. *FEMS Microbiology Letters* **305:**1-13.

Salama NR, Shepherd B, and Falkow S. 2004. Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. *Journal of Bacteriology* **186:**7926-7935.

Sandler SJ and Marians KJ. 2000. Role of PriA in replication fork reactivation in Escherichia coli. *Journal of Bacteriology* **182:**9-13.

Sandler SJ, Marians KJ, Zavitz KH, Coutu J, Parent MA et al. 1999. dnaC mutations suppress defects in DNA replication-and recombination-associated functions in

priB and priC double mutants in Escherichia coli K-12. *Molecular Microbiology* **34:**91-101.

Sassetti CM, Boyd DH, and Rubin EJ. 2001. Comprehensive identification of conditionally essential genes in mycobacteria. *Proceedings of the National Academy of Science USA* **98:**12712-12717.

Schaberg DR, Tompkins LS, and Falkow S. 1981. Use of agarose gel electrophoresis of plasmid deoxyribonucleic acid to fingerprint gram-negative bacilli. *Journal of Clinical Microbiology* **13:**1105-1108.

Schramm G, Zapatka M, Eils R, and Konig R. 2007. Using gene expression data and network topology to detect substantial pathways, clusters and switches during oxygen deprivation of Escherichia coli. *BMC Bioinformatics* **8:**149: doi:110.1186/1471-2105-1188-1149.

Scott F, Threlfall J, and Arnold C. 2002. Genetic structure of Salmonella revealed by fragment analysis. *International Journal of Systematic and Evolutionary Microbiology* **52:**1701-1713.

Selander RK, Beltran P, Smith NH, Helmuth R, Rubin FA et al. 1990. Evolutionary genetic relationships of clones of Salmonella serovars that cause human typhoid and other enteric fevers. *Infection and Immunity* **58:**2262-2275.

Selander RK and Levin BR. 1980. Genetic diversity and structure in Escherichia coli populations. *Science* **210:**545-547.

Selander RK, Li J, and Nelson K. 1996. Evolutionary Genetics of *Salmonella enterica*. In *EcoSal - Escherichia coli and Salmonella: cellular and molecular biology* (eds. A. Böck R. Curtiss III J.B. Kaper F.C. Neidhardt T. Nyström K.E. Rudd, and C.L. Squires). ASM Press, Washington, D.C.

Selkov E, Jr., Grechkin Y, Mikhailova N, and Selkov E. 1998. MPW: the Metabolic Pathways Database. *Nucleic Acids Research* **26:**43-45.

Seth-Smith HMB. 2008. SPI-7: Salmonella’s Vi-Encoding Pathogenicity Island. *Journal of Infection in Developing Countries* **2:**267-271.

Shakeri-Garakani A, Brinkkötter A, Schmid K, Turgut S, and Lengeler JW. 2004. The genes and enzymes for the catabolism of galactitol, D-tagatose, and related

carbohydrates in Klebsiella oxytoca M5a1 and other enteric bacteria display convergent evolution. *Molecular Genetics and Genomics* **271:**717-728.

Shamanna DK and Sanderson KE. 1979. Uptake and catabolism of D-xylose in Salmonella typhimurium LT2. *Journal of Bacteriology* **139:**64-70.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT et al. 2003. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Research* **13:**2498-2504.

Shea JE, Hensel M, Gleeson C, and Holden DW. 1996. Identification of a virulence locus encoding a second type III secretion system in Salmonella typhimurium. *Proceedings of the National Academy of Sciences USA* **93:**2593-2597.

Shelobolina ES, Sullivan SA, O'Neill KR, Nevin KP, and Lovley DR. 2004. Isolation, Characterization, and U(VI)-Reducing Potential of a Facultatively Anaerobic, Acid-Resistant Bacterium from Low-pH, Nitrate- and U(VI)-Contaminated Subsurface Sediment and Description of Salmonella subterranea sp. nov. *Applied and Environmental Microbiology* **70:**2959-2965.

Shi J, Blundell TL, and Mizuguchi K. 2001. FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties1. *Journal of Molecular Biology* **310:**243-257.

Shi L, Adkins JN, Coleman JR, Schepmoes AA, Dohnkova A et al. 2006. Proteomic Analysis of Salmonella enterica Serovar Typhimurium Isolated from RAW 264.7 Macrophages. *Journal of Biological Chemistry* **281:**29131-29140.

Shivaprasad HL. 2000. Fowl typhoid and pullorum disease. *Revue scientifique et technique (International Office of Epizootics)* **19:**405-424.

Shukla VK, Singh H, Pandey M, Upadhyay SK, and Nath G. 2000. Carcinoma of the gallbladder--is it a sequel of typhoid? *Digestive Diseases and Sciences* **45:**900-903.

Smith T and TenBroeck C. 1915. A note on the relation between B. pullorum (Rettger) and the fowl typhoid bacillus (Moore). *Journal of Medical Research* **31:**547-555.

Smoot JC, Barbian KD, Van Gompel JJ, Smoot LM, Chaussee MS et al. 2002. Genome sequence and comparative microarray analysis of serotype M18 group A

Streptococcus strains associated with acute rheumatic fever outbreaks. *Proceedings of the National Academy of Sciences USA* **99:**4668-4673.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Statistical Applications in Genetics and Molecular Biology* **3:**1027-1053.

St John-Brooks R and Rhodes M. 1923. The organisms of the fowl typhoid group. *Journal of Pathology and Bacteriology* **26:**433-439.

Sukhan A. 2000. The invasion-associated type III secretion system of Salmonella typhimurium: common and unique features. *Cellular and Molecular Life Sciences* **57:**1033.

Sydenham M, Douce G, Bowe F, Ahmed S, Chatfield S et al. 2000. Salmonella enterica Serovar Typhimurium surA Mutants Are Attenuated and Effective Live Oral Vaccines. *Infection and Immunity* **68:**1109-1115.

Tacket CO, Sztein MB, Losonsky GA, Wasserman SS, Nataro JP et al. 1997. Safety of live oral Salmonella typhi vaccine strains with deletions in htrA and aroC aroD and immune response in humans. *Infection and Immunity* **65:**452-456.

Thompson A, Rowley G, Alston M, Danino V, and Hinton JCD. 2006. Salmonella transcriptomics: relating regulons, stimulons and regulatory networks to the process of infection. *Current Opinion in Microbiology* **9:**109-116.

Thompson FL, Gevers D, Thompson CC, Dawyndt P, Naser S et al. 2005. Phylogeny and Molecular Identification of Vibrios on the Basis of Multilocus Sequence Analysis. *Applied and Environmental Microbiology* **71:**5107-5115.

Thomson N, Baker S, Pickard D, Fookes M, Anjum M et al. 2004. The role of prophage-like elements in the diversity of Salmonella enterica serovars. *Journal of Molecular Biology* **339:**279-300.

Thomson NR, Clayton DJ, Windhorst D, Vernikos G, Davidson S et al. 2008. Comparative genome analysis of Salmonella Enteritidis PT4 and Salmonella Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome Research* **18:**1624-1637.

Threlfall EJ, Hall ML, and Rowe B. 1992. Salmonella bacteraemia in England and Wales, 1981-1990. *Journal of Clinical Pathology* **45:**34-36.

Threlfall EJ, Rowe B, and Ward LR. 1989. Subdivision of Salmonella enteritidis phage types by plasmid profile typing. *Epidemiology and Infection* **102:**459-465.

Townsend SM, Kramer NE, Edwards R, Baker S, Hamlin N et al. 2001. Salmonella enterica Serovar Typhi Possesses a Unique Repertoire of Fimbrial Gene Sequences. *Infection and Immunity* **69:**2894-2901.

Tsolis RM, Baumler AJ, Heffron F, and Stojiljkovic I. 1996. Contribution of TonB-and Feo-mediated iron uptake to growth of Salmonella typhimurium in the mouse. *Infection and Immunity* **64:**4549-4556.

Tsui IS, Yip CM, Hackett J, and Morris C. 2003. The type IVB pili of Salmonella enterica serovar Typhi bind to the cystic fibrosis transmembrane conductance regulator. *Infection and Immunity* **71:**6049-6050.

Turner AK, Nair S, and Wain J. 2006. The acquisition of full fluoroquinolone resistance in Salmonella Typhi by accumulation of point mutations in the topoisomerase targets. *Journal of Antimicrobial Chemotherapy* **58:**733-740.

Uzzau S, Brown DJ, Wallis T, Rubino S, Leori G et al. 2000. Host adapted serotypes of Salmonella enterica. *Epidemiology and Infection* **125:**229-255.

Uzzau S, Hovi M, and Stocker BAD. 1999. Application of ribotyping and IS200 fingerprinting to distinguish the five Salmonella serotype O6, 7: c: 1, 5 groups: Choleraesuis sensu stricto, Choleraesuis var. Kunzendorf, Choleraesuis var. Decatur, Paratyphi C, and Typhisuis. *Epidemiology and Infection* **123:**37-46.

van Helden J, Naim A, Lemer C, Mancuso R, Eldridge M et al. 2001. From molecular activities and processes to biological function. *Briefings in Bioinformatics* **2:**81-93.

Vázquez-Boland JA, Domínguez-Bernal G, González-Zorn B, Kreft J, and Goebel W. 2001. Pathogenicity islands and virulence evolution in Listeria. *Microbes and Infection* **3:**571-584.

Virlogeux I, Waxin H, Ecobichon C, and Popoff MY. 1995. Role of the viaB locus in synthesis, transport and expression of Salmonella typhi Vi antigen. *Microbiology* **141:**3039-3047.

Vollaard AM, Ali S, van Asten HA, Widjaja S, Visser LG et al. 2004. Risk factors for typhoid and paratyphoid fever in Jakarta, Indonesia. *Journal of the American Medical Association* **291:**2607-2615.

Vugia DJ, Samuel M, Farley MM, Marcus R, Shiferaw B et al. 2004. Invasive Salmonella infections in the United States, FoodNet, 1996-1999: incidence, serotype distribution, and outcome. *Clinical Infectious Diseases* **38 S3:**S149-156.

Wain J, Diep TS, Ho VA, Walsh AM, Hoa NTT et al. 1998. Quantitation of bacteria in blood of typhoid fever patients and relationship between counts and clinical features, transmissibility, and antibiotic resistance. *Journal of Clinical Microbiology* **36:**1683-1687.

Wain J, House D, Parkhill J, Parry C, and Dougan G. 2002. Unlocking the genome of the human typhoid bacillus. *Lancet Infectious Diseases* **2:**163-170.

Wain J, House D, Zafar A, Baker S, Nair S et al. 2005. Vi Antigen Expression in Salmonella enterica Serovar Typhi Clinical Isolates from Pakistan. *Journal of Clinical Microbiology* **43:**1158-1165.

Wang JY, Hwang JJ, Hsu CN, Lin LC, and Hsueh PR. 2006. Bacteraemia due to ciprofloxacin-resistant Salmonella enterica serotype Choleraesuis in adult patients at a university hospital in Taiwan, 1996-2004. *Epidemiology and Infection* **134:**977-984.

Ward LR, De Sa JDH, and Rowe B. 1987. A phage-typing scheme for Salmonella enteritidis. *Epidemiology and Infection* **99:**291-294.

Wehland M and Bernhard F. 2000. The RcsAB box: Characterization of a new operator essential for the regulation of exopolysaccharide biosynthesis in enteric bacteria. *Journal of Biological Chemistry* **275:**7013-7020.

Wei J, Goldberg MB, Burland V, Venkatesan MM, Deng W et al. 2003. Complete genome sequence and comparative genomics of Shigella flexneri serotype 2a strain 2457T. *Infection and Immunity* **71:**2775-2786.

Whitfield C and Keenleyside WJ. 1995. Regulation of expression of group IA capsular polysaccharides in Escherichia coli and related extracellular polysaccharides in other bacteria. *Journal of Industrial Microbiology and Biotechnology* **15:**361-371.

Whittam TS, Ochman H, and Selander RK. 1983. Multilocus genetic structure in natural populations of Escherichia coli. *Proceedings of the National Academy of Sciences USA* **80:**1751-1755.

Winslow CE, Kligler IJ, and Rothberg W. 1919. Studies on the classification of the colon-typhoid with special reference to their fermentative reactions. *Journal of Bacteriology* **4:**429-503.

Wood MW, Jones MA, Watson PR, Siber AM, McCormick BA et al. 2000. The secreted effector protein of *Salmonella* dublin, SopA, is translocated into eukaryotic cells and influences the induction of enteritis. *Cellular Microbiology* **2:**293-303.

Woods CW, Murdoch DR, Zimmerman MD, Glover WA, Basnyat B et al. 2006. Emergence of Salmonella enterica serotype Paratyphi A as a major cause of enteric fever in Kathmandu, Nepal. *Transactions of the Royal Society of Tropical Medicine and Hygiene* **100:**1063-1067.

Woodward DL, Khakhria R, and Johnson WM. 1997. Human salmonellosis associated with exotic pets. *Journal of Clinical Microbiology* **35:**2786-2790.

Zerbino DR and Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18:**821-829.

Zhang H, Zhou Y, Bao H, and Liu H. 2006. Vi antigen biosynthesis in Salmonella typhi: characterization of UDP-N-acetylglucosamine C-6 dehydrogenase (TviB) and UDP-N-acetylglucosaminuronic acid C-4 epimerase (TviC). *Biochemistry* **45:**8163-8173.

Zhang S, Kingsley RA, Santos RL, Andrews-Polymenis H, Raffatellu M et al. 2003. Molecular Pathogenesis of Salmonella enterica Serotype Typhimurium-Induced Diarrhea. *Infection and Immunity* **71:**1-12.

Zhang XL, Tsui IS, Yip CM, Fung AW, Wong DK et al. 2000. Salmonella enterica serovar typhi uses type IVB pili to enter human intestinal epithelial cells. *Infection and Immunity* **68:**3067-3073.

Zhou D, Hardt WD, and Galan JE. 1999. Salmonella typhimurium encodes a putative iron transport system within the centisome 63 pathogenicity island. *Infection and Immunity* **67:**1974-1981.

Zhu M, Valdebenito M, Winkelmann G, and Hantke K. 2005. Functions of the siderophore esterases IroD and IroE in iron-salmochelin utilization. *Microbiology* **151:**2363-2372.

8 Appendices

8.1 Chapter 2

8.1.1 CD_001

Perl script genBank2pathologic.pl can be found upon the accompanying CD.

8.2 Chapter 3

8.2.1 Biolog phenotypes for PM-1

Substrate on plate PM1	Typhi	Gallinarum	Typhimurium
negative control	35	37	44
L-arabinose	68	**275**	**151**
N-acetyl-D-glucosamine	**132**	**302**	**196**
D-saccharic acid	25	29	**208**
succinic acid	42	59	**206**
D-galactose	**180**	**319**	**230**
L-aspartic acid	38	35	**223**
L-proline	**142**	**130**	**201**
D-alanine	37	**120**	75
D-trehalose	**182**	**291**	**203**
D-mannose	**152**	**310**	**152**
dulcitol	49	90	88
D-serine	**132**	**271**	**292**
D-sorbitol	**120**	27	**188**
glycerol	**151**	**264**	**115**
L-fucose	21	25	98
D-glucuronic acid	99	**268**	**253**
D-gluconic acid	**164**	**313**	**249**
D,L-alpha-glycerol-phosphate	80	**264**	**152**
D-xylose	85	40	50
L-lactic acid	**194**	**269**	**274**
formic acid	20	27	38
D-mannitol	**217**	**317**	**214**
L-glutamic acid	**103**	99	**245**
D-glucose-6-phosphate	**198**	**312**	**302**
D-galatonic acid-gamma-lactone	23	33	**141**
D,L-malic acid	37	18	**228**

D-ribose	68	**265**	**215**
tween 20	20	63	78
L-rhamnose	12	23	**146**
D-fructose	**130**	**297**	**177**
acetic acid	52	87	**172**
alpha-D-glucose	**150**	**299**	**188**
maltose	**115**	**221**	**226**
D-melibiose	**115**	39	**197**
thymidine	**120**	**314**	**209**
L-asparagine	54	47	**204**
D-aspartic acid	31	19	**155**
D-glucosaminic acid	20	22	**232**
1,2-propanediol	20	24	32
tween 40	18	35	72
alpha-keto-glutaric acid	18	21	31
alpha-keto-butyric acid	62	25	**196**
alpha-methyl-D-galactoside	**170**	22	**264**
alpha-D-lactose	25	29	44
lactulose	30	41	51
sucrose	**146**	34	**140**
uridine	**112**	**300**	**163**
L-glutamine	55	48	**190**
M-tartric acid	22	22	28
D-glucose-1-phosphate	**191**	**293**	**263**
D-fructose-6-phosphate	**162**	**263**	**238**
tween 80	24	39	94
alpha-hydroxy-glutaric acid-gamme-lactone	22	46	58
alpha-hydroxy butyric acid	**131**	39	**211**
beta-methyl-D-glucoside	**107**	35	**218**
adonitol	23	28	38
maltotriose	**178**	**289**	**229**
2-deoxy adenosine	**118**	**304**	**254**
adenosine	**150**	**321**	**273**
glycyl-L-aspartic acid	80	**140**	**257**
citric acid	26	34	77
M-inositol	62	22	41
D-threonine	21	39	77
fumaric acid	37	26	**197**
bromo succinic acid	28	24	**161**
propionic acid	95	**111**	**219**
mucic acid	26	32	**227**
glycolic acid	23	27	30
glyoxylic acid	24	34	29
D-cellibiose	77	48	**161**
inosine	**139**	**299**	**241**
glycyl-L-glutamic acid	97	**238**	**227**
tricarballylic acid	72	32	40
L-serine	**146**	**268**	**230**
L-threonine	95	34	72
L-alanine	29	**136**	**175**

L-alanyl-glycine	55	**204**	**212**
acetoactetic acid	25	43	34
N-acetyl-beta-D-mannosamine	**182**	**290**	**224**
mono methyl succinate	34	62	**114**
methyl pyruvate	**111**	**290**	**159**
D-malic acid	36	41	36
L-malic acid	57	81	**228**
glycyl-L-proline	**102**	**233**	**201**
p-hydroxy phenyl acetic acid	40	43	**152**
m-hydroxy phenyl acetic acid	36	37	**127**
tyramine	42	40	**119**
d-palacose	41	50	57
L-lyxose	62	73	79
glucuronamide	45	**117**	59
pyruvic acid	**119**	**233**	**184**
L-galactonic acid-gamma-lactone	36	48	47
D-galacturonic acid	71	77	**171**
phenylethylamine	39	53	51
2-aminoethanol	54	68	61

Substrates are listed in the order they appear in plate PM1 (wells A1-12, B1-12 etc). Values are the integer average of three replicates and those in bold exceed the 100 cutoff, indicating a positive metabolic phenotype for that substrate.

8.2.2 Biolog phenotypes for PM-2A

Substrate on plate PM2	Typhi	Gallinarum	Typhimurium
negative control	57	64	61
chondroitin sulfate C	47	47	52
alpha-cyclodextrin	39	45	42
beta-cyclodextrin	41	48	52
gamma-cyclodextrin	35	38	37
dextrin	54	**186**	**167**
gelatin	37	43	47
glycogen	42	62	83
inulin	37	44	47
laminarin	43	57	59
mannan	45	57	60
pectin	92	**122**	**128**
N-acetyl-D-galactosamine	47	53	55
N-acetyl-neuraminic acid	**152**	**296**	**223**
beta-D-allose	16	17	19
amygdalin	26	30	29
D-arabinose	32	41	39
D-arabitol	26	33	32
L-arabitol	23	26	26
arbutin	27	32	32
2-deoxy-D-ribose	56	**116**	40
L-erythritol	23	29	26

D-fucose	30	34	36
3-0-beta-D-galactopyranosyl-D-arabinose	50	52	79
gentiobiose	**141**	44	80
L-glucose	27	27	29
lactitol	19	20	22
D-melizitose	24	25	29
maltitol	24	31	32
alpha-methyl-D-glucoside	21	24	26
beta-methyl-D-galactoside	21	27	35
3-methyl glucose	23	25	32
beta-methyl-D-glucoronic acid	25	26	38
alpha-methyl-D-mannoside	26	32	35
beta-methyl-D-xyloside	26	28	35
palatinose	45	52	54
D-raffinose	88	44	86
salicin	22	23	26
deoheptulosan	18	18	24
L-sorbose	49	26	64
stachyose	19	21	26
D-tagatose	22	28	**185**
turanose	24	41	51
xylitol	18	21	24
N-acetyl-D-glucosaminitol	20	22	32
gamma-amino butyric acid	26	31	35
delta-amino valeric acid	23	29	27
butyric acid	42	44	44
capric acid	44	50	43
caproic acid	28	29	32
citraconic acid	20	23	26
citramelic acid	23	26	27
D-glucosamine	**219**	**323**	**233**
2-hydroxy benzoic acid	14	17	26
4-hydroxy benzoic acid	15	17	22
beta-hydroxy butyric acid	17	21	38
gamma-hydroxy butyric acid	24	26	27
alpha-keto valeric acid	24	33	32
itaconic acid	25	33	63
5-keto-D-gluconic acid	53	**231**	75
D-lactic acid methyl ester	**124**	55	**213**
malonic acid	24	21	33
melibionic acid	81	28	**196**
oxalic acid	24	29	27
oxalomalic acid	22	22	23
quinic acid	21	22	25
D-ribono-1,4-lactone	21	30	36
sabacic acid	18	22	25
sorbic acid	33	37	44
succinamic acid	24	51	63
D-tartric acid	29	33	**231**
L-tartric acid	45	47	58

acetamide	33	39	35
l-alaninamide	32	41	57
N-acetyl-L-glutamic acid	17	18	32
L-arginine	21	23	26
glycine	28	23	44
L-histidine	16	56	82
L-homoserine	16	22	23
Hydroxy-L-proline	21	22	35
L-isoleucine	23	27	27
L-leucine	19	24	35
L-lysine	24	27	33
L-methionine	45	50	60
L-ornithine	31	38	38
L-phenylalanine	24	26	26
L-pyroglutamic acid	22	22	24
L-valine	24	26	33
D,L-carnitine	16	20	18
sec-butylamine	7	6	10
D,L-octopamine	23	24	26
putrescine	22	21	28
dihydroxyacetone	77	90	83
2,3-butanediol	23	41	43
2,3-butanone	49	63	83
3-hydroxy 2-butanone	41	57	69

Substrates are listed in the order they appear in plate PM-2A (wells A1-12, B1-12 etc). Values are the integer average of three replicates and those in bold exceed the 100 cutoff, indicating a positive metabolic phenotype for that substrate.

8.3 Chapter 4

8.3.1 CD_002

Perl script removeTn5_new.pl can be found upon the accompanying CD.

8.3.2 CD_003

Excel table of TraDIS results for Typhi can be found on the accompanying CD.

8.3.3 Typhi essential genes

Total inserts	Total reads	Gene length	Insertion index	Gene name	Ty2 unique ID	CT18 unique ID	Start	End	Strand	LLR*	*P*-value for no inserts
0	0	683	0	-	t1702	STY1258	1767132	1767815	-	-174.95588	6.63E-23
0	0	1151	0	-	t1621	STY1343	1676999	1678150	-	-174.95588	4.22E-38
0	0	164	0	-	t1947	STY0989	1995908	1996072	-	-174.95588	4.73E-06
0	0	134	0	-	t0535	STY2559	609653	609787	+	-174.95588	4.45E-05
0	0	212	0	-	t1378	STY1610	1432097	1432309	-	-174.95588	1.31E-07
0	0	209	0	-	t3166	none	3257647	3257856	-	-174.95588	1.63E-07
0	0	1319	0	-	t3211	STY3472	3302434	3303753	-	-174.95588	1.48E-43
0	0	29	0	-	t2621	STY2853	2699223	2699252	+	-174.95588	0.1143721
0	0	995	0	-	t3128	STY3387	3215818	3216813	-	-174.95588	4.91E-33
0	0	707	0	-	t3236	STY3498	3323667	3324374	+	-174.95588	1.10E-23
0	0	1454	0	-	t0337	STY2764	387679	389133	+	-174.95588	6.11E-48
0	0	1082	0	-	t4513	STY4817	4661340	4662422	+	-174.95588	7.34E-36
0	0	941	0	accA	t0233	STY0255	269925	270866	+	-174.95588	2.78E-31
0	0	452	0	accB	t3294	STY3559	3389201	3389653	+	-174.95588	2.10E-15
0	0	362	0	acpS	t0280	STY2823	318329	318691	+	-174.95588	1.76E-12
0	0	626	0	adk	t2372	STY0532	2441400	2442026	-	-174.95588	4.71E-21
0	0	2612	0	alaS	t2728	STY2948	2810136	2812748	-	-174.95588	1.53E-85
0	0	1715	0	argS	t0968	STY2117	1057200	1058915	-	-174.95588	2.05E-56
0	0	1088	0	asd	t3981	STY4271	4120832	4121920	+	-174.95588	4.69E-36
0	0	1382	0	asnS	t1934	STY1004	1977691	1979073	+	-174.95588	1.33E-45
0	0	1754	0	aspS	t0976	STY2109	1064648	1066402	+	-174.95588	1.11E-57
0	0	614	0	cI	t4337	STY4644	4489204	4489818	+	-174.95588	1.15E-20

0	0	461	0	coaD	t3793	STY4069	3912702	3913163	-	-174.95588	1.07E-15
0	0	167	0	csrA	t2727	STY2947	2809716	2809883	-	-174.95588	3.78E-06
0	0	1367	0	cysS	t2324	STY0585	2387383	2388750	-	-174.95588	4.08E-45
0	0	806	0	dapD	t0214	STY0236	246298	247104	-	-174.95588	6.73E-27
0	0	1109	0	dapE	t0376	STY2721	445724	446833	-	-174.95588	9.75E-37
0	0	1205	0	dfp	t3788	STY4064	3909064	3910269	-	-174.95588	7.44E-40
0	0	1382	0	dnaA	t3681	STY3940	3790627	3792009	+	-174.95588	1.33E-45
0	0	719	0	dnaC	t4586	STY4896	4742290	4743009	-	-174.95588	4.50E-24
0	0	3464	0	dnaE	t0232	STY0254	266430	269894	+	-174.95588	3.30E-113
0	0	1727	0	dnaG	t3130	STY3389	3217510	3219237	+	-174.95588	8.34E-57
0	0	713	0	dnaQ	t2601	STY0285	2678132	2678845	-	-174.95588	7.04E-24
0	0	521	0	dnaT	t4587	STY4897	4743030	4743551	-	-174.95588	1.21E-17
0	0	437	0	dut	t3787	STY4063	3908628	3909065	-	-174.95588	6.45E-15
0	0	1178	0	dxr	t0221	STY0243	254731	255909	+	-174.95588	5.60E-39
0	0	716	0	fabG	t1725	STY1234	1791565	1792281	-	-174.95588	5.63E-24
0	0	437	0	fabZ	t0228	STY0250	263418	263855	+	-174.95588	6.45E-15
0	0	344	0	folB	t3126	STY3385	3214491	3214835	-	-174.95588	6.76E-12
0	0	650	0	folE	t0662	STY2427	747254	747904	+	-174.95588	7.82E-22
0	0	1244	0	ftsA	t0136	STY0152	155775	157019	+	-174.95588	4.03E-41
0	0	650	0	ftsE	t3952	STY4241	4091910	4092560	+	-174.95588	7.82E-22
0	0	1748	0	ftsI	t0126	STY0142	143141	144889	+	-174.95588	1.74E-57
0	0	1100	0	gcpE	t0333	STY2768	383234	384334	+	-174.95588	1.91E-36
0	0	560	0	gmhA	t2540	STY0355	2615917	2616477	-	-174.95588	6.54E-19
0	0	275	0	groES	t4381	STY4689	4534179	4534454	+	-174.95588	1.18E-09
0	0	1238	0	hemA	t1099	STY1902	1183829	1185067	+	-174.95588	6.31E-41
0	0	1046	0	hemE	t3464	STY3718	3552993	3554039	-	-174.95588	1.08E-34
0	0	527	0	hemG	t3311	STY3573	3409489	3410016	-	-174.95588	7.72E-18
0	0	605	0	hemM	t1098	STY1904	1182892	1183497	-	-174.95588	2.26E-20
0	0	2816	0	ileS	t0048	STY0055	53861	56677	+	-174.95588	3.63E-92
0	0	200	0	infA	t1980	STY0951	2039984	2040184	+	-174.95588	3.20E-07
0	0	524	0	infC	t1214	STY1777	1282586	1283110	+	-174.95588	9.66E-18
0	0	833	0	ipk	t1097	STY1905	1182044	1182877	-	-174.95588	8.93E-28
0	0	953	0	ispB	t3222	STY3484	3313375	3314328	+	-174.95588	1.13E-31
0	0	728	0	kdsB	t1946	STY0990	1995165	1995893	-	-174.95588	2.29E-24
0	0	2564	0	leuS	t2219	STY0699	2284624	2287188	+	-174.95588	5.53E-84
0	0	857	0	lgt	t2911	STY3143	3000970	3001827	-	-174.95588	1.49E-28
0	0	1997	0	ligA	t0431	STY2663	503096	505093	+	-174.95588	1.43E-65
0	0	596	0	lolA	t1973	STY0959	2028390	2028986	-	-174.95588	4.43E-20
0	0	770	0	lpxA	t0229	STY0251	263877	264647	+	-174.95588	9.93E-26
0	0	482	0	lspA	t0049	STY0056	56695	57177	+	-174.95588	2.23E-16
0	0	1064	0	mraY	t0129	STY0145	147730	148794	+	-174.95588	2.82E-35
0	0	1094	0	mrdB	t2228	STY0690	2294311	2295405	+	-174.95588	2.99E-36
0	0	1025	0	mreB	t3289	STY3554	3382621	3383646	-	-174.95588	5.21E-34
0	0	1730	0	msbA	t1950	STY0985	1998392	2000122	-	-174.95588	6.67E-57
0	0	1304	0	mukF	t1942	STY0994	1991051	1992355	-	-174.95588	4.54E-43
0	0	1457	0	murC	t0133	STY0149	152558	154015	+	-174.95588	4.88E-48

0	0	1298	0	murD	t0130	STY0146	148815	150113	+	-174.95588	7.11E-43
0	0	1469	0	murE	t0127	STY0143	144894	146363	+	-174.95588	1.99E-48
0	0	809	0	nadE	t1189	STY1803	1258005	1258814	-	-174.95588	5.38E-27
0	0	2267	0	nrdA	t0587	STY2506	663376	665643	-	-174.95588	2.44E-74
0	0	527	0	orn	t4402	STY4710	4554290	4554817	+	-174.95588	7.72E-18
0	0	1874	0	parE	t3102	STY3359	3186968	3188842	-	-174.95588	1.41E-61
0	0	530	0	pgsA	t0931	STY2153	1025903	1026433	+	-174.95588	6.17E-18
0	0	26	0	pheM	t1217	STY1774	1283949	1283975	+	-174.95588	0.1431315
0	0	965	0	pheS	t1218	STY1773	1284130	1285095	+	-174.95588	4.62E-32
0	0	2369	0	pheT	t1219	STY1772	1285129	1287498	+	-174.95588	1.19E-77
0	0	719	0	plsC	t3094	STY3350	3178960	3179679	-	-174.95588	4.50E-24
0	0	512	0	ppa	t4468	STY4773	4616745	4617257	-	-174.95588	2.37E-17
0	0	1064	0	prfA	t1100	STY1901	1185126	1186190	+	-174.95588	2.82E-35
0	0	1080	0	prfB	t2959	STY3197	3048772	3049852	-	-174.95588	8.52E-36
0	0	296	0	priB	t4443	STY4748	4589371	4589667	+	-174.95588	2.45E-10
0	0	929	0	prsA	t1096	STY1906	1180831	1181760	-	-174.95588	6.82E-31
0	0	1337	0	pssA	t0258	STY2845	295248	296585	-	-174.95588	3.85E-44
0	0	707	0	pyrH	t0219	STY0241	253160	253867	+	-174.95588	1.10E-23
0	0	920	0	ribF	t0047	STY0054	52878	53798	+	-174.95588	1.34E-30
0	0	452	0	ribH	t2446	STY0456	2517395	2517847	-	-174.95588	2.10E-15
0	0	803	0	rplB	t4068	STY4361	4219535	4220338	+	-174.95588	8.42E-27
0	0	521	0	rplE	t4077	STY4370	4223460	4223981	+	-174.95588	1.21E-17
0	0	515	0	rplF	t4080	STY4373	4224758	4225273	+	-174.95588	1.89E-17
0	0	347	0	rplL	t3475	STY3733	3571291	3571638	-	-174.95588	5.40E-12
0	0	410	0	rplM	t3261	STY3525	3351928	3352338	-	-174.95588	4.86E-14
0	0	353	0	rplN	t4075	STY4368	4222749	4223102	+	-174.95588	3.45E-12
0	0	416	0	rplO	t4084	STY4377	4226359	4226775	+	-174.95588	3.10E-14
0	0	392	0	rplP	t4072	STY4365	4221730	4222122	+	-174.95588	1.87E-13
0	0	335	0	rplR	t4081	STY4374	4225301	4225636	+	-174.95588	1.32E-11
0	0	338	0	rplT	t1216	STY1775	1283472	1283810	+	-174.95588	1.06E-11
0	0	293	0	rplU	t3221	STY3483	3312805	3313098	-	-174.95588	3.06E-10
0	0	314	0	rplV	t4070	STY4363	4220666	4220980	+	-174.95588	6.37E-11
0	0	284	0	rplW	t4067	STY4360	4219215	4219499	+	-174.95588	6.00E-10
0	0	296	0	rplX	t4076	STY4369	4223131	4223427	+	-174.95588	2.45E-10
0	0	239	0	rpmA	t3220	STY3482	3312528	3312767	-	-174.95588	1.73E-08
0	0	218	0	rpmB	t3790	STY4066	3911343	3911561	+	-174.95588	8.34E-08
0	0	173	0	rpmC	t4073	STY4366	4222140	4222313	+	-174.95588	2.41E-06
0	0	161	0	rpmD	t4083	STY4376	4226176	4226337	+	-174.95588	5.92E-06
0	0	122	0	rpmH	t3680	STY3939A	3789826	3789948	-	-174.95588	0.0001093
0	0	179	0	rpmI	t1215	STY1776	1283224	1283403	+	-174.95588	1.54E-06
0	0	971	0	rpoA	t4090	STY4383	4229869	4230840	+	-174.95588	2.95E-32
0	0	683	0	rpsC	t4071	STY4364	4221016	4221699	+	-174.95588	6.63E-23
0	0	602	0	rpsD	t4089	STY4382	4229223	4229825	+	-174.95588	2.83E-20
0	0	485	0	rpsE	t4082	STY4375	4225669	4226154	+	-174.95588	1.78E-16
0	0	377	0	rpsF	t4442	STY4747	4588969	4589346	+	-174.95588	5.73E-13
0	0	452	0	rpsG	t4058	STY4351	4211920	4212372	+	-174.95588	2.10E-15

0	0	374	0	rpsH	t4079	STY4372	4224353	4224727	+	-174.95588	7.17E-13
0	0	293	0	rpsJ	t4064	STY4357	4217629	4217922	+	-174.95588	3.06E-10
0	0	371	0	rpsK	t4088	STY4381	4228800	4229171	+	-174.95588	8.97E-13
0	0	356	0	rpsL	t4057	STY4350	4211450	4211806	+	-174.95588	2.75E-12
0	0	338	0	rpsM	t4087	STY4380	4228427	4228765	+	-174.95588	1.06E-11
0	0	287	0	rpsN	t4078	STY4371	4224014	4224301	+	-174.95588	4.79E-10
0	0	230	0	rpsP	t2631	STY2863	2707753	2707983	-	-174.95588	3.40E-08
0	0	236	0	rpsQ	t4074	STY4367	4222331	4222567	+	-174.95588	2.17E-08
0	0	260	0	rpsS	t4069	STY4362	4220373	4220633	+	-174.95588	3.61E-09
0	0	512	0	ssb	t4161	STY4451	4306069	4306581	+	-174.95588	2.37E-17
0	0	1910	0	thrS	t1213	STY1778	1280654	1282564	+	-174.95588	9.53E-63
0	0	623	0	tmk	t1720	STY1239	1787263	1787886	-	-174.95588	5.89E-21
0	0	833	0	tsf	t0218	STY0240	252164	252997	+	-174.95588	8.93E-28
0	0	1256	0	tyrS	t1317	STY1673	1377902	1379158	+	-174.95588	1.64E-41
0	0	737	0	ubiE	t3327	STY3589	3427124	3427861	-	-174.95588	1.17E-24
0	0	2837	0	valS	t4510	STY4814	4655999	4658836	-	-174.95588	7.55E-93
0	0	530	0	yfjA	t2630	STY2862	2707186	2707716	-	-174.95588	6.17E-18
0	0	461	0	ygbB	t2830	STY3054	2905705	2906166	-	-174.95588	1.07E-15
0	0	1460	0	yigC	t3319	STY3581	3420094	3421554	-	-174.95588	3.90E-48
1	2	2396	4E-04	gyrB	t3684	STY3943	3794432	3796828	+	-43.368975	1.58E-78
1	9	2240	4E-04	parC	t3095	STY3351	3179954	3182194	-	-42.651896	1.83E-73
1	1	2180	5E-04	priA	t3523	STY3775	3625523	3627703	+	-42.362453	1.63E-71
1	2	1844	5E-04	dxs	t2441	STY0461	2512409	2514253	+	-40.574479	1.32E-60
1	4	1628	6E-04	groEL	t4382	STY4690	4534516	4536144	+	-39.239302	1.37E-53
1	7	1628	6E-04	yidC	t3678	STY3938	3787580	3789208	-	-39.239302	1.37E-53
1	2	1499	7E-04	lysS	t2958	STY3196	3047245	3048744	-	-38.352224	2.11E-49
3	6	4448	7E-04	mukB	t1940	STY0996	1985900	1990348	-	-38.234347	3.68E-145
1	1	1409	7E-04	lpdA	t0160	STY0177	184153	185562	+	-37.685548	1.77E-46
1	1	1397	7E-04	dnaB	t4152	STY4442	4296374	4297771	+	-37.593361	4.34E-46
1	4	1352	7E-04	glmU	t3657	STY3916	3762297	3763649	+	-37.240239	1.25E-44
1	4	1340	7E-04	murF	t0128	STY0144	146378	147718	+	-37.144026	3.08E-44
3	6	4010	7E-04	rpoB	t3474	STY3732	3566945	3570955	-	-37.117142	6.14E-131
1	1	1334	7E-04	yaeL	t0224	STY0246	257880	259214	+	-37.095586	4.82E-44
1	6	1331	8E-04	accC	t3295	STY3560	3389682	3391013	+	-37.071281	6.03E-44
1	5	1313	8E-04	prlA	t4085	STY4378	4226801	4228114	+	-36.924257	2.32E-43
2	4	2618	8E-04	gyrA	t0592	STY2499	670713	673331	+	-36.891302	9.75E-86
1	1	1274	8E-04	serS	t1971	STY0961	2025486	2026760	-	-36.598437	4.28E-42
1	6	1241	8E-04	murA	t3224	STY3486	3314925	3316166	-	-36.314593	5.04E-41
1	1	1241	8E-04	rho	t3380	STY3638	3480453	3481694	-	-36.314593	5.04E-41
1	1	1226	8E-04	-	t1701	STY1259	1765888	1767114	-	-36.182984	1.55E-40
2	10	2336	9E-04	imp	t0096	STY0108	105614	107950	-	-35.657933	1.40E-76
1	1	1133	9E-04	ftsZ	t0137	STY0153	157098	158231	+	-35.327864	1.62E-37
1	29	1130	9E-04	lpxB	t0230	STY0252	264665	265795	+	-35.299082	2.03E-37
1	2	1112	9E-04	nrdB	t0586	STY2507	662133	663245	-	-35.124709	7.79E-37
1	1	1082	9E-04	dnaN	t3682	STY3941	3792032	3793114	+	-34.827474	7.34E-36
1	1	1061	9E-04	fba	t2987	STY3226	3078251	3079312	-	-34.614273	3.53E-35

1	1	1049	1E-03	murG	t0132	STY0148	151372	152421	+	-34.490466	8.65E-35
2	6	1964	0.001	rpoD	t3131	STY3390	3219270	3221234	+	-33.770948	1.68E-64
1	3	977	0.001	gapA	t1169	STY1825	1236581	1237558	-	-33.715218	1.88E-32
1	1	968	0.001	-	t0429	STY2664	502038	503006	+	-33.614149	3.69E-32
1	3	959	0.001	lpxK	t1949	STY0986	1997418	1998377	-	-33.512096	7.24E-32
1	1	956	0.001	hemB	t2492	STY0404	2564895	2565851	+	-33.477856	9.06E-32
1	2	944	0.001	hemH	t2371	STY0533	2440208	2441152	-	-33.339766	2.22E-31
1	5	911	0.001	fabD	t1726	STY1233	1792312	1793223	-	-32.950356	2.62E-30
3	9	2660	0.001	infB	t3204	STY3467	3294925	3297585	-	-32.653651	4.22E-87
1	11	881	0.001	ispA	t2440	STY0462	2511486	2512367	+	-32.583332	2.47E-29
1	6	860	0.001	dapA	t0370	STY2727	439626	440486	+	-32.318545	1.19E-28
1	4	860	0.001	yfjB	t2637	STY2869	2713443	2714303	+	-32.318545	1.19E-28
2	10	1700	0.001	proS	t0245	STY0269	280765	282465	-	-32.190065	6.28E-56
1	30	836	0.001	kdsA	t1104	STY1897	1188278	1189114	+	-32.007506	7.14E-28
1	10	836	0.001	rpoH	t3954	STY4243	4093872	4094708	+	-32.007506	7.14E-28
1	1	803	0.001	dapB	t0066	STY0073	74022	74825	+	-31.564172	8.42E-27
3	13	2402	0.001	plsB	t4141	STY4431	4285962	4288364	-	-31.532104	1.01E-78
3	6	2393	0.001	yaeT	t0225	STY0247	259264	261657	+	-31.490729	1.97E-78
2	12	1520	0.001	lnt	t2207	STY0711	2272626	2274146	+	-30.95685	4.40E-50
1	9	740	0.001	-	t0222	STY0244	256240	256980	+	-30.661925	9.35E-25
2	24	1457	0.001	ftsY	t3951	STY4240	4090432	4091889	+	-30.48851	4.88E-48
1	13	722	0.001	hemD	t3360	STY3622	3461587	3462309	+	-30.38921	3.59E-24
1	11	719	0.001	-	t2619	STY2852	2697772	2698491	+	-30.34306	4.50E-24
1	3	704	0.001	ybbF	t2326	STY0583	2389438	2390142	+	-30.109215	1.38E-23
1	1	644	0.002	yadF	t0176	STY0193	203598	204242	-	-29.117947	1.23E-21
1	1	638	0.002	pdxH	t1316	STY1674	1377119	1377757	+	-29.01347	1.92E-21
2	22	1256	0.002	hisS	t0334	STY2767	384463	385719	+	-28.836997	1.64E-41
3	19	1883	0.002	pbpA	t2227	STY0691	2292407	2294290	+	-28.831064	7.17E-62
1	1	623	0.002	ribE	t1294	STY1696	1357967	1358590	+	-28.747636	5.89E-21
1	7	614	0.002	-	t3620	STY3880	3721114	3721728	+	-28.584844	1.15E-20
1	1	611	0.002	rplC	t4065	STY4358	4217973	4218584	+	-28.530014	1.44E-20
1	1	602	0.002	yacE	t0146	STY0162	166354	166956	-	-28.363786	2.83E-20
2	15	1196	0.002	fabB	t0486	STY2609	559866	561062	+	-28.289054	1.46E-39
2	8	1190	0.002	sucB	t2139	STY0780	2200547	2201737	-	-28.232654	2.28E-39
1	2	587	0.002	rplD	t4066	STY4359	4218613	4219200	+	-28.080756	8.69E-20
2	2	1145	0.002	pgk	t2988	STY3227	3079432	3080577	-	-27.799715	6.61E-38
1	5	572	0.002	ribA	t1624	STY1340	1679593	1680165	+	-27.789888	2.67E-19
1	6	572	0.002	rlpB	t2220	STY0698	2287221	2287793	+	-27.789888	2.67E-19
2	8	1136	0.002	metK	t3002	STY3243	3092881	3094017	+	-27.710944	1.29E-37
1	12	557	0.002	rpoE	t0270	STY2833	309252	309809	+	-27.49074	8.19E-19
1	3	554	0.002	yrdC	t4102	STY4395	4239498	4240052	+	-27.429875	1.02E-18
5	18	2768	0.002	polA	t3621	STY3881	3722082	3724850	-	-27.421733	1.31E-90
2	6	1088	0.002	trmU	t1686	STY1274	1749012	1750100	+	-27.224407	4.69E-36
3	11	1619	0.002	pyrG	t2854	STY3082	2931600	2933219	-	-27.134108	2.68E-53
1	1	539	0.002	frr	t0220	STY0242	254032	254571	+	-27.120147	3.15E-18
2	12	1064	0.002	-	t4514	STY4818	4662440	4663504	+	-26.972424	2.82E-35

1	6	527	0.002	nusG	t3479	STY3737	3573930	3574457	-	-26.865623	7.72E-18
4	13	2096	0.002	fusA	t4059	STY4352	4212487	4214583	+	-26.801017	8.70E-69
4	10	2051	0.002	glyS	t3864	STY4144	3991761	3993812	+	-26.55515	2.51E-67
1	1	500	0.002	fabA	t1853	STY1088	1913358	1913858	+	-26.269371	5.81E-17
3	9	1475	0.002	mviN	t1750	STY1209	1814380	1815855	-	-26.078304	1.27E-48
1	1	491	0.002	fms	t4098	STY4391	4236714	4237205	-	-26.062867	1.14E-16
2	2	971	0.002	fepG	t2277	STY0635	2344426	2345397	+	-25.934647	2.95E-32
3	9	1415	0.002	waaE	t3120	STY3379	3205564	3206979	-	-25.605074	1.13E-46
4	8	1871	0.002	aceF	t0159	STY0176	182065	183936	+	-25.509857	1.76E-61
2	2	935	0.002	fabH	t1727	STY1232	1793257	1794192	-	-25.50375	4.35E-31
2	30	929	0.002	fmt	t4097	STY4390	4235751	4236680	-	-25.430187	6.82E-31
2	6	923	0.002	hemC	t3359	STY3621	3460649	3461572	+	-25.356108	1.07E-30
4	11	1829	0.002	secD	t2456	STY0445	2524726	2526555	-	-25.250262	4.07E-60
2	3	914	0.002	waaD	t3809	STY4085	3928396	3929310	-	-25.244002	2.09E-30
3	6	1352	0.002	purB	t1688	STY1272	1750850	1752202	+	-25.084175	1.25E-44
2	7	893	0.002	glyQ	t3863	STY4143	3990840	3991733	+	-24.977688	1.01E-29
1	1	443	0.002	yjeE	t4408	STY4714	4559146	4559589	+	-24.8874	4.12E-15
2	3	854	0.002	ubiA	t4140	STY4430	4284991	4285845	+	-24.464509	1.86E-28
3	117	1259	0.002	kdtA	t3794	STY4070	3913190	3914449	-	-24.264867	1.31E-41
2	2	839	0.002	cdsA	t0223	STY0245	257011	257850	+	-24.260285	5.70E-28
3	5	1250	0.002	folC	t0499	STY2596	571079	572329	+	-24.182093	2.57E-41
1	9	410	0.002	rplK	t3478	STY3736	3573344	3573754	-	-23.995793	4.86E-14
2	12	812	0.002	ftsQ	t0135	STY0151	154948	155760	+	-23.882413	4.30E-27
2	10	806	0.002	dapF	t3350	STY3612	3452074	3452880	-	-23.796571	6.73E-27
1	2	386	0.003	-	t1920	STY1020	1965700	1966086	+	-23.296494	2.92E-13
2	7	770	0.003	fabI	t1613	STY1352	1667618	1668388	+	-23.266327	9.93E-26
1	4	383	0.003	rbfA	t3203	STY3466	3294303	3294686	-	-23.205733	3.66E-13
5	11	1910	0.003	dnaX	t2376	STY0528	2445425	2447335	-	-23.175306	9.53E-63
5	10	1898	0.003	dnaK	t0012	STY0012	11603	13501	+	-23.10192	2.34E-62
7	11	2645	0.003	aceE	t0158	STY0175	179387	182032	+	-23.048307	1.30E-86
2	10	749	0.003	trmD	t2629	STY2861	2706374	2707123	-	-22.944262	4.77E-25
1	1	371	0.003	yadR	t0205	STY0226	237127	237498	+	-22.834688	8.97E-13
1	2	365	0.003	rplQ	t4091	STY4384	4230899	4231264	+	-22.644164	1.41E-12
1	5	365	0.003	secE	t3480	STY3738	3574477	3574842	-	-22.644164	1.41E-12
3	26	1085	0.003	ribD	t2447	STY0455	2517954	2519039	-	-22.536815	5.86E-36
2	21	710	0.003	ubiG	t0588	STY2505	666016	666726	-	-22.318791	8.81E-24
2	16	707	0.003	rpsB	t0217	STY0239	251181	251888	+	-22.269111	1.10E-23
1	2	347	0.003	ftsL	t0125	STY0141	142760	143107	+	-22.051091	5.40E-12
1	6	341	0.003	rnpA	t3679	STY3939	3789450	3789791	-	-21.845778	8.46E-12
2	20	677	0.003	-	t1057	STY1950	1141081	1141758	+	-21.759042	1.04E-22
5	9	1622	0.003	aarF	t3325	STY3587	3424872	3426494	-	-21.256089	2.14E-53
2	11	635	0.003	ribB	t3115	STY3373	3200858	3201493	-	-21.000987	2.40E-21
2	26	623	0.003	-	t2222	STY0696	2288844	2289467	+	-20.77406	5.89E-21
3	32	932	0.003	lytB	t0051	STY0058	57802	58734	+	-20.742155	5.45E-31
4	24	1223	0.003	cca	t3124	STY3383	3212226	3213449	+	-20.551938	1.94E-40
3	8	899	0.003	lpxC	t0138	STY0154	158350	159249	+	-20.311769	6.43E-30

3	11	896	0.003	accD	t0498	STY2597	570097	570993	+	-20.271765	8.04E-30
2	6	590	0.003	lexA	t4143	STY4433	4289030	4289620	+	-20.123779	6.95E-20
1	2	293	0.003	-	t2832	STY3056	2906913	2907206	-	-20.042182	3.06E-10
4	27	1157	0.003	yleB	t2202	STY0717	2267106	2268263	-	-19.88748	2.69E-38
5	14	1397	0.004	gltX	t0442	STY2654	512533	513930	+	-19.470078	4.34E-46
6	80	1649	0.004	glnS	t2189	STY0724	2256012	2257661	-	-19.270865	2.85E-54
7	60	1916	0.004	ftsH	t3213	STY3474	3304717	3306633	-	-19.221475	6.08E-63
10	30	2687	0.004	secA	t0140	STY0156	160107	162794	+	-18.99731	5.60E-88
5	35	1343	0.004	ffh	t2632	STY2864	2708345	2709688	-	-18.992791	2.46E-44
2	9	533	0.004	yfhC	t0289	STY2814	326228	326761	+	-18.89745	4.93E-18
1	1	266	0.004	rplY	t0630	STY2461	711619	711885	-	-18.874624	2.30E-09
4	16	1037	0.004	ftsX	t3953	STY4242	4092571	4093608	+	-18.56159	2.12E-34
3	11	776	0.004	map	t0216	STY0238	249937	250713	-	-18.534102	6.34E-26
3	5	776	0.004	thyA	t2910	STY3142	3000169	3000945	-	-18.534102	6.34E-26
2	24	512	0.004	fldA	t2181	STY0732	2247436	2247948	+	-18.407531	2.37E-17
5	16	1274	0.004	mesJ	t0238	STY0261	276262	277536	+	-18.350082	4.28E-42
4	21	1013	0.004	holA	t2221	STY0697	2287811	2288824	+	-18.275415	1.28E-33
4	10	1007	0.004	lpxD	t0227	STY0249	262287	263294	+	-18.202657	2.00E-33
1	3	251	0.004	rpsO	t3201	STY3464	3292939	3293190	-	-18.166092	7.07E-09
4	29	986	0.004	holB	t1719	STY1240	1786262	1787248	-	-17.944031	9.61E-33
5	12	1226	0.004	ftsW	t0131	STY0147	150131	151357	+	-17.879012	1.55E-40
3	45	722	0.004	-	t3415	STY3673	3511820	3512542	+	-17.649137	3.59E-24
4	10	959	0.004	-	t2444	STY0458	2515899	2516858	-	-17.602049	7.24E-32
4	27	956	0.004	lepB	t0275	STY2828	314016	314972	+	-17.563368	9.06E-32
4	21	950	0.004	psd	t4400	STY4708	4552071	4553021	-	-17.485583	1.42E-31
4	23	935	0.004	waaC	t3807	STY4083	3926394	3927329	-	-17.288614	4.35E-31
4	102	932	0.004	coaA	t3487	STY3740	3577233	3578165	+	-17.248782	5.45E-31
3	16	692	0.004	ygbP	t2831	STY3055	2906184	2906876	-	-17.123905	3.39E-23
4	11	887	0.005	era	t0276	STY2826	315941	316828	+	-16.632838	1.58E-29
3	20	659	0.005	mukE	t1941	STY0995	1990366	1991025	-	-16.514792	3.99E-22
1	8	218	0.005	acpP	t1724	STY1235	1791173	1791391	-	-16.419413	8.34E-08
6	53	1292	0.005	-	t1703	STY1257	1767826	1769118	-	-16.265024	1.11E-42
4	24	848	0.005	folD	t2321	STY0588	2385580	2386428	+	-16.068868	2.91E-28
5	25	1034	0.005	mreC	t3288	STY3553	3381504	3382538	-	-15.755546	2.66E-34
3	7	608	0.005	ftsJ	t3214	STY3475	3306755	3307363	-	-15.49983	1.81E-20
2	16	401	0.005	nusB	t2445	STY0457	2516955	2517356	-	-15.363324	9.52E-14
2	7	398	0.005	-	t3009	STY3250	3099855	3100253	+	-15.267798	1.19E-13
14	38	2783	0.005	sucA	t2140	STY0779	2201770	2204553	-	-15.254083	4.28E-91
1	1	197	0.005	rpsU	t3129	STY3388	3217059	3217256	+	-15.139102	4.01E-07
4	27	785	0.005	suhB	t0310	STY2792	355508	356293	-	-15.090443	3.23E-26
3	10	587	0.005	yigP	t3326	STY3588	3426509	3427096	-	-15.052445	8.69E-20
12	116	2336	0.005	lon	t2410	STY0492	2481066	2483402	-	-14.986994	1.40E-76
4	4	776	0.005	fepC	t2278	STY0634	2345412	2346188	+	-14.943138	6.34E-26
6	14	1160	0.005	visB	t2975	STY3213	3066494	3067654	-	-14.899104	2.15E-38
4	36	764	0.005	murI	t3494	STY3743	3586120	3586884	-	-14.743558	1.55E-25
2	4	374	0.005	rpsI	t3260	STY3524	3351520	3351894	-	-14.471595	7.17E-13

9	28	1655	0.005	rpsA	t1953	STY0981	2003111	2004766	-	-14.255247	1.82E-54
3	70	551	0.005	cI	t3402	STY3660	3502585	3503136	-	-14.239627	1.28E-18
6	50	1052	0.006	degS	t3265	STY3529	3355989	3357041	+	-13.63705	6.91E-35
2	6	338	0.006	yheM	t4055	STY4348	4210673	4211011	+	-13.155678	1.06E-11
10	16	1622	0.006	pgm	t2177	STY0736	2243696	2245318	-	-12.614041	2.14E-53
6	55	962	0.006	-	t2966	STY3204	3054944	3055906	+	-12.460539	5.78E-32
4	11	641	0.006	rpiA	t2981	STY3219	3072806	3073447	-	-12.453636	1.53E-21
8	64	1280	0.006	eno	t2853	STY3081	2930219	2931499	-	-12.4329	2.73E-42
3	12	473	0.006	mreD	t3287	STY3552	3381013	3381486	-	-12.23745	4.37E-16
2	3	311	0.006	-	t1835	STY1111	1896895	1897206	+	-12.052786	7.97E-11
13	29	1973	0.007	tktA	t2996	STY3236	3085841	3087814	-	-11.72735	8.58E-65
4	6	605	0.007	gmk	t3778	STY4052	3900085	3900690	-	-11.681284	2.26E-20
7	16	1043	0.007	recA	t2730	STY2950	2813619	2814662	-	-11.479405	1.36E-34
9	47	1340	0.007	rffT	t3369	STY3627	3468163	3469503	-	-11.469339	3.08E-44
3	22	440	0.007	yabB	t0123	STY0139	141362	141802	+	-11.266065	5.16E-15
1	7	146	0.007	-	t3157	STY3417	3244748	3244894	+	-11.204345	1.82E-05
4	23	572	0.007	grpE	t2636	STY2868	2712729	2713301	-	-10.922251	2.67E-19
7	14	992	0.007	ruvB	t0983	STY2102	1070786	1071778	+	-10.799114	6.14E-33
5	73	707	0.007	phoU	t3670	STY3930	3779243	3779950	+	-10.768804	1.10E-23
7	104	953	0.007	secF	t2455	STY0446	2523744	2524697	-	-10.249101	1.13E-31
1	6	131	0.008	-	t4531	STY4834	4677894	4678025	+	-9.716011	5.57E-05
16	34	2006	0.008	rep	t3384	STY3642	3485509	3487515	-	-9.102176	7.27E-66
4	13	500	0.008	-	t3235	STY3497	3323142	3323642	+	-9.0600298	5.81E-17
8	13	989	0.008	fepD	t2276	STY0636	2343419	2344408	+	-8.904122	7.68E-33
4	12	491	0.008	crr	t0424	STY2670	496734	497225	-	-8.8037667	1.14E-16
35	903	4205	0.008	rpoC	t3473	STY3731	3562645	3566850	-	-8.499316	2.86E-137
4	16	479	0.008	rplJ	t3476	STY3734	3571723	3572202	-	-8.4527162	2.79E-16
17	85	2015	0.008	metG	t0701	STY2384	786592	788607	-	-8.3066892	3.71E-66
8	29	932	0.009	miaA	t4411	STY4717	4562802	4563734	+	-8.0596268	5.45E-31
10	63	1154	0.009	-	t3218	STY3480	3310245	3311399	-	-7.9234155	3.37E-38
11	32	1262	0.009	hemL	t0203	STY0223	234219	235481	-	-7.8392958	1.05E-41
9	13	1028	0.009	waaF	t3808	STY4084	3927347	3928375	-	-7.7757289	4.16E-34
4	26	455	0.009	-	t2205	STY0714	2271097	2271552	+	-7.7159689	1.68E-15
2	3	227	0.009	yhhP	t3944	STY4233	4083902	4084129	+	-7.684203	4.26E-08
1	5	113	0.009	malY	t1332	STY1657a	1392721	1392834	-	-7.6204025	0.0002141
3	12	329	0.009	rplS	t2628	STY2860	2705986	2706315	-	-7.1856346	2.07E-11
11	58	1196	0.009	-	t0313	STY2789	358078	359274	+	-7.0600413	1.46E-39
10	48	1058	0.009	yjeQ	t4401	STY4709	4553131	4554189	-	-6.6595492	4.41E-35
4	23	419	0.01	holD	t4600	STY4907	4751767	4752186	+	-6.5123984	2.48E-14
2	23	209	0.01	rpsR	t4444	STY4749	4589690	4589899	+	-6.4770815	1.63E-07
12	54	1235	0.01	glyA	t0301	STY2802	345824	347059	+	-6.2508403	7.90E-41
9	70	923	0.01	yabC	t0124	STY0140	141822	142745	+	-6.1986095	1.07E-30
4	21	404	0.01	-	t3206	STY3469	3299158	3299562	-	-5.9708564	7.61E-14
9	37	902	0.01	htrB	t1765	STY1192	1827380	1828282	+	-5.8552635	5.13E-30
2	21	200	0.01	samA	t4339	STY4598	4457355	4457555	-	-5.8220224	3.20E-07
1	2	98	0.01	rpmJ	t4086	STY4379	4228164	4228262	+	-5.5184538	0.0006573

3	13	290	0.01	-	t1622	STY1342	1678175	1678465	-	-5.3115433	3.83E-10
10	36	938	0.011	fepB	t2274	STY0638	2341045	2341983	+	-4.8537461	3.48E-31
5	8	458	0.011	-	t2641	STY2873	2717001	2717459	-	-4.4897471	1.34E-15
1	2	89	0.011	-	t0694	STY2391	781098	781187	-	-4.0443813	0.0012883
14	37	1232	0.011	icdA	t1682	STY1278	1746072	1747304	-	-3.8684681	9.88E-41
17	75	1484	0.011	nusA	t3205	STY3468	3297628	3299112	-	-3.742686	6.49E-49
38	176	3185	0.012	rne	t1734	STY1226	1798647	1801832	+	-3.1024484	3.78E-104
8	15	665	0.012	cmk	t1954	STY0980	2004898	2005563	-	-2.9710588	2.55E-22
21	53	1703	0.012	cydC	t1978	STY0954	2037202	2038905	+	-2.5763481	5.02E-56
3	15	239	0.013	-	t2724	none	2809238	2809477	+	-2.2904746	1.73E-08
7	101	557	0.013	-	t3234	STY3496	3322562	3323119	+	-2.2711705	8.19E-19
4	34	314	0.013	yajC	t2457	STY0444	2526601	2526915	-	-2.0518214	6.37E-11

*LLR, log-likelihood ratio.

8.3.4 CD_004

Excel table of TraDIS results for Typhimurium can be found on the accompanying CD.

8.3.5 Typhimurium essential genes

Total inserts	Total reads	Gene length	Insertion index	Gene name	SL unique ID	CT18 unique ID	Ty2 unique ID	Start	End	Strand	LLR	P-value for no inserts
0	0	623	0	-	SL0633	STY0696	t2222	706265	706888	-	-174.35924	2.83E-31
0	0	323	0	-	SL0950	-	-	1061626	1061949	-	-174.35924	1.45E-16
0	0	683	0	-	SL1155	STY1258	t1702	1260022	1260705	+	-174.35924	3.23E-34
0	0	677	0	-	SL1749	STY1950	t1057	1875339	1876016	-	-174.35924	6.36E-34
0	0	959	0	-	SL2695	-	-	2877254	2878213	-	-174.35924	9.46E-48
0	0	743	0	-*	SL0916*	STY0979	t1955	1020636	1021379	+	-174.35924	3.69E-37
0	0	452	0	accB	SL3352	STY3559	t3294	3571524	3571976	+	-174.35924	6.86E-23
0	0	218	0	acpP	SL1133	STY1235	t1724	1236264	1236482	+	-174.35924	2.04E-11
0	0	362	0	acpS	SL2539	STY2823	t0280	2719788	2720150	-	-174.35924	1.78E-18
0	0	626	0	adk	SL0481	STY0532	t2372	545538	546164	+	-174.35924	2.02E-31
0	0	2612	0	alaS	SL2807	STY2948	t2728	2993956	2996568	-	-174.35924	8.36E-129
0	0	1715	0	argS	SL1844	STY2117	t0968	1960173	1961888	+	-174.35924	8.06E-85
0	0	557	0	cI	SL2708	STY4644	t4337	2886929	2887486	+	-174.35924	4.87E-28
0	0	1367	0	cysS	SL0530	STY0585	t2324	598632	599999	+	-174.35924	9.34E-68
0	0	1382	0	dnaA	SL3805	STY3940	t3681	4064947	4066329	-	-174.35924	1.72E-68
0	0	1082	0	dnaN	SL3804	STY3941	t3682	4063842	4064924	-	-174.35924	8.81E-54
0	0	521	0	dnaT	SL4475	STY4897	t4587	4824124	4824645	-	-174.35924	2.84E-26
0	0	1844	0	dxs	SL0416	STY0461	t2441	474125	475969	-	-174.35924	3.81E-91
0	0	1454	0	engA	SL2481	STY2764	t0337	2649182	2650636	-	-174.35924	5.06E-72
0	0	716	0	fabG	SL1132	STY1234	t1725	1235374	1236090	+	-174.35924	7.79E-36
0	0	770	0	fabI	SL1631	STY1352	t1613	1748998	1749768	+	-174.35924	1.75E-38
0	0	461	0	folA	SL0088	STY0102	t0090	100851	101312	+	-174.35924	2.48E-23
0	0	539	0	frr	SL0220	STY0242	t0220	257153	257692	+	-174.35924	3.72E-27
0	0	1133	0	ftsZ	SL0133	STY0153	t0137	155695	156828	+	-174.35924	2.78E-56
0	0	977	0	gapA	SL1225	STY1825	t1169	1325555	1326532	-	-174.35924	1.24E-48
0	0	1100	0	gcpE	SL2485	STY2768	t0333	2653981	2655081	-	-174.35924	1.15E-54
0	0	275	0	groES	SL4266	STY4689	t4381	4595364	4595639	+	-174.35924	3.28E-14
0	0	956	0	hemB	SL0367	STY0404	t2492	423173	424129	-	-174.35924	1.33E-47
0	0	605	0	hemM	SL1706	STY1904	t1098	1832886	1833491	+	-174.35924	2.16E-30
0	0	2816	0	ileS	SL0047	STY0055	t0048	53864	56680	+	-174.35924	8.31E-139
0	0	416	0	infC	SL1268	STY1777	t1214	1371104	1371520	+	-174.35924	4.00E-21
0	0	833	0	ipk	SL1707	STY1905	t1097	1833506	1834339	+	-174.35924	1.43E-41
0	0	836	0	kdsA	SL1700	STY1897	t1104	1827269	1828105	-	-174.35924	1.02E-41
0	0	728	0	kdsB	SL0925	STY0990	t1946	1031240	1031968	+	-174.35924	2.01E-36
0	0	2564	0	leuS	SL0636	STY0699	t2219	708544	711108	-	-174.35924	1.89E-126
0	0	770	0	lpxA	SL0229	STY0251	t0229	267000	267770	+	-174.35924	1.75E-38
0	0	1499	0	lysS	SL3016	STY3196	t2958	3219948	3221447	-	-174.35924	3.14E-74
0	0	1730	0	msbA	SL0921	STY0985	t1950	1027011	1028741	+	-174.35924	1.48E-85
0	0	530	0	pgsA	SL1875	STY2153	t0931	1993390	1993920	-	-174.35924	1.03E-26
0	0	965	0	pheS	SL1271	STY1773	t1218	1372540	1373505	+	-174.35924	4.81E-48
0	0	2369	0	pheT	SL1272	STY1772	t1219	1373539	1375908	+	-174.35924	6.88E-117
0	0	863	0	prfB	SL3017	STY3197	t2959	3221475	3222338	-	-174.35924	4.82E-43
0	0	1337	0	pssA	SL2616	STY2845	t0258	2793077	2794414	+	-174.35924	2.76E-66
0	0	707	0	pyrH	SL0219	STY0241	t0219	256281	256988	+	-174.35924	2.15E-35
0	0	572	0	ribA	SL1643	STY1340	t1624	1760971	1761543	+	-174.35924	8.96E-29

0	0	803	0	rplB	SL3404	STY4361	t4068	3615294	3616097	-	-174.35924	4.22E-40
0	0	521	0	rplE	SL3395	STY4370	t4077	3611651	3612172	-	-174.35924	2.84E-26
0	0	410	0	rplK	SL4088	STY3736	t3478	4384445	4384855	+	-174.35924	7.87E-21
0	0	347	0	rplL	SL4091	STY3733	t3475	4386565	4386912	+	-174.35924	9.66E-18
0	0	353	0	rplN	SL3397	STY4368	t4075	3612530	3612883	-	-174.35924	4.91E-18
0	0	416	0	rplO	SL3388	STY4377	t4084	3608857	3609273	-	-174.35924	4.00E-21
0	0	392	0	rplP	SL3400	STY4365	t4072	3613510	3613902	-	-174.35924	6.01E-20
0	0	365	0	rplQ	SL3381	STY4384	t4091	3604368	3604733	-	-174.35924	1.27E-18
0	0	257	0	rplR	SL3391	STY4374	t4081	3609996	3610253	-	-174.35924	2.50E-13
0	0	338	0	rplT	SL1270	STY1775	t1216	1371882	1372220	+	-174.35924	2.67E-17
0	0	314	0	rplV	SL3402	STY4363	t4070	3614652	3614966	-	-174.35924	4.01E-16
0	0	284	0	rplW	SL3405	STY4360	t4067	3616133	3616417	-	-174.35924	1.19E-14
0	0	296	0	rplX	SL3396	STY4369	t4076	3612205	3612501	-	-174.35924	3.06E-15
0	0	239	0	rpmA	SL3275	STY3482	t3220	3493927	3494166	-	-174.35924	1.91E-12
0	0	218	0	rpmB	SL3694	STY4066	t3790	3943549	3943767	-	-174.35924	2.04E-11
0	0	173	0	rpmC	SL3399	STY4366	t4073	3613319	3613492	-	-174.35924	3.29E-09
0	0	161	0	rpmD	SL3389	STY4376	t4083	3609295	3609456	-	-174.35924	1.27E-08
0	0	122	0	rpmH	SL3806	STY3939A	t3680	4067008	4067130	+	-174.35924	1.04E-06
0	0	179	0	rpmI	SL1269	STY1776	t1215	1371634	1371813	+	-174.35924	1.67E-09
0	0	683	0	rpsC	SL3401	STY4364	t4071	3613933	3614616	-	-174.35924	3.23E-34
0	0	485	0	rpsE	SL3390	STY4375	t4082	3609478	3609963	-	-174.35924	1.65E-24
0	0	374	0	rpsH	SL3393	STY4372	t4079	3610905	3611279	-	-174.35924	4.58E-19
0	0	314	0	rpsK	SL3384	STY4381	t4088	3606461	3606775	-	-174.35924	4.01E-16
0	0	356	0	rpsL	SL3415	STY4350	t4057	3623826	3624182	-	-174.35924	3.50E-18
0	0	338	0	rpsM	SL3385	STY4380	t4087	3606867	3607205	-	-174.35924	2.67E-17
0	0	272	0	rpsN	SL3394	STY4371	t4078	3611331	3611603	-	-174.35924	4.60E-14
0	0	260	0	rpsS	SL3403	STY4362	t4069	3614999	3615259	-	-174.35924	1.78E-13
0	0	1910	0	thrS	SL1267	STY1778	t1213	1369064	1370974	+	-174.35924	2.21E-94
0	0	1256	0	tyrS	SL1381	STY1673	t1317	1480348	1481604	+	-174.35924	2.59E-62
0	0	326	0	yadR	SL0205	STY0226	t0205	240169	240495	+	-174.35924	1.03E-16
0	0	860	0	yfjB	SL2655	STY2869	t2637	2833497	2834357	+	-174.35924	6.77E-43
1	2	3464	0.0002887	dnaE	SL0232	STY0254	t0232	269553	273017	+	-36.20768	1.40E-170
1	3	2396	0.0004174	yaeT	SL0225	STY0247	t0225	262384	264780	+	-33.37681	3.26E-118
1	1	2051	0.0004876	glyS	SL3620	STY4144	t3864	3861094	3863145	-	-32.17801	2.70E-101
1	2	1628	0.0006143	groEL	SL4267	STY4690	t4382	4595701	4597329	+	-30.39024	1.49E-80
1	2	1469	0.0006807	murE	SL0123	STY0143	t0127	143491	144960	+	-29.59172	9.30E-73
1	1	1457	0.0006863	murC	SL0129	STY0149	t0133	151155	152612	+	-29.52790	3.61E-72
1	23	1397	0.0007158	dnaB	SL4182	STY4442	t4152	4487892	4489289	+	-29.20046	3.16E-69
1	7	1382	0.0007236	asnS	SL0937	STY1004	t1934	1048055	1049437	-	-29.11634	1.72E-68
1	10	1352	0.0007396	glmU	SL3829	STY3916	t3657	4092373	4093725	-	-28.94525	5.08E-67
1	1	1319	0.0007582	-	SL3265	STY3472	t3211	3483121	3484440	-	-28.75248	2.11E-65
1	5	1298	0.0007704	murD	SL0126	STY0146	t0130	147412	148710	+	-28.62721	2.26E-64
1	3	1274	0.0007849	serS	SL0901	STY0961	t1971	1000368	1001642	+	-28.48146	3.39E-63
1	4	1244	0.0008039	ftsA	SL0132	STY0152	t0136	154372	155616	+	-28.29524	1.00E-61
1	6	1241	0.0008058	murA	SL3279	STY3486	t3224	3496321	3497562	-	-28.27636	1.41E-61
1	1	1241	0.0008058	rho	SL3876	STY3638	t3380	4147138	4148379	+	-28.27636	1.41E-61
1	5	1238	0.0008078	hemA	SL1705	STY1902	t1099	1831316	1832554	-	-28.25744	1.97E-61
2	8	2396	0.0008347	gyrB	SL3802	STY3943	t3684	4060178	4062574	-	-28.00050	3.26E-118
1	7	1193	0.0008382	tyrP	SL1870	STY2145	t0941	1987977	1989170	+	-27.96777	3.18E-59
1	4	1181	0.0008467	-	SL1154	STY1257	t1703	1258830	1260011	+	-27.88862	1.23E-58
1	6	1136	0.0008803	metK	SL3065	STY3243	t3002	3272404	3273540	+	-27.58421	1.98E-56
1	1	1109	0.0009017	dapE	SL2446	STY2721	t0376	2595419	2596528	+	-27.39553	4.18E-55

1	7	1088	0.0009191	asd	SL3506	STY4271	t3981	3729959	3731047	-	-27.24546	4.47E-54
1	1	1007	0.000993	lpxD	SL0227	STY0249	t0227	265410	266417	+	-26.63708	4.19E-50
2	23	1997	0.0010015	ligA	SL2390	STY2663	t0431	2536726	2538723	-	-26.57031	1.20E-98
1	15	971	0.0010299	rpoA	SL3382	STY4383	t4090	3604792	3605763	-	-26.35019	2.44E-48
2	6	1910	0.0010471	dnaX	SL0477	STY0528	t2376	540404	542314	+	-26.21912	2.21E-94
1	4	953	0.0010493	msbB	SL1823	STY2097	t0988	1941569	1942522	-	-26.20257	1.86E-47
1	6	929	0.0010764	prsA	SL1708	STY1906	t1096	1834623	1835552	+	-26.00117	2.80E-46
2	5	1829	0.0010935	rpoD	SL3185	STY3390	t3131	3397115	3398944	+	-25.87685	2.07E-90
1	34	911	0.0010977	fabD	SL1131	STY1233	t1726	1234432	1235343	+	-25.84653	2.14E-45
1	6	902	0.0011086	htrB	SL1092	STY1192	t1765	1198766	1199668	-	-25.76801	5.90E-45
1	1	893	0.0011198	glyQ	SL3621	STY4143	t3863	3863173	3864066	-	-25.68867	1.63E-44
5	29	4448	0.0011241	mukB	SL0931	STY0996	t1940	1036785	1041233	+	-25.65848	7.86E-219
1	10	860	0.0011628	dapA	SL2452	STY2727	t0370	2601768	2602628	-	-25.39045	6.77E-43
1	20	857	0.0011669	lgt	SL2980	STY3143	t2911	3180289	3181146	-	-25.36275	9.50E-43
1	1	809	0.0012361	nadE	SL1245	STY1803	t1189	1346981	1347790	-	-24.90520	2.14E-40
1	34	803	0.0012453	dapB	SL0065	STY0073	t0066	74029	74832	+	-24.84601	4.22E-40
1	2	776	0.0012887	thyA	SL2979	STY3142	t2910	3179488	3180264	-	-24.57383	8.90E-39
1	9	740	0.0013514	-	SL0222	STY0244	t0222	259360	260100	+	-24.19503	5.18E-37
1	5	719	0.0013908	dnaC	SL4474	STY4896	t4586	4823384	4824103	-	-23.96502	5.55E-36
1	1	719	0.0013908	plsC	SL3147	STY3350	t3094	3356666	3357385	-	-23.96502	5.55E-36
1	1	713	0.0014025	dnaQ	SL0259	STY0285	t2601	303402	304115	+	-23.89800	1.09E-35
1	9	710	0.0014085	ubiG	SL2245	STY2505	t0588	2378041	2378751	+	-23.86427	1.53E-35
4	12	2837	0.0014099	valS	SL4405	STY4814	t4510	4737605	4740442	-	-23.85581	7.76E-140
1	1	704	0.0014205	ybbF	SL0528	STY0583	t2326	597240	597944	-	-23.79635	3.02E-35
2	12	1406	0.0014225	lpdA	SL0154	STY0177	t0160	181020	182426	+	-23.78497	1.14E-69
1	3	698	0.0014327	-	SL1560	-	-	1677013	1677711	-	-23.72782	5.94E-35
2	28	1313	0.0015232	prlA	SL3387	STY4378	t4085	3607518	3608831	-	-23.23609	4.15E-65
2	2	1304	0.0015337	mukF	SL0929	STY0994	t1942	1034778	1036082	+	-23.18081	1.15E-64
1	1	650	0.0015385	folE	SL2170	STY2427	t0662	2288169	2288819	-	-23.15611	1.34E-32
1	4	623	0.0016051	ribE	SL1358	STY1696	t1294	1460413	1461036	+	-22.81455	2.83E-31
1	4	623	0.0016051	tmk	SL1137	STY1239	t1720	1239775	1240398	+	-22.81455	2.83E-31
1	1	611	0.0016367	rplC	SL3407	STY4358	t4065	3617048	3617659	-	-22.65768	1.10E-30
1	2	596	0.0016779	lolA	SL0899	STY0959	t1973	998142	998738	+	-22.45692	5.96E-30
1	7	587	0.0017036	rplD	SL3406	STY4359	t4066	3616432	3617019	-	-22.33387	1.65E-29
3	12	1754	0.0017104	aspS	SL1836	STY2109	t0976	1952687	1954441	-	-22.30165	9.86E-87
3	17	1748	0.0017162	ftsI	SL0122	STY0142	t0126	141738	143486	+	-22.27391	1.94E-86
2	3	1085	0.0018433	ribD	SL0410	STY0455	t2447	469339	470424	+	-21.69438	6.28E-54
2	6	1082	0.0018484	-	SL4409	STY4817	t4513	4742946	4744028	+	-21.67186	8.81E-54
3	4	1619	0.001853	pyrG	SL2932	STY3082	t2854	3120217	3121836	-	-21.65178	4.11E-80
2	14	1049	0.0019066	murG	SL0128	STY0148	t0132	149969	151018	+	-21.41962	3.66E-52
1	6	500	0.002	fabA	SL1007	STY1088	t1853	1114239	1114739	-	-21.02896	3.04E-25
3	19	1457	0.002059	ftsY	SL3536	STY4240	t3951	3760077	3761534	-	-20.79076	3.61E-72
4	18	1874	0.0021345	parE	SL3155	STY3359	t3102	3364675	3366549	-	-20.49531	1.29E-92
2	8	935	0.002139	fabH	SL1130	STY1232	t1727	1233463	1234398	+	-20.47774	1.42E-46
3	40	1397	0.0021475	gltX	SL2381	STY2654	t0442	2527888	2529285	-	-20.44543	3.16E-69
4	11	1811	0.0022087	glmS	SL3828	STY3917	t3658	4090355	4092166	-	-20.21378	1.58E-89
1	3	452	0.0022124	ribH	SL0411	STY0456	t2446	470531	470983	+	-20.20012	6.86E-23
1	4	452	0.0022124	rpsG	SL3414	STY4351	t4058	3623260	3623712	-	-20.20012	6.86E-23
3	15	1340	0.0022388	murF	SL0124	STY0144	t0128	144975	146315	+	-20.10220	1.97E-66
1	8	437	0.0022883	dut	SL3697	STY4063	t3787	3946045	3946482	+	-19.92146	3.73E-22
1	1	437	0.0022883	fabZ	SL0228	STY0250	t0228	266541	266978	+	-19.92146	3.73E-22
4	36	1727	0.0023162	dnaG	SL3184	STY3389	t3130	3395220	3396947	+	-19.82148	2.08E-85

4	27	1700	0.0023529	proS	SL0243	STY0269	t0245	282467	284167	-	-19.69097	4.38E-84
3	73	1256	0.0023885	hisS	SL2484	STY2767	t0334	2652596	2653852	-	-19.56645	2.59E-62
3	7	1226	0.002447	-	SL1156	STY1259	t1701	1260723	1261949	+	-19.36562	7.66E-61
2	4	815	0.002454	-	SL0707	STY0767	t2151	790305	791120	+	-19.34184	1.09E-40
3	8	1205	0.0024896	dfp	SL3696	STY4064	t3788	3944841	3946046	+	-19.22182	8.20E-60
10	41	4010	0.0024938	rpoB	SL4092	STY3732	t3474	4387248	4391258	+	-19.20798	2.36E-197
3	3	1178	0.0025467	dxr	SL0221	STY0243	t0221	257851	259029	+	-19.03287	1.73E-58
3	8	1157	0.0025929	yleB	SL0660	STY0717	t2202	735475	736632	+	-18.88261	1.85E-57
3	14	1139	0.0026339	ftsW	SL0127	STY0147	t0131	148815	149954	+	-18.75142	1.41E-56
1	7	377	0.0026525	rpsF	SL4324	STY4747	t4442	4651103	4651480	+	-18.69238	3.27E-19
1	5	365	0.0027397	secE	SL4086	STY3738	t3480	4383357	4383722	+	-18.42092	1.27E-18
3	24	1064	0.0028195	mraY	SL0125	STY0145	t0129	146327	147391	+	-18.17918	6.72E-53
3	8	1061	0.0028275	fba	SL3044	STY3226	t2987	3250240	3251301	-	-18.15537	9.43E-53
3	13	995	0.0030151	-	SL3182	STY3387	t3128	3393528	3394523	-	-17.61203	1.63E-49
5	29	1649	0.0030321	glnS	SL0668	STY0724	t2189	746841	748490	+	-17.56410	1.39E-81
3	17	959	0.0031283	lpxK	SL0922	STY0986	t1949	1028756	1029715	+	-17.29857	9.46E-48
2	12	635	0.0031496	ribB	SL3168	STY3373	t3115	3378565	3379200	-	-17.24058	7.30E-32
3	17	941	0.0031881	accA	SL0233	STY0255	t0233	273048	273989	+	-17.13690	7.22E-47
4	19	1241	0.0032232	accC	SL3353	STY3560	t3295	3572095	3573336	+	-17.04330	1.41E-61
3	6	920	0.0032609	ribF	SL0046	STY0054	t0047	52881	53801	+	-16.94388	7.73E-46
6	22	1829	0.0032805	secD	SL0402	STY0445	t2456	461826	463655	+	-16.89251	2.07E-90
2	32	605	0.0033058	gmk	SL3706	STY4052	t3778	3954421	3955026	+	-16.82664	2.16E-30
4	18	1196	0.0033445	fabB	SL2347	STY2609	t0486	2486347	2487543	-	-16.72675	2.26E-59
1	5	296	0.0033784	-	SL1264	STY1782	t1209	1365082	1365378	+	-16.64009	3.06E-15
8	27	2342	0.0034159	imp	SL0094	STY0108	t0096	105822	108164	-	-16.54509	1.45E-115
6	35	1622	0.0036991	aarF	SL3926	STY3587	t3325	4200212	4201834	+	-15.85577	2.93E-80
4	21	1064	0.0037594	prfA	SL1704	STY1901	t1100	1830193	1831257	-	-15.71509	6.72E-53
1	2	266	0.0037594	rplY	SL2201	STY2461	t0630	2324158	2324424	+	-15.71509	9.05E-14
2	12	530	0.0037736	rimM	SL2648	STY2862	t2630	2827270	2827800	-	-15.68226	1.03E-26
10	50	2618	0.0038197	gyrA	SL2241	STY2499	t0592	2371436	2374054	-	-15.57624	4.25E-129
2	5	515	0.0038835	rplF	SL3392	STY4373	t4080	3610359	3610874	-	-15.43145	5.59E-26
2	3	512	0.0039063	fldA	SL0676	STY0732	t2181	756548	757060	-	-15.38030	7.84E-26
6	52	1520	0.0039474	lnt	SL0655	STY0711	t2207	729592	731112	-	-15.28851	2.94E-75
5	19	1265	0.0039526	aroA*	SL0915*	STY0978	t1956	1019210	1020475	+	-15.27696	9.37E-63
9	77	2240	0.0040179	parC	SL3148	STY3351	t3095	3357660	3359900	-	-15.13306	1.46E-110
1	5	248	0.0040323	minE	SL1744	STY1946	t1061	1870496	1870744	+	-15.10159	6.91E-13
3	13	737	0.0040706	ubiE	SL3924	STY3589	t3327	4198845	4199582	+	-15.01837	7.27E-37
1	2	236	0.0042373	rpsQ	SL3398	STY4367	t4074	3613065	3613301	-	-14.66373	2.68E-12
3	15	707	0.0042433	-	SL3291	STY3498	t3236	3505063	3505770	+	-14.65120	2.15E-35
4	26	932	0.0042918	coaA	SL4083	STY3740	t3487	4380034	4380966	-	-14.55027	2.00E-46
1	5	230	0.0043478	rpsP	SL2649	STY2863	t2631	2827837	2828067	-	-14.43513	5.27E-12
1	3	230	0.0043478	ssaI	SL1342	STY1712	t1276	1446893	1447123	+	-14.43513	5.27E-12
5	27	1145	0.0043668	pgk	SL3045	STY3227	t2988	3251421	3252566	-	-14.39637	7.17E-57
5	60	1130	0.0044248	lpxB	SL0230	STY0252	t0230	267788	268918	+	-14.27890	3.90E-56
3	27	677	0.0044313	-	SL1967	-	-	2072253	2072930	+	-14.26574	6.36E-34
4	25	887	0.0045096	era	SL2542	STY2826	t0276	2721651	2722538	-	-14.10938	3.21E-44
6	79	1280	0.0046875	eno	SL2931	STY3081	t2853	3118836	3120116	-	-13.76231	1.72E-63
6	58	1274	0.0047096	wbaX	SL2065	STY2297	t0785	2166349	2167623	-	-13.72002	3.39E-63
4	22	839	0.0047676	cdsA	SL0223	STY0245	t0223	260131	260970	+	-13.60971	7.25E-42
4	21	836	0.0047847	-	SL0831	-	-	925884	926720	+	-13.57738	1.02E-41
4	42	833	0.0048019	murI	SL4080	STY3743	t3494	4371346	4372179	+	-13.54492	1.43E-41
4	24	833	0.0048019	tsf	SL0218	STY0240	t0218	255285	256118	+	-13.54492	1.43E-41

4	30	806	0.0049628	dapD	SL0214	STY0236	t0214	249294	250100	-	-13.24654	3.01E-40
3	47	602	0.0049834	coaE	SL0140	STY0162	t0146	163220	163822	-	-13.20890	3.03E-30
3	28	602	0.0049834	rpsD	SL3383	STY4382	t4089	3605807	3606409	-	-13.20890	3.03E-30
1	3	200	0.005	infA	SL0891	STY0951	t1980	986893	987093	-	-13.17866	1.56E-10
8	34	1556	0.0051414	mviN	SL1107	STY1209	t1750	1211114	1212670	+	-12.92454	5.04E-77
7	91	1343	0.0052122	ffh	SL2650	STY2864	t2632	2828399	2829742	-	-12.79943	1.40E-66
11	132	2096	0.0052481	fusA	SL3413	STY4352	t4059	3621049	3623145	-	-12.73658	1.68E-103
4	44	761	0.0052562	ssaT	SL1355	STY1699	t1289	1456516	1457277	+	-12.72235	4.84E-38
5	16	947	0.0052798	lipA	SL0621	STY0683	t2234	694359	695306	-	-12.68128	3.67E-47
5	8	944	0.0052966	hemH	SL0482	STY0533	t2371	546411	547355	+	-12.65216	5.15E-47
3	15	566	0.0053004	pth	SL1711	STY1909	t1093	1837979	1838545	+	-12.64567	1.76E-28
7	35	1250	0.0056	folC	SL2334	STY2596	t0499	2475079	2476329	-	-12.13821	5.09E-62
3	13	533	0.0056285	yfhC	SL2530	STY2814	t0289	2711718	2712251	-	-12.09108	7.32E-27
6	59	1064	0.0056391	-	SL4410	STY4818	t4514	4744046	4745110	+	-12.07365	6.72E-53
14	115	2402	0.0058285	plsB	SL4172	STY4431	t4141	4477481	4479883	-	-11.76599	1.66E-118
1	1	167	0.005988	csrA	SL2806	STY2947	t2727	2993536	2993703	-	-11.51303	6.48E-09
16	78	2660	0.006015	infB	SL3259	STY3467	t3204	3474324	3476984	-	-11.47074	3.70E-131
6	41	986	0.0060852	holB	SL1138	STY1240	t1719	1240413	1241399	+	-11.36162	4.49E-49
1	5	164	0.0060976	-	SL1659	STY1323	t1640	1782420	1782584	+	-11.34249	9.09E-09
6	46	968	0.0061983	-	SL2391	STY2664	t0429	2538813	2539781	-	-11.18773	3.43E-48
6	21	959	0.0062565	-	SL0413	STY0458	t2444	471520	472479	+	-11.09933	9.46E-48
6	35	932	0.0064378	sifB	SL1532	STY1462	t1511	1649043	1649975	+	-10.82802	2.00E-46
3	18	461	0.0065076	coaD	SL3691	STY4069	t3793	3941947	3942408	+	-10.72512	2.48E-23
3	12	461	0.0065076	ispF	SL2908	STY3054	t2830	3092836	3093297	-	-10.72512	2.48E-23
5	28	749	0.0066756	trmD	SL2647	STY2861	t2629	2826458	2827207	-	-10.48112	1.88E-37
3	9	443	0.006772	yjeE	SL4291	STY4714	t4408	4620476	4620919	+	-10.34321	1.90E-22
4	20	590	0.0067797	lexA	SL4174	STY4433	t4143	4480549	4481139	+	-10.33234	1.17E-29
2	16	293	0.0068259	rpsJ	SL3408	STY4357	t4064	3617710	3618003	-	-10.26677	4.29E-15
5	14	731	0.0068399	ssaJ	SL1343	STY1711	t1277	1447138	1447869	+	-10.24699	1.43E-36
5	44	719	0.0069541	-	SL2623	STY2852	t2619	2807081	2807800	+	-10.08698	5.55E-36
12	80	1655	0.0072508	rpsA	SL0918	STY0981	t1953	1022367	1024022	+	-9.68054	7.05E-82
9	80	1223	0.007359	cca	SL3177	STY3383	t3124	3389935	3391158	+	-9.53552	1.07E-60
19	110	2579	0.0073672	topA	SL1646	STY1336	t1627	1766654	1769233	-	-9.52454	3.47E-127
31	405	4205	0.0073722	rpoC	SL4093	STY3731	t3473	4391353	4395558	+	-9.51791	6.47E-207
5	22	674	0.0074184	-	SL1069	STY1168	t1789	1176998	1177672	+	-9.45654	8.93E-34
7	65	938	0.0074627	hemC	SL3893	STY3621	t3359	4166603	4167541	-	-9.39802	1.01E-46
3	34	401	0.0074813	nusB	SL0412	STY0457	t2445	471022	471423	+	-9.37350	2.17E-20
9	57	1193	0.007544	mesJ	SL0237	STY0261	t0238	278756	279949	+	-9.29125	3.18E-59
1	1	131	0.0076336	-	SL1203	STY1853	t1146	1306483	1306614	-	-9.17465	3.77E-07
6	46	785	0.0076433	suhB	SL2508	STY2792	t0310	2682161	2682946	+	-9.16206	3.22E-39
6	49	776	0.007732	map	SL0216	STY0238	t0216	253057	253833	-	-9.04782	8.90E-39
7	66	899	0.0077864	lpxC	SL0134	STY0154	t0138	156947	157846	+	-8.97811	8.28E-45
7	42	896	0.0078125	accD	SL2335	STY2597	t0498	2476415	2477311	-	-8.94488	1.16E-44
4	7	512	0.0078125	ppa	SL4347	STY4773	t4468	4675331	4675843	-	-8.94488	7.84E-26
5	44	629	0.0079491	-	SL1564	STY1437	t1537	1681521	1682150	+	-8.77210	1.44E-31
3	9	377	0.0079576	-	SL1785	STY1988	t1022	1910995	1911372	+	-8.76151	3.27E-19
9	26	1121	0.0080285	cydB	SL0723	STY0787	t2135	810533	811654	+	-8.67270	1.08E-55
9	61	1115	0.0080717	-	SL0703	STY0761	t2155	785494	786609	+	-8.61894	2.12E-55
11	51	1352	0.0081361	purB	SL1170	STY1272	t1688	1276349	1277701	-	-8.53927	5.08E-67
4	35	491	0.0081466	def	SL3373	STY4391	t4098	3598427	3598918	+	-8.52626	8.40E-25
2	12	245	0.0081633	rpsT	SL0044	STY0052	t0045	52289	52534	-	-8.50578	9.70E-13
7	49	848	0.0082547	folD	SL0535	STY0588	t2321	602199	603047	-	-8.39364	2.62E-42

9	30	1088	0.0082721	trmU	SL1172	STY1274	t1686	1278451	1279539	-	-8.37248	4.47E-54
10	39	1193	0.0083822	kdtA	SL3690	STY4070	t3794	3940727	3941920	+	-8.23883	3.18E-59
13	69	1550	0.0083871	cydA	SL0722	STY0786	t2136	808949	810499	+	-8.23296	9.93E-77
2	6	236	0.0084746	-	SL0514	-	-	581902	582138	+	-8.12782	2.68E-12
6	34	707	0.0084866	rpsB	SL0217	STY0239	t0217	254302	255009	+	-8.11348	2.15E-35
15	67	1760	0.0085227	-	SL0706	STY0765	t2152	788334	790094	+	-8.07029	5.01E-87
14	155	1628	0.0085995	yidC	SL3809	STY3938	t3678	4067748	4069376	+	-7.97906	1.49E-80
9	90	1046	0.0086042	hemE	SL4106	STY3718	t3464	4405485	4406531	+	-7.97350	5.13E-52
8	44	929	0.0086114	fmt	SL3374	STY4390	t4097	3598952	3599881	+	-7.96497	2.80E-46
6	30	692	0.0086705	ispD	SL2909	STY3055	t2831	3093315	3094007	-	-7.89522	1.17E-34
2	4	230	0.0086957	-	SL1480	-	-	1583701	1583931	-	-7.86567	5.27E-12
5	31	572	0.0087413	rlpB	SL0635	STY0698	t2220	707939	708511	-	-7.81221	8.96E-29
4	149	449	0.0089087	-	SL2593	STY2066	-	2768375	2768824	+	-7.61769	9.63E-23
9	149	1010	0.0089109	murB	SL4081	STY3742	t3489	4378012	4379022	+	-7.61515	2.99E-50
18	105	2015	0.008933	metG	SL2132	STY2384	t0701	2247464	2249479	+	-7.58967	1.57E-99
20	82	2180	0.0091743	priA	SL4044	STY3775	t3523	4324808	4326988	-	-7.31454	1.27E-107
6	36	641	0.0093604	rpiA	SL3039	STY3219	t2981	3245506	3246147	-	-7.10600	3.71E-32
9	62	953	0.0094439	ispB	SL3277	STY3484	t3222	3494773	3495726	+	-7.01340	1.86E-47
11	115	1160	0.0094828	visB	SL3033	STY3213	t2975	3239194	3240354	-	-6.97046	1.32E-57
5	20	527	0.0094877	hemG	SL3940	STY3573	t3311	4216690	4217217	+	-6.96506	1.44E-26
2	5	209	0.0095694	-	SL1341	STY1713	t1275	1446654	1446863	+	-6.87530	5.65E-11
8	35	836	0.0095694	rpoH	SL3533	STY4243	t3954	3757258	3758094	-	-6.87530	1.02E-41
4	18	410	0.0097561	rplM	SL3317	STY3525	t3261	3534240	3534650	-	-6.67227	7.87E-21
10	87	1013	0.0098717	holA	SL0634	STY0697	t2221	706908	707921	-	-6.54799	2.13E-50
17	67	1703	0.0099824	cydC	SL0894	STY0954	t1978	988172	989875	-	-6.42992	3.12E-84
9	27	896	0.0100446	pagO	SL1793	STY2000	t1016	1917084	1917980	+	-6.36393	1.16E-44
4	21	398	0.0100503	-	SL3072	STY3250	t3009	3279377	3279775	+	-6.35800	3.05E-20
10	57	983	0.0101729	wbaV	SL2064	STY2296	t0786	2165343	2166326	-	-6.22886	6.30E-49
7	24	686	0.0102041	mukE	SL0930	STY0995	t1941	1036081	1036767	+	-6.19626	2.30E-34
10	146	977	0.0102354	-	SL2633	-	-	2816812	2817789	-	-6.16352	1.24E-48
7	37	680	0.0102941	mtn	SL0208	STY0229	t0208	242067	242747	-	-6.10239	4.54E-34
6	26	581	0.010327	-	SL3948	STY3880	t3620	4230699	4231280	-	-6.06823	3.24E-29
7	31	662	0.010574	rnc	SL2543	STY2827	t0277	2722553	2723215	-	-5.81424	3.46E-33
10	90	932	0.0107296	lytB	SL0050	STY0058	t0051	57805	58737	+	-5.65638	2.00E-46
7	67	650	0.0107692	ftsE	SL3535	STY4241	t3952	3759406	3760056	-	-5.61644	1.34E-32
16	64	1460	0.0109589	yigC	SL3932	STY3581	t3319	4205152	4206612	+	-5.42663	2.57E-72
9	140	812	0.0110837	ftsQ	SL0131	STY0151	t0135	153545	154357	+	-5.30295	1.53E-40
30	247	2687	0.0111649	secA	SL0136	STY0156	t0140	158702	161389	+	-5.22309	1.76E-132
6	35	536	0.011194	-	SL3290	STY3497	t3235	3504502	3505038	+	-5.19448	5.22E-27
6	49	527	0.0113852	nusG	SL4087	STY3737	t3479	4383742	4384269	+	-5.00820	1.44E-26
10	64	878	0.0113895	-	SL0702	STY0760	t2156	784585	785463	+	-5.00401	8.87E-44
20	69	1748	0.0114416	cydD	SL0895	STY0955	t1977	989894	991642	-	-4.95361	1.94E-86
11	43	956	0.0115063	lepB	SL2544	STY2828	t0275	2723507	2724463	-	-4.89133	1.33E-47
4	28	344	0.0116279	folB	SL3179	STY3385	t3126	3392200	3392544	-	-4.77477	1.36E-17
5	26	422	0.0118483	-	SL0032	STY0038	t0033	34385	34807	-	-4.56562	2.03E-21
2	9	164	0.0121951	-	SL0924	STY0989	t1947	1031061	1031225	+	-4.24187	9.09E-09
23	102	1883	0.0122146	mrdA	SL0628	STY0691	t2227	701442	703325	-	-4.22391	4.66E-93
6	28	482	0.0124481	lspA	SL0048	STY0056	t0049	56698	57180	+	-4.00954	2.32E-24
12	85	962	0.012474	-	SL3024	STY3204	t2966	3227647	3228609	+	-3.98595	6.75E-48
6	46	479	0.0125261	rplJ	SL4090	STY3734	t3476	4386001	4386480	+	-3.93859	3.26E-24
7	67	554	0.0126354	yrdC	SL3369	STY4395	t4102	3595580	3596134	-	-3.83962	6.84E-28
12	56	944	0.0127119	birA	SL4082	STY3741	t3488	4379037	4379981	+	-3.77070	5.15E-47

6	19	458	0.0131004	-	SL2659	STY2873	t2641	2837055	2837513	-	-3.42470	3.49E-23
16	174	1208	0.013245	sseJ	SL1561	STY1439a	-	1678155	1679363	+	-3.29767	5.84E-60
10	44	743	0.013459	-	SL1563	STY1438	t1536	1680753	1681496	+	-3.11139	3.69E-37
4	15	293	0.0136519	rplU	SL3276	STY3483	t3221	3494204	3494497	-	-2.94503	4.29E-15
26	91	1871	0.0138963	aceF	SL0153	STY0176	t0159	178931	180802	+	-2.73643	1.81E-92
19	92	1355	0.0140221	-	SL1628	STY1353	t1612	1747019	1748374	-	-2.62996	3.62E-67
12	73	854	0.0140515	ubiA	SL4171	STY4430	t4140	4476510	4477364	+	-2.60519	1.33E-42
15	155	1061	0.0141376	wbaG	SL2068	STY2301	t0781	2169990	2171051	-	-2.53280	9.43E-53
7	29	491	0.0142566	crr	SL2396	STY2670	t0424	2544385	2544876	+	-2.43318	8.40E-25
11	75	770	0.0142857	envF	SL1179	-	-	1285051	1285821	-	-2.40891	1.75E-38
3	34	209	0.0143541	-	SL2549	-	-	2730709	2730918	+	-2.35201	5.65E-11
18	123	1250	0.0144	-	SL0742	-	-	827815	829065	+	-2.31387	5.09E-62
5	52	347	0.0144092	ftsL	SL0121	STY0141	t0125	141357	141704	+	-2.30622	9.66E-18
9	36	623	0.0144462	lipB	SL0623	STY0686	t2232	696785	697408	-	-2.27556	2.83E-31
16	70	1094	0.0146252	mrdB	SL0627	STY0690	t2228	700327	701421	-	-2.12796	2.27E-54

*SL0916 and *aroA* are (partly) deleted in the strain of Typhimurium used

8.3.6 Shared essential genes

Total inserts Tm	Total reads Tm	Total inserts Ty	Total reads Ty	Unique Tm ID	Unique Ty Ty2 ID	Unique Ty CT18 ID	Gene name	Gene length Tm	Gene length Ty
3	6	0	0	SL0046	t0047	STY0054	ribF	920	920
0	0	0	0	SL0047	t0048	STY0055	ileS	2816	2816
6	28	0	0	SL0048	t0049	STY0056	lspA	482	482
10	90	3	32	SL0050	t0051	STY0058	lytB	932	932
1	34	1	1	SL0065	t0066	STY0073	dapB	803	803
8	27	2	10	SL0094	t0096	STY0108	imp	2342	2336
5	52	1	2	SL0121	t0125	STY0141	ftsL	347	347
3	17	0	0	SL0122	t0126	STY0142	ftsI	1748	1748
1	2	0	0	SL0123	t0127	STY0143	murE	1469	1469
3	15	1	4	SL0124	t0128	STY0144	murF	1340	1340
3	24	0	0	SL0125	t0129	STY0145	mraY	1064	1064
1	5	0	0	SL0126	t0130	STY0146	murD	1298	1298
3	14	5	12	SL0127	t0131	STY0147	ftsW	1139	1226
2	14	1	1	SL0128	t0132	STY0148	murG	1049	1049
1	1	0	0	SL0129	t0133	STY0149	murC	1457	1457
9	140	2	12	SL0131	t0135	STY0151	ftsQ	812	812
1	4	0	0	SL0132	t0136	STY0152	ftsA	1244	1244
0	0	1	1	SL0133	t0137	STY0153	ftsZ	1133	1133
7	66	3	8	SL0134	t0138	STY0154	lpxC	899	899
30	247	10	30	SL0136	t0140	STY0156	secA	2687	2687
3	47	1	1	SL0140	t0146	STY0162	yacE	602	602
26	91	4	8	SL0153	t0159	STY0176	aceF	1871	1871
2	12	1	1	SL0154	t0160	STY0177	lpdA	1406	1409
0	0	1	1	SL0205	t0205	STY0226	yadR	326	371
4	30	0	0	SL0214	t0214	STY0236	dapD	806	806
6	49	3	11	SL0216	t0216	STY0238	map	776	776
6	34	2	16	SL0217	t0217	STY0239	rpsB	707	707
4	24	0	0	SL0218	t0218	STY0240	tsf	833	833
0	0	0	0	SL0219	t0219	STY0241	pyrH	707	707
0	0	1	1	SL0220	t0220	STY0242	frr	539	539
3	3	0	0	SL0221	t0221	STY0243	dxr	1178	1178
1	9	1	9	SL0222	t0222	STY0244	-	740	740
4	22	2	2	SL0223	t0223	STY0245	cdsA	839	839
1	3	3	6	SL0225	t0225	STY0247	yaeT	2396	2393
1	1	4	10	SL0227	t0227	STY0249	lpxD	1007	1007
1	1	0	0	SL0228	t0228	STY0250	fabZ	437	437
0	0	0	0	SL0229	t0229	STY0251	lpxA	770	770
5	60	1	29	SL0230	t0230	STY0252	lpxB	1130	1130
1	2	0	0	SL0232	t0232	STY0254	dnaE	3464	3464
3	17	0	0	SL0233	t0233	STY0255	accA	941	941
9	57	5	16	SL0237	t0238	STY0261	mesJ	1193	1274
4	27	2	10	SL0243	t0245	STY0269	proS	1700	1700
1	1	0	0	SL0259	t2601	STY0285	dnaQ	713	713
0	0	1	1	SL0367	t2492	STY0404	hemB	956	956
6	22	4	11	SL0402	t2456	STY0445	secD	1829	1829
8	137	7	104	SL0403	t2455	STY0446	secF	953	953

2	3	3	26	SL0410	t2447	STY0455	ribD	1085	1085
1	3	0	0	SL0411	t2446	STY0456	ribH	452	452
3	34	2	16	SL0412	t2445	STY0457	nusB	401	401
6	21	4	10	SL0413	t2444	STY0458	-	959	959
0	0	1	2	SL0416	t2441	STY0461	dxs	1844	1844
2	6	5	11	SL0477	t2376	STY0528	dnaX	1910	1910
0	0	0	0	SL0481	t2372	STY0532	adk	626	626
5	8	1	2	SL0482	t2371	STY0533	hemH	944	944
1	1	1	3	SL0528	t2326	STY0583	ybbF	704	704
0	0	0	0	SL0530	t2324	STY0585	cysS	1367	1367
7	49	4	24	SL0535	t2321	STY0588	folD	848	848
16	70	0	0	SL0627	t2228	STY0690	mrdB	1094	1094
23	102	3	19	SL0628	t2227	STY0691	pbpA	1883	1883
0	0	2	26	SL0633	t2222	STY0696	-	623	623
10	87	4	21	SL0634	t2221	STY0697	holA	1013	1013
5	31	1	6	SL0635	t2220	STY0698	rlpB	572	572
0	0	0	0	SL0636	t2219	STY0699	leuS	2564	2564
6	52	2	12	SL0655	t2207	STY0711	lnt	1520	1520
3	8	4	27	SL0660	t2202	STY0717	yleB	1157	1157
5	29	6	80	SL0668	t2189	STY0724	glnS	1649	1649
2	3	2	24	SL0676	t2181	STY0732	fldA	512	512
1	3	0	0	SL0891	t1980	STY0951	infA	200	200
17	67	21	53	SL0894	t1978	STY0954	cydC	1703	1703
1	2	0	0	SL0899	t1973	STY0959	lolA	596	596
1	3	1	1	SL0901	t1971	STY0961	serS	1274	1274
12	80	9	28	SL0918	t1953	STY0981	rpsA	1655	1655
0	0	0	0	SL0921	t1950	STY0985	msbA	1730	1730
3	17	1	3	SL0922	t1949	STY0986	lpxK	959	959
2	9	0	0	SL0924	t1947	STY0989	-	164	164
0	0	0	0	SL0925	t1946	STY0990	kdsB	728	728
2	2	0	0	SL0929	t1942	STY0994	mukF	1304	1304
7	24	3	20	SL0930	t1941	STY0995	mukE	686	659
5	29	3	6	SL0931	t1940	STY0996	mukB	4448	4448
1	7	0	0	SL0937	t1934	STY1004	asnS	1382	1382
1	6	1	1	SL1007	t1853	STY1088	fabA	500	500
1	6	9	37	SL1092	t1765	STY1192	htrB	902	902
8	34	3	9	SL1107	t1750	STY1209	mviN	1556	1475
2	8	2	2	SL1130	t1727	STY1232	fabH	935	935
1	34	1	5	SL1131	t1726	STY1233	fabD	911	911
0	0	0	0	SL1132	t1725	STY1234	fabG	716	716
0	0	1	8	SL1133	t1724	STY1235	acpP	218	218
1	4	0	0	SL1137	t1720	STY1239	tmk	623	623
6	41	4	29	SL1138	t1719	STY1240	holB	986	986
1	4	6	53	SL1154	t1703	STY1257	-	1181	1292
0	0	0	0	SL1155	t1702	STY1258	-	683	683
3	7	1	1	SL1156	t1701	STY1259	-	1226	1226
11	51	3	6	SL1170	t1688	STY1272	purB	1352	1352
9	30	2	6	SL1172	t1686	STY1274	trmU	1088	1088
0	0	1	3	SL1225	t1169	STY1825	gapA	977	977
1	1	0	0	SL1245	t1189	STY1803	nadE	809	809
0	0	0	0	SL1267	t1213	STY1778	thrS	1910	1910
0	0	0	0	SL1268	t1214	STY1777	infC	416	524

0	0	0	0	SL1269	t1215	STY1776	rpmI	179	179
0	0	0	0	SL1270	t1216	STY1775	rplT	338	338
0	0	0	0	SL1271	t1218	STY1773	pheS	965	965
0	0	0	0	SL1272	t1219	STY1772	pheT	2369	2369
1	4	1	1	SL1358	t1294	STY1696	ribE	623	623
0	0	0	0	SL1381	t1317	STY1673	tyrS	1256	1256
0	0	2	7	SL1631	t1613	STY1352	fabI	770	770
0	0	1	5	SL1643	t1624	STY1340	ribA	572	572
0	0	1	30	SL1700	t1104	STY1897	kdsA	836	836
4	21	0	0	SL1704	t1100	STY1901	prfA	1064	1064
1	5	0	0	SL1705	t1099	STY1902	hemA	1238	1238
0	0	0	0	SL1706	t1098	STY1904	hemM	605	605
0	0	0	0	SL1707	t1097	STY1905	ipk	833	833
1	6	0	0	SL1708	t1096	STY1906	prsA	929	929
0	0	2	20	SL1749	t1057	STY1950	-	677	677
3	12	0	0	SL1836	t0976	STY2109	aspS	1754	1754
0	0	0	0	SL1844	t0968	STY2117	argS	1715	1715
0	0	0	0	SL1875	t0931	STY2153	pgsA	530	530
18	105	17	85	SL2132	t0701	STY2384	metG	2015	2015
1	1	0	0	SL2170	t0662	STY2427	folE	650	650
1	2	1	1	SL2201	t0630	STY2461	rplY	266	266
10	50	2	4	SL2241	t0592	STY2499	gyrA	2618	2618
1	9	2	21	SL2245	t0588	STY2505	ubiG	710	710
7	35	3	5	SL2334	t0499	STY2596	folC	1250	1250
7	42	3	11	SL2335	t0498	STY2597	accD	896	896
4	18	2	15	SL2347	t0486	STY2609	fabB	1196	1196
3	40	5	14	SL2381	t0442	STY2654	gltX	1397	1397
2	23	0	0	SL2390	t0431	STY2663	ligA	1997	1997
6	46	1	1	SL2391	t0429	STY2664	-	968	968
7	29	4	12	SL2396	t0424	STY2670	crr	491	491
1	1	0	0	SL2446	t0376	STY2721	dapE	1109	1109
1	10	1	6	SL2452	t0370	STY2727	dapA	860	860
0	0	0	0	SL2481	t0337	STY2764	-	1454	1454
3	73	2	22	SL2484	t0334	STY2767	hisS	1256	1256
0	0	0	0	SL2485	t0333	STY2768	gcpE	1100	1100
6	46	4	27	SL2508	t0310	STY2792	suhB	785	785
3	13	2	9	SL2530	t0289	STY2814	yfhC	533	533
0	0	0	0	SL2539	t0280	STY2823	acpS	362	362
4	25	4	11	SL2542	t0276	STY2826	era	887	887
11	43	4	27	SL2544	t0275	STY2828	lepB	956	956
0	0	0	0	SL2616	t0258	STY2845	pssA	1337	1337
5	44	1	11	SL2623	t2619	STY2852	-	719	719
5	28	2	10	SL2647	t2629	STY2861	trmD	749	749
2	12	0	0	SL2648	t2630	STY2862	yfjA	530	530
1	5	0	0	SL2649	t2631	STY2863	rpsP	230	230
7	91	5	35	SL2650	t2632	STY2864	ffh	1343	1343
0	0	1	4	SL2655	t2637	STY2869	yfjB	860	860
6	19	5	8	SL2659	t2641	STY2873	-	458	458
0	0	0	0	SL2708	t4337	STY4644	cI	557	614
1	1	0	0	SL2806	t2727	STY2947	csrA	167	167
0	0	0	0	SL2807	t2728	STY2948	alaS	2612	2612
3	12	0	0	SL2908	t2830	STY3054	ygbB	461	461

6	30	3	16	SL2909	t2831	STY3055	ygbP	692	692
6	79	8	64	SL2931	t2853	STY3081	eno	1280	1280
3	4	3	11	SL2932	t2854	STY3082	pyrG	1619	1619
1	2	3	5	SL2979	t2910	STY3142	thyA	776	776
1	20	0	0	SL2980	t2911	STY3143	lgt	857	857
0	0	1	2	SL3016	t2958	STY3196	lysS	1499	1499
0	0	0	0	SL3017	t2959	STY3197	prfB	863	1080
12	85	6	55	SL3024	t2966	STY3204	-	962	962
11	115	6	14	SL3033	t2975	STY3213	visB	1160	1160
6	36	4	11	SL3039	t2981	STY3219	rpiA	641	641
3	8	1	1	SL3044	t2987	STY3226	fba	1061	1061
5	27	2	2	SL3045	t2988	STY3227	pgk	1145	1145
1	6	2	8	SL3065	t3002	STY3243	metK	1136	1136
4	21	2	7	SL3072	t3009	STY3250	-	398	398
1	1	0	0	SL3147	t3094	STY3350	plsC	719	719
9	77	1	9	SL3148	t3095	STY3351	parC	2240	2240
4	18	0	0	SL3155	t3102	STY3359	parE	1874	1874
2	12	2	11	SL3168	t3115	STY3373	ribB	635	635
9	80	4	24	SL3177	t3124	STY3383	cca	1223	1223
4	28	0	0	SL3179	t3126	STY3385	folB	344	344
3	13	0	0	SL3182	t3128	STY3387	-	995	995
4	36	0	0	SL3184	t3130	STY3389	dnaG	1727	1727
2	5	2	6	SL3185	t3131	STY3390	rpoD	1829	1964
16	78	3	9	SL3259	t3204	STY3467	infB	2660	2660
1	1	0	0	SL3265	t3211	STY3472	-	1319	1319
0	0	0	0	SL3275	t3220	STY3482	rpmA	239	239
4	15	0	0	SL3276	t3221	STY3483	rplU	293	293
9	62	0	0	SL3277	t3222	STY3484	ispB	953	953
1	6	1	6	SL3279	t3224	STY3486	murA	1241	1241
6	35	4	13	SL3290	t3235	STY3497	-	536	500
3	15	0	0	SL3291	t3236	STY3498	-	707	707
4	18	0	0	SL3317	t3261	STY3525	rplM	410	410
0	0	0	0	SL3352	t3294	STY3559	accB	452	452
4	19	1	6	SL3353	t3295	STY3560	accC	1241	1331
7	67	1	3	SL3369	t4102	STY4395	yrdC	554	554
4	35	1	1	SL3373	t4098	STY4391	fms	491	491
8	44	2	30	SL3374	t4097	STY4390	fmt	929	929
0	0	1	2	SL3381	t4091	STY4384	rplQ	365	365
1	15	0	0	SL3382	t4090	STY4383	rpoA	971	971
3	28	0	0	SL3383	t4089	STY4382	rpsD	602	602
0	0	0	0	SL3384	t4088	STY4381	rpsK	314	371
0	0	0	0	SL3385	t4087	STY4380	rpsM	338	338
2	28	1	5	SL3387	t4085	STY4378	prlA	1313	1313
0	0	0	0	SL3388	t4084	STY4377	rplO	416	416
0	0	0	0	SL3389	t4083	STY4376	rpmD	161	161
0	0	0	0	SL3390	t4082	STY4375	rpsE	485	485
0	0	0	0	SL3391	t4081	STY4374	rplR	257	335
2	5	0	0	SL3392	t4080	STY4373	rplF	515	515
0	0	0	0	SL3393	t4079	STY4372	rpsH	374	374
0	0	0	0	SL3394	t4078	STY4371	rpsN	272	287
0	0	0	0	SL3395	t4077	STY4370	rplE	521	521
0	0	0	0	SL3396	t4076	STY4369	rplX	296	296

0	0	0	0	SL3397	t4075	STY4368	rplN	353	353
1	2	0	0	SL3398	t4074	STY4367	rpsQ	236	236
0	0	0	0	SL3399	t4073	STY4366	rpmC	173	173
0	0	0	0	SL3400	t4072	STY4365	rplP	392	392
0	0	0	0	SL3401	t4071	STY4364	rpsC	683	683
0	0	0	0	SL3402	t4070	STY4363	rplV	314	314
0	0	0	0	SL3403	t4069	STY4362	rpsS	260	260
0	0	0	0	SL3404	t4068	STY4361	rplB	803	803
0	0	0	0	SL3405	t4067	STY4360	rplW	284	284
1	7	1	2	SL3406	t4066	STY4359	rplD	587	587
1	1	1	1	SL3407	t4065	STY4358	rplC	611	611
2	16	0	0	SL3408	t4064	STY4357	rpsJ	293	293
11	132	4	13	SL3413	t4059	STY4352	fusA	2096	2096
1	4	0	0	SL3414	t4058	STY4351	rpsG	452	452
0	0	0	0	SL3415	t4057	STY4350	rpsL	356	356
1	7	0	0	SL3506	t3981	STY4271	asd	1088	1088
8	35	1	10	SL3533	t3954	STY4243	rpoH	836	836
7	67	0	0	SL3535	t3952	STY4241	ftsE	650	650
3	19	2	24	SL3536	t3951	STY4240	ftsY	1457	1457
1	1	4	10	SL3620	t3864	STY4144	glyS	2051	2051
1	1	2	7	SL3621	t3863	STY4143	glyQ	893	893
10	39	3	117	SL3690	t3794	STY4070	kdtA	1193	1259
3	18	0	0	SL3691	t3793	STY4069	coaD	461	461
0	0	0	0	SL3694	t3790	STY4066	rpmB	218	218
3	8	0	0	SL3696	t3788	STY4064	dfp	1205	1205
1	8	0	0	SL3697	t3787	STY4063	dut	437	437
2	32	4	6	SL3706	t3778	STY4052	gmk	605	605
2	8	1	2	SL3802	t3684	STY3943	gyrB	2396	2396
0	0	1	1	SL3804	t3682	STY3941	dnaN	1082	1082
0	0	0	0	SL3805	t3681	STY3940	dnaA	1382	1382
0	0	0	0	SL3806	t3680	STY3939A	rpmH	122	122
14	155	1	7	SL3809	t3678	STY3938	yidC	1628	1628
1	10	1	4	SL3829	t3657	STY3916	glmU	1352	1352
1	1	1	1	SL3876	t3380	STY3638	rho	1241	1241
7	65	2	6	SL3893	t3359	STY3621	hemC	938	923
3	13	0	0	SL3924	t3327	STY3589	ubiE	737	737
6	35	5	9	SL3926	t3325	STY3587	aarF	1622	1622
16	64	0	0	SL3932	t3319	STY3581	yigC	1460	1460
5	20	0	0	SL3940	t3311	STY3573	hemG	527	527
6	26	1	7	SL3948	t3620	STY3880	-	581	614
20	82	1	1	SL4044	t3523	STY3775	priA	2180	2180
4	42	4	36	SL4080	t3494	STY3743	murI	833	764
4	26	4	102	SL4083	t3487	STY3740	coaA	932	932
1	5	1	5	SL4086	t3480	STY3738	secE	365	365
6	49	1	6	SL4087	t3479	STY3737	nusG	527	527
0	0	1	9	SL4088	t3478	STY3736	rplK	410	410
6	46	4	16	SL4090	t3476	STY3734	rplJ	479	479
0	0	0	0	SL4091	t3475	STY3733	rplL	347	347
10	41	3	6	SL4092	t3474	STY3732	rpoB	4010	4010
31	405	35	903	SL4093	t3473	STY3731	rpoC	4205	4205
9	90	0	0	SL4106	t3464	STY3718	hemE	1046	1046
12	73	2	3	SL4171	t4140	STY4430	ubiA	854	854

14	115	3	13	SL4172	t4141	STY4431	plsB	2402	2402
4	20	2	6	SL4174	t4143	STY4433	lexA	590	590
1	23	1	1	SL4182	t4152	STY4442	dnaB	1397	1397
0	0	0	0	SL4266	t4381	STY4689	groES	275	275
1	2	1	4	SL4267	t4382	STY4690	groEL	1628	1628
3	9	1	1	SL4291	t4408	STY4714	yjeE	443	443
1	7	0	0	SL4324	t4442	STY4747	rpsF	377	377
4	7	0	0	SL4347	t4468	STY4773	ppa	512	512
4	12	0	0	SL4405	t4510	STY4814	valS	2837	2837
2	6	0	0	SL4409	t4513	STY4817	-	1082	1082
6	59	2	12	SL4410	t4514	STY4818	-	1064	1064
1	5	0	0	SL4474	t4586	STY4896	dnaC	719	719
0	0	0	0	SL4475	t4587	STY4897	dnaT	521	521

Ty, Typhi; Tm, Typhimurium.

8.3.7 Core gene functions in Typhimurium

Biological process	Sub-process	Essential genes	Non-essential genes
Cell division		**ftsALQWYZ**, minE, **mukB**, **SL2391**	**ftsHJ**K*N**X***, minCD, sdiA, cedA, sulA
DNA replication	Polymerases I, II and III	**dnaENQX**, **holAB**	**polA**B, holC**DE**
	Supercoiling	**gyrAB**, **parCE**	
	Primosome-associated	**dnaBCGT**, **priA**	**priB**C, **rep**, **ssb***
Transcription	RNA polymerase	**rpoABC**	
	Sigma, elongation, anti- and termination factors	**nusBG**, **rpoDH**, **rho**	**nusA**, **rpoE**NS
Translation	tRNA-synthetases	**alaS**, **argS**, **asnS**, **aspS**, **cysS**, **glnS**, **gltX**, **glyQS**, **hisS**, **ileS**, **leuS**, **lysS**, **metG**, **pheST**, **proS**, **serS**, **thrS**, **tyrS**, **valS**,	trpS, trpS2
	Ribosome components	**rplBCDEFJKLMNOPQRTUVWXY**, **rpmABCDHI**, **rpsABCDEFGH**I**JKLMNPQS**T	rplAI**S***, rpmEE2, rpmFG**J**J2, rps**I*ORU***V
	Initiation, elongation and peptide chain release factors	**fusA**, **infABC**, **prfAB**, **tsf**, **yrdC**	efp, prfCH, selB, tuf
Biosynthetic pathways			
Peptidoglycan		**murA**B**CDEFGI**,	ddl, ddlA
Fatty acids		**accABCD**, **fabABDGHIZ**	-

Gene names in bold are also essential in Typhi. *Unassigned gene due to LR between -2 and 2.

8.3.8 Essential genes omitted from comparison

Five essential Typhi were omitted from the comparison with Typhimurium. Two (t1217 and t2621) were too short (26bp and 29bp, respectively) for accurate comparison. A further two (t2724 and t3166) are annotated in the Ty2 genome as coding sequences, but comparison with SL1344 indicates that these actually represent RNA genes, and therefore

were not relevant to the analysis. Lastly, t4339 is a gene only present in Typhi that has been disrupted by the insertion of a phage; it is therefore not an essential gene in Typhi.

8.3.9 Putative essential genes

8.3.9.1 Typhi

Total inserts Ty	Total reads Ty	Total inserts Tm	Total reads Tm	Unique Ty Ty2 ID	Unique Ty CT18 ID	Unique Tm ID	Gene name	Gene length Ty	Gene length Tm
4	23	59	408	t3807	STY4083	SL3678	waaC	935	935
5	73	76	604	t3670	STY3930	SL3820	phoU	707	707
7	11	76	336	t0158	STY0175	SL0152	aceE	2645	2645
9	70	61	537	t0124	STY0140	SL0120	yabC	923	923
1	2	49	250	t0586	STY2507	SL2247	nrdB	1112	1112
3	7	42	263	t3214	STY3475	SL3269	ftsJ	608	608
7	60	64	393	t3213	STY3474	SL3268	ftsH	1916	1916
5	10	35	231	t0012	STY0012	SL0012	dnaK	1898	1898
17	75	45	424	t3205	STY3468	SL3260	nusA	1484	1484
0	0	42	193	t1621	STY1343	SL1640	-	1151	1097
2	8	44	207	t2139	STY0780	SL0719	sucB	1190	1190
2	3	28	166	t3809	STY4085	SL3676	waaD	914	914
14	37	41	248	t1682	STY1278	SL1176	icdA	1232	1232
12	54	62	260	t0301	STY2802	SL2517	glyA	1235	1235
2	10	28	152	t3350	STY3612	SL3901	dapF	806	809
3	12	24	144	t3287	STY3552	SL3344	mreD	473	473
0	0	9	116	t4161	STY4451	SL4192	ssb	512	512
1	1	22	117	t1316	STY1674	SL1380	pdxH	638	638
0	0	27	114	t3289	STY3554	SL3346	mreB	1025	1025
4	21	20	157	t3206	STY3469	SL3261	-	404	404
2	6	25	114	t4055	STY4348	SL3417	yheM	338	338
1	11	23	122	t2440	STY0462	SL0417	ispA	881	881
9	47	21	190	t3369	STY3627	SL3887	rffT	1340	1340
1	13	12	121	t3360	STY3622	SL3892	hemD	722	722
0	0	23	89	t2540	STY0355	SL0306	gmhA	560	560
1	3	12	94	t3201	STY3464	SL3256	rpsO	251	251
4	16	17	118	t3953	STY4242	SL3534	ftsX	1037	1037
5	25	31	132	t3288	STY3553	SL3345	mreC	1034	1034
7	14	26	111	t0983	STY2102	SL1828	ruvB	992	911
4	34	17	146	t2457	STY0444	SL0401	yajC	314	314
2	3	15	89	t3944	STY4233	SL3543	yhhP	227	227
4	21	16	119	t4400	STY4708	SL4285	psd	950	950
1	7	13	91	t3157	STY3417	SL3209	-	146	113
10	63	25	187	t3218	STY3480	SL3273	-	1154	1154

11	58	22	165	t0313	STY2789	SL2505	-	1196	1196
1	4	14	69	t3203	STY3466	SL3258	rbfA	383	383
0	0	7	61	t4443	STY4748	SL4325	priB	296	296
38	176	58	328	t1734	STY1226	SL1122	rne	3185	3185
3	13	12	75	t1622	STY1342	SL1641	-	290	290
4	23	18	88	t4600	STY4907	SL4485	holD	419	419
4	23	9	84	t2636	STY2868	SL2654	grpE	572	572
2	23	5	82	t4444	STY4749	SL4326	rpsR	209	209
7	101	20	194	t3234	STY3496	SL3289	-	557	557
1	2	6	43	t2832	STY3056	SL2910	-	293	293
8	15	11	59	t1954	STY0980	SL0917	cmk	665	665
3	12	6	54	t2628	STY2860	SL2646	rplS	329	329
1	6	7	45	t3679	STY3939	SL3807	rnpA	341	341
1	2	6	34	t4086	STY4379	SL3386	rpmJ	98	98
0	0	10	30	t4402	STY4710	SL4287	orn	527	527
2	4	6	35	t3260	STY3524	SL3316	rpsI	374	374
1	5	4	35	t1332	STY1657a	SL4528	malY	113	113
1	1	12	28	t0176	STY0193	SL0172	yadF	644	644
1	2	3	26	t0694	STY2391	SL2139	-	89	89
2	3	6	22	t1835	STY1111	SL1024	-	311	311
1	1	3	18	t3129	STY3388	SL3183	rpsU	197	197
4	26	10	47	t2205	STY0714	SL0657	-	455	455
0	0	3	15	t0535	STY2559	SL2298	-	134	134
10	48	16	70	t4401	STY4709	SL4286	yjeQ	1058	1058

Ty, Typhi; Tm, Typhimurium

8.3.9.2 Typhimurium

Total inserts Ty	**Total reads Ty**	**Total inserts Tm**	**Total reads Tm**	**Unique Ty Ty2 ID**	**Unique Ty CT18 ID**	**Unique Tm ID**	**Gene name**	**Gene length Ty**	**Gene length Tm**
17	104	0	0	t0090	STY0102	SL0088	folA	461	461
65	413	15	155	t0781	STY2301	SL2068	rfbG	1061	1061
23	129	7	37	t0208	STY0229	SL0208	mtn	680	680
19	107	9	26	t2135	STY0787	SL0723	cydB	1121	1121
14	145	12	56	t3488	STY3741	SL4082	birA	944	944
61	216	19	110	t1627	STY1336	SL1646	topA	2579	2579
17	61	1	7	t0941	STY2145	SL1870	tyrP	1193	1193
13	274	9	149	t3489	STY3742	SL4081	murB	1010	1010
16	53	5	16	t2234	STY0683	SL0621	lipA	947	947
8	47	3	15	t1093	STY1909	SL1711	pth	590	566
29	114	20	69	t1977	STY0955	SL0895	cydD	1748	1748
5	22	1	5	t1061	STY1946	SL1744	minE	248	248
4	19	2	12	t0045	STY0052	SL0044	rpsT	245	245
15	37	7	31	t0277	STY2827	SL2543	rnc	662	662

Ty, Typhi; Tm, Typhimurium

8.3.10 Attempted construction of a Typhi *recA* mutant

This work was performed by Jana Haase.

The suicide vector allelic-exchange method involves introducing, by homologous recombination, mutant sequences ligated to a suicide vector into the target gene to generate partial merodiploids. In this particular case, the structure of the *recA*-suicide vector construct was such that, depending on the point of cross-over, the merodiploids may be either wild-type or mutant for *recA*, and for a non-essential gene, one would expect to obtain approximately equal numbers of each. During the attempted construction of the *recA* mutant in Typhi, 19 out of 19 partial merodiploids were recombinants of the type possessing a wild-type copy of *recA*, a strong indication that recA is essential. The next stage involves processing of the *recA*-suicide vector recombinant for loss of the suicide vector which occurs by homologous recombination between the merodiploid sequences, resulting in either reversion or incorporation of the mutation. Again, depending on the point of crossover this process would be expected to generate approximately equal numbers of wild-type (revertant) and mutant derivatives, for a non-essential gene. In this particular case, on three independent occasions a total of 39 out of 39 derivatives were *recA* revertants, providing additional strong evidence that the *recA* gene in Typhi is essential.

8.4 Chapter 5

8.4.1 Typhimurium genes essential for macrophage infection

Inserts (input)	Reads (input)	Inserts (p.i.)	Reads (p.i.)	ID	Gene	Start	End	Strand	Log_2 FC	*P*-value
38	1850	0.35	26	SL0093	surA	104482	105750	-	-7.67	6.01E-10
30	1307	0.00	0	SL2482	-	2650773	2651933	-	-7.61	8.37E-10
11	1149	0.04	8	SL1860	flhD	1979293	1979625	-	-7.37	2.72E-09
13	1011	0.00	0	SL1622	sapA	1740541	1742172	+	-7.24	5.14E-09
29	1374	0.44	24	SL1774	prc	1898711	1900741	-	-7.16	7.75E-09
16	1190	0.31	17	SL0275	-	319216	319734	+	-7.09	1.07E-08
14	723	0.00	0	SL1904	fliL	2020640	2021089	+	-6.76	4.80E-08
19	793	0.15	26	SL1517	-	1629447	1630490	-	-6.69	6.72E-08
11	702	0.04	8	SL1466	hyaC2	1568164	1568889	-	-6.67	7.31E-08
16	1323	0.96	47	SL1772	-	1896215	1897570	+	-6.66	7.65E-08
13	1334	0.98	46	SL1082	-	1188296	1188604	+	-6.66	7.69E-08
38	1270	0.92	63	SL1428	dmsA2	1531281	1533698	-	-6.63	8.71E-08
17	896	0.38	17	SL0185	pcnB	213879	215279	-	-6.60	9.66E-08
16	966	0.55	16	SL1374	-	1475447	1476325	-	-6.54	1.26E-07
15	1150	0.86	18	SL1884	fliZ	2001714	2002247	-	-6.53	1.35E-07
8	636	0.04	8	SL1180	msgA	1286328	1286549	-	-6.53	1.36E-07
42	1738	2.10	55	SL1898	fliF	2014116	2015780	+	-6.39	2.47E-07
14	587	0.07	16	SL1634	-	1753404	1755368	+	-6.36	2.75E-07
2	537	0.00	0	SL4326	rpsR	4651824	4652033	+	-6.34	3.06E-07
17	735	0.37	17	SL1123	rluC	1228282	1229223	+	-6.33	3.13E-07
13	837	0.59	26	SL0996	gtgF	1099854	1100027	-	-6.30	3.50E-07
38	1431	1.75	58	SL3287	-	3502388	3503356	+	-6.28	3.84E-07
30	1327	1.56	25	SL0318	proA	366917	368149	+	-6.27	3.99E-07
10	574	0.12	9	SL1108	flgN	1212770	1213174	-	-6.27	4.03E-07
2	496	0.00	0	SL3543	yhhP	3767840	3768067	-	-6.22	4.91E-07
14	944	0.90	18	SL3454	aroK	3663755	3664258	-	-6.21	5.10E-07
9	510	0.04	8	SL1769	-	1894496	1894621	-	-6.21	5.17E-07
1	491	0.00	0	SL1210	-	1310000	1310230	+	-6.21	5.21E-07
19	656	0.35	18	SL1857	motB	1976774	1977685	-	-6.19	5.75E-07
22	481	0.00	0	SL3433	cap	3637137	3637751	+	-6.18	5.88E-07
24	1130	1.34	51	SL2872	spaM	3058841	3059266	-	-6.17	6.05E-07
2	475	0.00	0	SL0707	-	790305	791120	+	-6.16	6.33E-07
10	490	0.04	8	SL3809	yidC	4067748	4069376	+	-6.15	6.54E-07
23	1114	1.38	41	SL1885	fliA	2002324	2003025	-	-6.13	7.23E-07
2	469	0.04	8	SL1439	-	1544465	1544650	-	-6.09	8.43E-07
7	450	0.00	0	SL4281	yjeA	4606685	4607644	+	-6.08	8.67E-07
25	1137	1.52	38	SL1964	-	2069158	2069958	-	-6.08	9.00E-07
9	757	0.72	55	SL1478	-	1582867	1583238	-	-6.05	1.01E-06
13	584	0.33	8	SL2232	eco	2357784	2358260	+	-6.04	1.02E-06
2	408	0.00	0	SL0810	-	901541	901996	+	-5.95	1.52E-06

14	489	0.21	29	SL0746	-	832602	833282	+	-5.93	1.61E-06
8	522	0.30	21	SL1595	-	1714751	1715335	-	-5.92	1.67E-06
5	404	0.04	8	SL1354	ssaS	1456249	1456497	+	-5.88	1.97E-06
12	633	0.62	38	SL1146	nagZ	1248758	1249765	+	-5.88	1.99E-06
9	398	0.04	8	SL1111	flgB	1214398	1214796	+	-5.86	2.14E-06
13	406	0.08	9	SL1585	-	1704171	1705088	+	-5.83	2.40E-06
25	666	0.77	24	SL2843	sitC	3030688	3031530	+	-5.82	2.48E-06
9	504	0.38	21	SL1987	-	2090681	2090989	-	-5.79	2.83E-06
11	816	1.25	24	SL2207	-	2329217	2329489	+	-5.77	3.06E-06
12	358	0.00	0	SL1312	pykF	1416787	1418127	-	-5.76	3.13E-06
12	588	0.63	54	SL1621	pspF	1739449	1740411	+	-5.76	3.14E-06
12	569	0.59	18	SL0780	moaC	871428	871895	+	-5.75	3.25E-06
9	474	0.33	16	SL2603	rseA	2779168	2779800	-	-5.75	3.29E-06
28	1404	2.97	71	SL4323	-	4650826	4651014	+	-5.72	3.60E-06
9	362	0.04	8	SL2835	hypB	3022808	3023662	+	-5.72	3.60E-06
11	1300	2.68	77	SL1874	-	1992357	1992671	-	-5.72	3.63E-06
8	400	0.15	8	SL1297	-	1401165	1401506	-	-5.72	3.65E-06
11	599	0.73	17	SL0666	-	743551	744171	-	-5.70	3.93E-06
4	342	0.00	0	SL3134	-	3341448	3341579	-	-5.70	4.02E-06
11	340	0.00	0	SL1577	ldhA	1696891	1697862	+	-5.69	4.14E-06
24	554	0.63	29	SL1900	fliH	2016779	2017468	+	-5.68	4.31E-06
13	500	0.47	38	SL2571	-	2753096	2753281	-	-5.67	4.35E-06
6	333	0.00	0	SL1584	-	1703916	1704101	+	-5.66	4.63E-06
11	918	1.73	75	SL0978	-	1081781	1082161	+	-5.66	4.67E-06
26	1541	3.65	80	SL0536	fimA	603627	604151	+	-5.63	5.16E-06
21	951	1.90	42	SL2052	hisH	2151782	2152357	+	-5.62	5.41E-06
21	711	1.18	70	SL0684	kdpE	765759	766418	-	-5.62	5.44E-06
13	423	0.31	9	SL1163	potA	1268181	1269299	-	-5.60	5.71E-06
8	677	1.10	48	SL2502	hscB	2677785	2678282	-	-5.60	5.86E-06
10	708	1.21	51	SL1946	-	2055079	2055261	-	-5.59	6.05E-06
18	820	1.56	38	SL2148	yohD	2265082	2265660	+	-5.58	6.11E-06
11	469	0.48	16	SL2337	truA	2478187	2478981	-	-5.58	6.18E-06
17	324	0.04	8	SL2123	thiD	2237986	2238768	-	-5.57	6.52E-06
8	912	1.90	46	SL1257	-	1359202	1359852	-	-5.56	6.76E-06
12	425	0.37	25	SL1651	-	1772666	1773523	-	-5.54	7.08E-06
2	388	0.26	21	SL0473	ybaM	538432	538581	-	-5.54	7.10E-06
5	307	0.00	0	SL1625	sapD	1744029	1745003	+	-5.54	7.13E-06
2	303	0.00	0	SL1738	-	1866793	1866912	-	-5.52	7.64E-06
36	1170	2.86	32	SL1848	flhA	1965788	1967848	-	-5.50	8.29E-06
4	310	0.04	9	SL2116	-	2231538	2231636	-	-5.50	8.33E-06
9	735	1.45	29	SL1866	-	1985499	1986119	-	-5.49	8.55E-06
2	304	0.04	8	SL1107	mviN	1211114	1212670	+	-5.48	9.09E-06
8	310	0.06	13	SL1827	yebI	1945767	1946534	+	-5.47	9.29E-06
5	302	0.04	8	SL2191	-	2311383	2311937	+	-5.47	9.41E-06
19	759	1.61	8	SL2671	-	2860782	2861279	-	-5.45	1.02E-05
1	285	0.00	0	SL1264	-	1365082	1365378	+	-5.44	1.05E-05
8	757	1.64	38	SL1252	-	1353174	1353914	+	-5.43	1.09E-05
2	282	0.00	0	SL0714	sdhC	797412	797783	+	-5.42	1.11E-05
12	599	1.10	9	SL0528	ybbF	597240	597944	-	-5.42	1.11E-05
18	1022	2.63	83	SL3119	hybE	3328524	3328994	-	-5.40	1.22E-05

102	2437	7.73	178	SL3810	thdF	4069524	4070882	+	-5.38	1.30E-05
13	633	1.29	38	SL1607	-	1726810	1727697	-	-5.38	1.31E-05
5	485	0.77	30	SL1921	yedJ	2033410	2034087	-	-5.37	1.35E-05
5	309	0.14	21	SL0730	tolB	815384	816658	+	-5.36	1.37E-05
16	1437	4.23	113	SL4400	-	4733509	4733907	+	-5.36	1.40E-05
14	649	1.37	21	SL1530	pcgL	1646816	1647568	-	-5.36	1.40E-05
2	267	0.00	0	SL3039	rpiA	3245506	3246147	-	-5.35	1.46E-05
8	471	0.76	46	SL1511	narV	1623522	1624184	+	-5.33	1.54E-05
8	264	0.00	0	SL0853	-	946829	947287	+	-5.33	1.55E-05
21	1133	3.21	76	SL4327	rplI	4652093	4652524	+	-5.33	1.56E-05
17	262	0.00	0	SL1537	-	1654400	1655050	-	-5.32	1.61E-05
6	252	0.00	0	SL1208	-	1309044	1309667	+	-5.26	1.96E-05
4	252	0.00	0	SL1236	-	1337315	1337713	-	-5.26	1.96E-05
5	250	0.00	0	SL2883	-	3069469	3069558	+	-5.25	2.03E-05
14	660	1.62	59	SL1064	phoH	1170647	1171483	+	-5.24	2.13E-05
14	256	0.04	8	SL2473	xseA	2623804	2625135	+	-5.24	2.17E-05
13	443	0.78	17	SL3295	-	3508234	3509070	+	-5.23	2.21E-05
2	263	0.07	16	SL1989	-	2091713	2091862	+	-5.22	2.27E-05
4	244	0.00	0	SL1186	-	1291954	1292208	+	-5.22	2.29E-05
18	794	2.21	62	SL1648	-	1769947	1770975	-	-5.21	2.34E-05
12	342	0.40	16	SL3883	rffC	4154664	4155194	+	-5.21	2.40E-05
20	689	1.80	33	SL0658	ybeZ	732656	733723	-	-5.20	2.41E-05
21	410	0.68	25	SL1339	sseG	1445615	1446286	+	-5.20	2.44E-05
15	381	0.58	43	SL1589	-	1709704	1710249	-	-5.19	2.52E-05
7	495	1.05	25	SL1009	-	1116772	1117206	+	-5.19	2.57E-05
28	864	2.56	63	SL0812	ybiS	903167	904069	-	-5.18	2.59E-05
13	991	3.08	49	SL2113	-	2229221	2230249	+	-5.18	2.61E-05
7	269	0.13	29	SL1057	-	1161306	1161515	-	-5.18	2.68E-05
9	897	2.73	29	SL1350	ssaO	1453856	1454215	+	-5.17	2.73E-05
3	244	0.04	8	SL1979	-	2082182	2082637	+	-5.17	2.75E-05
12	492	1.06	16	SL1510	narW	1622830	1623507	+	-5.17	2.75E-05
6	324	0.38	21	SL2162	-	2278668	2278895	+	-5.16	2.83E-05
11	357	0.52	33	SL2110	-	2224455	2225153	+	-5.16	2.86E-05
2	233	0.00	0	SL1207	-	1308632	1308796	+	-5.15	2.87E-05
10	322	0.39	9	SL1401	tus	1502262	1503173	-	-5.13	3.11E-05
27	1016	3.35	93	SL0277	-	321096	322382	+	-5.13	3.18E-05
15	573	1.48	38	SL1841	-	1957394	1958347	+	-5.12	3.23E-05
5	227	0.00	0	SL1090	msyB	1196941	1197297	-	-5.12	3.26E-05
4	226	0.00	0	SL2389	ypeB	2536506	2536715	-	-5.11	3.33E-05
3	224	0.00	0	SL0012	dnaK	11602	13500	+	-5.10	3.48E-05
4	233	0.04	9	SL0462	rpmE2	525849	526091	+	-5.10	3.48E-05
18	753	2.31	16	SL4431	-	4776169	4776408	-	-5.09	3.56E-05
14	809	2.58	49	SL1050	yccD	1155025	1155312	-	-5.08	3.69E-05
3	221	0.00	0	SL4295	hfq	4625165	4625455	+	-5.08	3.71E-05
11	711	2.18	29	SL4193	-	4499158	4499772	+	-5.07	3.90E-05
13	557	1.51	59	SL1633	rnb	1751208	1753124	+	-5.06	3.97E-05
25	565	1.56	50	SL3952	glnG	4234304	4235695	-	-5.05	4.06E-05
1	249	0.15	8	SL0800	ybiI	890271	890519	-	-5.05	4.11E-05
16	312	0.43	27	SL1307	-	1413046	1414248	+	-5.05	4.15E-05
32	1134	4.15	97	SL3220	-	3435924	3437051	-	-5.04	4.27E-05

8	424	0.96	25	SL1336	sseE	1443918	1444316	+	-5.03	4.45E-05
19	704	2.25	33	SL2534	yfhH	2715314	2716144	+	-5.02	4.53E-05
10	211	0.00	0	SL1134	fabF	1236586	1237809	+	-5.02	4.62E-05
40	1406	5.56	96	SL0441	tig	501160	502440	+	-5.00	4.87E-05
1	247	0.18	8	SL1433	-	1538169	1538492	-	-4.99	4.98E-05
15	675	2.21	46	SL0756	modE	843705	844475	-	-4.98	5.20E-05
18	548	1.62	35	SL1105	-	1209277	1209906	+	-4.98	5.30E-05
16	599	1.87	33	SL0852	rimK	945830	946714	+	-4.97	5.41E-05
3	204	0.00	0	SL1340	ssaG	1446398	1446595	+	-4.97	5.42E-05
12	453	1.19	33	SL2190	-	2310250	2310951	+	-4.97	5.47E-05
8	350	0.70	8	SL3678	rfaC	3928620	3929555	+	-4.97	5.47E-05
6	349	0.69	38	SL1617	pspD	1737726	1737926	-	-4.96	5.49E-05
1	203	0.00	0	SL0221	dxr	257851	259029	+	-4.96	5.55E-05
14	504	1.46	29	SL0541	fimF	609189	609689	+	-4.95	5.82E-05
25	803	2.91	33	SL2194	yejB	2315759	2316835	+	-4.94	5.88E-05
4	253	0.25	29	SL3535	ftsE	3759406	3760056	-	-4.94	5.90E-05
17	875	3.28	46	SL2120	-	2235264	2236250	+	-4.94	6.04E-05
15	1038	4.13	113	SL2214	-	2335549	2336862	+	-4.92	6.39E-05
10	444	1.22	29	SL1751	-	1878133	1878459	+	-4.92	6.45E-05
12	550	1.75	52	SL1199	aadA	1303458	1304228	+	-4.91	6.61E-05
3	243	0.24	9	SL2334	folC	2475079	2476329	-	-4.91	6.62E-05
7	218	0.11	17	SL0478	ybaB	542378	542689	+	-4.91	6.64E-05
12	405	1.06	16	SL1551	-	1665918	1666475	-	-4.89	7.03E-05
6	321	0.65	17	SL1183	pagD	1288038	1288283	-	-4.89	7.13E-05
43	1988	9.13	117	SL4377	-	4709292	4710014	+	-4.87	7.49E-05
2	190	0.00	0	SL4090	rplJ	4386001	4386480	+	-4.87	7.53E-05
9	344	0.78	29	SL3170	glgS	3380028	3380219	-	-4.87	7.55E-05
7	230	0.22	8	SL2234	alkB	2360206	2360838	-	-4.85	8.03E-05
8	186	0.00	0	SL2290	nuoI	2428088	2428612	-	-4.84	8.30E-05
3	184	0.00	0	SL0719	sucB	803981	805171	+	-4.82	8.72E-05
8	546	1.90	46	SL1080	csgB	1187282	1187719	+	-4.82	8.73E-05
23	828	3.38	79	SL2944	-	3136548	3136979	-	-4.82	8.79E-05
10	183	0.00	0	SL1871	yecA	1989263	1989910	-	-4.82	8.94E-05
12	512	1.74	47	SL0789	ybhP	877614	878354	-	-4.82	8.97E-05
13	361	0.99	16	SL3900	-	4173576	4173761	+	-4.78	0.000101
11	266	0.48	8	SL1018	-	1126241	1126639	+	-4.78	0.000101
9	178	0.00	0	SL1276	btuD	1378056	1378787	+	-4.78	0.000101
2	177	0.00	0	SL2394	ptsH	2542311	2542550	+	-4.77	0.000104
6	392	1.17	42	SL1698	chaB	1825618	1825830	-	-4.77	0.000104
8	229	0.29	29	SL2834	hypA	3022383	3022721	+	-4.77	0.000105
10	440	1.46	26	SL0492	cueR	560053	560451	+	-4.76	0.000109
18	724	3.03	68	SL2147	yohC	2264363	2264932	-	-4.75	0.000111
3	210	0.20	9	SL1531	ugtL	1648108	1648488	-	-4.75	0.000111
5	181	0.04	8	SL1890	fliS	2008397	2008786	+	-4.75	0.000111
8	696	2.89	38	SL2946	yqcC	3137798	3138109	-	-4.75	0.000113
12	598	2.35	81	SL0445	hupB	507600	507854	+	-4.74	0.000113
5	173	0.00	0	SL0892	-	987039	987404	-	-4.74	0.000115
4	172	0.00	0	SL1897	fliE	2013585	2013881	-	-4.73	0.000118
6	172	0.00	0	SL4114	purD	4412501	4413772	-	-4.73	0.000118
15	618	2.50	50	SL1126	-	1230642	1231208	-	-4.73	0.000119

3	341	0.96	30	SL1202	-	1306156	1306386	-	-4.72	0.000123
6	170	0.00	0	SL3813	-	4073660	4074577	+	-4.71	0.000124
23	763	3.40	29	SL1115	flgF	1217191	1217928	+	-4.70	0.00013
9	876	4.05	68	SL3026	-	3229566	3229859	-	-4.70	0.00013
12	353	1.06	32	SL1935	-	2046003	2046518	-	-4.70	0.000132
16	698	3.06	49	SL1993	cobS	2096124	2096849	-	-4.69	0.000134
11	687	3.00	77	SL0888	-	984225	984884	-	-4.69	0.000136
3	262	0.55	17	SL0121	ftsL	141357	141704	+	-4.69	0.000137
7	452	1.66	72	SL3839	atpI	4101496	4101858	-	-4.68	0.00014
3	163	0.00	0	SL0152	aceE	176253	178898	+	-4.66	0.00015
8	438	1.63	21	SL1895	yedF	2012581	2012796	+	-4.65	0.000152
4	160	0.00	0	SL1275	btuE	1377505	1378038	+	-4.63	0.000162
11	259	0.59	17	SL1157	-	1261996	1262889	+	-4.63	0.000162
27	613	2.72	59	SL1119	flgJ	1220604	1221536	+	-4.63	0.000163
21	520	2.20	24	SL3885	-	4156349	4157581	+	-4.61	0.000171
13	707	3.33	41	SL4303	yjfI	4635045	4635431	+	-4.61	0.000171
23	347	1.15	42	SL1903	fliK	2019318	2020517	+	-4.61	0.000172
9	188	0.18	16	SL0398	yajB	458380	458943	-	-4.61	0.000172
15	459	1.83	26	SL2005	cbiF	2105484	2106239	-	-4.61	0.000173
9	893	4.54	72	SL1939	-	2050978	2051286	-	-4.59	0.000183
7	205	0.30	13	SL2025	pduN	2122653	2122910	+	-4.59	0.000183
25	1055	5.54	159	SL0811	ybiR	902011	903105	+	-4.59	0.000183
13	167	0.07	8	SL0215	glnD	250148	252802	-	-4.59	0.000185
14	393	1.47	38	SL2017	pduD	2117357	2118013	+	-4.59	0.000186
17	2254	13.06	100	SL2745	-	2924055	2925911	+	-4.58	0.000191
4	153	0.00	0	SL1308	-	1414279	1414677	+	-4.57	0.000197
24	240	0.55	8	SL4505	serB	4853813	4854763	+	-4.56	0.000199
19	212	0.38	39	SL1452	sotB	1555764	1556936	-	-4.56	0.0002
4	152	0.00	0	SL2903	rpoS	3088064	3089038	-	-4.56	0.000202
5	150	0.00	0	SL1801	-	1922089	1922334	-	-4.54	0.000214
4	178	0.18	8	SL3289	-	3503958	3504515	+	-4.54	0.000218
15	983	5.37	41	SL0546	-	613247	613573	-	-4.53	0.000222
1	148	0.00	0	SL1370	-	1471645	1471866	+	-4.52	0.000226
9	147	0.00	0	SL3020	xerD	3225154	3226032	-	-4.51	0.000233
19	577	2.84	90	SL3051	-	3256422	3257060	+	-4.50	0.000244
17	304	1.04	55	SL2040	phsB	2135806	2136366	-	-4.50	0.000244
2	145	0.00	0	SL1415	-	1518921	1519187	-	-4.49	0.000247
7	174	0.20	9	SL3411	yheA	3619414	3619590	-	-4.49	0.000252
9	337	1.28	21	SL0251	yafC	295893	296789	-	-4.49	0.000253
7	507	2.41	29	SL0382	yaiI	439700	440137	+	-4.49	0.000254
24	949	5.37	148	SL3019	dsbC	3224417	3225073	-	-4.48	0.000259
14	511	2.49	55	SL1383	gst	1482601	1483188	-	-4.46	0.000272
15	512	2.51	55	SL1811	-	1927969	1928607	+	-4.46	0.000276
21	430	1.97	17	SL1396	-	1496062	1497552	-	-4.45	0.000281
11	669	3.61	51	SL2857	iagB	3044120	3044584	+	-4.45	0.000286
22	877	5.03	71	SL1734	nhaB	1863690	1865216	+	-4.45	0.000287
21	1179	7.11	106	SL3042	-	3248318	3248935	-	-4.44	0.00029
9	494	2.43	35	SL1306	-	1411778	1413031	+	-4.44	0.000292
12	439	2.05	51	SL2786	-	2971894	2972214	+	-4.44	0.000292
19	1097	6.57	76	SL2195	yejE	2316853	2317860	+	-4.44	0.000294

15	492	2.44	63	SL2258	-	2393765	2394550	-	-4.43	0.0003
3	138	0.00	0	SL2528	-	2709889	2710104	+	-4.43	0.000304
9	175	0.26	16	SL4004	sodA	4287986	4288588	+	-4.43	0.000304
7	270	0.92	24	SL1850	cheZ	1969204	1969830	-	-4.42	0.000306
6	370	1.62	26	SL1926	umuD	2038740	2039141	-	-4.42	0.000311
13	532	2.75	82	SL1289	-	1392599	1393471	+	-4.42	0.000313
12	385	1.73	30	SL1932	-	2044512	2045015	-	-4.41	0.000316
7	136	0.00	0	SL1541	rimL	1657904	1658425	-	-4.41	0.000322
21	1043	6.39	109	SL1106	mviM	1209926	1210831	+	-4.40	0.000329
4	135	0.00	0	SL0637	-	711414	711686	+	-4.40	0.000332
28	1113	6.99	91	SL1230	sppA	1329772	1331610	-	-4.38	0.000349
19	457	2.33	64	SL4226	phnA	4551468	4551785	-	-4.37	0.000359
14	347	1.57	17	SL0301	-	349161	349358	+	-4.36	0.000376
4	160	0.22	21	SL1386	rnfE	1486053	1486727	-	-4.35	0.000386
12	475	2.52	29	SL0279	-	323381	323794	+	-4.34	0.000388
9	419	2.12	29	SL1052	scsA	1156485	1156829	+	-4.34	0.000392
7	226	0.71	18	SL2222	narP	2344427	2345056	+	-4.34	0.000392
7	254	0.92	17	SL1256	-	1358465	1359046	-	-4.33	0.0004
8	824	5.13	29	SL4251	-	4581471	4582328	-	-4.33	0.000403
1	158	0.22	8	SL2931	eno	3118836	3120116	-	-4.33	0.000409
5	138	0.07	8	SL0529	ppiB	597965	598441	-	-4.32	0.000411
1	184	0.42	21	SL2332	cvpA	2473616	2474086	-	-4.32	0.000416
4	461	2.48	63	SL1809	-	1927119	1927481	-	-4.32	0.000417
9	334	1.54	80	SL4379	relB	4712040	4712264	+	-4.32	0.000418
4	146	0.15	8	SL2504	nifU	2678848	2679216	-	-4.31	0.000433
7	223	0.73	24	SL2271	-	2408414	2408653	+	-4.30	0.000442
6	146	0.16	9	SL1127	-	1231424	1231927	+	-4.29	0.00045
14	404	2.15	29	SL1416	-	1519564	1520799	-	-4.28	0.000472
11	210	0.67	17	SL3450	rpe	3659523	3660182	-	-4.27	0.000483
5	174	0.40	16	SL1217	-	1314579	1315079	-	-4.26	0.000499
10	200	0.61	38	SL0933	-	1043697	1044227	+	-4.26	0.000499
42	1024	7.01	109	SL0371	sbmA	429956	431056	+	-4.26	0.0005
17	1229	8.64	62	SL1095	yceJ	1201582	1202136	-	-4.25	0.000509
17	862	5.79	122	SL1762	-	1888853	1889293	+	-4.25	0.000513
3	184	0.50	17	SL2247	nrdB	2381523	2382635	+	-4.24	0.000523
6	232	0.88	34	SL2020	pduH	2120399	2120731	+	-4.24	0.000526
7	357	1.86	38	SL0575	ybdZ	644744	644944	+	-4.24	0.000529
16	541	3.32	59	SL0349	-	400355	400801	+	-4.24	0.000533
17	1218	8.70	122	SL2400	-	2547535	2548236	+	-4.23	0.000542
9	462	2.73	72	SL1743	minD	1869680	1870474	+	-4.22	0.00055
4	441	2.58	38	SL1330A	-	1366420	1366605	+	-4.22	0.000565
4	118	0.00	0	SL3699	pyrE	3947282	3947905	-	-4.21	0.00057
17	563	3.58	109	SL1715	-	1842805	1844580	+	-4.21	0.000574
22	1358	10.00	64	SL4440	-	4784233	4784661	+	-4.20	0.000583
12	554	3.52	55	SL0137	mutT	161559	161936	+	-4.20	0.000584
5	212	0.77	38	SL1235	-	1337080	1337337	+	-4.20	0.000586
16	469	2.87	105	SL0928	smtA	1033982	1034767	+	-4.19	0.000604
5	116	0.00	0	SL2021	pduJ	2120768	2121025	+	-4.19	0.00061
15	511	3.31	42	SL0500	ybbA	565936	566604	+	-4.16	0.000665
5	387	2.28	55	SL1566	-	1682921	1683583	+	-4.16	0.000668

7	325	1.77	55	SL3181	-	3393352	3393606	+	-4.15	0.000677
8	219	0.91	30	SL4392	pyrL	4725533	4725616	-	-4.14	0.000708
6	419	2.60	29	SL1697	-	1825185	1825520	+	-4.13	0.000714
6	475	3.09	29	SL1367	gloA	1469311	1469700	-	-4.13	0.000721
7	380	2.31	51	SL2588	-	2764956	2765408	-	-4.12	0.000742
23	702	5.10	85	SL3490	-	3708204	3708518	-	-4.11	0.000761
13	484	3.23	38	SL2037	gyrI	2133129	2133578	-	-4.11	0.000768
3	141	0.27	9	SL0620	tatE	694070	694255	+	-4.11	0.000771
5	109	0.00	0	SL1909	fliQ	2023650	2023901	+	-4.10	0.000775
11	241	1.16	29	SL4100	thiS	4400264	4400446	-	-4.09	0.000797
29	1274	10.14	140	SL2796	proX	2980865	2981842	+	-4.09	0.000798
18	776	5.91	41	SL0804	glnP	894731	895372	-	-4.07	0.000844
5	131	0.22	8	SL2790	nrdH	2974442	2974669	+	-4.07	0.000856
1	106	0.00	0	SL2508	suhB	2682161	2682946	+	-4.07	0.000862
20	375	2.40	38	SL4011	pfkA	4293777	4294721	+	-4.06	0.000873
6	230	1.11	17	SL1726	-	1854130	1854846	+	-4.06	0.000874
16	772	5.94	93	SL2154	-	2271357	2272040	-	-4.06	0.000877
31	1656	13.82	101	SL2177	lysP	2296613	2298064	-	-4.06	0.000879
8	402	2.68	60	SL2902	-	3087783	3088001	+	-4.04	0.000914
14	787	6.19	89	SL3380	yhdN	3603892	3604242	-	-4.04	0.000937
31	1959	16.83	122	SL3229	-	3446539	3447327	+	-4.03	0.000943
2	115	0.11	8	SL0136	secA	158702	161389	+	-4.03	0.000962
6	441	3.09	25	SL0670	-	750410	750724	+	-4.02	0.00097
22	1524	13.00	101	SL1996	cbiO	2098923	2099720	-	-4.02	0.000973
18	567	4.27	71	SL4277	frdD	4603069	4603410	-	-4.02	0.000988
4	102	0.00	0	SL1059	-	1162535	1162684	+	-4.01	0.000996

p.i. post infection; reads and inserts from post infection are the combined values of the 30 minute, 2 hour and 4 hour timepoints, normalised for sequencing yield per sample.

8.4.2 Typhi genes required for macrophage infection (original assay)

Inserts (input)	Reads (input)	Inserts (p.i.)	Reads (p.i.)	ID	Gene	Start	End	Strand	Log_2 FC	*P*-value
14	117	0.27	5	t3623	dsbA	3726177	3726782	-	-3.79	7.96E-09
18	184	1.01	10	t1975	lrp	2033312	2033788	-	-3.75	1.07E-08
19	87	0.00	0	t0336	-	386382	387542	+	-3.74	1.22E-08
32	120	0.59	13	t2980	serA	3071308	3072522	-	-3.50	8.23E-08
87	740	10.35	81	t3645	gidA	3750936	3752807	+	-3.22	6.60E-07
11	47	0.00	0	t3214	ftsJ	3306755	3307363	-	-2.94	4.74E-06
58	1173	21.08	93	t1033	prc	1116361	1118391	+	-2.92	5.29E-06
9	71	0.55	5	t2131	tolQ	2192191	2192865	-	-2.84	9.02E-06
5	104	1.22	19	t1068	-	1149801	1149956	+	-2.83	9.78E-06
12	41	0.00	0	t3505	ppc	3598965	3601598	+	-2.77	1.43E-05
18	86	1.00	9	t4411	miaA	4562802	4563734	+	-2.72	1.93E-05
6	282	5.64	20	t4378	-	4531819	4531923	-	-2.63	3.43E-05
6	32	0.00	0	t4623	yjjX	4776185	4776502	-	-2.47	8.68E-05
21	83	1.42	21	t3372	rffA	3471846	3472958	-	-2.40	0.0001282

11	160	3.51	13	t1070	dsbB	1150495	1151007	-	-2.39	0.0001331
4	40	0.27	9	t2628	rplS	2705986	2706315	-	-2.39	0.0001358
5	50	0.59	9	t3210	secG	3301879	3302193	-	-2.34	0.000175
5	213	5.17	24	t1054	-	1138452	1138613	+	-2.34	0.0001819
75	561	15.15	59	t3797	waaP	3917060	3917839	+	-2.32	0.0002015
4	28	0.00	0	t0879	-	981038	981181	+	-2.31	0.0002074
5	36	0.23	4	t4579	-	4733209	4733493	+	-2.31	0.0002075
6	53	0.73	9	t4643	ssrA	2734390	2734838	+	-2.30	0.0002231
13	87	1.71	17	t0732	-	813181	813642	+	-2.30	0.0002248
9	79	1.50	9	t2916	-	3007291	3007491	+	-2.29	0.000239
1	44	0.49	10	t1719	holB	1786262	1787248	-	-2.28	0.000246
7	44	0.52	12	t2365	-	2432203	2432664	+	-2.25	0.0002899
26	80	1.60	8	t0317	hscA	360783	362615	+	-2.25	0.0002942
3	25	0.00	0	t0045	rpsT	52286	52531	-	-2.18	0.0004114
2	24	0.00	0	t2130	tolR	2191759	2192169	-	-2.14	0.0005204
11	110	2.88	17	t0887	vsr	988540	988992	+	-2.10	0.0006378
8	30	0.23	4	t1312	slyA	1374300	1374716	+	-2.10	0.0006429
7	23	0.00	0	t1974	ftsK	2029146	2033159	-	-2.09	0.0006605
6	30	0.26	6	t0606	napG	691360	692037	+	-2.06	0.0007736

p.i. post infection; reads and inserts from post infection are the combined values of the 30 minute, 2 hour and 4 hour timepoints, normalised for sequencing yield per sample.