

**Characterisation of the Transcriptome and  
Proteome of *Salmonella enterica* subspecies  
*enterica* serovar Typhi.**

**Thesis submitted to the University of Cambridge for the degree of Doctor of  
Philosophy**

**2<sup>nd</sup> January 2009**

**Timothy Trevor Perkins  
Jesus College**

# Declaration

I hereby declare that this thesis consists of work derived entirely of my own work. Due to the nature of this work, and interdisciplinary nature of biological sciences, it was not possible and impractical to perform all of these techniques, however, it was possible to design all experiments associated with this thesis. Work that was done by other persons is clearly stated in the Materials and Methods section.

This thesis is no longer than 300 pages as required by the Department of Biological Sciences.

# Abstract

*Salmonella enterica* subspecies *enterica* serovar Typhi (*S. Typhi*) is the cause of human typhoid. A combination of approaches were utilised to analyse the transcriptome and proteome of this human-restricted and highly clonal pathogen.

A novel method for strand-specific RNA-seq was developed, facilitating a whole genome analysis of the *S. Typhi* Ty2 BRD948 transcriptome. These data were validated using novel analyses methods and compared to RNA derived from an *ompR* mutant derivative. The data was used to reinterpret a previous annotation of the *S. Typhi* Ty2 genome. Mass Spectrometry sequenced peptides were mapped back to the genome to further enhance the quality of the annotated data of the *S. Typhi* genome.

DNA microarray analysis was used to quantify differences in gene expression between wild type and a *S. Typhi* Ty2 *ompR* mutant and identify genes, which were significantly differentially expressed in both transcriptome sequencing and microarray experiments. A subset of genes that were OmpR-regulated, hypothetical and absent from *E. coli* were identified to evaluate their virulence potential using *S. Typhimurium* mutant derivatives and *in vitro* assays. Protein homology searches were used to explore the potential function of several novel OmpR-regulated proteins. Finally, applying new DNA sequencing technology allowed development of a chIP-seq protocol to identify OmpR-binding sites.

“The last thing I'll say to the people that don't believe in cycling – the cynics, the sceptics, I'm sorry for you. I'm sorry you can't dream big and I'm sorry you don't believe in miracles.”

Lance Armstrong, 2005.

# Acknowledgements

First and foremost I would like to thank my parents, Jenny and Phil, who encouraged me throughout my formative years to believe in myself and to take whichever opportunities arose, no matter how large or small. Their guidance and counsel gave me the confidence to accept this challenge offered by Professor Gordon Dougan (Doog).

To Doog, many thanks for introducing me to molecular microbiology, training me and above all, persisting with me, I hope I have upheld your belief.

Most importantly I would like to thank my best friend and wife, Kura who has stood by me for all these years, always believing. Kura has been a rock and inspiration to me since we met and has always offered unconditional support to all of my pursuits.

During my PhD, I have received substantial guidance from Dr Robert Kingsley (Dr Bob). He has honed my scientific skills over the four years I've been at the Sanger Institute, many thanks Dr Bob. I would also like to thank rest of Doog's team, those who have directly helped with experiments or those who have just put things away I forgetfully left around the lab.

Finally, many thanks to the Wellcome Trust who have underwritten this body of work.

Declaration .....	1
Abstract .....	2
Acknowledgements.....	4
Tables.....	10
Figures.....	11
Abbreviations .....	13
1 Introduction .....	15
1.1 The genus <i>Salmonella</i> .....	16
1.1.1 Classification and nomenclature .....	16
1.2 Disease syndromes caused by <i>Salmonella</i> .....	18
1.3 <i>Salmonella</i> pathogenicity .....	21
1.3.1 The biology of <i>Salmonella</i> infection .....	22
1.3.2 Typhoid vaccines.....	25
1.4 The genomics of invasive <i>Salmonella</i> .....	27
1.5 Genetics of <i>Salmonella</i> disease.....	33
1.5.1 TTSSs .....	33
1.5.2 The virulence (Vi) antigen.....	39
1.5.3 Horizontally acquired DNA.....	40
1.5.4 Adherence .....	41
1.5.5 Motility .....	42
1.5.6 Two-component regulators and virulence .....	42
1.6 Regulation of bacterial gene expression.....	44
1.6.1 Transcriptional activation .....	45
1.6.2 Post-transcriptional control.....	49
1.7 DNA sequencing technologies.....	49
1.7.1 DNA microarrays .....	49
1.7.2 High-capacity DNA sequencing .....	51
1.8 <i>In silico</i> - software and tools.....	52
1.8.1 Artemis and Artemis Comparison Tool (ACT) .....	53
1.8.2 BLAST and other packages .....	53
1.9 Aims of the Thesis .....	53
2 Materials and Methods.....	54
2.1 General chemicals, reagents and buffers.....	55
2.2 Bacterial strains.....	55
2.3 Plasmids.....	55
2.4 Microbiological media and techniques .....	56
2.4.1 Bacterial culture .....	56
2.4.1.1 Standard methods .....	56
2.4.1.2 Minimal media .....	58
2.4.2 Harvesting of bacteria from broth cultures.....	58
2.4.3 Growth curves.....	58
2.4.4 Identification of bacteria.....	59
2.5 Molecular biology techniques.....	59
2.5.1 Bacterial genomic DNA extraction .....	59
2.5.2 Plasmid DNA extraction.....	60
2.5.3 Agarose gel electrophoresis .....	60
2.5.4 Quantification of DNA and RNA by spectroscopy.....	61

2.5.5 DNA manipulation techniques.....	61
2.5.5.1 DNA restriction .....	61
2.5.5.2 Oligonucleotides.....	61
2.5.5.3 Polymerase chain reaction (PCR).....	64
2.5.5.3.1 PCR supermix (Invitrogen) .....	64
2.5.5.3.2 PFU polymerase high fidelity PCR.....	65
2.5.5.3.3 Real time PCR .....	65
2.5.5.3.4 Overlap extension PCR .....	66
2.5.5.4 Purification of PCR products .....	66
2.5.5.5 Cloning.....	66
2.5.5.5.1 Cloning PCR amplicons.....	66
2.5.5.5.2 Sub-cloning sequences into pWT12 .....	67
2.5.6 Precipitation of DNA.....	67
2.5.7 DNA sequencing .....	67
2.5.7.1 Capillary-based sequencing .....	67
2.5.7.2 Illumina sequencing.....	68
2.5.8 DNA transformation.....	69
2.5.8.1 Transformation of chemically competent cells .....	69
2.5.8.1.1 Preparation of chemically competent cells.....	69
2.5.8.1.2 Chemical transformation .....	69
2.5.8.2 Transformation by electroporation .....	70
2.5.8.2.1 Preparation of electrocompetent cells.....	70
2.5.8.2.2 Electroporation .....	70
2.5.8.3 Screening for transformed DNA .....	71
2.5.8.3.1 Screening for non recombinant DNA .....	71
2.5.8.3.2 Screening for recombinant DNA .....	71
2.5.9 Bacterial conjugation.....	71
2.5.10 Transduction.....	72
2.5.10.1 Infection of donor strain with P22 phage.....	72
2.5.10.2 Isolation of P22 phage containing genomic DNA.....	72
2.5.10.3 Infection of recipient strain with P22 phage .....	72
2.5.11 Mutagenesis .....	73
2.5.11.1 Red recombinase mutagenesis.....	73
2.5.11.2 FLP recombinase mutagenesis .....	74
2.5.11.3 Allelic exchange mediated by suicide vector constructs .....	74
2.5.12 RNA methodologies .....	75
2.5.12.1 RNA stabilisation .....	75
2.5.12.2 RNA isolation.....	75
2.5.12.2.1 RNA isolation for microarray.....	75
2.5.12.2.2 RNA isolation for Illumina sequencing .....	76
2.5.12.2.3 rRNA removal .....	76
2.5.12.3 Removal of DNA from RNA samples.....	76
2.5.12.4 Reverse transcription of RNA .....	77
2.5.12.4.1 Reverse transcription of RNA for aminoallyl labelling .....	77
2.5.12.4.2 Reverse transcription of RNA for Illumina sequencing and real time PCR.....	77
2.5.12.5 Labelling of RNA for microarray.....	77
2.5.12.5.1 Purification of cDNA .....	77
2.5.12.5.2 Coupling of CyDye™ ester to dUTP .....	78
2.5.12.5.3 Purification of labelled cDNA .....	78

2.5.13 DNA microarray.....	79
2.5.13.1 Hybridisation and scanning.....	79
2.5.13.2 Gridding and feature extraction.....	79
2.5.14 Protein methodologies .....	80
2.5.14.1 Preparation of protein .....	80
2.5.14.1.1 Whole cell lysate.....	80
2.5.14.1.2 Cellular fractionation .....	80
2.5.14.1.2.1 Cell lysis .....	80
2.5.14.1.2.2 Precipitation of the soluble fraction .....	81
2.5.14.1.2.3 Separation of the membrane fractions.....	81
2.5.14.1.3 Heat shock .....	82
2.5.14.1.4 Preparation of proteins for mass spectroscopy analysis.....	82
2.5.14.2 One-dimensional SDS-PAGE gel electrophoresis of proteins.....	82
2.5.14.3 Protein visualisation.....	82
2.5.14.3.1 Coomassie blue protein stain.....	82
2.5.14.3.2 Western blot.....	83
2.5.14.4 Immunoprecipitation for chIP-seq.....	83
2.5.14.5 chIP-seq data mapping and analysis .....	83
2.6 Murine model of Salmonellosis.....	84
2.6.1 Competitive infections.....	84
2.7 LC-MS analysis .....	85
2.8 Bioinformatics tools.....	87
2.8.1 Artemis and ACT .....	87
2.8.2 Mapping software.....	89
2.8.2.1 Transcriptome mapping software.....	89
2.8.2.1 Protein mapping scripts.....	91
2.8.3 Perl scripts.....	91
2.8.3.1 Pileup to stranded plot - maqpileup2depth.pl.....	91
2.8.3.2 Local plots.....	91
2.8.3.3 Extracting expression data from plot.....	91
2.8.4 R .....	92
2.8.4.1 Limma analysis.....	92
2.8.5 Genespring .....	93
2.8.6 Graph pad prism 5.0 .....	93
3 A Method for RNA-seq in prokaryotes.....	94
3.1 Aims of the work described in this Chapter .....	95
3.2 Introduction.....	95
3.3 Results .....	96
3.3.1 Directional Sequencing of Single Stranded cDNA .....	96
3.3.2 Sequencing of the <i>S. Typhi</i> transcriptome .....	100
3.3.3 Mapping and visualising data .....	103
3.4 Discussion and conclusion.....	105
4 Deep sequencing of the <i>S. Typhi</i> Ty2 transcriptome.....	107
4.1 Aims of the work described in this Chapter .....	108
4.2 Introduction.....	108
4.3 Results .....	109
4.3.1 Mapping Sequence Reads to the Annotated Ty2 Genome .....	109
4.3.2 Coding Sequences .....	112
4.3.2.1 General Features.....	112
4.3.2.2 Virulence Genes .....	118

4.3.2.3 Phage and putative cargo genes.....	121
4.3.2.4 Pseudogenes .....	122
4.3.2.5 Hypothetical Genes.....	123
4.3.2.6 A <i>S. Typhi</i> Ty2 Specific Insertion.....	123
4.3.3 Non-Coding Sequences .....	124
4.3.4 Transcriptome comparison with <i>ompR</i> -null mutant.....	129
4.3.4.1 Quality control.....	129
4.3.4.2 General features of the BRD948 <i>ompR</i> transcriptome .....	133
4.3.4.3 Quantified differences.....	134
4.3.4.4 Quantified differences pre-Benjamini Hochberg correction.....	137
4.4 Discussion.....	137
4.4.1 General transcriptome results.....	137
4.4.2 OmpR comparison.....	139
4.5 Conclusion.....	141
5 Global transcriptome analysis of an <i>S. Typhi ompR</i> -null Mutant. ....	142
5.1 Aims of this chapter .....	143
5.2 Introduction.....	143
5.3 Results .....	145
5.3.1 Analysis of the <i>S. Typhi</i> OmpR regulon, preliminary work.....	145
5.3.2 Microarray analysis .....	145
5.3.2.1 Biological conditions, quality control and analysis.....	145
5.3.2.2 The OmpR regulon during exponential growth (OD <sub>600</sub> =0.6) .....	148
5.3.2.3 The OmpR regulon during early stationary phase (OD <sub>600</sub> =1.1).....	150
5.3.2.4 The OmpR regulon grown under SPI-2 inducing conditions.....	151
5.3.3 Real time PCR confirmation of expression profile .....	152
5.3.4 Succinate dehydrogenase enzyme assay .....	153
5.3.5 chIP-seq to determine ompR-binding domains.....	155
5.3.5.1 Rabbit anti-OmpR.....	155
5.3.5.2 Construction of an OmpR:3xFLAG .....	155
5.3.5.3 chIP-Seq Quality Control.....	156
5.3.5.4 chIP-Seq using the Illumina sequencing platform.....	157
5.3.5.6 Comparison with genes differentially transcribed in microarray data .....	159
5.4 Discussion.....	160
5.4.1 Microarray Data .....	160
5.4.1.1 Regulators .....	161
5.4.1.2 Membrane transport.....	163
5.4.1.3 Respiration .....	165
5.4.1.4 Chemotaxis and motility .....	167
5.4.1.5 Genes of unknown function .....	167
5.4.1.6 Indirect comparison with RNA-seq data.....	168
5.4.2 chIP-seq .....	170
5.5 Conclusion.....	171
6 Characterisation of Hypothetical Genes in the OmpR Regulon.....	172
6.1 Aims of this chapter .....	173
6.2 Introduction.....	173
6.3 Results .....	174
6.3.1 <i>In silico</i> analyses .....	174
6.3.1.1 t1787-t1793 exhibited increased expression in the <i>ompR</i> mutant .....	175
6.3.1.2 t4498-4501 exhibited decreased expression in the <i>ompR</i> mutant .....	177

6.3.2 Identification of proteins using LC-MS.....	178
6.3.3 Construction of recombinant <i>S. Typhimurium</i> mutants in which homologues of t1787-1790, t1791-1793 and t4498-4501 are deleted .....	180
6.3.4 <i>in vitro</i> Phenotyping of Mutants .....	182
6.3.4.1 LPS silver stained gel .....	182
6.4.4.2 Growth on different carbon sources .....	183
6.4.5 <i>in vivo</i> phenotyping .....	184
6.4.5.1 Competitive infection assays.....	184
6.5 Discussion.....	187
6.5.1 RAK103 and RAK105 characterisation .....	187
6.5.2 TT56.1 phenotype .....	190
6.6 Conclusion.....	190
7 General Discussion .....	192
8 References .....	196
9 Appendices .....	216
9.1 Maqdeh2pileup.pl .....	216
9.2 extractLines.pl.....	218
9.3 tram.pl.....	220
9.4 LIMMA Analysis for Illumina data .....	221
9.5 Hypothetical gene sequence coverage.....	227
9.6 Non-coding RNA details .....	281
9.7 Differential expression pre-Benjamini-Hochberg correction .....	288
9.8 Microarray Differences .....	297
9.9 chIP data .....	302

# Tables

Table 1.1 Salmonella nomenclature in use at CDC, 2000 [4].....	17
Table 1.2 Three years cumulative risk of typhoid fever (RR, 95% CI) [66] .....	27
Table 1.3 Genes unique to pairs of genomes [82] .....	31
Table 1.4. Effectors of the SPI1- and SPI2-encoded type III secretion systems [108]	36
Table 1.5 Major families of bacterial transcription factors [164] .....	46
Table 2.1 Strains used in this study .....	55
Table 2.2 Plasmids used in this study .....	56
Table 2.3 Antibiotic concentration and solvents .....	58
Table 2.4 Primers used in this study .....	62
Table 3.1 Read mapping for different depletion methods .....	102
Table 4.1 Analysis of sequences mapped to the Ty2 genome .....	110
Table 4.2. Sequence data mapped for each experiment.....	130
Table 4.3 Statistically different genes in the <i>ompR</i> mutant .....	135
Table 5.1 Genes differentially expressed in microarray experiments and with chIP-seq enrichment sites mapped upstream .....	159
Table 5.2 Experimentally verified and published genes in the OmpR regulon of <i>S.</i> <i>Typhi</i> or <i>S. Typhimurium</i> identified as being differentially regulated in the microarray experiments (2-fold, $p < 0.05$ ) .....	161
Table 6.1 Differentially expressed hypothetical genes and the published annotation .....	174
Table 6.2 Differentially expressed hypothetical genes re-annotated according to functional identity derived using pBLAST and Pfam database searches.....	178
Table 6.3 Growth on specific sole carbon sources .....	184

# Figures

Figure 1.1 Disease syndromes and subspecies [7].	19
Figure 1.2 WHO estimates of the global burden of typhoid fever and its geographical distribution [8].	20
Figure 1.3 The horizontal transfer of virulence-associated genes within the genus <i>Salmonella</i> [78].	29
Figure 1.4 Analysis of the P4-like bacteriophage ST46 [89].	32
Figure 1.5 Model for substrate recognition and delivery of proteins by TTSS machines [96].	34
Figure 1.6 Cellular phenotypes associated with the function of SPI-2 [113].	38
Figure 1.7 Genetic structure of the <i>viaB</i> locus in <i>S. Typhi</i> Ty2.	40
Figure 1.8 Genetic content of horizontally acquired regions in <i>S. Typhi</i> CT18.	41
Figure 1.9 Some of the different combinations of histidine protein kinase (HPK) and aspartate response regulator (RR) domains in histidine–aspartate phosphorelay (HAP) systems [166].	48
Figure 1.10 Schematic of Illumina sequencing technology ( <a href="http://www.illumina.com/">http://www.illumina.com/</a> ).	52
Figure 3.1 The hypotheses proposed to account for the attachment of Illumina adapted dimers to <i>sscDNA</i> .	98
Figure 3.2 Single stranded sequencing	99
Figure 3.3 Quantitative comparison of terminator exonuclease and oligonucleotide hybridisation rRNA depletion techniques.	103
Figure. 3.4 Representation of transcriptomic sequence read coverage plot relative to genome annotation, as displayed in Artemis.	105
Figure 4.1 Circular plot of mapped sequence data.	111
Figure 4.2 Artemis representation of transcriptome plot.	112
Figure 4.3 Defined deletions in BRD948.	112
Figure 4.4 Sense and Antisense outliers.	114
Figure 4.5. Transcripts mapping to well characterised regions.	115
Figure 4.6 Consistency between sequencing experiments.	115
Figure 4.7 “Mosaic coding”.	116
Figure 4.8. Putative error in the annotation.	116
Figure 4.9. Functional classification of sequenced transcripts.	117
Figure 4.10. Attenuation.	118
Figure 4.11 Coverage for SPI-1.	119
Figure 4.12 Coverage across SPI-2.	120
Figure 4.13. Sequenced transcripts identifying cargo genes within <i>Salmonella</i> prophage.	122
Figure 4.14. Third Party Annotation of the Ty2-specific Insertion.	124
Figure 4.15 Coverage of putative non-coding RNAs.	126
Figure 4.16. SPI-1 Sequence coverage	128
Figure 4.17. Paralogues of RUF107c.	129
Figure 4.18 Quality control.	132
Figure 4.19 Assigning transcriptome to functional class.	133
Figure 4.20. Start of the <i>viaB</i> locus.	134
Figure 4.21 Outer membrane porin C.	134
Figure 5.1 Quality control for expression microarrays.	147

Figure 5.2 Numbers of genes differentially transcribed by functional class.....	149
Figure 5.3 Circular plot of the <i>S. Typhi</i> genome and genes differentially transcribed in the <i>ompR</i> mutant. ....	150
Figure 5.4 Comparison of expression values determined by real-time PCR and microarray.....	153
Figure 5.5 Succinate dehydrogenase assay. ....	154
Figure 5.6 Schematic of overlap extension PCR to incorporate 3XFLAG peptide into the C-terminal region of OmpR.....	156
Figure 5.7 Western blot of immunoprecipitated OmpR:3XFLAG.....	157
Figure 5.8 Transporter transcript changes.....	164
Figure 5.9 Respiration.....	166
Figure 5.10 Genes of unknown function.....	168
Figure 5.11 Expression ratio for genes found to be differentially expressed in microarray and deep-sequencing experiments.....	169
Figure 6.1. Comparative alignment of regions predicted to be OmpR regulated in <i>S. Typhi</i> . ....	176
Figure 6.2 The chemical structure of related acetylated aminosugars.....	177
Figure 6.3 Sequenced peptides mapped to regions predicted to be OmpR-regulated. ....	180
Figure 6.4 Details of mutagenesis for further phenotypic characterisation. ....	182
Figure 6.5 LPS silver stained gel.....	183
Figure 6.6 Control competitive infections. ....	186
Figure 6.7 Comparison of mutants RAK103 and TT56.1 with RAK113 (SL1344 $\Delta$ <i>phoN</i> ).....	186

# Abbreviations

Acronym	Definition
ACT	Artemis comparison tool
ADS	arginine deiminase
AM	arithmetic mean per base-pair
ATP	adenosine tri-phosphate
BLAST	basic local alignment search tool
bp	base-pair
CDC	Centre for Disease Control
cDNA	complementary DNA
CDS	coding sequence
CFU	colony forming units
CGH	comparative genome hybridisation
chIP	chromatin immunoprecipitation
CL3	containment level 3
DEPC	diethylpyrocarbonate
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	ethylenediaminetetraacetic acid
EGTA	ethylene glycol tetraacetic acid
FDR	false discovery rate
FNR	false non-discovery rate
gDNA	genomic DNA
HAP	histidine aspartate phosphorelay
HIV	human immunodeficiency virus
HPK	histidine protein kinase
JCICSB	Judicial Commission of the International Committee of Systematic Bacteriology
LASER	light amplification by stimulated emission of radiation
LB	Luria-Bertani
LC-MS	liquid chromatography mass spectroscopy
LPS	lipopolysaccharide
M cells	Microfold cells
MOPS	3-(N-morpholino)propanesulfonic acid
mRNA	messenger RNA
NAG	N-acetyl glucosamine
NAM	N-acetylmuramic acid
NB	Note
nc	nucleotide
NCBI	National Center for Biotechnology Information
ncRNA	non-coding RNA
Nramp-1	natural resistance-associated macrophage protein one
NTS	Non-typoidal Salmonella
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline

PCR	polymerase chain reaction
PMN	polymorphonuclear
RES	reticuloendothelial system
RNA	ribonucleic acid
RR	response regulator
rRNA	ribosomal RNA
RUF	RNA of unknown function
<i>S.</i>	<i>Salmonella enterica</i> serovar <i>enterica</i>
SCV	<i>Salmonella</i> containing vacuole
SDS	sodium dodecyl sulfate
SPI	<i>Salmonella</i> Pathogenicity Island
STM	signature tagged mutagenesis
TCA	trichloroacetic acid
tRNA	transfer RNA
TTSS	type three secretion system
UDP	uridine diphosphate
USA	United States of America
UTR	untranslated region
UV	ultraviolet
WHO	World Health Organisation
WT	wild-type
WTSI	Wellcome Trust Sanger Institute

# 1 Introduction

## 1.1 The genus *Salmonella*

The genus *Salmonella* incorporates Gram-negative, facultative anaerobic rod shaped bacilli that are members of the family *Enterobacteriaceae*. This genus, which is estimated to have diverged from *Escherichia coli* (*E. coli*) approximately 100 to 150 million years ago [1], has adapted to colonise many different niches. For example, *Salmonella* can be found as both as a commensal and pathogen in warm and cold-blooded animals and is capable of surviving free in the environment. Presently *Salmonella* is more commonly associated with food poisoning in developed countries. However host adapted *Salmonella* are historically important as the causative agent of typhoid fever, for which there is still a severe disease burden in many developing countries. In this chapter I discuss aspects of the syndromes and diseases caused by *Salmonella*, as well as the biology of host interactions.

### 1.1.1 Classification and nomenclature

Various methods for the classification and nomenclature of the *Salmonella* species, subspecies, subgenera and serotypes have been proposed over the years. However, the Judicial Commission of the International Committee of Systematic Bacteriology (JCICSB) has not so far adopted a universal system. Initially *Salmonella* isolates were defined primarily by their clinical role, biochemical characteristics or genomic relatedness. The isolation of many more diverse isolates moved Kauffmann [2] to propose classification by serological identification of O (somatic), H (flagella) and K (capsular) antigens with each serotype considered a different species. This classification would currently define ~2,500 species of *Salmonella*. In 1973, Crosa *et al* [3] reported that DNA-DNA hybridisations indicated all serotypes and subgenera

from subspecies I, II and IV belonged to a single species, which excluded *S. bongori*, previously known as subspecies V.

The current classification system used by the Centre for Disease Control (CDC), Georgia, USA, based on recommendations from the World Health Organisation (WHO), is summarised in table 1.1(a) [4]. This system was rejected by the JCICSB due to the opinion that the status of *S. Typhi* was not adequately addressed and consequently may be overlooked by physicians if reported as *S. enterica* subsp. *enterica* serotype Typhi.

Table 1.1 Salmonella nomenclature in use at CDC, 2000 [4]

Taxonomic position	Nomenclature
Genus (italics) .....	<i>Salmonella</i>
Species (italics) .....	- <i>enterica</i> which includes subspecies I, II, IIIa, IIIb, IV, and VI - <i>bongori</i> (formerly subspecies V)
Serotype (capitalized, not italicized) <sup>b</sup> .....	- The first time a serotype is mentioned in the text; the name should be preceded by the word "serotype" or "ser." - Serotypes are named in subspecies I and designated by antigenic formulae in subspecies II to IV, and VI and <i>S. bongori</i> - Members of subspecies II, IV, and VI and <i>S. bongori</i> retain their names if named before 1966

<sup>a</sup> In 1984 Farmer et al. (10) updated the reporting system used at CDC for *Salmonella*. The major changes that CDC made and that result in a difference from the 1984 reporting system are (i) capitalization of the serotype name, (ii) inclusion of subspecies VI and *S. bongori* and (iii) adoption of the type species name *S. enterica*

<sup>b</sup> Examples of serotype designations are *Salmonella* serotype (ser.) Typhimurium, *Salmonella* II 50:b:z<sub>6</sub>, *Salmonella* IIIb 60:k:z, and *Salmonella* ser. Marina (IV 48:g,z<sub>51</sub>:—).

The genus *Salmonella* bifurcates to species, *S. enterica* and *S. bongori*, each of which contains multiple serotypes. *S. enterica* is divided into 6 subspecies, each referred to by a Roman numeral and differentiated by biochemical characteristics and genomic

phylogeny. A majority of the currently classified isolates fall into subspecies I, since they are predominantly responsible for human clinical diseases. For serotypes in subspecies I, historical names are retained, e.g. Enteritidis, Typhimurium. The geographical location where first isolated was commonly used to name serotypes. For subspecies II, IV, VI and *S. bongori*, antigenic formulas are assigned for serotypes described post 1966. To stress the fact that named serotypes are not regarded as separate species, the name is not italicised and is written in title case. For the initial citation of a serotype, the genus name is followed by serotype (or ser.) and then the serotype name, e.g. *Salmonella* serotype Typhi and subsequent citations written as *S.* Typhi. In modern literature on *Salmonella* the terms serotype and serovar are used almost interchangeably.

## 1.2 Disease syndromes caused by *Salmonella*

Infection by *Salmonella* serotypes can result in varied clinical syndromes or disease states. The outcome of an interaction between a *Salmonella* strain and a potential host is dependent on many factors including serovar, host-type, infecting dose, immunological competence and gut flora. The spectrum of disease ranges from a potentially fatal febrile illness (typhoid fever) to asymptomatic carriage. *Salmonella enterica* subspecies I *enterica* is the principal cause of salmonellosis in mammals. This group is generally described as causing either typhoidal or non-typhoidal salmonellosis (NTS) (figure 1.1). *Salmonella* infections have occasionally caused meningitis [5] and osteomyelitis [6].

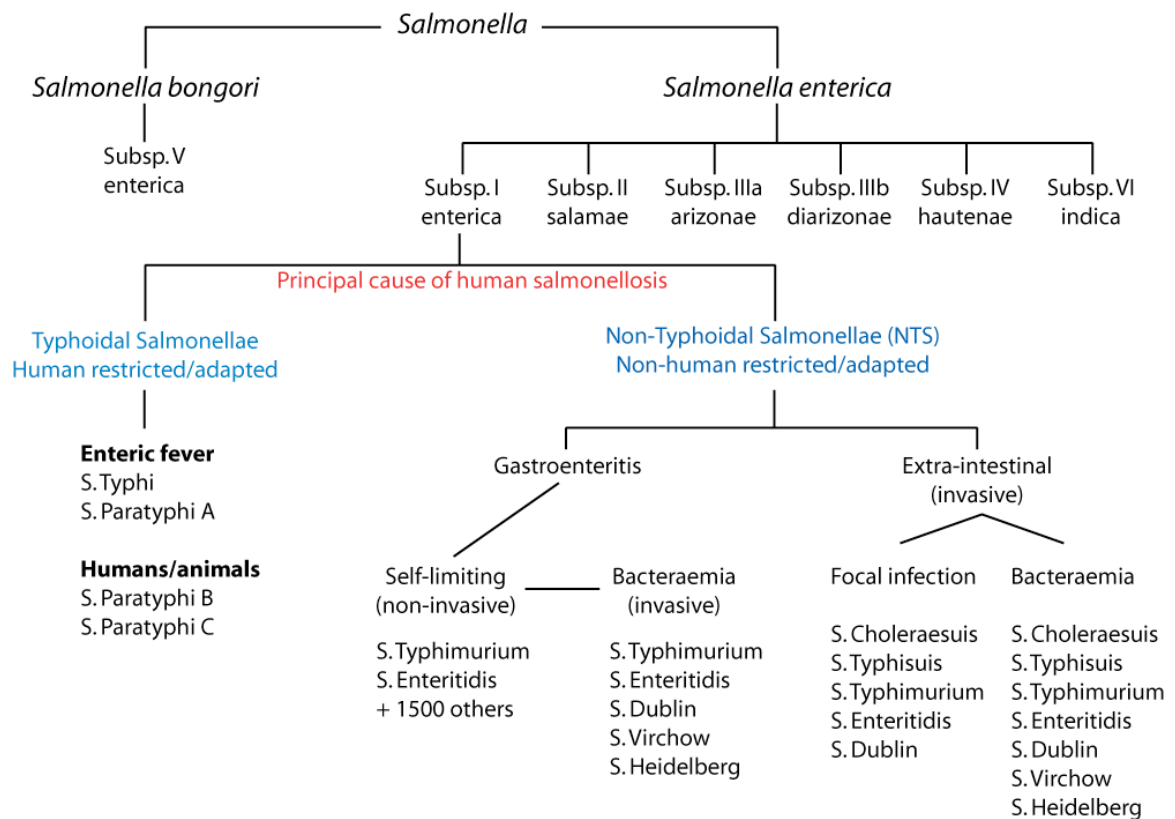


Figure 1.1 Disease syndromes and subspecies [7].

Historically a major disease burden in all parts of the world, typhoid fever remains endemic in developing countries with limited public health infrastructure (figure 1.2). The estimated annual burden is approximately 16 million cases, resulting in ~600,000 deaths [8]. Typhoidal *Salmonella* include *S. Typhi* and *S. Paratyphi A*, which are human restricted and *S. Paratyphi B* and *C* that are generally regarded as being host-adapted, as some isolates can be pathogenic in animals. The symptoms of enteric fever generally appear 7-28 days after ingestion and commonly include fever, nausea, abdominal cramp, vomiting and chills. If untreated the liver and spleen will enlarge and further complications may result. such as perforation of the large intestine [9] an outcome that is generally fatal in the absence of surgical intervention.

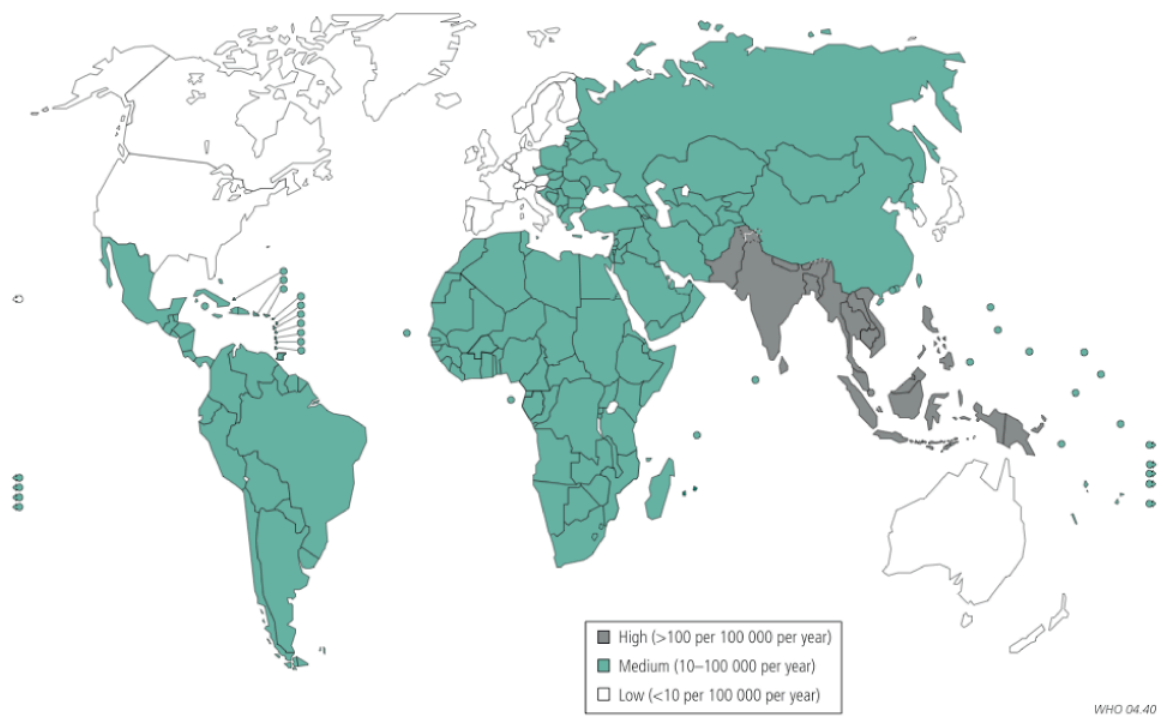


Figure 1.2 WHO estimates of the global burden of typhoid fever and its geographical distribution [8]

Some Non-typhoidal *Salmonella* (NTS) serotypes are broad host-range pathogens capable of infecting more than one host species (figure 1.1). These NTS generally cause a self-limiting gastroenteritis associated with abdominal pain, vomiting and inflammatory diarrhoea. Occasionally these serovars or specific strains within a serovar can cause bacteraemia or extra-intestinal invasive infection. Invasive disease associated with NTS is regarded as a relatively rare event in the developed world [10]. However, NTS bacteraemia has emerged as a significant public health problem in sub-Saharan Africa, associated with HIV in adults and malnutrition, anaemia, malaria and HIV in children.

Although *S. Typhi* can establish febrile systemic infection, in some individuals colonisation may be asymptomatic leading to a chronically persistent carrier state. Perhaps the most famous human typhoid carrier was Typhoid Mary, a US citizen who

regularly prepared food and was consequently responsible for infecting hundreds of people. Asymptomatic carriage is frequently characterised by shedding of high numbers of *S. Typhi* in the faeces, a phenomenon that may result in subsequent contamination of food or other communal intermediates, leading to the infection of other hosts. It is not uncommon for whole families to be infected by an asymptomatic carrier within that family. The carrier state is thought to be a mechanism through which a pathogen survives inter-epidemic periods when the number of susceptible individuals is not great enough to maintain the pathogen in the population. This is a common property of pathogens with a restricted host range but may not be required for more promiscuous zoonotic serotypes.

### 1.3 *Salmonella* pathogenicity

*Salmonella* can infect a wide range of host species. However, due to the human-restricted nature of *S. Typhi*, there is no direct animal model suitable for studying this organism other than exploiting volunteers. Many non-typhoidal serovars harbour isolates capable of infecting animals. Some of these are promiscuous in the sense that they can infect several different host species. Some can cause diseases in animals and spread to humans as zoonotic infections. Isolates from some serovars, such as serovar Typhimurium can cause an invasive and systemic typhoid-like disease in susceptible mice [11]. This murine Typhimurium model has been exploited as a surrogate for typhoid with some success, at least in terms of being used to identify *Salmonella* genes involved in general pathogenicity. It is clear from such studies that there are many components that contribute to *Salmonella* pathogenesis. To date, over 100 genes have been implicated and these are not limited to recently horizontally acquired regions of DNA but also include core regions of the genome. A number of regions

harbouring multiple pathogenicity genes have been defined as Pathogenicity Islands. Two of these, *Salmonella* Pathogenicity Island I (SPI-1) and SPI-2, encode Type III secretion systems (TTSS) key to the invasive [12] and persistence phenotype [13]. Although SPI-1 and SPI-2 are essential for *Salmonella* pathogenesis, many other classes of genes, for example genes involved in metabolism or biosynthesis, are also required [14,15,16,17]. During infection the bacterium may be starved of essential amino acids that are in short supply in host tissues. Consequently, *Salmonella* auxotrophic for these limiting metabolites may be attenuated in terms of their ability to cause infection.

### 1.3.1 The biology of *Salmonella* infection

In order to understand the molecular basis of *Salmonella* infection, including the ability to colonise and survive within a host, it is important to understand the macro biological stages of infection. Much of our current understanding of *Salmonella* infection and the data used to formulate our ideas about mechanisms are derived from studies using the murine typhoid model. Studies in the mouse have been complemented by the exploitation of other infection models including cattle and birds as well as clinical observations on human disease.

In natural infection *Salmonella* are typically acquired from the environment by oral ingestion of contaminated water or food, or by contact with a carrier. Interestingly, the environmental reservoir often remains unknown. *S. Typhi* is particularly difficult to culture from the environment. However, in a recent campaign designed to encourage indigenous people of the Mekong delta, Vietnam, to boil drinking water, a significant reduction of typhoid cases were reported indicating a water borne route of transmission [18].

Following ingestion of sufficient numbers of salmonellae, a proportion of the inoculum survive the low pH environment of the stomach to enter the lower small intestine where infection can be established. Conditions that increase the pH of the stomach can decrease the infective dose. However, *Salmonella* do have an adaptive acid tolerance response, which may aid survival in this environment [19]. To colonise the small intestine, *Salmonella* must gain access to the epithelium and avoid the neutralising effects of the innate immune system, including antimicrobial peptides [20], Immunoglobulin A [21] and chemical barriers such as bile salts [22]. Efficient adhesion to the epithelial layer is a prerequisite for invasion [23,24,25,26,27] and adherence to the apical membrane surface of the epithelial cell is mediated by adhesions such as fimbriae [28]. *Salmonella* can mediate direct invasion of the epithelial enterocyte cells, however, at this point in the infection process there are several possible routes towards systemic invasion. There is evidence that some *Salmonella* have a preference to exploit the microfold (M) cells, which are specialised epithelial cells that sample the antigenic content of the gut via pinocytosis [29]. After accessing M cells *Salmonella* can then gain access to the lymphoid cells and Peyer's patch. Alternatively, the bacterium may be phagocytosed by CD18-positive dendritic cells, which reach through the epithelial barrier and pull the bacteria down into the sub-epithelial layers [30]. Another mechanism may involve the ability of *Salmonella* to disrupt tight junctions, thus depositing the epithelial layer's capacity to control ionic balance and immune cell localisation [31].

At the present time relatively little is known about the mechanisms which prevent most *Salmonella* serotypes from going systemic. Clearly active T cell immunity is one requirement as HIV-positive individuals are more susceptible to bacteraemia caused by NTS isolates [32,33]. However, it is likely that the lack of a strong

polymorphonuclear cell (PMN) influx in typhoid facilitates systemic spread through some stealth mechanisms involving a targeted intracellular location. There is evidence that the Vi capsule of *S. Typhi* is anti-inflammatory [34] possibly masking access to pattern recognition molecules by innate receptors [35]. Whatever, typhoidal *Salmonella* are phagocytosed surreptitiously by intestinal macrophages and are disseminated through the reticuloendothelial system [36,37]. In contrast non-typhoidal *Salmonella* induce a localised inflammatory response in immune competent individuals, provoking a massive influx of polymorphonuclear leukocytes to the intestinal lumen and diarrhoea.

Systemic infection therefore requires a combination of an efficient transfer across the epithelial layer of the gut, combined with the ability to remain undetected by the immune system, referred to above as a stealth infection. In order to spread systemically, *Salmonella* undergoes either passive or active macropinocytosis [37,38,39] to gain entry to the macrophage. Once internalised, the bacteria subverts the normal maturation of the phagosome to form a *Salmonella* containing vacuole (SCV) [40] permissive for survival, persistence and eventually replication [41,42,43,44,45,46,47,48,49]. In the murine model, *Salmonella* mutants that are defective in mechanisms required for survival within macrophages are avirulent in mice [50,51,52,53,54]. Many auxotrophs are also attenuated, as tryptophan, tyrosine, phenylalanine, purines and pyrimidines are essential for bacterial replication within the SCV and possibly as extracellular bacteria [14,15,55,56]. Conversely, macrophages that are deficient in natural resistance associated macrophage protein 1 (Nramp-1) are extremely susceptible to *Salmonella* infection and cannot control replication as efficiently as wild type cells [57]. Nramp-1 may remove ions from the SCV, a process predicted to restrict bacterial replication [58]. Mice harbouring null

mutation in Nramp-1 are hyper susceptible to *S. Typhimurium* infection. These data, combined with the observation that *Salmonella* are typically isolated from the lymphatic tissues and RES organs during infection, further support the hypothesis that survival within macrophages is essential for efficient systemic infection. Following *Salmonella* invasion bacterial adaptive responses (section 1.6), sensing the phagosomal milieu, low pH and magnesium ion content, tightly regulate expression of virulence determinants (section 1.5.1) to optimise survival in this endosomal compartment [45,59].

### 1.3.2 Typhoid vaccines

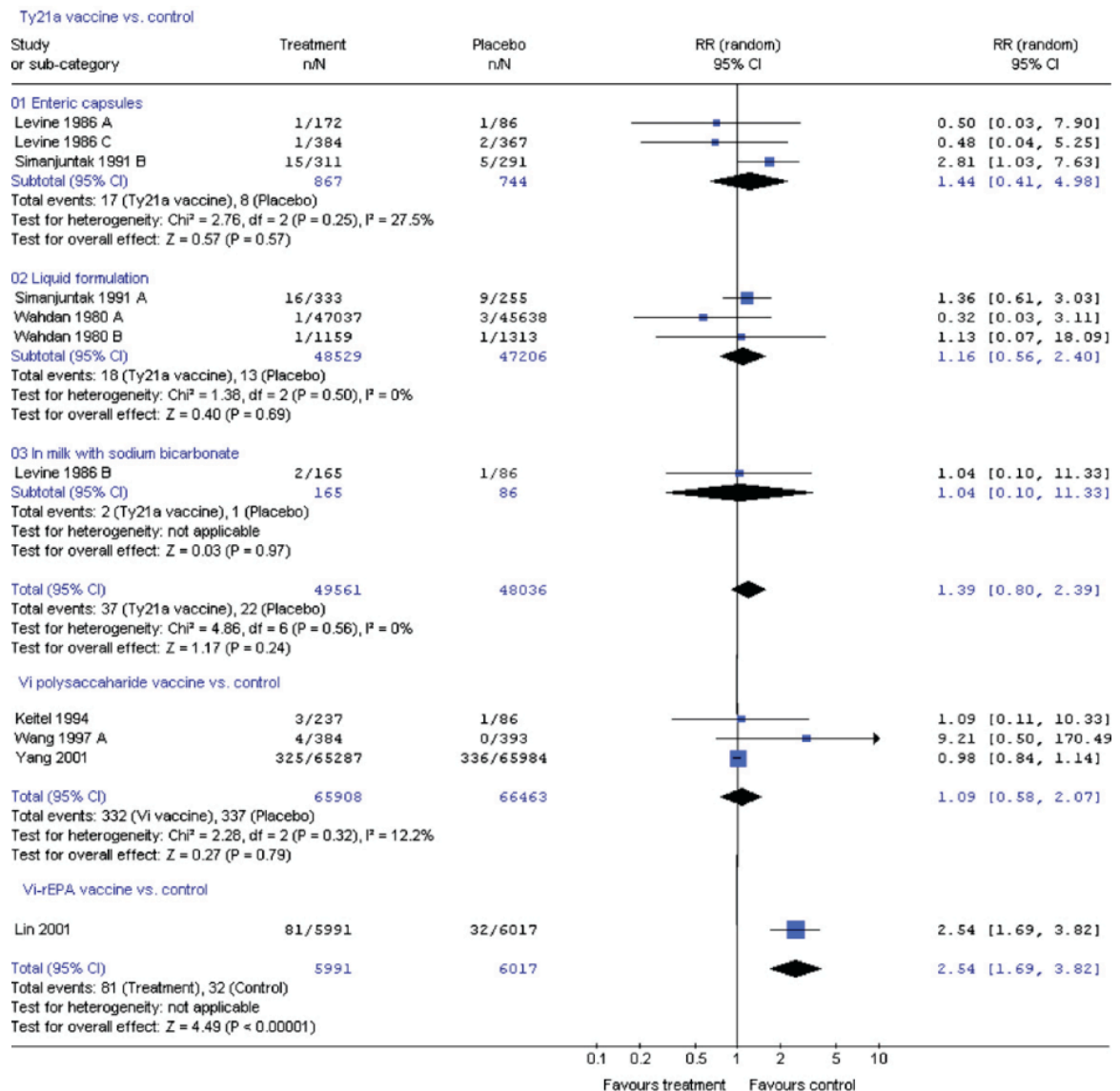
The first typhoid vaccine, introduced in 1896, was based on a whole-cell inactivated *S. Typhi* preparation. The efficacy of the vaccine has been questioned but during controlled trials in 1960 the efficacy after two parenterally administered doses was 73% over a 3-year period [60]. However, 10% of recipients reacted badly to the vaccine and were incapable of normal daily routine the following day. These factors eventually led to the use of the whole cell typhoid vaccine being generally abandoned.

There are currently two licensed typhoid fever vaccines. One, which is composed of purified Vi exopolysaccharide antigen, is administered parenterally in one dose, as it is non-boostable [61]. The other is a live oral vaccine, Ty21a, based on an attenuated *S. Typhi* Ty2 mutant derivative [62,63,64,65]. Ty2 vaccine, which is available in a liquid or capsule form, requires two or more oral doses to induce limited protection and is currently not in general use.

Fraser *et al* [66] reviewed a range of field studies involving Ty21a and Vi in which efficacy values were derived for each vaccine after a 3-year period (table 1.2).

Interestingly, these results showed a higher efficacy for the liquid compared to other formulation of Ty21a. This review highlighted the need for a more efficacious vaccine. However, the funds required to obtain another licensed vaccine are prohibitive and generally not commercially attractive as this disease affects mainly the developing world. Rationally designed live oral typhoid vaccines based on modified *S. Typhi* harbouring multiple defined deletions in genes such as those involved in aromatic compound synthesis (*aro*) [67,68], intracellular survival (SPI-2) [69], virulence regulators (*ompR*, *phoP*) [70,71] and serine proteases (*htrA*) [72] have been constructed. Some of these are currently in clinical trial. Due to the potent nature of *Salmonella* immunogenicity, attenuated strains are also being exploited to deliver heterologous antigens for different pathogens to the immune system [73,74,75].

Table 1.2 Three years cumulative risk of typhoid fever (RR, 95% CI) [66]



## 1.4 The genomics of invasive *Salmonella*

The whole genome sequence and annotation of *S. Typhi* isolate CT18 [76] and *S. Typhimurium* isolate LT2 [77] were published in 2001. These sequence data further confirm that the *Salmonella* are related to *E. coli*, as all three genomes are predominantly co-linear and syntenic. However, both *S. Typhi* and *S. Typhimurium* harbour hundreds of deletions and insertions relative to *E. coli*, apparently driving

vast phenotypic diversity. Genome diversity within the *Salmonella* is responsible for the 7 subspecies and 2,465 currently classified serovars [4]. The fact that the genomes of both *Salmonella* and *E. coli* are related facilitates genome comparisons between these two species. One of the most striking differences between *E. coli* and *Salmonella* is the acquisition of SPI-1 and SPI-2 by *S. enterica*. SPI-1 is essential for host cell invasion [12] and the discovery of this invasion specifying locus provided a significant advance in elucidating the molecular basis of *Salmonella* pathogenicity and evolution. The horizontal acquisition of SPI-1 associated DNA, combined with other factors such as the adhesions *fim* and *lpf*, approximately 120-160 million years ago is perceived to be key to the divergent speciation step between *E. coli* and *Salmonella* [78] (figure 1.3). The SPI-1 region encodes a TTSS that translocates effector proteins into the host cell and mediates cell invasion. The ability of *Salmonella* to invade tissue is a key property associated with driving *Salmonella* into a distinct ecological niche outside of the lumen of the intestine. The effector proteins encoded within the SPI-1 region, together with others encoded elsewhere and perhaps acquired independently, mediate endocytosis involving actin rearrangement and the formation of a protective SCV [79]. The mechanisms associated with SPI-1 mediated endocytosis are discussed further in section 1.5.

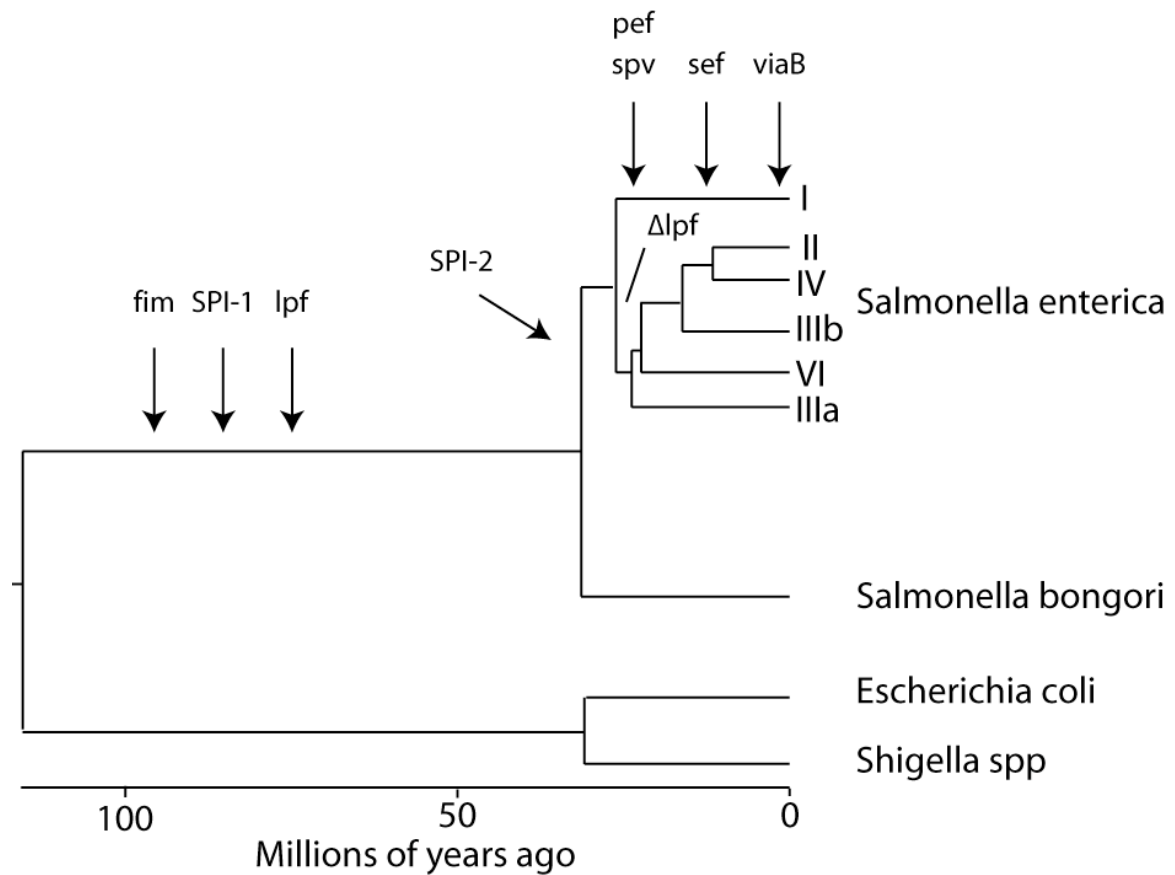


Figure 1.3 The horizontal transfer of virulence-associated genes within the genus *Salmonella* [78].

Deletion of the *lpf* operon is indicated by  $\Delta$ . Arrows mark acquisition of genes by horizontal transfer. Abbreviations: *fim*, type 1 fimbriae; *lpf*, long polar fimbriae; *pef*, plasmid encoded fimbriae; *sef*, *S. enterica* serotype Enteritidis fimbriae; *spv*, *Salmonella* plasmid virulence; *viaB*, Vi capsular antigen locus.

The murine model of salmonellosis has proven to be a fertile ground for the development of novel molecular tools. Signature tagged mutagenesis, which was developed using this model was originally used to identify *Salmonella* genes essential for intracellular and systemic survival in mice, and led to the characterisation of SPI-2 [13]. SPI-2 is a 35kb island that encodes a TTSS apparatus, effector proteins and a tetrathionate catabolism locus (see section 1.5). The insertion of the SPI-2 locus into a tRNA gene theoretically marks the divergence from *S. bongori* and the speciation of *S. enterica*. The acquisition of SPI-1, SPI-2 and adhesins such as *fim* [80] and *lpf* [81] permitted *Salmonella* to fill an intracellular and systemic niche, removing it from the

nutrient competitive environment of the gut. Some of the serovars of *S. enterica* subspecies I are further evolved from this ancestor in the sense that they are host-adapted, a process likely to involve both the acquisition and loss of specific genes.

What are the features of a *Salmonella* serovar (and the associated genomes) that contribute to host restriction? Sequencing of the *S. Typhi*, *S. Paratyphi* and *S. Gallinarum* genomes has highlighted possible roles for both gene acquisition and gene loss through deletion and pseudogene formation [76,82,83]. Of the 601 genes present in the host restricted *S. Typhi* CT18 (excluding plasmids) that are absent from the host promiscuous *S. Typhimurium* LT2, perhaps the most striking are those encoded by the insertion known as SPI-7. SPI-7 is a 134kb composite island that encodes the Vi biosynthetic genes (*viaB* locus), the SopE phage and a type IVB pilus operon [84,85]. Several studies have identified this region as being directly associated with *S. Typhi* host-adapted virulence. The Vi antigen itself may have immunomodulatory effects that help *S. Typhi* move through epithelial cell layers without inducing inflammation [35,86]. However, the *viaB* locus is not present in *S. Paratyphi* [82] and is missing from some of the *S. Typhi* clinical isolates sequenced by Holt *et al* [87]. Importantly, the purified Vi antigen is currently one of two licensed *S. Typhi* vaccines, which relies on humoral recognition of the antigen encoded by this unstable region [88]. Instability and absence of this entire insertion from some invasive *Salmonella* reduces the possibility of host-adapted virulence phenotype being entirely dependent on Vi expression, the SopE phage or the type IVB pilus operon. Comparisons by McClelland *et al* (2004) between the genomes of *E. coli*, *S. Typhi*, *S. Typhimurium* and *S. Paratyphi* A reveal the number of genes that are unique to each pair [82] (table 1.3). Interestingly, *S. Typhi* and *S. Paratyphi* A harbour ~172 genes, which are not present in either *E. coli* or *S. Typhimurium* suggesting some of their function may

contribute to systemic spread and typhoid and paratyphoid fever. However, host restriction and systemic spread are not likely to be explained simply by acquisition or loss of the SPI-7 region.

Table 1.3 Genes unique to pairs of genomes [82]

	STY	STM	ECO
SPA	172	53	0
STY		60	15
STM			48

Number of genes shared by a pair of genomes but not the other two genomes, comparing Paratyphi A 9150 (SPA), Typhi CT18 (STY), Typhimurium LT2 (STM) and E. coli K12 (ECO). Shared genes: 4 95% identity in a 100-bp window, except for E. coli comparison (4 75% in a 100-bp window).

Comparative genomics of bacterial genomes and bacteriophages illustrates the importance of lateral gene transfer for virulence and the evolution of pathogens [89]. Bacterial species such as *Vibrio cholerae* [90] and *Corynebacterium diphtheriae* [91] each rely on the acquisition of specific toxin-encoding prophages to cause disease. However, bacterial species such as *Streptococcus pyogenes* [92] and *S. enterica* harbour several prophage that encode virulence-associated determinants called “cargo genes” [89]. Indeed, *S. Typhi* CT18 harbours around seven prophage-related elements and many of them encode putative uncharacterised cargo genes (figure 1.4). For example, the P2-phage family includes ST27, ST35 and the SopE phage. The SopE phage is encoded within SPI-7 and is also found elsewhere in the genome in *S. Paratyphi* A and *S. Typhimurium* LT2 [77,82]. The *sopE* moron is one of the best-known examples of phage encoded cargo genes in *S. enterica*. The *sopE* gene is expressed during log phase within the predominantly transcriptionally silent prophage [93]. Lambda-like bacteriophages ST10 and ST18 are also found in *S. Typhi* as well as a chimeric phage, ST15. ST46 is a P4-like bacteriophage, which also encodes

predicted cargo. Comparison of these prophage-like elements in *S. Typhi* with *S. Typhimurium* LT2 perhaps reflect the divergent selective pressures such viral elements impose or facilitate in closely related and recently diverged serovars [89]. Interestingly, diversity within the recently emerged *S. Typhi* (circa 50,000 years ago) has remained extremely limited with only one variable phage region ST20 currently recorded [87]. This region has three known allelic forms in this insertion locus.

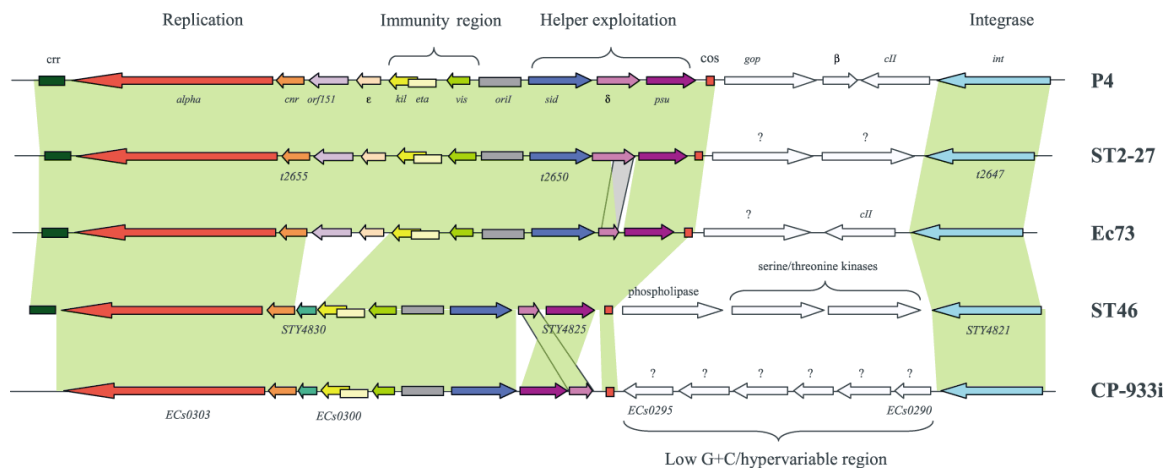


Figure 1.4 Analysis of the P4-like bacteriophage ST46 [89].

Alignment of the bacteriophage ST46 with other P4 family bacteriophages including ST2-27 (Ty2) (NC\_00463; this study) Retron 73 (*E. coli* M64113), CP-933i (*E. coli* 0157:H7) (NC\_002655) and P4 (X51522). Regions of significant amino acid homology are indicated via a shaded connection. Orthologous genes have the same colour. Genes in the low GC region are coloured white. A question mark indicates that the gene product has no known function.

Pseudogenes are CDSs that are potentially inactivated by mutations. Such mutations can result in a frameshift involving an insertion, deletion, or truncation of the reading frame leading to the introduction of a STOP codon. Pseudogenes are sometimes identified by sequence comparison with orthologous genes in related species that are not inactivated. Genome degradation, in part due to pseudogene formation, is a recurring theme in the evolution of host-restricted pathogens from an ancestral broad host range pathogen. Over 4% (~220 genes) of the CDS of *S. Typhi* are potentially

inactivated and it has been proposed that this gene loss was important in the adaptation of this pathogen to a restricted host range. Surface exposed proteins are important mediators of the interaction of bacterium with host-cells. Extracellular fimbriae mediate adhesion to host cells and are expressed on the bacterial cell surface. *S. Typhi* and *S. Paratyphi* share 11 clusters of genes associated with fimbrial biosynthesis and many of these loci exhibit some form of gene degradation, or are specific to each serovar [82].

## 1.5 Genetics of *Salmonella* disease

### 1.5.1 TTSSs

TTSSs are predominantly protein based delivery systems that are assembled across the inner and outer membranes of the bacterial wall to deliver effector proteins to the host-cell through a “syringe and needle” complex. The *S. enterica* genome encodes two characterised TTSS, which are vital components of two distinct infection processes, cellular invasion and systemic survival.

SPI-1 is a ~40kb island encoding a TTSS and associated effector proteins that are essential for the bacterial-mediated endocytotic invasion of epithelial cells [12]. SPI-1 consists of 4 divergently transcribed regions, which are expressed in a co-ordinated fashion under the control of the endogenous regulators *invF*, *hilA*, *hilC* and *hilD* [94]. TTSSs are believed to have evolved from ancestral flagella export and motility complexes and are present in many different species of bacteria. The structure, which resembles a syringe needle complex as viewed using electron microscopy, consists of approximately 20 different structural proteins [95]. As with flagella, the basal body is constructed across the inner and outer membranes with the hollow needle-like

complex protruding to form a channel to the host (figure 1.5). Once in place, a translocon is expressed to breach the host-cell membrane and form a pore for efficient host-cell delivery of effector proteins, which require specific chaperones to stabilise and aid translocation through the hollow TTSS needle [79]. Effector proteins are not all encoded within the SPI-1 locus but the genes for some are distributed around the chromosome. Upon entering the cell cytoplasm, effector proteins induce membrane rearrangements by targeting many different signalling pathways (table 1.4).

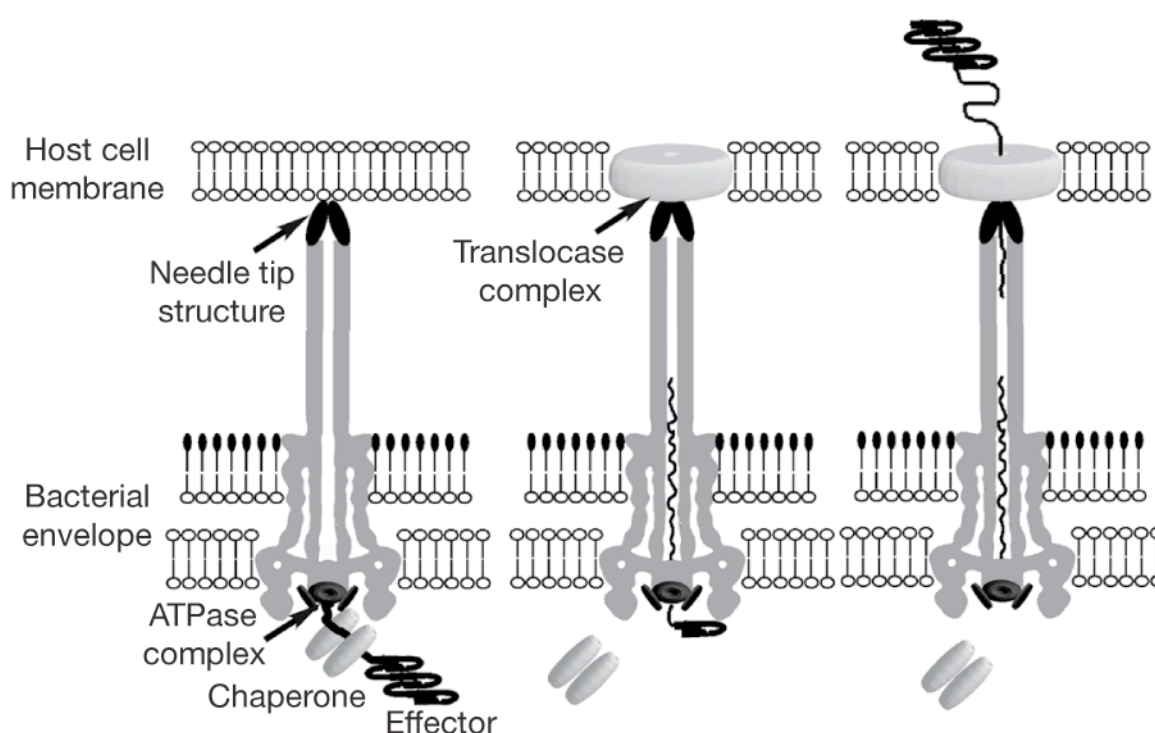


Figure 1.5 Model for substrate recognition and delivery of proteins by TTSS machines [96].

The effector–chaperone complex is recognized by the secretion machinery, including a TTSS-associated ATPase. The ATPase ‘strips’ the chaperones from the complex, which remains within the bacterial cell, and mediates the unfolding and ‘threading’ of the effector proteins through the central channel of the needle complex. A ‘translocator complex’ made up of proteins also secreted by the TTSS is assembled on the host cell membrane and mediates the passage of the effector proteins through the target cell membrane. The translocated effectors re-fold within the host cell to carry out their function.

SPI-1 mediates *S. Typhimurium* invasion in part by inducing actin rearrangement of epithelial cells, with at least 5 effector proteins involved. SopE, SopE2 and SopB induce actin cytoskeletal rearrangement, membrane ruffling and macropinocytosis and SipA and SipC directly control actin dynamics [97,98,99,100,101,102,103,104,105,106]. SopB is crucial during the early stages of invasion, effectively driving the macropinocytosis. Immunoprecipitation studies have identified cdc42 as possibly the only host protein bound by SopB during invasion [38] and a deletion in *sopB* attenuates intracellular growth *in vivo*. In *S. Typhimurium*, SopE and SopE2 have a 69% identity with the latter predicted to be a pseudogene in *S. Typhi*, with a stop codon 71bp from the GTG start codon [99]. The SipC protein forms functional domains to nucleate actin polymerization and bundle actin filaments (F-actin) and SipA independently binds F-actin to inhibit filament depolymerization [107].

Table 1.4. Effectors of the SPI1- and SPI2-encoded type III secretion systems [108]

Effector	Cellular function	Host-cell target
SPI1 T3SS		
AvrA	Inhibits nuclear factor (NF)- $\kappa$ B signalling and interleukin (IL)-8 production; also prevents ubiquitination of $\beta$ -catenin	Unknown
SipA or SspA	Decreases the critical concentration of G-actin and increases the stability of F-actin; also induces PMN transepithelial migration and disrupts tight junctions	F-actin; T-plastin
SipB or SspB*	Binds and activates caspase-1 and induces autophagy in macrophages	Caspase-1; cholesterol
SipC or SspC*	Nucleates and bundles actin	F-actin; cytokeratin-8 and cytokeratin-18
SopA	Stimulates PMN transmigration by HECT-like E3 ubiquitin ligase activity	Unknown
SopB or SigD	Activates Cdc42, RhoG, AktA and chloride secretion through its inositol phosphatase activity and disrupts tight junctions	Unknown
SopD	Stimulates fluid accumulation in bovine ligated ileal loops and contributes to diarrhoea in calves and systemic disease in mice	Unknown
SopE	Activates Cdc42, Rac1 and RhoG by its GEF activity and disrupts tight junctions	Cdc42, Rac1 and Rab5
SopE2	Activates Cdc42, Rac1 and RhoG by its GEF activity and disrupts tight junctions	Cdc42 and Rac1
SptP	Inhibits Cdc42 and Rac1 by its GAP activity and MAPK signalling and IL-8 secretion through its tyrosine phosphatase activity	Rac1
SPI2 T3SS		
GogB	Unknown	Unknown
PipB	Unknown	Unknown
PipB2	Contributes to Sif formation	Kinesin-1
SifA	Induces Sif formation, maintains integrity of the SCV and downregulates kinesin recruitment to the SCV	SKIP and Rab7
SifB	Unknown	Unknown
SopD2	Contributes to Sif formation	Unknown
SpIC*	Interferes with endosomal trafficking	Hook3
SpvB <sup>‡</sup>	Actin-specific ADP-ribosyltransferase and downregulates Sif formation	Actin
SseF	Contributes to Sif formation and microtubule bundling	Unknown
SseG	Contributes to Sif formation and microtubule bundling	Unknown
SseI or SrfH	Contributes to host-cell dissemination	Filamin and TRIP6
SseJ	Maintains integrity of the SCV and has deacylase activity	Unknown
SseK1	Unknown	Unknown
SseK2	Unknown	Unknown
SseL	Deubiquitinase	Ubiquitin
SspH2	Inhibits the rate of actin polymerization and contributes to virulence in calves	Filamin and profilin
SteA	Unknown	Unknown
SteB	Unknown	Unknown
SteC	Unknown	Unknown
SPI1 and SPI2 T3SS		
SlrP	Contributes to virulence in calves	Unknown
SspH1	Inhibits NF- $\kappa$ B signalling and IL-8 secretion, contributes to virulence in calves and has E3 ubiquitin ligase activity	PKN1

\* Also a component of the secretion apparatus. <sup>‡</sup> Has not been definitively shown to be an SPI2 T3SS effector. GAP, GTPase-activating protein; GEF, guanine nucleotide exchange factor; HECT, homologous to E6-AP carboxyl terminus; MAPK, mitogen-activated protein kinase; PMN, polymorphonuclear leukocyte; SCV, Salmonella-containing vacuole; Sif, Salmonella-induced filament; SPI, Salmonella pathogenicity island.

SPI-2 also encodes a TTSS that is important for *Salmonella* intracellular proliferation and persistence within the hostile SCV of macrophages [13]. This region was originally identified through a signature tagged mutagenesis screen in mice [13]. Two

operons encode 31 genes whose products form a classical TTSS, which is expressed in the SCV [50]. The expression of this region, which also encodes an endogenous two-component regulator *ssrAB* requires functional OmpR and PhoP [109].

The SPI-2 translocon and effector proteins required for intracellular survival have been characterised in detail in the murine model. The delivery of effector proteins requires three proteins, SseBCD, that form the translocon, and the chaperone SseA [110]. Mutation in the genes *sseBCD* attenuates the bacterium to the same level as an apparatus mutant and prevents delivery of effector proteins. Functional analyses of these proteins suggests they perform similar roles and all three proteins have been found attached to the bacterial cell surface *in vitro* [111]. There are currently ~13 known effector proteins translocated by the SPI-2 TTSS (table 1.4) and 10 of these loci map outside of SPI-2 [112]. Within this set of 13 effector proteins, there are two distinct groups, one group of eight proteins encoding an N-terminal peptide signal sequence required for translocation by the SPI-2 TTSS. The three effector proteins encoded within SPI-2 (table 1.4) and PipB and PipB2 do not require this secretion domain [112]. Surprisingly, no single defined deletion in any of the effector proteins induces significant attenuation in mice.

Tissue culture models are commonly used to characterise the SPI-2 TTSS mediated intracellular lifestyle of *Salmonella*. A variety of murine and human cell lines, including epithelial cells and macrophages, have been studied, yielding many apparently conflicting results. Figure 1.6 [113] represents at best a consensus of the mechanisms of each effector protein.

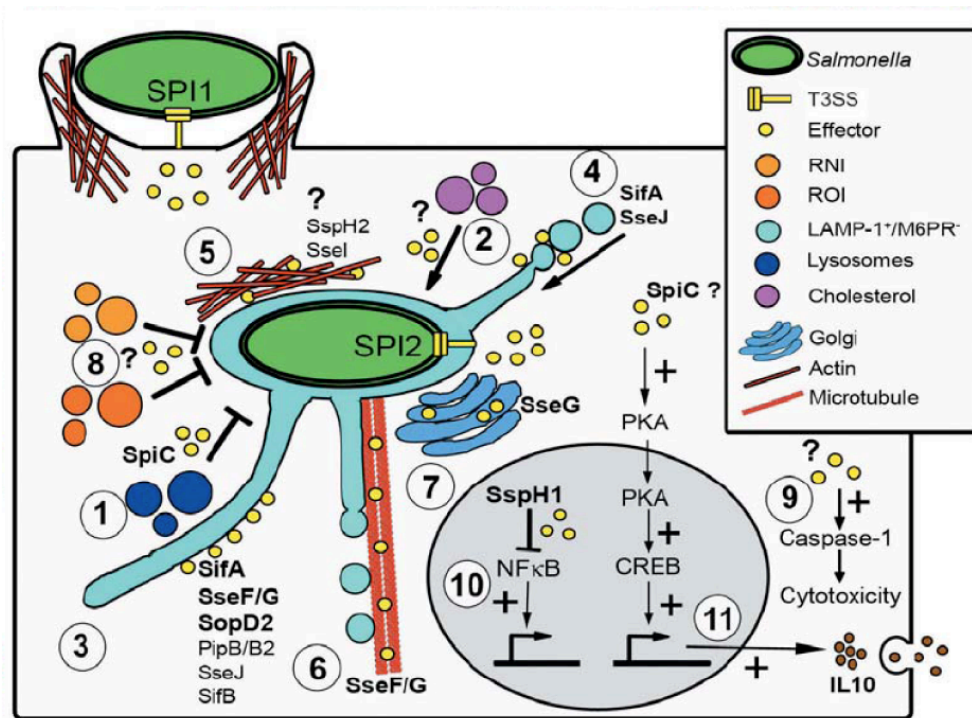


Figure 1.6 Cellular phenotypes associated with the function of SPI-2 [113].

The interactions of intracellular *S. enterica* with host cell functions via the SPI-2-encoded TTSS are shown in a model. While SPI-1 is activated by extracellular bacteria and triggers invasion, the SPI-2 system is activated by bacteria residing within a SCV. SPI-2 function interferes with a variety of different host cell processes: (1) modification of cellular trafficking and alteration of SCV maturation; (2) recruitment of cholesterol to the SCV; (3) formation of Salmonella-induced filaments (SIF) in HeLa cells (indicated by arrows). Infection with an *ssaV* or *sifA* strain did not induce SIF, while infection with the *sseF* mutant strain results in ‘pseudo-SIF’ formation; (4) maintenance of SCV integrity by combined fusion and scission events. In the presence of SseJ, a *sifA*-deficient strain escapes the SCV and is killed or replicates in the cytoplasm of macrophages or HeLa cells, respectively; (5) actin accumulation in the vicinity of the SCV; (6) bundling of microtubules and associated SIF-formation; (7) association of the trans-Golgi network (TGN) with the SCV; (8) inhibition of, and delivery of, reactive oxygen (ROI) and nitrogen intermediates (RNI) to the SCV; (9) delayed cell death; (10) inhibition of NFκB-dependent gene expression and (11) induction of interleukin 10 expression. Bold typeface indicates involvement of SPI-2 effectors in the respective phenotype; regular typeface indicates localization of the effector to the respective compartment.

The effects of SPI-2 effector proteins on the host cells represent a multi-pronged approach, which includes modification of the endosomal system and intracellular

transport, actin rearrangement, microtubule rearrangement and apoptotic dysfunction, all which at some point are redundant components yet derive such omnipotence for intracellular survival over the host-cell.

### 1.5.2 The virulence (Vi) antigen

The Virulence (Vi) antigen was first described as a capsular polysaccharide on the cell surface of *S. Typhi* by Felix and Pitt [114] and is encoded by the *viaB* locus in SPI-7. It is the current target of the purified parenteral Vi polysaccharide vaccine. The *viaB* locus harbours 10 genes, with 5 of these encoding proteins that are devoted to biosynthesis (*tviABCDE*) and 5 genes functional in exporting the mature polysaccharide to form the capsule (*vexABCDE*) [85] (figure 1.7). SPI-7 is an unstable region in *S. Typhi* and this region is not present in 2 of the 21 *S. Typhi* strains sequenced since [87]. The Vi antigen is also expressed by some isolates of the cattle adapted serovar *S. Dublin* and by most *S. Paratyphi C* [115]. Vi is also expressed on the surface of some isolates of the Gram-negative bacillus *Citrobacter freundii* [115]. Conservation of ViaB-associated gene synteny is observed across these species. The Vi capsule is known to mask serotyping by the O-somatic antigen and reduce invasion of tissue culture cells [35]. The Vi antigen has been associated with increased virulence, however, fully defining its effect on the severity of the disease has been limited due to the host-adapted nature *S. Typhi*. The instability of this locus *in vitro* has prompted speculation about the contribution of Vi to *Salmonella* virulence. However, it is unreasonable to assume this region confers no selective advantage during typhoid fever pathogenesis even though the antigen is not expressed by *S. Paratyphi A* or *B*.

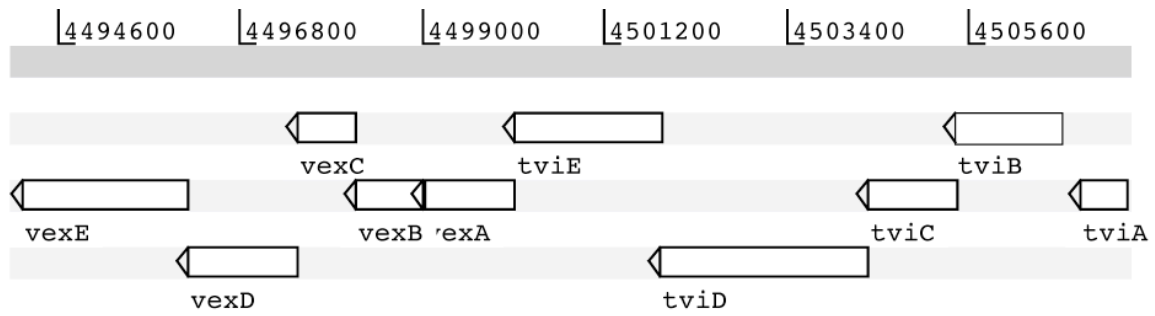


Figure 1.7 Genetic structure of the *viaB* locus in *S. Typhi* Ty2.

Expression of Vi antigen *in vitro* is regulated by the osmolarity of the growth media and is dependent on both functional *ompR-envZ* [116] and *rscBC* loci [117,118]. This, combined with the direct influence on SPI-2 by OmpR, suggests regulation of this locus is an important component of *S. Typhi* pathogenesis (see section 1.5.6).

### 1.5.3 Horizontally acquired DNA

Aside from the heavily characterised and virulence essential SPI-1, SPI-2 and SPI-7, *Salmonellae* encode other virulence-associated regions including SPI-3, SPI-4, SPI-5 and CS54. All of these regions exhibit some form of degradation in *S. Typhi*, based on functional homologies in other *Salmonellae* [76] (figure 1.8). SPI-3 is associated with intracellular survival and replication [119], SPI-4 is involved in colonisation in some animal models [120] and SPI-5 encodes effector proteins delivered by the SPI-2 TTSS [121]. Mutations in genes encoded on the CS54 island reduce colonisation and shedding in the murine model [122].

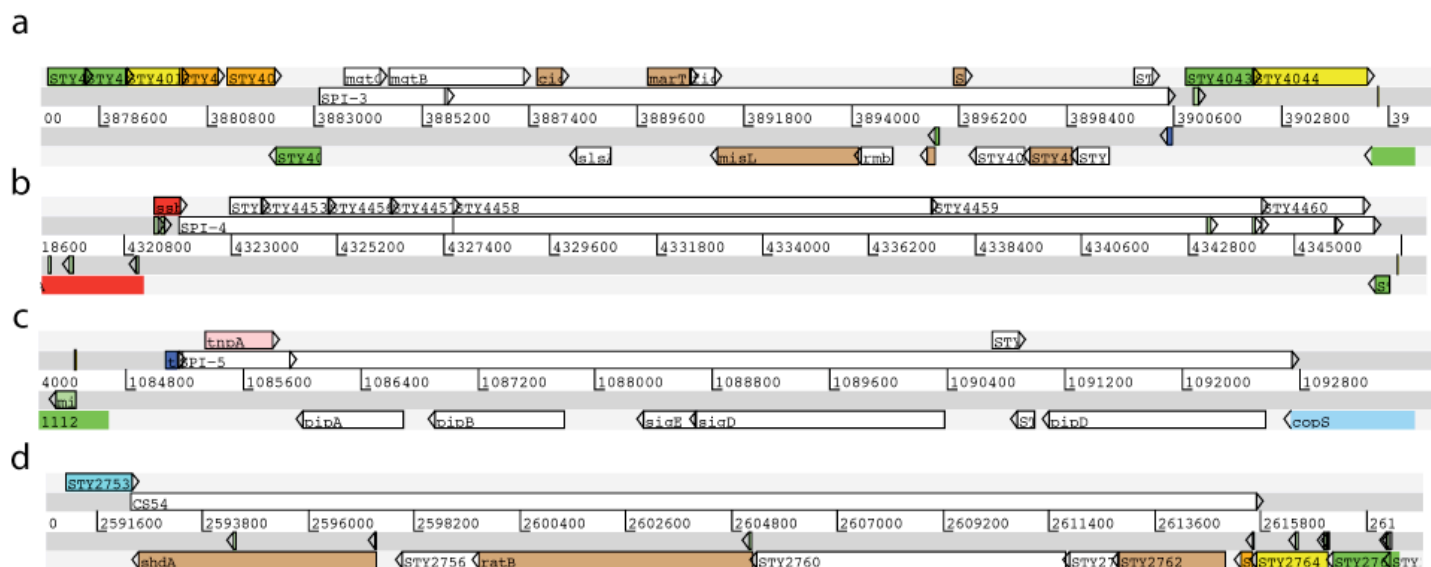


Figure 1.8 Genetic content of horizontally acquired regions in *S. Typhi* CT18.

(a) SPI-3 (b) SPI-4 (c) SPI-5 (d) CS54. Colour of ORFs represents functional class. Pseudogene, brown; virulence associated, white; outer membrane and structural, green; regulator, blue; conserved hypothetical, orange; stable RNA, red; yellow, metabolism; pink, phage/IS elements.

### 1.5.4 Adherence

The whole genome sequence and annotation of *S. Typhi* CT18 and Ty2 predict that 13 loci encode for fimbriae including one nucleating (*csg*) and 12 chaperone-usher fimbrial families [76]. Fimbriae are important for efficient adhesion to host-cells [123,124,125,126] and systemic persistence [127]. Comparison with orthologous genes identified putative pseudogenes in 5 of these loci (*sef*, *bcf*, *ste*, *stg* and *sth*) in *S. Typhi*, which are not predicted to be inactivated in *S. Typhimurium* [77]. *S. Typhi* harbours two fimbrial loci, *tcf* and *sta*, that are not encoded within the *S. Typhimurium* genome and a type IVb pili operon in SPI-7 that binds to the cystic fibrosis transmembrane conductance regulator [128,129].

Surface exposed structures that can interact directly with the immune system are candidates for exhibiting increased genetic diversity. However, 5 of the 13 putative

fimbrial operons encode potentially inactivating pseudogenes suggesting that a specific adhesion repertoire has been selected for and surplus adhesion factors selectively lost during *S. Typhi* evolution. This specificity supports the current model that a “stealth” invasion is required and prompts the hypothesis that *S. Typhi* preferentially targets a susceptible cell type that remains immunologically silent following invasion by the bacteria.

### 1.5.5 Motility

*S. Typhi* is monophasic (*fliC*), whereas *S. Typhimurium* is diphasic as it has the capacity to express two forms of flagellin, FljB and FliC. Flagellin may play a major role in invasive *Salmonella* pathogenesis (our unpublished data). There is a hypervariable region in flagellin protein and this harbors the antigenic epitope of the flagellin filament. Flagellin may interact directly with the host cells on the bacterial cell surface and it has been speculated that flagellin can be translocated to the host-cell cytosol via the SPI-1 TTSS [130].

### 1.5.6 Two-component regulators and virulence

An adaptive bacterial response is critical for the co-ordinated gene expression of virulence determinants during various stages of pathogenesis. Global regulators are conserved in ancestral *E. coli* yet have adapted to control essential virulence mediators in tandem through regulons. The two-component sensor, *phoPQ*, is a classic example of a conserved regulator controlling cellular  $Mg^{2+}$  homeostasis [131]. The host cell starves intracellular bacteria of  $Mg^{2+}$ , a signal that activates regulation of bacterial genes required to control  $Mg^{2+}$  content through PhoPQ. This system has evolved to regulate the expression of virulence genes required for intracellular

survival [132]. Mutations in the *phoPQ* operon inhibit the ability of *Salmonella* bacteria to survive in macrophages and increases susceptibility to acid pH [133], bile resistance [134] and antimicrobial peptides [135].

The outer membrane porin regulator, OmpR, is a transcriptional regulator conserved throughout enteric bacteria including closely related species *E. coli*, *S. Typhi* and *S. Typhimurium* [77,136,137]. OmpR responds to changes in osmolarity, via phosphoryl transfer from the histidine kinase sensor EnvZ, and is also highly conserved throughout Enterobacteriaceae. Transcription of both *ompC* and *ompF* in *E. coli* and *S. Typhimurium* is altered by changes in osmolarity [138,139], through OmpR, however, low pH medium also alters transcription of *ompC* and *ompF*, independently of changes in osmolarity [140,141]. *S. Typhi* OmpR does not reciprocally control *ompC* and *ompF* expression [142] and has incorporated into its regulon two porins, *ompS1* and *ompS2* that are not present in *E. coli* or *S. Typhimurium* [143,144]. The acquired regulation of these genes, and the alteration in controls of *ompF* from that in *E. coli* and *S. Typhimurium*, may be due to *S. Typhi* and its host-restriction. Such adaptation would limit the possible number of environmental changes that *S. Typhi* is exposed to during its lifestyle and may reduce the dynamic range of permeability required in its outer membrane provided by both the specific porin, OmpC and the more generalised porin, OmpF. OmpS1 and OmpS2 may provide a more specialised permeability required for survival within the host.

Transcription of *ompR* is autoinduced by acid shock and a mutation in *ompR* reduces the bacterium's capacity to withstand low pH conditions during stationary phase growth [145]. A null mutation in *ompR* does not alter expression of the master flagella regulator, *flhDC* [146], the curli, *csg* [147], and *agf* [148] fimbrial operons and the

SPI-1 master regulator, *hilA* [149], are regulated by OmpR. Both adhesion and invasion are required for colonisation of the host gastrointestinal tract suggesting that OmpR plays a crucial role in adaptive response during this disease phase. Furthermore, in *S. Typhi*, OmpR tightly regulates Vi antigen expression [116], which is thought to increase the virulence of *S. Typhi* [150,151,152]. Mutations in *ompR* and *envZ* attenuate *Salmonella*-induced filament formation in HeLa cells [153] and OmpR is a positive regulator of the anaerobically induced *tppB* gene [154], which encodes a major outer membrane permease protein. OmpR also regulates the *aas* gene that encodes a 2-acylglycerolphosphoethanolamine acyltransferase that is induced inside macrophages [155,156]. Growth within the SCV requires a functional OmpR to activate transcription of the SPI-2 endogenous regulatory locus *ssrAB*, which regulates expression of proteins required for macrophage killing and intracellular survival [157]. OmpR is a crucial global regulatory protein required to regulate gene expression in of genes required for adhesion, invasion and intracellular survival.

## 1.6 Regulation of bacterial gene expression

Bacterial gene expression is a two-part process involving transcription of messenger RNA (mRNA) from template DNA and the translation of the nascent mRNA to protein. Regulation of this mechanism is controlled initially by transcriptional activation or repression and secondarily by alternative strategies such as termination, riboswitching, ribozyme activity, RNA polymerase and ribosome stalling (attenuation), or antisense RNA.

### 1.6.1 Transcriptional activation

RNA polymerase is a quaternary structure of proteins that has the capacity to bind alternative sigma factors, thus directing gene expression by conferring different consensus sequence specificities [158]. *E. coli* and *Salmonella* are predicted to encode 7 alternative sigma factors, controlling transcription of nitrogen uptake [159], flagellar [159], stationary phase [160], nutrient limitation [161] heat shock [162] and housekeeping genes. There are many families of gene expression regulatory proteins (table 1.5) [163]. However, in this study we are mainly concerned with two-component systems and how they control their regulon.

Table 1.5 Major families of bacterial transcription factors [164]

The AraC family	
Examples	E. coli AraC, MelR, RhaS, RhaR, SoxS Many homologous to other organisms
Domain structure	N-terminal domain concerned with triggering by small ligand. C-terminal domain carries two helix-turn-helix motifs responsible for operator binding
Main properties	Transcription activators that overlap -35 region Bind to ~18bp in absence and presence of ligand
The LysR family	
Examples	E. coli LysR, OxyR, MetR, CysB Many homologues to other organisms
Domain structure	N-terminal domain carries helix-turn-helix motif responsible for operator binding C-terminal domain concerned with triggering
Main properties	Co-inducer responsive transcription activators Bind in absence and presence of ligand
The CRP family	
Examples	E. coli CRP and FNR Homologues in many other organisms
Domain structure	C-terminal DNA binding domain carries helix-turn-helix. N-terminal domain concerned with triggering
Main properties	Transcription activators Binding to target is ligand dependent Variety of promoter architectures
The MerR family	
Examples	E. coli SoxR and transposon encoded MerR Many homologues to other organisms
Domain structure	No evidence for domains N-terminal carries helix-turn-helix and C-terminal domain concerned with triggering
Main properties	Transcription activators Binding to target is ligand independent
The response regulator family	
Examples	E. coli NarL, UhpA, OmpR, PhoB Many homologues to other organisms
Domain structure	N-terminal domain (the response domain) triggered by phosphorylation. C-terminal domain for operator binding carries helix-turn-helix motif that binds DNA. There are two types of domain; the OmpR module found in OmpR and PhoB etc, and the LuxR module found in NarL, NarP and UhpA etc. The LuxR DNA binding module is also found in some activators (eg LuxR and MaltT) and in sigma factors.
Main properties	Transcription activators that bind a variety of positions in target promoters.
The sigma-54 bacterial enhancerbinding family	
Examples	E. coli FlhA, Klebsiella pneumoniae NifA and NtrC Homologues in many organisms
Domain structure	N-terminal domain responsible for triggering (some N-terminal domains are related to the response domains of the response regulator family). C-terminal part carries helix-turn-helix motif that binds DNA and a segment that binds E-sigma-54.
Main properties	Transcription enhancer-like proteins that can bind well upstream of target promoters.
The Lac repressor family	
Examples	E. coli LacI, GalR, PurR, CytR
Domain structure	N-terminal carries helix-turn-helix motif that binds DNA. C-terminal carries segment responsible for triggering
Main properties	Transcription repressors. Proteins bind as dimer but some can form tetramer.
The metJ repressor family	
Examples	E. coli metJ, phage P22, Arc and Mnt repressors
Domain structure	Single domain
Main properties	Transcription repressors. Contact DNA via beta-strand.

Two-component regulatory systems consist of a transmembrane sensing protein domain and a DNA-binding protein, which binds to specific sequences and directs RNA polymerase activity. This adaptive response mechanism permits bacteria to sense their environment and tightly regulate gene expression, properties that have facilitated the colonisation of various niches. Two-component phosphorylation systems were first characterised in bacteria and have since been identified in archaea and eukaryotes [165].

In prokaryotes, the system involves information transfer via the exchange of a phosphoryl group from a donor (sensing protein) to a receiver (transcriptional activator). A phosphate group is acquired by autophosphorylation, via formation of an ATP binding site in the cytoplasmic histidine residue. This occurs in response to conformational change of the sensing protein under specific conditions. Subsequently, the phosphate group is transferred to an aspartate residue in an acidic pocket of the receiver domain in the transcriptional activator (figure 1.9) [166]. The phosphotransfer enzyme activity is housed within the receiver domain leading to an unstable phosphorylation, whose half-life is dependent upon the configuration of the recipient protein. Donor domains may harbour phosphatase activities that control the phosphorylation status of their cognate recipient domains. Due to the conserved nature of two-component systems, cross reactivity and phosphorylation can occur under excess conditions [165].

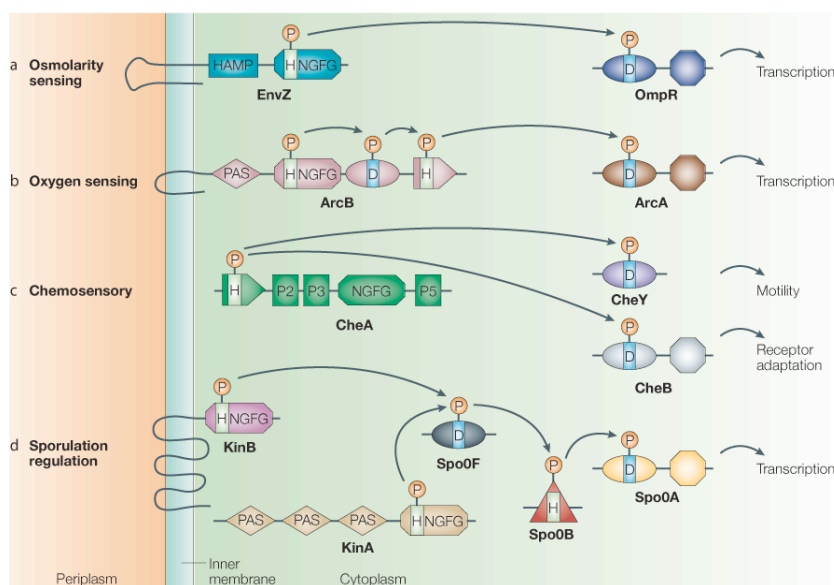


Figure 1.9 Some of the different combinations of histidine protein kinase (HPK) and aspartate response regulator (RR) domains in histidine–aspartate phosphorelay (HAP) systems [166].

(a) The EnvZ/OmpR pathway of *E. coli*, which is involved in regulating the expression of the two outer-membrane porins OmpF and OmpC. A membrane-bound HPK (EnvZ) controls the activity of the RR OmpR in response to changes in osmolarity. (b) The complex ArcB–ArcA HAP system of *E. coli*. The membrane-bound HPK ArcB senses changes in the redox state of components of the respiratory electron-transport chain through its PAS (PER, ARNT, SIM) domain. The phosphoryl group is then passed from the conserved His in the ArcB kinase domain to a fused RR domain, then to a fused histidine-containing phosphotransfer (HPT) domain and finally to a DNA-binding RR ArcA. ArcA regulates microaerophilic gene expression. (c) The chemosensory pathway of *E. coli*. The soluble HPK chemotaxis protein (Che)A has five domains per monomer that are designated P1–P5 from the N terminus to the C terminus. CheA senses changes through transmembrane chemoreceptors, which induce the trans-autophosphorylation of dimeric CheA on a His residue of the HPT domain. Two RRs compete for this phosphoryl group: CheY, a single-domain, motor-binding protein, which controls flagellar motor switching, and CheB, which controls the adaptation of the chemoreceptors. (d) Part of the complex system that regulates sporulation in *Bacillus subtilis*. A single-domain RR, Spo0F, is regulated by two HPKs, one of which has numerous transmembrane domains [134], the other of which is soluble with numerous PAS domains (KinA). Spo0F indirectly phosphorylates a DNA-binding RR, Spo0A, by way of a His residue in Spo0B. Throughout this figure, light-green rectangles highlight conserved, phosphorylatable His residues, light-blue rectangles highlight conserved, phosphorylatable Asp residues, and orange circles highlight phosphoryl (P) groups. NGFG represents the kinase domain and, with the exception of CheA, the conserved His residue that precedes the kinase domain is contained within the dimerisation domain. Despite being dimeric in nature, HPKs are shown here as monomers for simplicity, and the HAMP domain ('histidine kinases, adenylyl cyclases, methyl-binding proteins and phosphatases' domain) is a linker domain.

## 1.6.2 Post-transcriptional control

Various mechanisms exist in order to control protein abundance once a transcript has been synthesised. Nascent mRNA can interact with inhibitory proteins, ribosomes, antisense RNA, metabolites and itself, to regulate translation. These interactions are generally mediated by, or related to, the product of the translated protein.

## 1.7 DNA sequencing technologies

DNA sequencing technology has been exploited to determine the whole genome sequence of hundreds of prokaryote and eukaryote species. The availability of such sequence information has facilitated gene identification, transcriptomic studies and underpins experiments to link genotype to phenotype. Many novel DNA interrogation techniques have since been designed to exploit these sequence data on a high-throughput, genome-wide scale.

### 1.7.1 DNA microarrays

Microarrays consist of defined DNA sequences printed at high-density on a glass slide. They are used to interrogate vast number of nucleotide sequences by fluorescently labelling the DNA population prior to hybridisation [167] (methods have now been developed that are not dependent on DNA labelling). Microarray analysis relies on the complementary nature of DNA strands, much like northern and Southern blotting techniques. Comparative genome hybridisation (CGH), differential transcriptome and chIP-on-chip profiling can be performed using DNA microarrays to interrogate prokaryotic systems.

DNA expression arrays require the isolation of mRNA from the bacteria, reverse transcription of the mRNA to cDNA and coupling of a fluorescent or other type of label. For spotted arrays a control is also needed and both samples are labelled, usually either green or red. Samples are hybridised in the presence of a control and are excited by LASER during scanning. Subsequently, the fluorescence intensity of each spot is determined and normalised and the ratio of fluorescence can then be interpreted as differences in gene expression. This technique provides a vast amount of data for transcriptional differences across an entire genome. It is possible to use mRNA or genomic DNA (gDNA) as a reference to determine the fluorescence ratio.

Comparative genome hybridisation is very similar and compares labelled genomic DNA isolated from a target strain to genomic DNA from a sequenced isolate. This technique is effective but limited to measuring the presence or absence of sequences present on the array and cannot detect novel DNA sequences. Spots will also bind highly similar sequences that will represent false positives. However, this approach has been used effectively to identify genome variation and architecture between *Salmonella* serovars [89].

Enriched protein-bound DNA sequences obtained by immunoprecipitation can be labelled and compared to a control using microarray analysis. Recent increases in density permit genome tiling of sufficient sequence density to identify regulatory binding sites. This technique has been used recently to identify genes regulated by H-NS [168].

### 1.7.2 High-capacity DNA sequencing

Recent advances in DNA sequencing technologies has facilitated the determination of nucleotide sequence with a genomic read depth several orders of magnitude greater than was previously possible [169]. Several novel approaches have been developed including 454 (pyrosequencing) and Solexa or Illumina sequencing. Illumina sequencing involves sequencing millions of short (~36-55bp) reads using a slide based system for capturing DNA (figure 1.10). It is possible to computationally map each read to the previously sequenced genome as a reference and identify genomic DNA content. This method has been employed by Holt *et al* [87] to determine genome variation in 21 different *S. Typhi* isolates and more recently similar methods have been applied to sequence the transcriptome of *S. pombe* [170] and immunoprecipitated mRNA bound to the antisense RNA directing protein, Hfq [170].

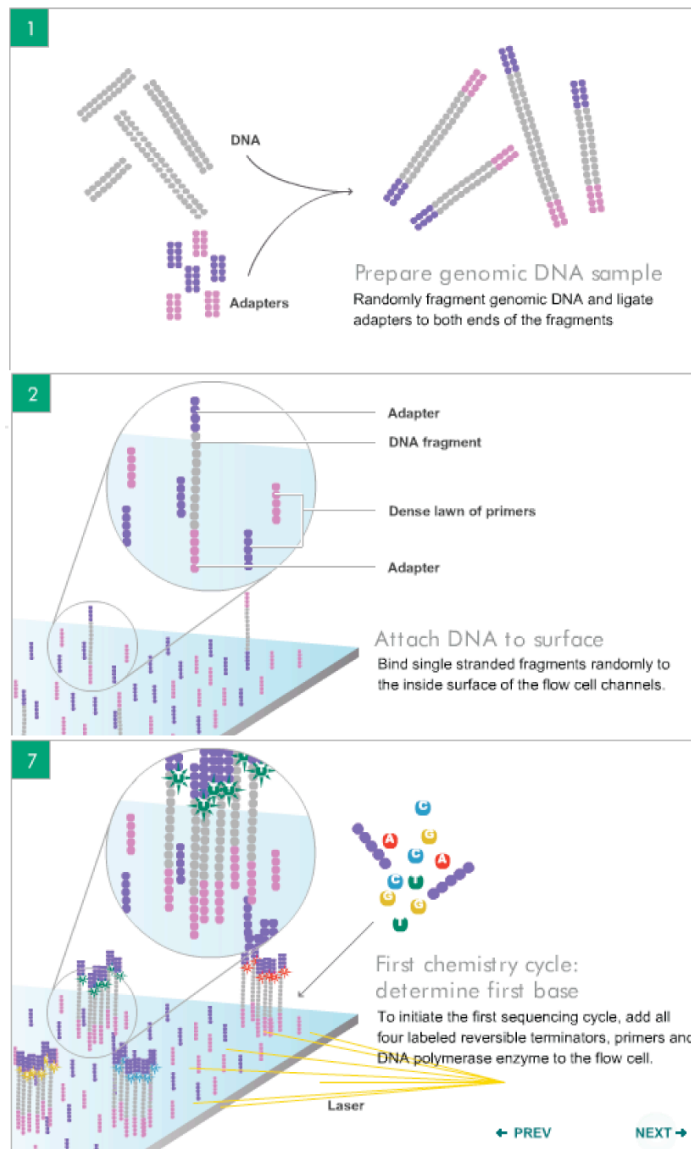


Figure 1.10 Schematic of Illumina sequencing technology (<http://www.illumina.com/>).

## 1.8 *In silico* - software and tools

High-throughput sequencing produces so much information that integrated genome software and bioinformatics tools are required for analyses and for data interpretation.

### 1.8.1 Artemis and Artemis Comparison Tool (ACT)

Artemis and ACT were developed for genome browsing and annotation and are important tools used in this study making it possible to annotate and manipulate data without detailed knowledge of script writing [171]. Sequenced transcriptome data was displayed in Artemis as a plot mapped back to the entire genome and it is possible to represent independent DNA strands as different coloured plots.

ACT is based on Artemis and is used for whole genome alignment and comparison for related species [172,173].

### 1.8.2 BLAST and other packages

Basic local alignment search tool (BLAST) is a tool provided by NCBI that finds regions of similarity in biological function between DNA sequences. It is possible to compare nucleotide, translated nucleotide and protein sequence data with every submitted sequence [174,175].

## 1.9 Aims of the Thesis

The aims of this thesis were to apply the new sequencing capacity and technology available at the WTSI to analyse the transcriptome of *S. Typhi*. To achieve this, a novel transcriptome sequencing approach, RNA-Seq, was developed based on Illumina sequencing linked to Artemis and ACT for analysis. The information gleaned from RNA-Seq and microarray analysis of *S. Typhi*, combined with proteomics, biochemistry and mutagenesis, was used to further define the genome annotation and the *ompB* regulon of *S. Typhi*.

## **2 Materials and Methods**

## 2.1 General chemicals, reagents and buffers

General laboratory chemicals were purchased from Sigma unless otherwise stated. Buffers were prepared as aqueous solutions in distilled water according to standard methods [176]. Solutions were sterilised by autoclaving or by filtration (Millipore, 0.22µm).

## 2.2 Bacterial strains

Bacterial strains in this study were a combination of laboratory strains and genetically modified *Salmonella* and *Escherichia coli* organisms. The strains from all studies are outlined in table 2.1.

## 2.3 Plasmids

Plasmids were isolated from cultures using miniprep plasmid purification kit (Qiagen) according to manufacturer's instructions (section 2.5.2) and are outlined in table 2.2.

Table 2.1 Strains used in this study

Strain	Species or serovar	Parent Strain	Genotype	Selection marker	Source	Reference
BRD948	S. Typhi	Ty2	aroC, aroD, htrA	none	Gordon Dougan	[177]
TT10	S. Typhi	BRD948	ompR::kan	aph	This Study	-
SL1344	S. Typhimurium	-		L-His	Rob Kingsley	[178]
RAK83	S. Typhimurium	SL1344	ompR::kan	aph	This Study	-
RAK103	S. Typhimurium	SL1344	STM1128-29::kan	aph	This Study	-
RAK105	S. Typhimurium	SL1344	STM1130-33::kan	aph	This Study	-
RAK113	S. Typhimurium	SL1344	phoN::CmR	CmR	Rob Kingsley	[122]
TT56.1	S. Typhimurium	SL1344	STM4464-4467::kan	aph	This Study	-

T53.8	S. Typhi	BRD948	ompR3XFLAG	None	This Study	-
TOP10	E. coli	E. coli	lacZΔM15	None	Invitrogen	-
Ctr18λpir	E. coli	E. coli		None	Keith Turner	[179]

Table 2.2 Plasmids used in this study

Plasmids	Function	Source	Reference
pCR2.1	Cloning vector	Invitrogen	-
pWT12	Suicide Vector	Keith Turner	-
pKD13	Kanamycin template	Anne Bishop	[180]
pKD46	Red recombinase	Anne Bishop	[180]
pKD3	Chloramphenicol Template	Stephen Baker	[180]
pBAD202	Protein expression and tagging	Invitrogen	-
p/c/1	ssaG::lacZ reporter	Derek Pickard	[181]

## 2.4 Microbiological media and techniques

### 2.4.1 Bacterial culture

#### 2.4.1.1 Standard methods

Bacterial strains were grown in Luria-Bertani (LB) medium, 1% w/v tryptone (Beckton Dickinson), 0.5% w/v yeast extract (Beckton Dickinson) 0.5% NaCl in ddH<sub>2</sub>O, with antibiotic selection when required, from a single colony. Stocks of bacteria were stored at -70° C in cryovials in a minimum of 10% glycerol (v/v). Microbiological agar plates or broth media, 10ml, were inoculated using a plastic loop (VWR - Leicestershire) from a cryovial. Plate cultures were stored at 4°C for up to a month wrapped in Parafilm. Broth cultures grown from plate cultures were inoculated from a single colony using a plastic loop.

Bacterial strains were routinely cultured in LB medium. Bacterial strains were routinely grown on LB agar plates, 1.5% Bacto-agar (Beckton Dickinson) in LB broth. Liquid cultures were grown in 50ml Falcon tubes or 20ml universal containers in a New Brunswick shaking incubator (225 rpm) at 37°C unless otherwise stated. LB-agar plate cultures were incubated statically at 37°C. When required, media were supplemented with antibiotics at the concentrations and solvents shown in table 2.3. All antibiotics were filter sterilised at the requisite stock concentration unless dissolved in 100% ethanol where required.

Table 2.3 Antibiotic concentration and solvents

<b>Antibiotic</b>	<b>Solvent</b>	<b>Concentration (<math>\mu\text{gml}^{-1}</math>)</b>
Kanamycin	Water	40
Chloramphenicol	Ethanol	30
Ampicillin	Water	100

For the culture of *S. Typhi* BRD948 and derivatives, an auxotroph of *S. Typhi* Ty2, all media were supplemented with 1% (v/v) aro mix (40mgL<sup>-1</sup> phenylalanine, 40mgL<sup>-1</sup> tryptophan, 10mgL<sup>-1</sup> *para*-amino benzoic acid and 10mgL<sup>-1</sup> dihydroxybenzoic acid) and 1% (v/v) tyrosine mix (40mgL<sup>-1</sup> tyrosine disodium salt).

#### 2.4.1.2 Minimal media

Minimal medium, 42mM Na<sub>2</sub>HPO<sub>4</sub>, 24mM KH<sub>2</sub>PO<sub>4</sub>, 9mM NaCl, 19mM NH<sub>4</sub>Cl, 0.2mM MgSO<sub>4</sub>, 0.1mM CaCl<sub>2</sub>, 1.0% glucose (w/v), was used as defined medium and supplemented as required. The carbon source is glucose and the nitrogen source is NH<sub>4</sub>Cl. These were replaced with other sources for some minimal medium assays.

#### 2.4.2 Harvesting of bacteria from broth cultures

Bacterial cultures smaller than 50ml and greater than 1.5ml were harvested by centrifugation at 4600 rpm for 10 mins (Sorvall legend RT). Cultures smaller than 1.5ml were harvested in a bench top microcentrifuge at 13,000 rpm for 3 mins.

#### 2.4.3 Growth curves

Growth curves of bacterial strains in this study were performed to determine if growth of recombinant strains was comparable to the wild-type. Overnight culture was diluted 1/100 into 30ml of room temperature LB in a 50ml Falcon tube and grown at

37°C. Samples were taken hourly and the OD<sub>600nm</sub> measured in 1ml cuvettes using sterile LB as the reference. If the mutation was known to affect the outer membrane or capsule serial dilutions were performed and density calculated to determine if there was any difference between the number of colony forming units (CFU) and OD<sub>600nm</sub> readings.

## 2.4.4 Identification of bacteria

The identification of *Salmonella* serovars was performed using standard laboratory slide agglutinations according to manufacturer's instructions. O antigen and Virulence (Vi) antigen agglutination tests were used to confirm the serology for each serovar. Briefly, 10µl of antisera was pipetted into a Petri dish and mixed with approximately half a loop of fresh bacterial plate culture. A positive result was recorded if bacteria clumped together in discrete groups within 60 seconds of mixing. Antisera for alternative antigens were used as a negative control.

## 2.5 Molecular biology techniques

### 2.5.1 Bacterial genomic DNA extraction

Genomic DNA was isolated using Wizard SV Genomic DNA (Promega) kit according to the manufacturer's instructions. Briefly, cultures were harvested, the supernatant removed and the cells resuspended in 600µl of nuclei lysis solution then incubated at 80°C for 5 mins. RNA was then digested using 3µl of RNase at 37°C for 15mins and the sample cooled to room temperature. Proteins were then precipitated with 200µl of protein precipitation solution and vortexed vigorously for 20sec before incubating on ice for 5mins. Precipitated proteins were then pelleted by 3 mins

centrifugation at 13,000xg and the DNA containing supernatant transferred to a clean tube containing 600µl of isopropanol to precipitate the DNA. DNA was then harvested by centrifugation (13,000xg for 2mins) and the supernatant discarded. The pellet was then washed with 70% ethanol (ice-cold) and harvested once more. The supernatant was discarded and the pellet allowed to dry before being resuspended in nuclease free water to the desired volume.

## 2.5.2 Plasmid DNA extraction

Plasmid DNA was isolated from 5ml of overnight culture using Qiagen miniprep kit (Qiagen) according to the manufacturer's instructions. Briefly, the bacterial culture is harvested and lysed by high alkaline conditions. The plasmid DNA is then adsorbed on a QIAprep membrane, washed then eluted with nuclease free water.

## 2.5.3 Agarose gel electrophoresis

Genomic DNA, plasmid DNA, PCR products, restriction enzyme digests and RNA were routinely separated by agarose gel electrophoresis. Agarose concentration varied from 0.8% to 2% (w/v) depending on the predicted size of DNA fragments. Agarose (Sigma ultra pure grade) was suspended in TAE at the requisite concentration and boiled. Cooled, molten agarose was then supplemented with Ethidium Bromide (Sigma) to a concentration of  $1\mu\text{gml}^{-1}$  and poured into a mould containing the requisite number of wells. DNA samples were mixed with DNA loading buffer to a ratio of 4:1 and loaded into the wells of the TAE submerged, solid agarose gel. All samples were compared to Hyperladder I™ (Bioline) or Hyperladder IV™ depending on the estimated size of the products. Application of a constant voltage across the gel

resulted in electrophoresis of the DNA and was visualised on a UV transilluminator. Photographic records were taken using UVtech hardware and software.

## 2.5.4 Quantification of DNA and RNA by spectroscopy

DNA and RNA preparations were quantified using the ND-1000 (NanoDrop Technologies). Briefly, the NanoDrop method determines absorbance at 260nm. A conversion factor of 50 for DNA and 40 for RNA for every unit of absorbance represents an estimated  $50\mu\text{g}\mu\text{l}^{-1}$  and  $40\mu\text{g}\mu\text{l}^{-1}$ , respectively.

## 2.5.5 DNA manipulation techniques

### 2.5.5.1 DNA restriction

New England Biolabs supplied all restriction endonucleases unless otherwise stated. Determination of appropriate endonuclease was done *in silico* using Macvector© (Accelrys). Restriction enzyme digests were performed according to the manufacturer's instructions. Briefly, 20 units of enzyme were used in a reaction mixture containing appropriate buffer, 10 $\mu\text{l}$  of DNA and incubated at 37°C for 1-2 hours in a water bath. Digests used for sub-cloning were dephosphorylated with 2 $\mu\text{l}$  Antarctic Shrimp Phosphatase (Invitrogen) to prevent re-ligation.

### 2.5.5.2 Oligonucleotides

All oligonucleotides used in this study were synthesised by Invitrogen™ and resuspended in nuclease free H<sub>2</sub>O (Ambion) to a concentration of 100pmol $\mu\text{l}^{-1}$ . Table 2.4 outlines the sequences requested for each oligonucleotide. Primers for general PCR were between 18 and 25 nucleotides. For design of general PCR primers each

cytosine or guanine was assigned four units and each adenine or thymine was assigned 2 units. Optimal primers contained a total of 64 units with a 5' and 3' terminal cytosine or guanine. Using Macvector© (Accelrys) software and the DNA sequence file for each *Salmonella* serovar, *in silico* analysis was performed to check for predicted non-specific products. Mutagenesis primers were between 48 and 55 bases long. Real Time PCR primers were between 18 and 25 bases designed by Profinder 2.4 (Roche). Overlap extension PCR primers were between 25 and 69 bases.

Table 2.4 Primers used in this study

**cDNA sequencing mechanism primers**

AACATCTGCAAG[(N)19]CAGCGACGCATC[(N)5]	DNA primer
---------------------------------------	------------

GAUGCGUCGUG	RNA primer
-------------	------------

<b>Mutagenesis</b>	Strain
--------------------	--------

GGATCGTCTGCTGACCCGTGAATCTTTCCATCTCATGGGTGTAGGCTGGAGCTGCTTC	delta ompR
--	------------

GTCTGAATATAACGCGGATGCGCCGGATCTTCTTCCACATTCCGGGGATCCGTCGACC	delta ompR
--	------------

AAAACGGAGTAAACTTCAAAATATATAAGGCGGAATGGGTGTAGGCTGGAGCTGCTTCG	RAK103
---	--------

GTCACCGTGTGCTGTGTCGGTATAGCGTGGTATCATGAAAATTCCGGGGATCCGTCGACC	RAK103
--	--------

TCACTAATGATGAAGCTTTACTCCAGTTGTATTTCTTCGCGTGTAGGCTGGAGCTGCTTCG	RAK105
---	--------

AGCGCCACCGGCCAATAACACCACCATCCGGCTTTAAATTATTCCGGGGATCCGTCGACC	RAK105
--	--------

TCAAGCGATTAATGCATGATTTACTCATCGCAACCGTTTATTCCGGGGATCCGTCGACC	TT56.1
---	--------

ACTCCTTCTTTATTCTTGTAATTATGTAAAAGGTATAATGGTGTAGGCTGGAGCTGCTTCG	TT56.1
---	--------

**Mutagenesis checking primers**

TGATGACGATATGCGTCTGCG	delta ompR
-----------------------	------------

GGATCTTCTTCCACCATAACGGC	delta ompR
-------------------------	------------

AACACTCCACAACATAATAT	RAK103
----------------------	--------

TTGTCGCCCCGCGCGTTGGC	RAK103
----------------------	--------

TGAAATTATGCACCATAAGA  
TATCCAGTCTACATAAGCGC  
TACTCAGCTCTGTTTTGGGA  
CACGCTAAATATTTGCAGCC

RAK105  
RAK105  
TT56.1  
TT56.1

# **Real Time PCR**

GGATCTCGGTGGTGCTCA  
GGCACACAGAAGCTGATGC  
CTCTATCCGTTCCCGCATAA  
ATGGACGTGAGCGTATGGTT  
GCGTCACCCCTGAAGAGAT  
CCTGTTTTAATTCGTCGCTCA  
AAGCCTTGCCCCTATGCT  
CCCTCGACTTTTACGCTGAC  
TCGTCAGTGAGCGTCTGG  
GGCGTTATCAGCGACCTTAAT  
CGTACCTGCGGATGACTTAAC  
TTGCATCTAAGTGCGCAAAG  
TGCGGCGCTTAATGATTT  
CGTATGTTTTTACATAGCGTAATGTTT  
CCTGAAGAATCAGTGAACAGTCA  
AGCCGGATCATCATTACCTG  
GCACCAGCCTCTCTTCTCC  
TTAACGATGAAGCCGTCCAT  
ACGTATGGGTTGACTACCGTTT  
CGTAGCTGGAGCTGTAGTCGT  
TTTCTGGCGCTATTCCTGAC  
AACGGAGATCATCTGGAAGG  
GTGATATGGCTGCATGGTCTT  
CGCAGAGCGAATGAAAGATT  
TTTCAACTCCGCTATCACCA  
CTACGGGCAGAAGACAGGTT  
AGCGTGAGCGGTGAAAAC  
GTAACCGTCGCCGTTCTG  
TGGTATCGACAAAACCAAAGC  
AGGCGCGTTTTAAATTCAGTG  
TCAGCTACTGATTGAAAGTTATACCAA  
GCGTTGAGAATCCAGCAGTT  
AACTGTTGCCATCCTTCAA  
GAAGCTGGCAATGGTCTTGT  
CGGTATAGACGAGGTAAAAATCG  
ATTCAACTGCCAGAGGAAGC

# **Gene**

STY0002  
STY0002  
STY0779  
STY0780  
STY2281  
STY2281  
degQ  
degQ  
STY3913  
STY3913  
STY3941  
STY3941  
waaY  
waaY  
phoN  
phoN  
STY0775  
STY0775  
STY1002  
STY1002  
STY1287  
STY1287  
STY1523  
STY1523  
fliC  
fliC  
ompC  
ompC  
STY2806  
STY2806  
slsA  
slsA  
STY4402  
STY4402  
tviB  
tviB

# **Universal Probe**

82  
82  
4  
4  
70  
70  
27  
27  
19  
19  
65  
65  
62  
62  
77  
77  
40  
40  
62  
62  
23  
23  
62  
62  
61  
61  
40  
40  
40  
40  
4  
4  
29  
29  
23  
23

# **Overlap extension PCR**

CGTCAGGCAAACGAACTGCC	5' to 3' ompR bases
CCGTCATGGTCTTTGTAGTCTGCTTTAGAACC	(364:383)
GTCCGGTA	Full reverse primer
GACTACAAAGACCATGACGGTGATTATAAAG	sequence 5' to 3'
ATCATGATATCGATTACAAGGATGACGATGA	Concatenated primers
CAAGTAGGTACCGGACGGTTCTAAAGC	are 5' to 3' Forward
	(1:69 FLAG + 1:20)
CGAAACGCAGGCGGCACG	Reverse for envZ is 5'
	to 3' (213:230)

### 2.5.5.3 Polymerase chain reaction (PCR)

PCR was used to amplify specific regions of DNA. For general PCR amplification of template DNA for colony screening or molecular cloning, Supermix (Invitrogen) was used. For high fidelity PCR, PFU polymerase (Promega) was used. For real time PCR Taqman Master Mix (Roche) was used. All PCR reactions were done in 0.2ml thin wall PCR tubes (ABgene) on a DNA engine DYAD thermal cycler (MJ Research) except for real time PCR where the Applied Biosystems ABI7900 was used. Negative controls were the same reaction without the addition of DNA template.

#### 2.5.5.3.1 PCR supermix (Invitrogen)

DNA amplifications using PCR Supermix™ were performed as per manufacturer's instructions in a total volume of 25µl. Each reaction mix contained 22.5µl of Supermix, 0.5µl of each forward and reverse primer and 10-100ng of template DNA. Template DNA was omitted for all PCR reactions as a negative control.

DNA was amplified using the cycle parameters of one stage at 95°C for 120s then 30 cycles of 95°C for 30s, 55°C for 30s† and 72°C for 90s†† followed by an elongation step at 72°C for 300s.

† Annealing temperature is dependent on melting temperature of primers.

†† Elongation step was altered depending on the predicted size of product. 60s/kb is general rule.

#### 2.5.5.3.2 PFU polymerase high fidelity PCR

For PCR amplicons where high fidelity of the template is required the PFU enzyme is used. This dual enzyme is a 5' to 3' polymerase and 3' to 5' proof reading exonuclease. The following components and the final concentrations were combined in a total volume of 25µl: Pfu DNA polymerase 10x buffer, 10mM dNTP mix, 200µM each, upstream primer, 1.0µM, downstream primer, 1.0µM, DNA template <0.5µg, Pfu DNA polymerase, 0.625u. Template DNA was omitted for all PCR reactions as a negative control.

DNA was amplified as in 2.5.5.3.2

#### 2.5.5.3.3 Real time PCR

To determine DNA and cDNA template copy number fluorescent probes were used. Briefly, to make a probe the reverse complement DNA is ligated to a fluorophore and a quencher that anneals to the target DNA during PCR cycles. The Taq polymerase cleaves the quenching substrate during polymerisation of dNTPs. The fluorophore is then excitable by its specific frequency and the fluorescence is measured. When this fluorescence is plotted on a graph an exponential phase can be identified and related to the cycle number of the PCR called the cycle threshold,  $C_t$  value.

We used the Universal Probe Library (Roche) and derived amplicons and primers by submitting the target DNA sequence to ProbeFinder version 2.4 (<https://www.roche-applied-science.com/sis/rtpcr/upl/index.jsp?id=UP030000>).

#### **2.5.5.3.4 Overlap extension PCR**

Overlap extension PCR was used to construct large DNA fragments from two smaller DNA fragments by a ligase-independent methodology. For example, to add a peptide to the end of protein, overlap extension PCR incorporates the existing sequence at the start of a primer and appends the designed peptide sequence at the end of the primer. During PCR this single strand of DNA is extended by Taq polymerase and included in the amplicon. A further round of PCR was then used to join two such amplicons if they include a 20-25bp homologous region. This amplicon is cloned and sequenced as described in 2.5.5.5 and 2.5.7.

#### **2.5.5.4 Purification of PCR products**

PCR amplicons required for further applications were purified to remove excess primer and nucleotides using a PCR purification kit (Qiagen) according to the manufacturer's instructions. Products excised from Agarose gels were purified using a Gel Extraction kit (Qiagen) according to the manufacturer's protocol. Briefly, the agarose gel was dissolved and the DNA bound to a membrane. This was then washed and eluted in nuclease free water.

#### **2.5.5.5 Cloning**

##### **2.5.5.5.1 Cloning PCR amplicons**

PCR amplicons were cloned for further sequencing or subcloning using the TOPO® TA Cloning Kits (Invitrogen) according to the manufacturer's instructions. Briefly, 2µl of fresh PCR, 1µl of salt solution (Invitrogen) and 1µl of vector mix in a final

volume of 6µl were incubated at room temperature for five minutes. DNA was then transformed as described in 2.5.8.1.

#### 2.5.5.5.2 Sub-cloning sequences into pWT12

Rapid DNA Ligation Kit (Roche) is used to ligate previously cloned sequences into the suicide vector pWT12. Both plasmids were digested with restriction endonucleases XbaI and SacIII, these sites flank the cloned amplicon and also flank the p15A in the suicide vector. The fragments were separated and purified from an agarose gel. Ligated DNA was then transformed into *E. coli* host strain C118λpir [179] as described in 2.5.8.2. Constructs were then isolated, screened and sequenced as described previously.

### 2.5.6 Precipitation of DNA

To precipitate DNA, sodium acetate is added to a final concentration of 300mM. 2.5 volumes of 96-100% ethanol is added, mixed well and centrifuged for 15 mins at a minimum of 13,000rpm at 4°C. The DNA pellet is washed with 70% ethanol at 0°C, spun at 13,000 rpm. The supernatant is discarded and the pellet air-dried.

### 2.5.7 DNA sequencing

#### 2.5.7.1 Capillary-based sequencing

All capillary-based sequencing was done by Team 41, Wellcome Trust Sanger Institute (WTSI), however all experimental design and sample preparation was done by Tim Perkins.

DNA sequencing was done using big dye terminator reactions (PerkinElmer/ABI) in 0.2ml thin wall PCR tubes. 100ng of plasmid DNA was added to 4pmol of M13 forward primer, 4pmol of M13 reverse primer and 4µl of big dye reaction mixture in a total volume of 10µl. Dideoxynucleotides were incorporated by PCR during 35 cycles of 96°C for 10s, 50°C for 5s and 60°C for 240s. DNA was then precipitated using the method previously described. The air-dried pellets were loaded onto an ABI 3700 capillary sequencer. Sequence reads were checked by eye using DNA visualisation software MacVector (Accelrys).

### **2.5.7.2 Illumina sequencing**

All experimental design and biological sample preparation was done by Tim Perkins, however, Illumina sequencing was submitted to the project co-ordinator Theresa Feltwell, Team 81 (WTSI) and libraries were prepared under the direction of Mike Quail (WTSI). Libraries were then submitted to the Illumina sequencing team (WTSI) and the quality-controlled data sent to a repository for further analysis.

Sequencing libraries for the Illumina GA platform were constructed by shearing the enriched cDNA by nebulisation (35psi, 6min) followed by end-repair with klenow polymerase, T4 DNA polymerase and T4 polynucleotide kinase (to blunt-end the DNA fragments). A single 3' adenosine moiety was added to the cDNA using klenow exo- and dATP. The Illumina adapters (containing primer sites for sequencing and flowcell surface annealing) were ligated onto the repaired ends on the cDNA. Gel-electrophoresis was used to separate ligated cDNA fragments from unligated adapters by selecting cDNA fragments between 200-250 bps in size. cDNA fragments were recovered by gel extraction at room temperature to ensure representation of AT rich sequences. Libraries were amplified by 18 cycles of PCR with Phusion polymerase.

Sequencing libraries were denatured with sodium hydroxide and diluted to 3.5 pM in hybridisation buffer for loading onto a single lane of an Illumina GA flowcell. Cluster formation, primer hybridisation and single-end, 36 cycle sequencing were performed using proprietary reagents according to manufacturer's recommended protocol (<https://icom.illumina.com/>).

The efficacy of each stage of library construction was ascertained in a quality control step that involved measuring the adapter-cDNA on a Agilent DNA 1000 chip. A final dilution of 2nM of the library was submitted to the sequencing machine.

## 2.5.8 DNA transformation

### 2.5.8.1 Transformation of chemically competent cells

#### 2.5.8.1.1 Preparation of chemically competent cells

Cultures were grown to an OD<sub>600</sub> of 0.4 and chilled on ice for 5min. Cells were harvested by centrifugation and resuspended in 40ml of ice cold Tbf1 (30mM of potassium acetate, 100mM potassium chloride, 10mM calcium chloride, 50mM manganese chloride, 15% (v/v) glycerol) then incubated on ice for 5min. Harvested cells were resuspend in 4ml of ice cold Tbf2 (10mM MOPS, 75mM calcium chloride, 10mM potassium chloride, 15% glycerol) and incubated on ice for 15mins. Cells were either used immediately in 100µl aliquots or snap frozen (dry ice and methanol bath).

#### 2.5.8.1.2 Chemical transformation

Frozen cells were thawed on wet ice and 2µl of plasmid DNA was added to one aliquot of cells for 30 minutes. Cells were then placed in a water bath at 42°C for 45 s

and returned to ice for 2 minutes before the addition of 400µl of recovery media SOC (10mM MgCl<sub>2</sub>, 10mM MgSO<sub>4</sub> and 20mM glucose in LB). Cells were incubated at 37°C for 1-2 hours. 100µl and 20µl aliquots were plated onto selective agar plates and incubated overnight at 37°C. An aliquot of cells without addition of DNA was used as a negative control.

## **2.5.8.2 Transformation by electroporation**

### **2.5.8.2.1 Preparation of electrocompetent cells**

Electroporation was used for transformation of linear and plasmid DNA into *Salmonella* or *E. coli*. Bacterial cultures were grown overnight under selection in LB, diluted 1/100 into 30ml of LB and grown to OD<sub>600</sub> = 0.3 at 37°C. Strains containing heat sensitive plasmids were grown at 30°C. Cells were harvested at 4°C and washed three times in 30ml of 0°C sterile 10% glycerol. The pellet was resuspended in 0°C 10% glycerol to 1/1000 of the culture volume, 300µl, to give a cell density of circa 10<sup>10</sup> CFU/ml.

### **2.5.8.2.2 Electroporation**

Electroporation cuvettes (Equibio) were chilled on ice for 15 mins before the addition of 60µl of competent cells. 10-100ng of plasmid DNA or 100-1000ng of linear DNA was added and the total volume made up to 66µl. The cuvettes were dried and immediately electroporated using a Bio-Rad Gene Pulser set at 25µF, 600Ω and 2.4kV. Room temperature SOC, supplemented with Aro mix for BRD948, was added to the cuvettes immediately after electroporation and cells were incubated shaking at

37°C for 1-2 hours. Cells were then incubated overnight on selective LB agar plates. Cells with the addition of 6µl of nuclease-free water were used as a negative control.

### **2.5.8.3 Screening for transformed DNA**

#### **2.5.8.3.1 Screening for non recombinant DNA**

Colonies were grown up individually in 10ml of LB and plasmid DNA was isolated using methods previously described and DNA was screened by restriction endonuclease profiles by agarose gel electrophoresis.

#### **2.5.8.3.2 Screening for recombinant DNA**

Colonies were initially screened by PCR with primers designed upstream and downstream of the predicted recombination site. Further screening was carried out on positive clones using primers internal, both forward and reverse strand, to the transformed DNA in combination with the upstream and downstream primers.

### **2.5.9 Bacterial conjugation**

The donor and recipient strains were grown overnight in 10ml selective LB at 37°C. Cultures were then mixed and grown statically for 3 hours in a water bath at 37°C. Cells were then plated on agar plates selecting for both the donor and recipient. Colonies were screened as previously described in 2.5.8.3.

## **2.5.10 Transduction**

### **2.5.10.1 Infection of donor strain with P22 phage**

The P22 phage lysate 10-fold dilutions were prepared to a minimal dilution of 1/1000. 10µl of each dilution was mixed with 100µl overnight culture of the donor strain and incubated at 37°C for 20mins. Liquid top agar (0.75% agar (w/v), LB) less than 45°C was added to the broth, poured over a warm LB agar plate and incubated at 37°C until confluent plaques formed. Optimal titre was generally when confluent plaques formed by 4 hours.

### **2.5.10.2 Isolation of P22 phage containing genomic DNA**

Top agar was scraped off the LB with a glass slide and collected in a tube with 3ml of LB. 100µl of chloroform is added to lyse the cells, vortexed and incubated at room temperature for 20mins. The supernatant was then collected after centrifugation and sterilised with 50µl of chloroform. Lysates were stored at 4°C.

### **2.5.10.3 Infection of recipient strain with P22 phage**

The lysate from 2.5.10.2 was serially diluted as in 2.5.10.1 and mixed with 100µl of an overnight culture of the recipient strain and incubated statically at 37°C for 15mins. 1ml of LB supplemented with 10mM ethylene glycol tetraacetic acid (EGTA) was then added and incubated at 37°C for 1hour. Cells were harvested and resuspended in 100µl and plated onto selective LB agar supplemented with 10mM EGTA and incubated overnight at 37°C. P22 lysate was plated as a negative control. Colonies were screened as previously described.

## 2.5.11 Mutagenesis

### 2.5.11.1 Red recombinase mutagenesis

Allelic exchange used in this study was based on the method described by Datsenko and Wanner [180]. The basis of this methodology is as follows. The temperature sensitive plasmid, pKD46, carries an arabinose inducible lambda Red recombinase gene, and confers ampicillin resistance with  $\beta$ -lactamase. The recombinase mediates efficient recombination between short homologous sequences flanking an antibiotic cassette that is amplified by extension PCR. Briefly, pKD46 was transformed into the strain to be mutated and grown at 30°C. Electrocompetent cells were made as previously described with the supplementation of arabinose to a final concentration of 0.2% (w/v) one hour after dilution.

Homologous recombination was used to insert the cassette and remove the target region of DNA, to do this two primers were designed to append these homologous sequences. Firstly, the forward primer, is designed to encode 45-51bp upstream of the target DNA sequence and a region homologous to the first 18-20 bases of the resistance cassette. The reverse primer encodes homologous regions to the opposite end of the target DNA and is reverse complemented. Subsequently, PCR was used to amplify the cassette from plasmid DNA, either pKD3 (chloramphenicol) or pKD13 (kanamycin) using these primers. The resulting amplicon encoded the cassette with each end containing regions homologous to each end of the target DNA. This amplicon was then purified and electroporated as previously described. Cells were plated out and colonies were screened by colony PCR as previously described in

2.5.8.3.2. If no colonies grew the remaining broth left at room temperature overnight was plated as before and colonies screened.

The genotype was screened by PCR and desired colonies were grown in selective LB at 43°C, a non-permissive temperature for replication of pKD46. Broth was plated as previously described and clones that did not grow on ampicillin plates were maintained.

### **2.5.11.2 FLP recombinase mutagenesis**

To remove the antibiotic resistance cassette, in order to generate an unmarked, non-polar mutation, FLP recombinase recognition sites were incorporated into the flanking regions of the resistance gene. The heat sensitive, ampicillin resistant and FLP recombinase expressing plasmid, pCP20, was transformed into the mutant and plated onto ampicillin. Colonies were selected if they were sensitive to the previously resistant antibiotic and grown at 43°C overnight to inhibit replication of pCP20. Clones sensitive to ampicillin after heat treatment were screened as previously described.

### **2.5.11.3 Allelic exchange mediated by suicide vector constructs**

The vector unable to replicate in *Salmonella* (suicide vector) was transferred to the host strain by bacterial conjugation as described in 2.5.9. Appropriate isogenic hosts were chosen. The recombinant strain was then grown overnight and diluted 1/10 into LB supplemented with chloramphenicol and incubated at 37°C for 10 minutes. Chloramphenicol is bacteriostatic so any cells having lost the suicide vector do not grow. Ampicillin was then added to a final concentration of 500 µgml<sup>-1</sup>, as it only

lyses growing cells, and incubated until cells were lysed. 1ml of LB was then washed in fresh LB twice and resuspended in 200 $\mu$ l. This was then plated onto LB without NaCl and supplemented with sucrose and grown overnight. Colonies were screened as previously described.

## 2.5.12 RNA methodologies

### 2.5.12.1 RNA stabilisation

RNA was stabilised using RNAlprotect Bacteria Reagent (Qiagen) according to the manufacturer's instructions. Briefly, 1 volume of culture was diluted in 2 volumes of RNAlprotect and vortexed to precipitate RNase enzymes. The bacteria were harvested at 4600rpm for 25 minutes in Sorvall bench top centrifuge, supernatant discarded and pellet air-dried.

### 2.5.12.2 RNA isolation

#### 2.5.12.2.1 RNA isolation for microarray

RNA was isolated from bacterial cultures using RNeasy RNA purification kit (Qiagen) according to the manufacturer's instructions with the following modifications. A concentration of 1mgml<sup>-1</sup> of lysozyme was used and RNA eluted using 30 $\mu$ l of DEPC treated water. RNA was quantified as previously described. Briefly, cells were lysed in TE containing 1mgml<sup>-1</sup> lysozyme and incubated for 10mins with occasional vortexing. Cells were then lysed in buffer RLT, mixed with ethanol and bound to a membrane. The membrane was then washed and RNA eluted in DEPC water.

#### 2.5.12.2.2 RNA isolation for Illumina sequencing

RNA was isolated from bacterial cultures using SV RNA isolation purification kit (Promega) according to the manufacturer's instructions with the following modifications. A concentration of  $1\text{mgml}^{-1}$  of lysozyme was used and RNA eluted using  $100\mu\text{l}$  of DEPC treated water. RNA was quantified as previously described.

#### 2.5.12.2.3 rRNA removal

The 16S and 23S rRNA were removed from the total RNA population, from a final concentration of  $0.83\text{ mgml}^{-1}$  by either oligo depletion (MicrobExpress, Ambion) according to manufacturers instructions or by incubating  $10\mu\text{g}$  RNA/Unit of Terminator Exonuclease (Epibio) for 60mins at  $30^{\circ}\text{C}$ . Briefly, to oligo deplete the rRNA, capture oligos were annealed to 16s and 23s sequences then annealed biotinylated oligos. Streptavidin-coated magnetic beads were then bound to the biotinylated motif and removed by magnetic capture. The beads were then washed and the RNA precipitated.

#### 2.5.12.3 Removal of DNA from RNA samples

DNA was removed from RNA using Amplification Grade DNase I (Invitrogen) according to the manufacturer's instructions. Briefly,  $1\mu\text{g}$  of RNA was added to  $1\mu\text{l}$  of DNase Reaction Buffer and 1 unit of DNase in a final volume of  $10\mu\text{l}$  and incubated at room temperature for 15mins. DNase was inactivated by the addition of EDTA to a final concentration of  $2.5\text{mM}$  and incubating the sample at  $65^{\circ}\text{C}$  for 10mins. Samples were screened by PCR using to check for undigested DNA.

## **2.5.12.4 Reverse transcription of RNA**

### **2.5.12.4.1 Reverse transcription of RNA for aminoallyl labelling**

RNA was reverse transcribed to complementary DNA (cDNA) incorporating dUTP using Superscript III Reverse Transcriptase (Invitrogen). 16µg of RNA was incubated at 70°C for 10 mins with 1µl of pd(N)<sub>6</sub> random hexamers (GE Biotech) in a total volume of 27.7µl then cooled on ice. For the reaction, 9.0µl of First Strand buffer, 4.5µl of DTT, 1.8µl of aa-dNTP (5µl each of dATP, dCTP, dGTP (100mM; Promega), 2µl dTTP (100mM), 6µl aa-dUTP and 17µl SDW) and 2µl of Superscript III were added and incubated at 42°C for 3 hours. The RNA was then hydrolysed with 15µl of 1M NaOH, 15µl of 0.5M EDTA (pH 8.0) and incubated at 65°C for 15mins. This reaction mix was then neutralised with 15µl of 1M HCl.

### **2.5.12.4.2 Reverse transcription of RNA for Illumina sequencing and real time PCR**

Real time PCR was used to determine copy number of mRNA to verify microarray data. RNA was reverse transcribed to cDNA as described in 2.5.12.4.1 with the following modification. A final concentration of 10mM of dTTP, dATP, dCTP and dGTP each were used instead of the aa-dNTP mix.

## **2.5.12.5 Labelling of RNA for microarray**

### **2.5.12.5.1 Purification of cDNA**

For purification of cDNA, MinElute Columns (Qiagen) were used with a modified protocol. 450µl of Buffer PB (Qiagen) was added to the reaction mixture and spun through a MinElute column at 13,000rpm in a microcentrifuge. The flow through was discarded and column washed with 750µl of phosphate wash buffer (4.75mM  $K_2HPO_4$ , 0.25mM  $KH_2PO_4$  and 84.4% Ethanol) then eluted twice in 10µl each of phosphate elution buffer (0.019mM  $K_2HPO_4$  and 0.01mM  $KH_2PO_4$ ).

#### 2.5.12.5.2 Coupling of CyDye™ ester to dUTP

CyDye™ esters were coupled to purified cDNA add 1M  $Na_2CO_3$  (pH9.0) to a final concentration of 100mM. The reference dye was resuspended in CyDye™ Cy3 (Amersham) in 10µl of dimethyl sulfoxide (DMSO) and 1µl was added to the cDNA sample. For the target sample, CyDye™ Cy5 (Amersham) was resuspended in 10µl of DMSO, 1µl of this was added to the cDNA sample. Samples were then incubated at room temperature in the dark for 60mins.

#### 2.5.12.5.3 Purification of labelled cDNA

Labelled cDNA was purified using MinElute Columns (Qiagen) according to the following protocol. Hydroxylamine hydrochloride was added to each reaction to a final concentration of 500mM and incubated in the dark for 30mins at room temperature. Both labelled targets were combined and 48µl of  $dH_2O$  added to 500µl of PB buffer (Qiagen) and applied to a MinElute column. Samples were spun at 13,000rpm in a microcentrifuge and washed with PE Buffer (Qiagen), the column was dried and the samples eluted twice with 10µl of phosphate elution buffer.

## 2.5.13 DNA microarray

### 2.5.13.1 Hybridisation and scanning

Microarray slides were designed and printed at the WTSI. Generation III *pan-Salmonella* array design and the protocol for hybridisation and scanning was previously published by Doyle *et al* [182]. The microarray used in this study comprises specific unique PCR products (200–500 bases) of 4097 CT18 genes, plus 345 gene segments representing the LT2-specific (relative to CT18) loci. Appropriate positive and negative controls were also printed on the arrays. The PCR were carried out using specific primers (Sigma-Genosys) in a two-step protocol. Three slides were used for each isolate, with dye reversal. The washing procedures were stringent and included 200 ml of 2x SSC at room temperature for 5 min, two washes in 200 ml of 0.1x SSC-0.1% sodium dodecyl sulfate at 65°C with gentle agitation for 30 min, and two washes in 200 ml of 0.1x SSC at 65°C with gentle agitation for 30 min (1x SSC is 0.15 M NaCl plus 0.015 M sodium citrate). Slides were scanned using a Genepix 4000B scanner (Axon Instruments [now Molecular Devices], California), and every spot was assessed with Genepix Pro software (Axon Instruments).

### 2.5.13.2 Gridding and feature extraction

Spots were gridded and extracted by aligning the grid and applying the FLAG features filter in Genepix with the following parameters.

[Flags] < > [Not found] And

[F532CV] > 135 Or

[F635CV] > 135 Or

[F532 % Sat.] > 30 And

[F635 % Sat.] > 30

The information was then analysed in Genespring or using LIMMA in R(section 2.8).

## 2.5.14 Protein methodologies

### 2.5.14.1 Preparation of protein

#### 2.5.14.1.1 Whole cell lysate

Cultures were harvested and resuspended in 200µl of PBS then mixed with 200µl of 2× SDS loading buffer (90mM Tris-HCl pH 8.45, 24% (v/v) glycerol, 4% (w/v) sodium dodecyl sulfate (SDS), 0.2% (v/v) 2-mercaptoethanol, 0.001% bromophenol blue). Samples were then boiled for 10mins and cooled on ice.

#### 2.5.14.1.2 Cellular fractionation

Cellular fractionation was performed according to protocol previously described by Hantke *et al.* [183] and is described in detail below.

##### 2.5.14.1.2.1 Cell lysis

Cellular fractionation is used to separate the two membrane components and the soluble cytosol for further analysis. Harvested cells were resuspended in 500µl of 100mM Tris.HCl (pH 8.0) supplemented with Complete (Roche) protease inhibitor cocktail. 1ml of 1M Sucrose dissolved in 100mM Tris.HCl (pH 8.0) was added then

100 $\mu$ l of 10mM EDTA and the solution mixed. 100 $\mu$ l of 2mgml<sup>-1</sup> lysozyme dissolved in Tris.HCl (pH 8.0) was added before 3.2ml of dH<sub>2</sub>O. Cells were incubated for 10mins at room temperature to allow formation of spheroplasts. 12.5ml of dH<sub>2</sub>O was added to lyse cells. To the clear solution 3.5mls of 1M MgCl<sub>2</sub> was added with 100 $\mu$ l of 4mgml<sup>-1</sup> DNase 1 (Roche) and the solution incubated at room temperature with gentle agitation for 5mins. The solution is then centrifuged at 18,000rpm for 60mins at 4°C to separate the soluble fraction from the membrane fractions.

#### **2.5.14.1.2.2 Precipitation of the soluble fraction**

Soluble proteins were present in the supernatant and were separated from the precipitated membrane fractions. The supernatant was removed and the soluble proteins precipitated with trichloroacetic acid (TCA) to a final concentration of 10% (v/v) and centrifuged at 20,000 rpm for 30mins at 4°C. The pellet was then washed with dH<sub>2</sub>O and resuspended in the requisite buffer.

#### **2.5.14.1.2.3 Separation of the membrane fractions**

The cytoplasmic membrane was solubilised from the remaining pellet by resuspension in 5ml of 2% Triton X-100 dissolved in 50mM Tris.HCl (pH 8.0) and centrifuged at 20,000 rpm for 30mins at 4°C and the supernatant removed. This was repeated and the supernatants were pooled leaving the precipitated outer membrane fraction in the pellet. The cytoplasmic membrane proteins, contained in the soluble fraction, were precipitated by mixing well with 10ml of chloroform, 20ml of methanol and incubated on ice for 30min. After centrifugation at 20,000 rpm for 30mins at 4°C the pelleted precipitated proteins were resuspended in the requisite buffer.

#### 2.5.14.1.3 Heat shock

Samples were washed with PBS, resuspended in 1/1000 the volume of PBS and heat shocked for 10 mins at 60°C then diluted in an equal volume of SDS loading buffer.

#### 2.5.14.1.4 Preparation of proteins for mass spectroscopy analysis

Samples were reduced in SDS loading buffer with 1mM dithiothreitol (DTT) at 70°C for 10mins, cooled to room temperature then alkylated for 30 minutes with 2mM iodoacetamide in the same buffer.

### 2.5.14.2 One-dimensional SDS-PAGE gel electrophoresis of proteins

Protein preparations were routinely separated by electrophoresis in denaturing 0.1% SDS polyacrylamide gels (PAGE). Polyacrylamide concentration affects the speed at which different sized proteins run. Suitable concentration or concentration gradient was chosen to optimise resolution of the target protein. All pre-cast gels (Invitrogen) were run at 150V for 55 mins.

#### 2.5.14.3 Protein visualisation

##### 2.5.14.3.1 Coomassie blue protein stain

To visualise proteins resolved by SDS-PAGE, proteins were stained with Coomassie Blue reagent. Proteins were fixed for 1 hour in 40% methanol, 2% acetic acid then stained for 1-4 hours with Brilliant Blue G (Sigma) and prepared according to the

manufacturer's instructions. The gel was rinsed for 1 min in 25% methanol, 5% acetic acid then destained in 25% methanol until the bands were visible.

#### **2.5.14.3.2 Western blot**

Resolved proteins were transferred to a nitrocellulose membrane by electrophoresis using Trans-blot semi-dry (Bio-rad). The nitrocellulose membrane was probed with a specific primary antibody and detected with Opti-4CN (Bio-rad) kit which included secondary antibody. This was done according to the manufacturer's instructions.

#### **2.5.14.4 Immunoprecipitation for chIP-seq**

Cultures were fixed with 1% formaldehyde to crosslink DNA-bound proteins and incubated at the same temperature for 20mins. The reaction was then quenched with glycine pH7 to a final concentration of 0.5M. Cells were harvested and washed twice in TBS (0.137M NaCl, 0.27mM KCl, 0.25mM Tris, pH7.5) and lysed according to cellular fractionation protocol. DNA was then sheared 3 times, 20 seconds on and 15 seconds off on ice at a 20% amplitude using a VibraCell sonicator (Sonics, Newton, USA) with a tapered probe (Model CV33). The target protein was immunoprecipitated (IP) using Protein G Immunoprecipitation Kit (Sigma) according to manufacturers instructions and uncrosslinked using pronase ( $0.8\text{mgml}^{-1}$ ) at  $65^{\circ}\text{C}$  overnight. DNA was then purified using Qiagen PCR purification kit. DNA was sequenced as per standard Illumina sequencing protocol.

#### **2.5.14.5 chIP-seq data mapping and analysis**

To generate a plot for use in Artemis the data were mapped using MAQ according to the same parameters as the transcriptome data mapping. The plots were then z-score

normalised and the data from the WT culture deducted from the TT53.8 (ompR:3XFLAG). This plot was then loaded into Artemis using Graph, Add User plot. Peak finder in Artemis was used to generate “misc\_binding” features from all peaks that were 50bp in length with a score greater than 2. These data were manually annotated and most peaks not proximal to the start of a gene were discarded unless there were multiple sites close by. As the data do not align to a particular strand the direction of the binding site was estimated according to the closest CDS. The sequence of these sites were used to determine if there were any over-represented motifs using YMF [184] an online motif finding package. The parameters for the search were motif size = 8, Maximum of spacers in middle = 0, maximum of 2 degenerate symbols = 2. Degenerate symbols allowed in a motif are R (purine - A or G), Y (pyrimidine - C or T), W (A or T), and S (C or G).

## **2.6 Murine model of Salmonellosis**

### **2.6.1 Competitive infections**

Experimental design, culture, organ homogenisation and organ plating and CFU counting was done by Tim Perkins in these murine experiments. Infection by oral gavage and vivisection was performed by Robert Kingsley of the Molecular Pathogenesis group (WTSI).

Competitive infections were performed as described by previously (Kingsley *et al*, 2003). Briefly, 6 to 8 week-old female BALB/c (ByJ; Jackson Lab) mice were used. Bacteria were routinely cultured statically overnight prior to infection. In all experiments, the bacterial titre of the inoculum was determined by spreading serial 10-fold dilutions on agar plates containing the appropriate antibiotics and determining

the number of CFU. Groups of 5 mice were infected by oral gavage with an approximate 1:1 mixture of mutant and isogenic parents at a dose of approximately  $10^7$  CFU/strain/mouse. The caecum, 3 Peyer's patches of the terminal ileum adjacent to the caecum, the mesenteric lymph nodes, the liver and the spleen were harvested aseptically and homogenized in 5 ml of phosphate-buffered saline, pH 7.4. Dilutions of homogenized organs were plated on LB plates kanamycin and X-phos (Sigma) to distinguish between colonies expressing PhoN and colonies that were PhoN negative (RAK113). Data were normalized by dividing the output ratio (CFU of the mutant/CFU of the wild type) by the input ratio (CFU of the mutant/CFU of the wild type). In case only one bacterial strain was recovered from fecal pellets, the limit of detection was determined for the missing strain and used to calculate the minimum mutant-to-wild type ratio. All data were converted logarithmically prior to the calculation of averages and statistical analysis. A Student's *t*-test was used to determine whether the log value of the mutant-to-wild type ratio recovered from infected organs was significantly different ( $p < 0.05$ ) from the log value of the mutant-to-wild type ratio present in the inoculum.

## 2.7 LC-MS analysis

Experimental design, sample preparation, resolving of proteins, excision of bands and digestion of proteins was performed by Tim Perkins. Protein sequencing and database searching was done by Team 17 (WTSI).

Reduced and alkylated proteins were resolved as in 2.5.14.2. Protein bands were excised, destained completely, and digested with trypsin (sequencing grade; Roche). Peptides were extracted from gel with 5% formic acid–50% acetonitrile, dried. The

extracted peptides were redissolved in 0.5% formic acid (FA) and analyzed with on-line nano LC-MS/MS on an Ultimate 3000 Nano/Capillary LC System (Dionex) coupled to a LTQ FT Ultra mass spectrometer (ThermoElectron) equipped with a nanoelectrospray ion source (NSI). Samples were first loaded and desalted on a PepMap C18 trap (0.3 mm id x 5 mm, Dionex) at  $25\ \mu\text{Lmin}^{-1}$  with 0.1% FA for 5 min then separated on a BEH C18 analytical column (75  $\mu\text{m}$  id x 10 cm) (Waters) over a 30 or 45 or 60 min linear gradient of 4-32%  $\text{CH}_3\text{CN}/0.1\%$  FA based on the gel band's size and intensity. The LTQ FT Ultra mass spectrometer was operated in the standard data dependent acquisition mode controlled by Xcalibur 2.0. The survey scans ( $m/z$  400-1500) were acquired on the FT-ICR at a resolution of 100,000 at  $m/z$  400 and one microscan per spectrum. The three most abundant doubly and triply charged ions ( $2+$  and  $3+$ ) with a minimal intensity at 1000 counts were subject to MS/MS in the linear ion trap at an isolation width of 2 Th. Precursor activation was performed with an activation time of 30 msec and the activation Q at 0.25. The normalised collision energy was set at 35%. The dynamic exclusion width was set at  $\pm 10$  ppm with 1 repeats within 45 sec and excluding for 60 sec. To achieve high mass accuracy, the automatic gain control (AGC) target value was regulated at  $1 \times 10^6$  for FT and  $1 \times 10^4$  for the ion trap, with maximum injection time at 1000 msec for FT, and 250 msec for ion trap respectively. The instrument was externally calibrated using the standard calibration mixture of caffeine, MRFA and Ultramark 1600

The Raw files were processed by BioWorks 3.3 and then submitted to a database search in Mascot server 2.2 ([www.MatrixScience.com](http://www.MatrixScience.com)) against an in-house built Typhi Ty2 genomic 6-frame translated database using following search parameters: trypsin/P with 2 mis-cleavage, 20 ppm for MS, 0.5 Da for MS/MS, with 9 variable modification of Acetyl (N-term), Carbamidomethyl (C), Deamidated (NQ),

Dioxidation (M), Formyl (N-term), Gln->pyro-Glu (N-term Q), Glu->pyro-Glu (N-term E), Methyl (E) and Oxidation (M).

Database search results were post-processed with an in-house tool that employs Percolator, a machine learning method for MS data analysis. This aids improved sensitivity and provides robust error measures using a random database as a null model. All peptides with a posterior error probability (probability that an individual peptide was identified by chance alone) of 1% or less were accepted for subsequent analysis, resulting in an overall false discovery rate of 0.076%.

## **2.8 Bioinformatics tools**

### **2.8.1 Artemis and ACT**

Artemis is a genome browsing tool developed for genome annotation and analysis at the Wellcome Trust Sanger Institute (WTSI) and the manual is available at <http://www.sanger.ac.uk/Software/Artemis/>. Important features of Artemis that made this study possible are outlined below. A new entry can be created and features copied, moved to or created within this feature. Features can be classified accordingly and examples include coding sequences, islands, misc\_RNA, binding sites. Feature can also be selected by the details contained within the notes annotated to each feature. The published annotation for each feature usually includes these details and each feature can be easily edited. It is also possible to assign systematic gene names to each selected feature and select features on the forward or reverse strand. Keys and IDs can also be altered easily. In the CT18 annotation each feature was assigned a colour according to its functional class and these details can be easily searched for. An example of an annotated feature is below.

CDS 337..2799

/gene="thrA"

/locus\_tag="t0002"

/EC\_number="2.7.2.4"

/EC\_number="1.1.1.13"

/note="multifunctional homotetrameric enzyme that catalyzes the phosphorylation of aspartate to form aspartyl-4-phosphate as well as conversion of aspartate semialdehyde to homoserine; functions in a number of amino acid biosynthetic pathways"

/codon\_start=1

/transl\_table=11

/product="bifunctional aspartokinase I/homoserine dehydrogenase I"

/protein\_id="NP\_803887.1"

/db\_xref="GI:29140545"

/db\_xref="GeneID:1066974"

/translation="MRVLKFGGTSVANAERFLRVADILESNSRQGQVATVLSAPAKIT  
NHLVAMIEKTIGGQDALPNISDAERIFSDLLAGLASAQPGEPLARLKMVVEQE  
FAQIKHVLHGISLLGQCPDSINAALICRGEKMSIAIMAGLLEARGHRVTVIDPV  
EKLLAVGHYLESTVDIAESTRRIAASQIPADHMILMAGFTAGNEKGELVVLGR  
NGSDYSAAVLAACLRADCCEIWTDVDGVYTCDPRQVPDARLLKSMSYQEA

MELSYFGAKVLHPRTITPIAQFQIPCLIKNTGNPQAPGTLIGASSDDDNLPVKGI  
SNLNNMAMFSVSGPGMKGMIGMAARVFAAMSRAGISVVLITQSSSEYSISFC  
VPQSDCARARRAMQDEFYLELKEGLLEPLAVTERLAIISVVGDMRTLRGISA  
KFFAALARANINIVAIAQGSSERSISVVVNDDATTGVRVTHQMLFNTDQVIE  
VFIGVGGVGGALLEQLKRQQTWLKNKHIDLRVCGVANSKALLTNVHGLNL  
DNWQAELAQANAPFNLGRLIRLVKEYHLLNPVIVDCTSSQAVADQYADFLRE  
GFHVVTNKKANTSSMDYYHQLRFAAAQSRRKFLYDTNVGAGLPVIENTLQN  
LLNAGDELQKFSGILSGSLSFIFGKLEEGMSLSQATALAREMGYTEPDPRDDL  
SGMDVARKLLILARETGRELELSDIVIEPVLDEFDASGDVTAFAHLPQLDD  
AFAARVAKARDEGKVLRYVGNIEEDGVCRVKIAEVDGNDPLFKVKNGENAL  
AFYSHYYQPLPLVLRGYGAGNDVTAAGVFADLLRTLWSWKLGV"

An important feature for this study is the graph function. It is possible to plot transcription data or chIP-seq data and it is similar to the GC plot function. Using the menu graph->Add user plot the file needs to have one value per base and every base must be represented, right clicking on the plot enables scaling of the local plot to the highest peak.

## 2.8.2 Mapping software

### 2.8.2.1 Transcriptome mapping software

The WTSI developed a data analysis pipeline for managing the enormous datasets produced by high-throughput sequencing. Each file can contain in the region of 12.5 million reads consisting of 450 million bp. For this study we map the sequence reads to the Ty2 genome using the programme MAQ [185]. The sequence data file (fastq)

and the reference genome file (fasta) were converted to binary format using the maq fastq2bfq and maq fasta2bfa command line respectively.

```
% maq fastq2bfq [-n nreads] <in.fastq> <out.prefix>|<out.bfq>
```

```
% maq fasta2bfa <in.fasta> <out.bfa>
```

Then maq map command was used to map the bfq file to the reference bfa file.

```
% maq map [options] <out.map> <chr.bfa> <reads_1.bfq> [reads_2.bfq]
```

These data were then “piled up” using maq -q 30 pileup by aligning each of the sequence reads to the genome, assigning a score to each base-pair and a description of the strand it is aligned to.

```
% maq pileup [options] <chr.bfa> <align.map>
```

The quality (-q) option was used for all mapping and was stipulated at 30. An example of the pileup output is below.

```
Forward strand      all_bases    7887   G    45   @.....
```

```
Reverse strand          all_bases    914   G    6   @,,,,,
```

```
Overlapping Strands all_bases    7690   G    38   @,,,,,,,,,,,,,,,,,,,,,
```

To produce a plot readable into Artemis, we used the unix command

```
% awk '{print $4;}' filename.pileup > filename.plot.
```

### **2.8.2.1 Protein mapping scripts**

The protein mapping script was written by Keith James, microarray facility (WTSI) for general use.

### **2.8.3 Perl scripts**

#### **2.8.3.1 Pileup to stranded plot - maqpileup2depth.pl**

To produce a stranded plot for Artemis we derived strand information from the pileup plot using the script maqpileup2depth.pl (written by Miao He, WTSI, appendix 9.1). STDOUT is a filename.plot containing two columns separated by a space representing reverse (column 1) and forward strand (column 2).

```
% ./maqpileup2depth STDIN.pileup > STDOUT
```

#### **2.8.3.2 Local plots**

Due to the sheer size of each plot use the extractLines.pl (written by Kathryn Holt, WTSI, appendix 9.2) to extract specific regions of the Artemis plot.

```
% ./extractLines.pl -i filename.plot -start -stop > STDOUT
```

#### **2.8.3.3 Extracting expression data from plot**

To extract data for any feature we used a script called tram.pl (written by Sammy Assefa, WTSI, appendix 9.3) This script quantifies sequence data aligned to each feature and take into the uniqueness of the sequence in each feature. Uniqueness is

determined by mapping the entire genome back to itself, thus it filtered out any repetitive sequences.

```
% ./tram.pl <GFF file> < Specific strand plot file > < Uniqueness plot > 25 1 1 >
```

STDOUT

Example output

		Start	End	Len.	Uniq	No-cov	nonU	%Ulen	GM	AM	Med	Max	Min
Misc													
RNA	rfam_100.1	1	189	189	148	16	24	78.31	3.51	47.99	20	195	0
Misc													
RNA	Thr_leader	191	311	121	83	38	0	68.6	2.48	6.89	8	21	0
Misc													
RNA	rfam_107c	5014	5123	110	77	33	0	70	1.03	0.76	1	2	0

Start, first base; stop, last base; Len., number of bases in feature, Uniq, number of uniquely covered bases; No-cov, number of bases with no coverage; nonU, number of bases with coverage above zero but less than the uniqueness cutoff; %Ulen, percentage of bases that are uniquely covered; GM, geometric mean; AM, arithmetic mean; Med, median; Max, maximum nucleotides one base has mapped; Min, minimum number of nucleotides one base has mapped.

## 2.8.4 R

R is a freeware package for statistical analysis [186]. Various analysis libraries are available freely from published packages.

### 2.8.4.1 Limma analysis

For differential analysis of transcriptome data we used the script outlined in Appendix 9.4 [187].

### 2.8.5 Genespring

For differential analysis of spotted microarray analysis we used Genespring GX (Agilent).

### 2.8.6 Graph pad prism 5.0

To plot and analyse data we used the programme Graph Pad Prism 5.0a for OS X. This is a self-contained package and contains standard comprehensive statistical analysis.

### **3 A Method for RNA-seq in prokaryotes**

### 3.1 Aims of the work described in this Chapter

Recent advances in DNA sequencing technology have transformed the amount of sequencing information that can be obtained in single sequencing runs by orders of magnitude. The aim of this chapter is to apply such technology to sequence cDNA made from mRNA, thus developing a novel approach to assess gene expression profiles in bacteria.

### 3.2 Introduction

The advent of high-throughput sequencing technologies has permitted new approaches to exploring functional genomics, including the direct sequencing of complementary DNA (cDNA), an approach that has been named RNA-seq. Recent publications have described the exploitation of this high-density, high-resolution technology to build an accurate picture of the transcriptional patterns within eukaryotic organisms [188,189]. These studies have demonstrated a number of advantages over microarray-based techniques, including greater sensitivity, increased dynamic range, reduced background noise, single nucleotide resolution and the capacity to map data to the entire genome sequence [190]. Furthermore, the results are not biased by array design as most expression arrays have limited density and features are normally designed on the basis of classical *in silico* genome annotation. Consequently, RNA-seq has led to the discovery of novel genetic features [191,192].

One of the drawbacks of the initial RNA-seq studies, relative to microarray work, was the lack of strand-specificity in the data, as these protocols sequence double stranded cDNA, which masks directionality [188,189]. However, two recent studies have

demonstrated that directionality can be retained through asymmetrically modifying the 5' and 3' ends of the RNA molecules prior to reverse transcription, either by attaching RNA linkers [193] or switch strand PCR [194], allowing the transcripts to be mapped back to the reference genomes in a strand specific manner. This is crucial for resolving overlapping genetic features, detecting antisense transcription and assigning the sense strand for non-coding RNA (ncRNA).

This chapter details the development of a directional RNA-seq protocol that eliminates both the need for second strand cDNA synthesis and the modification of transcripts prior to reverse transcription. This method was applied to the rRNA depleted RNA population of *S. Typhi* allowing capture of an unbiased view of the coding and non-coding RNA transcriptome. In combination with this method, the teams at the WTSI have also developed a computational pipeline based on Artemis ([www.sanger.ac.uk/Projects/Pathogen/Transcriptome/](http://www.sanger.ac.uk/Projects/Pathogen/Transcriptome/)) that facilitates mapping of the transcriptome data [195,196]. Together, these methods should greatly enhance the understanding of microbial genome's transcriptional content.

## 3.3 Results

### 3.3.1 Directional Sequencing of Single Stranded cDNA

Sequencing using the Illumina platform requires the ligation of adapters onto either end of a DNA molecule (<http://icom.illumina.com>), which are necessary for PCR amplification, flow cell attachment and sequencing reaction priming. As DNA ligase only works efficiently on DNA duplexes, samples are prepared as double stranded DNA and subjected to an end repair reaction, using a mixture of enzymes to repair 3' overhangs and extend from recessed 3' giving blunt ended products. DNA molecules

are subsequently 3' monoadenylated and ligated to adapter dimers with a 5' monothymidine overhang. During the course of development of a high-throughput RNA-seq method for *S. Typhi*, in cases where the second strand cDNA synthesis failed or was omitted, the Illumina sequencing team were still able to generate data using the standard Illumina sample preparation methods but unlike the double stranded cDNA method, the transcript data retained directional fidelity.

Four hypotheses were possible in explaining the ligation of linkers to single stranded cDNA and subsequent processing to generate sequence data and maintain directionality (figure 3.1). The first hypothesis required the ligation of single stranded DNA adaptors to the single stranded cDNA molecules (figure 3.1a). This is possible because T4 DNA ligase can ligate single stranded DNA molecules, albeit at low efficiency [195] and directionality would be maintained because the second strand is never synthesised (figure 3.1a). The alternate possibilities involved the formation of duplexes during the end repair reaction (figure 3.1b-d). Either annealed RNA fragments (the remains of transcripts that served as templates in the reverse transcription reaction) or inter- or intra-molecular hybridisation of cDNA was suggested to prime complementary strand synthesis, leading to the formation of blunt ended, double stranded constructs that could then function as the substrate for the efficient ligation of adapters. If complementary strand synthesis were primed by annealed RNA fragments, this strand would be composed of both RNA and DNA, which cannot be amplified and sequenced by DNA-dependent DNA polymerases. Consequently only the original single stranded cDNA strand would be sequenced. If complementary strand synthesis were primed by intra- or inter-molecular cDNA annealing, then 3' end processing would produce a reverse complement of the annealed cDNA's 5' end. Hence sequences with sense and antisense orientations

would be segregated into the 3' and 5' regions of the cDNA strands respectively, so by sequencing only the 5' end, all sequence reads maintain the same orientation relative to the original direction of transcription.

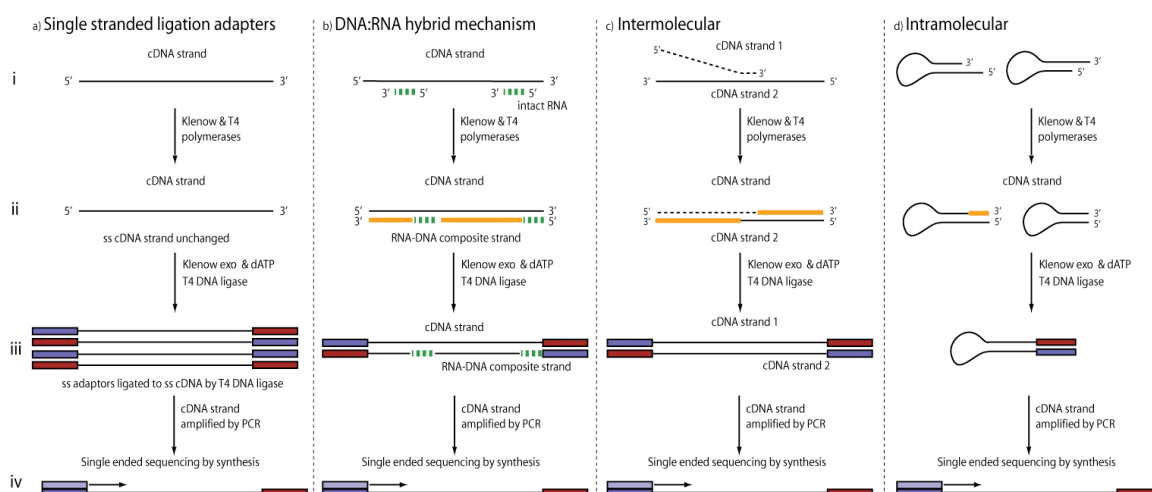


Figure 3.1 The hypotheses proposed to account for the attachment of Illumina adapted dimers to ss cDNA

a) attachment of adapters to ss cDNA, and priming of second strand synthesis by b) RNA fragments c) intermolecular cDNA annealing d) intramolecular cDNA annealing. Dark blue rectangles represent linker sequence and red, the reverse complement of the linker sequence. Light blue rectangles represent sequencing primer used on the Illumina slide during sequencing.

To test these hypotheses, a 48-mer DNA oligonucleotide was designed, consisting of a defined sequence tag, an RNA oligonucleotide-binding site and two stretches of random sequence (figure 3.2a). Solutions containing either this DNA oligonucleotide alone, or in the presence of a 12 nucleotide RNA oligonucleotide complementary to the binding site, were subjected to standard Illumina sample preparation and sequencing reactions. If RNA primed the second strand synthesis, then resection of the overhanging 5 nucleotide random sequence at the 3' end would be observed and libraries would not be generated in the absence of the RNA oligonucleotide (figure 3.1b). Conversely, if inter- or intra-molecular cDNA annealing were the dominant mechanism, then the reverse complement of the known 5' sequence tag would be

observed at the 3' end of the oligo (figure 3.1c and d). However, if the mechanism was simple annealing of linkers directly to single stranded-cDNA, the DNA oligonucleotide would remain unaltered (figure 3.1a).

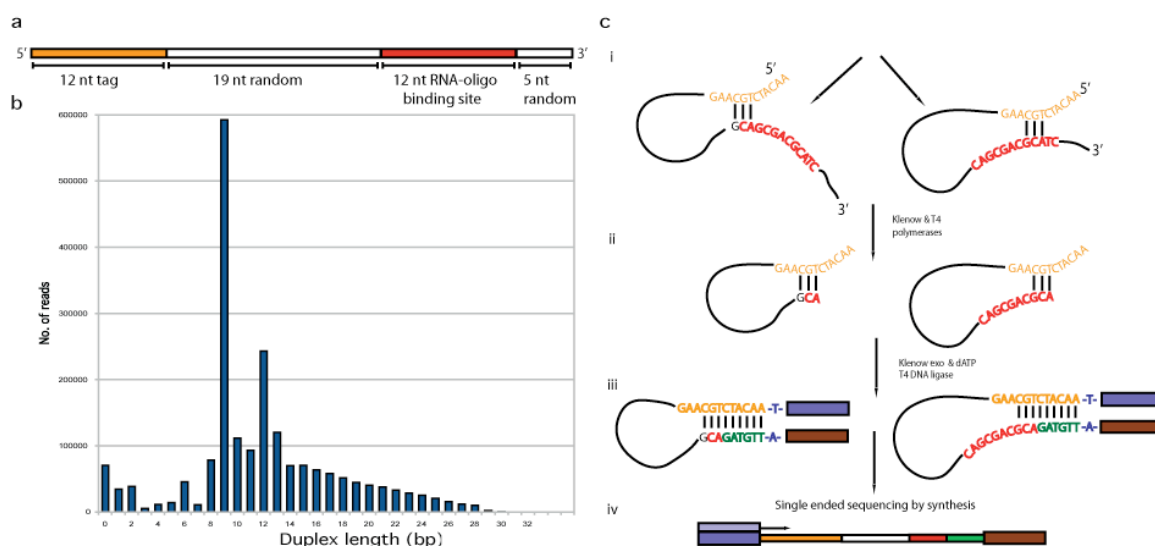


Figure 3.2 Single stranded sequencing

a) Schematic representation of the DNA oligonucleotides from which Illumina libraries were generated. b) Distribution of duplex lengths amongst a sequenced sample of single stranded DNA oligonucleotides. Reads were extracted corresponding to the oligonucleotide by searching the output data for the 12 nt known sequence tag. Duplex lengths were then calculated by counting the number of bases at the 3' end and found to be the reverse complement of those at the 5' end. This revealed a smooth distribution of values over a range of sizes, with large peaks at 12 bp (likely resulting from intermolecular annealing) and 9 bp (probably the consequence of a 3 bp duplex that can form between the known sequence tag and RNA binding site). c) Two proposed mechanisms for the formation of species with a 9 nt of reverse complementarity between the 5' and 3' ends, the most common duplex length observed. The vast majority of these were found to have 3 bp “seed duplexes” formed by base pairing between –CGT– in the sequence tag and either the –GCA– in the 3' half of the RNA oligonucleotide binding site, or the CA– dinucleotide at the start of the binding site when the preceding nucleotide (the last of the 19 nt random sequence) was G.

Following the standard Illumina preparation protocol, libraries were successfully generated both from mixtures of the DNA and RNA oligonucleotides, as well as the DNA oligonucleotide alone (but not the RNA oligonucleotide alone), suggesting RNA molecules were not required for priming of second strand synthesis. Illumina

sequence reads were obtained from a library constructed from a solution of single stranded DNA oligonucleotides. In 88% of the sequences, the RNA binding site sequence had been at least partially altered to the reverse complement of the known sequence tag, implying intra- or inter-molecular annealing had occurred and primed second-strand synthesis to give blunt-ended duplexes. The most common species (29% of the sequenced population, figure 3.2b) had 9 nucleotides of reverse complement of the 5' tag at the 3' end (a 9 bp “duplex length”), which is likely to have arisen from the scenarios outlined in Figure. 3.2c. The second most common species (12% of the sequenced population) have a 12 bp duplex length. This is likely an artefact of inter-strand annealing: because the 19 nucleotide random sequences will be different in the two annealed strands, only the 12 nucleotide tags will match between the two ends of the repaired duplex. However, a third of the duplex lengths are longer than 12 bp, indicating that intra-strand annealing also occurs to a detectable extent (because the complementary nature of the two ends extends beyond the known tags into the random sequence unique to each oligonucleotide molecule). Hence, this has demonstrated that Illumina libraries can be constructed from single stranded DNA using standard techniques, where both inter- and intra-strand annealing occur to a comparable extent and make a significant contribution to the formation of double stranded cDNA during end repair.

### 3.3.2 Sequencing of the *S. Typhi* transcriptome

The total RNA population was extracted from *S. Typhi* BRD948 [177], single stranded cDNA generated by a reverse transcription reaction and the cDNA was used as the substrate for standard library construction reactions and Illumina sequencing. This project generated a dataset of 5.4 million 36 nucleotide reads that was mapped

back to the genome in a strand-specific manner. However, as the total RNA population was sampled, the majority of reads mapped to the rRNA operons, suggesting removal of these molecules prior to reverse transcription would greatly increase the sensitivity of the technique.

Two alternative methods were successful in reducing the amount of highly abundant 16S and 23S rRNA molecules, thus enriching for mRNA and non-coding RNA (ncRNA) transcripts: Reduction through hybridisation of capture oligonucleotides and subsequent annealing to biotinylated oligonucleotides and removal by magnetic streptavidin coated beads or degradation using a terminator exonuclease specific for 5' monophosphorylated transcripts [197]. Both approaches depleted the rRNA and increased the number of uniquely mapping sequences (table 3.1). Plotting the arithmetic mean coverage per base of each annotated coding sequence in the *S. Typhi* genome in biological replicates using the two different techniques yielded a correlation coefficient of  $R=0.90$ , suggesting the two different methods give comparable results (figure 3.3). No clear outliers were evident in the correlation plots, suggesting there is no bias for a subset of transcripts specific to either method; however, depletion through hybridisation to oligonucleotides is likely to be the more specific technique, as any cleaved transcripts or RNA species generated through endonucleolytic cleavage of a larger precursor would be removed by the exonuclease treatment.

Table 3.1 Read mapping for different depletion methods

<b>Depletion Method</b>	Undepleted	Oligonucleotide hybridisation	Terminator exonuclease
<b>Amount RNA (µg)</b>	100	100	100
<b>Amount cDNA (µg)</b>	20	20	20
<b>Size fraction (bp)</b>	200-250	200-250	200-250
<b>Read Length (bp)</b>	36	36	36
<b>No. Reads</b>	5372456	7183969	7043338
<b>Proportion of Reads Mapped</b>	94%	91%	94%
<b>Proportion of Reads Mapped Uniquely</b>	14%	41%	51%

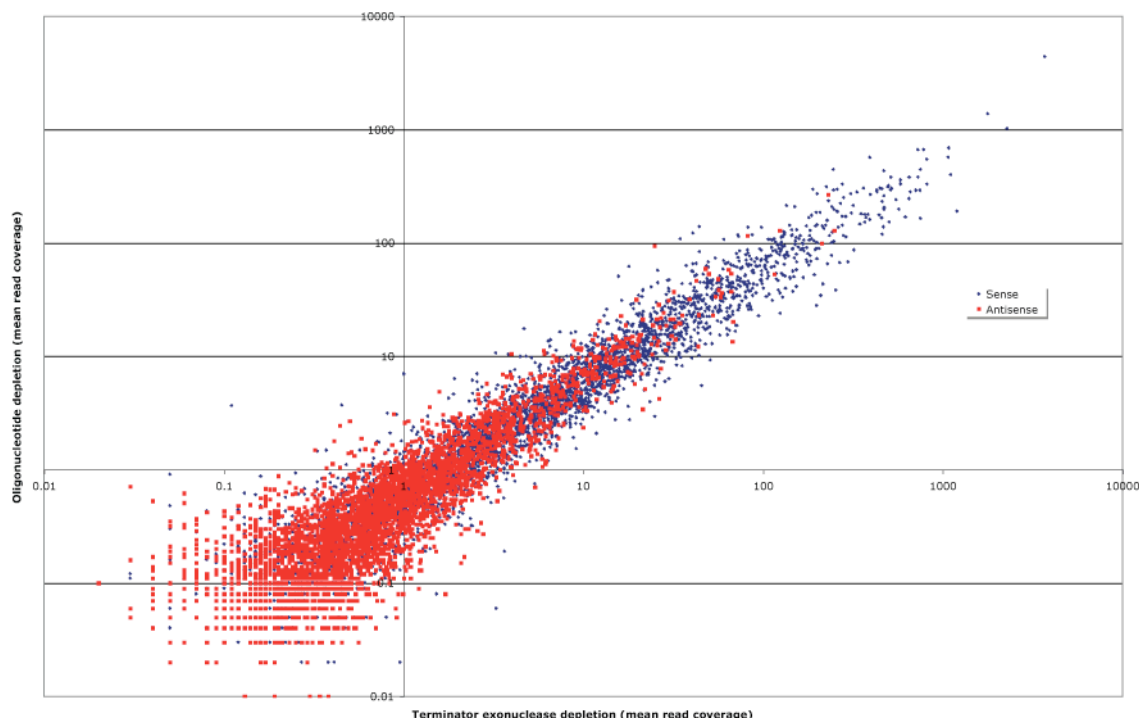


Figure 3.3 Quantitative comparison of terminator exonuclease and oligonucleotide hybridisation rRNA depletion techniques.

This log-log plot shows the mean read coverage in the sense and antisense directions for all annotated CDS features in the *S. Typhi* Ty2 genome in datasets depleted by the two different techniques. The Pearson  $R^2$  and Spearman correlation coefficients of the results of the two methods are  $>0.80$  in both directions, indicating that these results are robust and reproducible. The absence of large numbers of outliers demonstrates that the terminator exonuclease does not remove a large number of non-rRNA transcripts.

### 3.3.3 Mapping and visualising data

In order to sample a wide range of different transcriptional patterns, *S. Typhi* cultures were grown shaking in LB to three different time points ( $OD_{600}=0.3, 0.45$  and  $0.6$ ) in 50ml falcon tubes and in 15ml volumes. Illumina sequencing of the pooled samples generated 45 million  $\sim 36$  nucleotide reads from 7 independent experiments. 4.8 million reads mapped to annotated coding sequences (CDS). Strand specific mapping allowed the unambiguous assignment of the 'sense' strand for transcriptional units,

including ncRNA, and permitted deconvolution of overlapping genetic features on opposite strands.

These data were used to assist development of a computational pipeline at the WTSI (composed of tools freely available from <http://www.sanger.ac.uk/Projects/Pathogens/Transcriptome/>) that allow the visualisation of sequence read depth mapped to the annotated genome sequence within the freeware program Artemis (figure 3.4). The programme MAQ [185], an algorithm that retains non-uniquely mapping reads within the dataset, was used to calculate a fold coverage value for every nucleotide in the genome in both the forward and reverse directions. A second script, `maqpileuptodepth.pl`, is able to produce an analogous output from data mapped with MAQ. Using these tools, patterns of transcription were observed within *S. Typhi* to single nucleotide resolution.

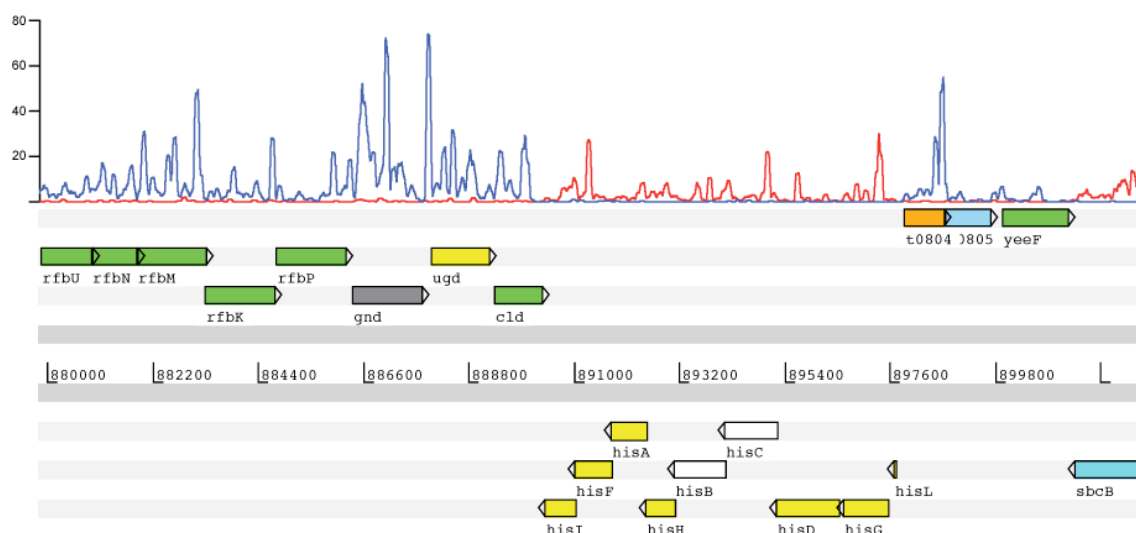


Figure. 3.4 Representation of transcriptomic sequence read coverage plot relative to genome annotation, as displayed in Artemis

A histidine biosynthesis gene cluster in *S. Typhi* Ty2. The graph lines represent the mean number of sequence reads (averaged over a 100 nt sliding window) mapping to the forward (blue) or reverse (red) strand.

### 3.4 Discussion and conclusion

This chapter summarises the development of a high-throughput RNA-seq method for RNA isolated from *S. Typhi*. This method identifies the template strand for the transcriptome effectively deconvoluting the sequence data. This approach includes a novel method for the isolation of a total RNA sample, depleted of rRNA, suitable for high-throughput sequencing, and a new approach for retaining directional fidelity in transcriptomic data by sequencing single-stranded cDNA, a method that is actually simpler than the original RNA-seq protocols as it abrogates the need for second strand cDNA synthesis. These sequence data have assisted in the development of the Illumina sequencing and analysis pipeline, which includes custom-built programmes for mapping and visualising the output data in context with the reference genome.

This method has the capacity to further our understanding of the *S. Typhi* genome expression.

Datasets produced in this manner allow the detection of ncRNA, operon structures and 5' and 3' untranslated regions, features crucial for gene regulation that are difficult to predict from genome sequences *de novo*, across the entire chromosome. Hence as well as measuring transcriptional activity it is clear that this approach will prove to be of great value for complementing and refining current genome annotations in prokaryotes.

## 4 Deep sequencing of the *S. Typhi* Ty2 transcriptome

## 4.1 Aims of the work described in this Chapter

To survey the *S. Typhi* transcriptome using the RNA-seq protocol developed in Chapter 3.

## 4.2 Introduction

DNA sequencing has been exploited to determine the whole genome sequence of hundreds of prokaryotic and eukaryotic species [76,198,199]. The availability of whole genome sequence has facilitated gene identification, transcriptomic studies and underpins experiments to link genotype to phenotype. To date genome-wide analysis of the transcriptome has relied, to a significant degree, on the use of DNA microarrays. However, recent advances in DNA sequencing technologies has facilitated the determination of nucleotide sequence with a genomic read depth several orders of magnitude greater than was previously possible. In this chapter, this technology is applied to characterise the transcriptome of *S. Typhi* by mapping sequence reads of cDNA prepared by reverse transcription of total cellular RNA depleted of ribosomal RNA.

Bacterial genomes are relatively small and have a high density of coding sequence in comparison to most eukaryotes. For example, the genome of *S. Typhi* is ~4.8 Mbp in length, with ~4,700 open reading frames currently defined in available annotation [76,136]. As outlined in detail in Chapter 1, *S. Typhi* is interesting in that, unlike most *Salmonella* serotypes that have a broad host range and cause localised gastroenteritis, this pathogen is highly host adapted and causes systemic typhoid fever only in higher primates (host-restricted). The genome of *S. Typhi* harbours novel features including horizontally acquired genetic islands specific to this serotype and ~220 pseudogenes,

genes that are potentially inactivated in this pathogen but intact in related species such as *S. Typhimurium* [77]. Of particular note, *S. Typhi* also expresses a polysaccharide, known as the Vi capsule, associated with increased virulence.

Illumina-based high throughput sequencing was exploited to characterise the transcriptome of *S. Typhi* and bioinformatics approaches were developed and used to identify key advantages of the approach over microarray analysis. Further, this analysis was performed in a strand-specific manner, allowing overlapping transcripts encoded on opposite strands to be readily identified. In addition to confirming previous transcriptional data, this method has been able to identify novel transcripts, including many potential non-coding small RNAs, putative *cis*-acting RNA elements and previously hypothetical genes. This study has also defined the transcription of pseudogenes under these conditions and annotated a *S. Typhi* Ty2 specific island which maps significant transcript data. By mapping transcripts to the whole genome this approach has identified expressed regions of *S. Typhi* prophages that encode putative cargo genes and potentially identified antisense expression of known coding sequences. This method has putatively mapped a significant amount of sequence data to the 5' untranslated regions of genes.

## 4.3 Results

### 4.3.1 Mapping Sequence Reads to the Annotated Ty2

#### Genome

The protocol devised in Chapter 3 was used in to perform an entire survey of the *S. Typhi* transcriptome. RNA was prepared from three replicates of *S. Typhi* Ty2

BRD948 grown to mid-log phase in LB broth (OD<sub>600</sub>=0.6). This material was pooled, reverse transcribed and subjected to Illumina sequencing. The sequence reads were then mapped to the Ty2 reference genome (table 4.1). To achieve the transcript plot, each nucleotide of the genome was assigned a value derived by a pileup of ~36bp nucleotide sequence reads generated from the Typhi cDNA (figure 4.1). Sequence reads were then mapped to each strand of the *S. Typhi* Ty2 BRD948 genome and the sequence coverage per base plotted and visualised using Artemis software (figure 4.2). Importantly, no sequence data mapped to *aroC*, *aroD* or *htrA*, which all harbour large deletions in this attenuated strain (figure 4.3).

Table 4.1 Analysis of sequences mapped to the Ty2 genome

	Sequencing Data Table		
Flowcell/Lane	876/2	1104/2	1354/1
Strain	BRD948	BRD948	BRD948
rRNA Depletion	Oligo	Oligo	Oligo
Mass of Total RNA	300	100	100
Read Length	36	36	36
No Reads	5608589	7183969	5848604
Total Mapped	5438270	6513814	5356994
Total Mapped (%)	96.9	90.7	91.5
Total Uniquely Mapped	1493759	2942477	2326287
Total Mapped Uniquely (%)	0.266334189	0.409589323	0.397750814
Reads Mapped to CDS	1235932	2212650	1937241
Reads Mapped to NC sequences	257827	729827	389046
Reads mapped to pseudogenes	12403	45771	36678
Reads mapped to hypothetical genes	131242	266139	264871
GC content (ALL)	0.519742418	0.41840995	0.453535147
GC content (UNIQUE)	0.50318224	0.44267108	0.471412408
Coding Sense	1124620	1732827	1565156
Antisense Coding	111461	480504	372642

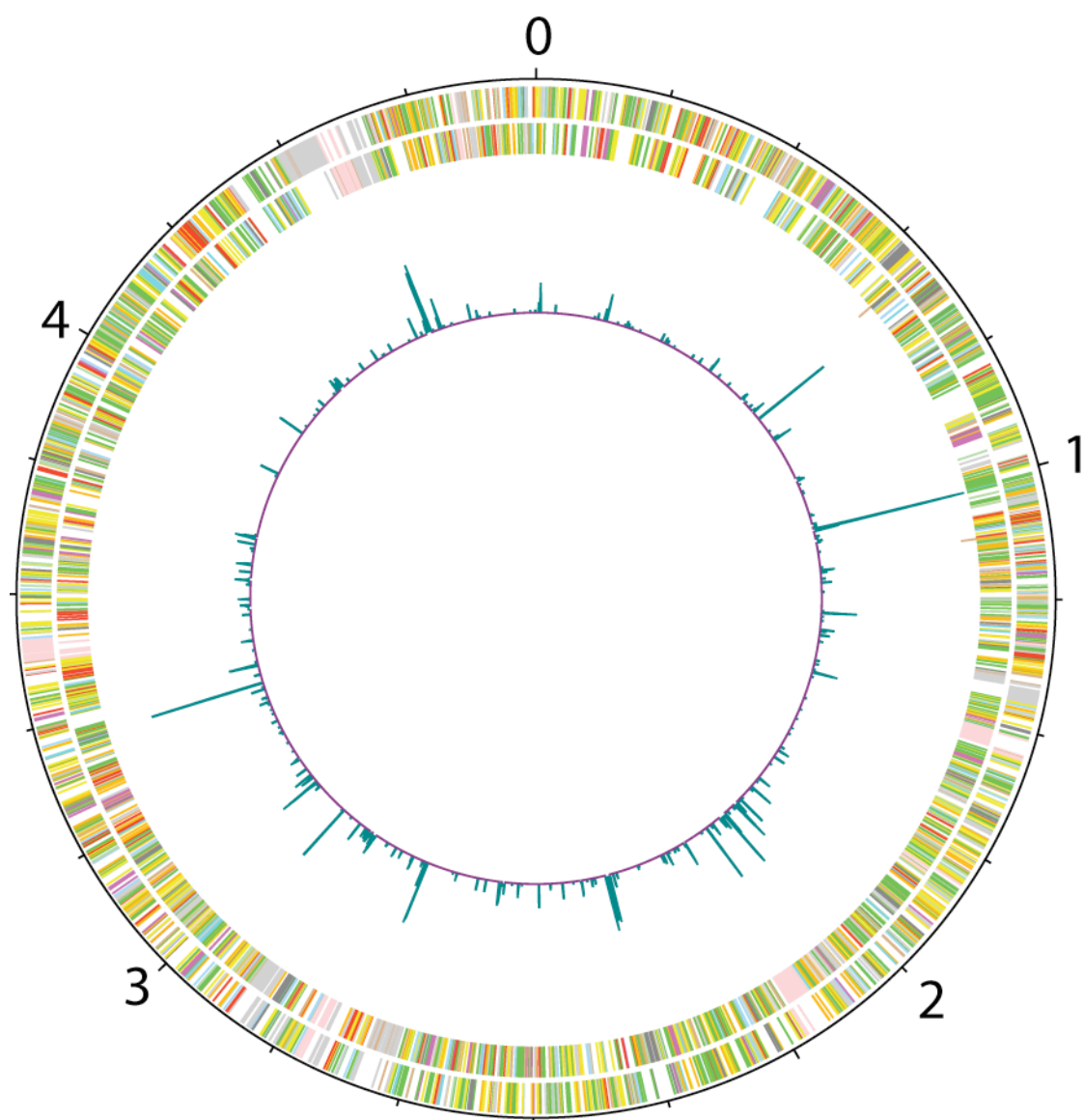


Figure 4.1 Circular plot of mapped sequence data.

Circles are described from outer to inner. The outermost represents base number in megabases, next outermost represents CDS annotated on the forward strand. The circle inside that represents CDS on the reverse strand and the innermost circle represents the plot of sequence data aligning to both strands. Each gene is coloured according to the original CT18 annotation [76] and represent different gene classes.

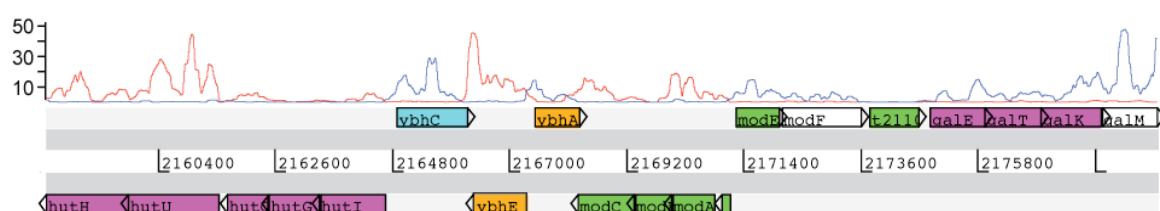


Figure 4.2 Artemis representation of transcriptome plot.

Blue plot represents forward strand and red, reverse strand. y-axis depth coverage of sequence reads.



Figure 4.3 Defined deletions in BRD948.

Coding strand traces for three RNA-seq sequencing experiments with no mapping of sequenced cDNA to deleted loci. Coloured traces represent each sequencing run, red, first; green, second; blue, third.

## 4.3.2 Coding Sequences

### 4.3.2.1 General Features

The methodology allowed a strand specific alignment of sequence, which has not previously been performed using RNA-seq. To test the genome-wide strand specificity of the sequence information, the arithmetic mean (AM) for mapped

sequence reads was determined for the coding strand of the genes currently annotated on the *S. Typhi* Ty2 genome. This value was then plotted against the AM for the putative non-coding strand (figure 4.4). 91% of the reads mapped to previously annotated *S. Typhi* Ty2 coding strand providing supporting evidence for a successful deconvolution of the strands. Further detailed evaluation of the plots revealed regions with high numbers of mapped reads consistent with annotated coding sequences. Examples of such analysis are shown in figure 4.5(a and b). Although the sequence coverage varied across each coding sequence, indicated by peaks and troughs, the profile was remarkably consistent between experiments (figure 4.6). However, many intergenic regions and 34% of the annotated coding sequences had few ( $AM < 1$ ) or no mapped reads. For sequence data mapping to a region where CDS orientation is highly “mosaic” the plots align to the annotation (figure 4.7), further illustrating the strand-specific nature of the RNA-seq data. Sequence reads that mapped to non-coding strands may represent transcriptionally active but previously unannotated features of the genome. Indeed, these data enabled identification of putative errata in the annotation of a number of, mostly hypothetical genes (figure 4.4) such as the hypothetical locus t2145. This predicted CDS mapped significant sequence data to the opposing strand, which is proximal to the 5' region of the gene *gltA* (figure 4.8). These data are also consistent with mapping of sequenced transcripts upstream of known riboswitch encoding genes such as *btuB* [200,201] and *glmS* [202].

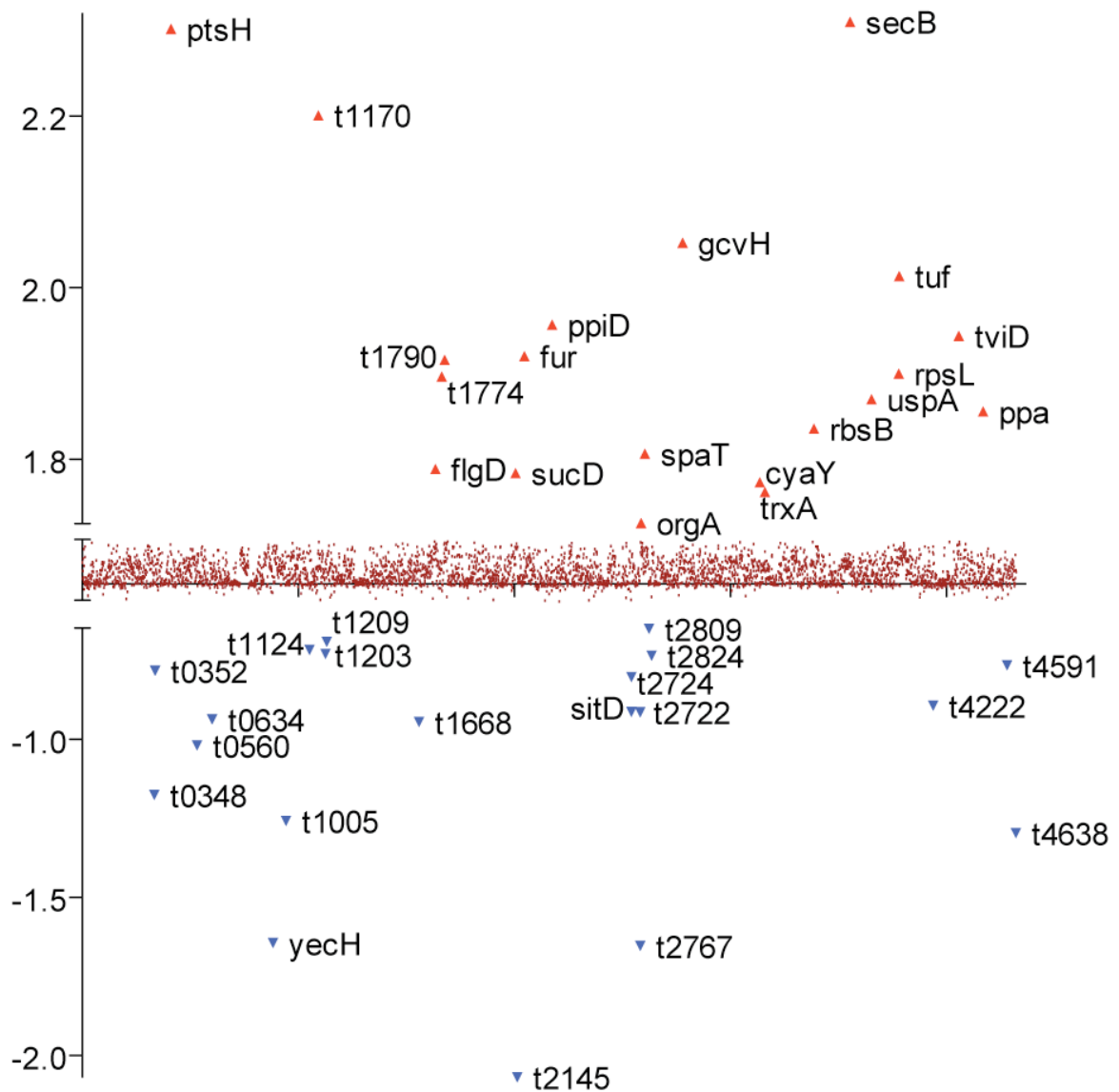


Figure 4.4 Sense and Antisense outliers.

The AM for both the coding strand and non-coding strand was determined for each annotated gene. This plot represent the  $\log(\text{AM}+1)_{\text{sense}} - \log(\text{AM}+1)_{\text{antisense}}$ . The highest and lowest 20 genes are plotted on separate y-axis scales and genes in between as plotted as burgundy.

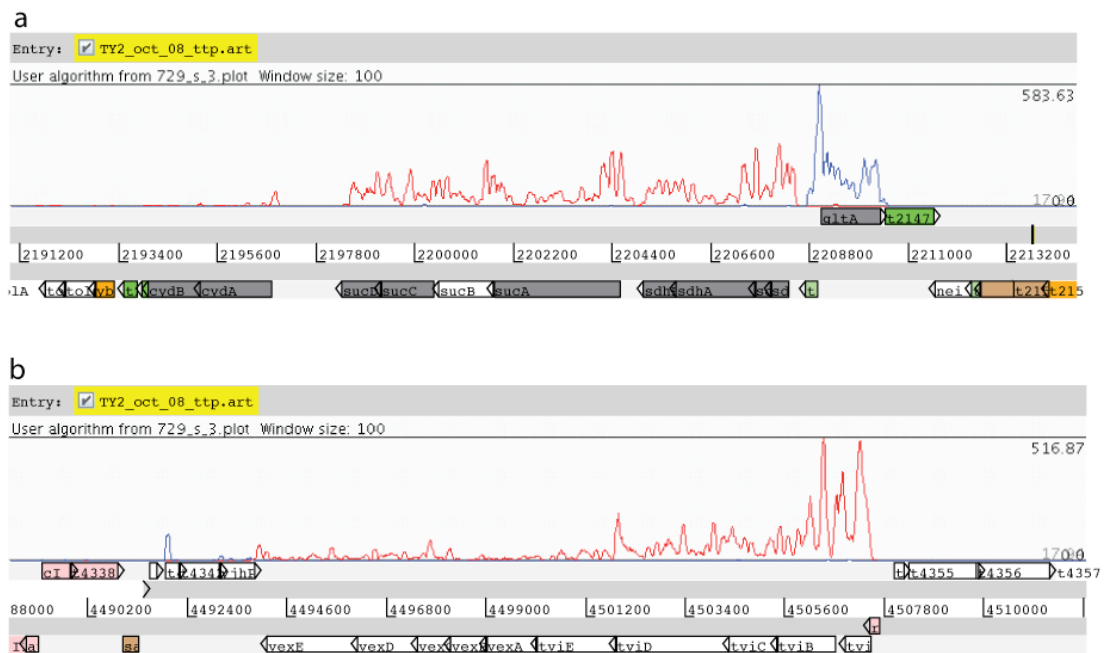


Figure 4.5. Transcripts mapping to well characterised regions.

Directional transcripts generally map to the coding strand for (a) succinate dehydrogenase operon (b) *viaB* locus. Red plot represents the reverse strand and blue, forward, window size=100bp.

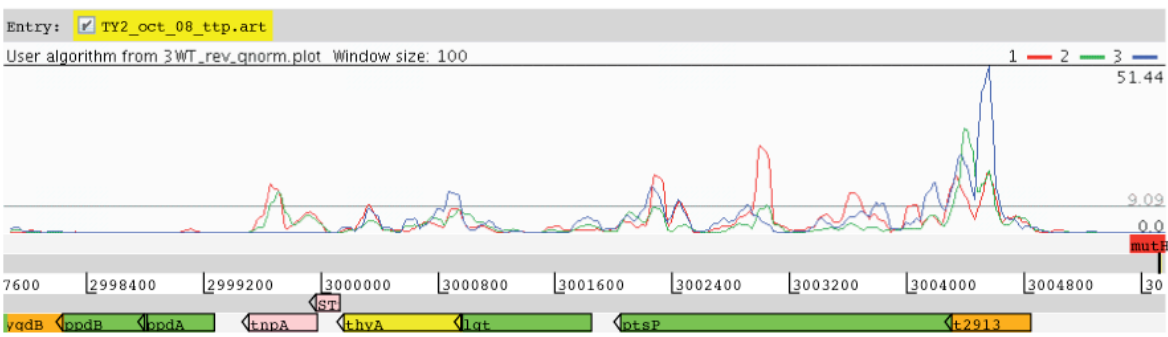


Figure 4.6 Consistency between sequencing experiments.

Colours represent sequencing run samples mapped to the reverse strand with each plot coloured, red represents the first sequencing run; green, second; blue, third.

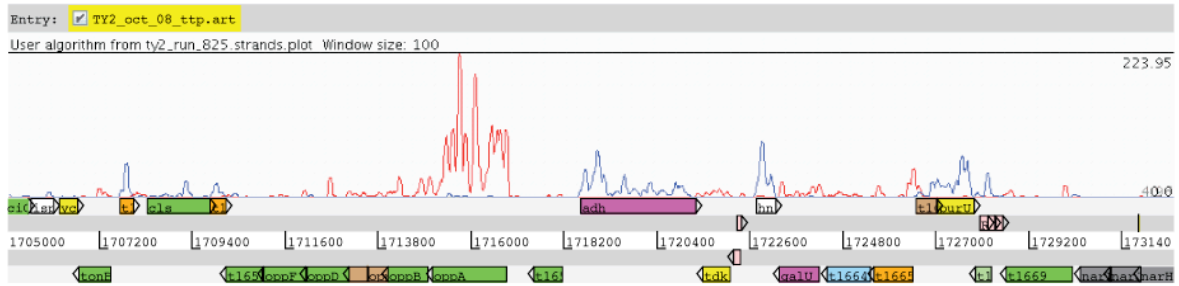


Figure 4.7 “Mosaic coding”.

Region where coding orientation is mosaic the directional sequence data is consistent with the annotation. Reverse strand, red; forward strand, blue.

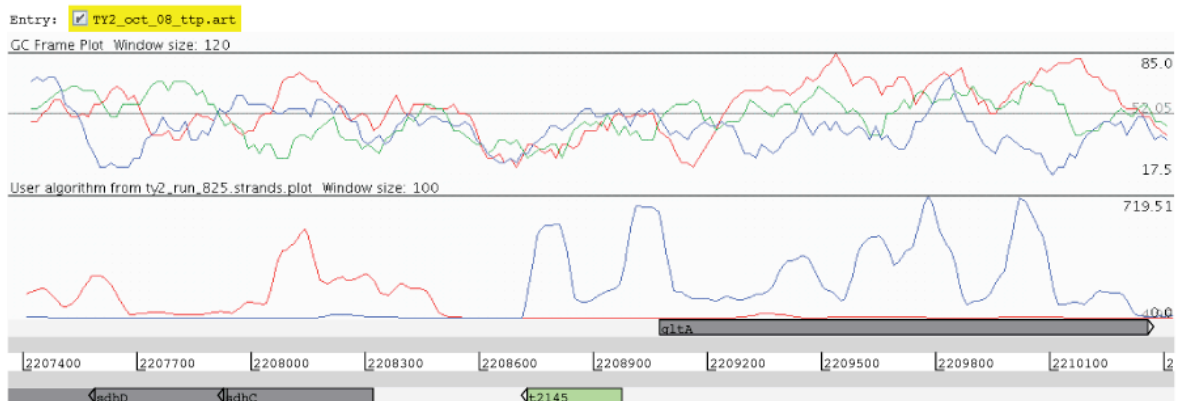


Figure 4.8. Putative error in the annotation.

Directional RNA-seq data highlights putative errata in the previous annotation. GC frame plot indicates ORF annotated as t2145 is not coding and sequence data maps to the 5' region of *gltA*, indicating a putative riboswitch.

In order to provide an overview of the gene classes identified in the genome-wide transcriptome, the AM for each CDS was determined and compared to the previously assigned functional group classification [76]. Each functional class, represented as a percentage of the total genome-wide predicted number of CDSs, was compared with proportion of each in the entire transcriptome (figure 4.9). This approach effectively identified transcriptionally active classes expressed in the mRNA populations. A ratio of  $>1$  represents a highly transcriptionally active class. The ratio for outer membrane/surface structures, regulators, conserved hypotheticals and central

intermediary metabolism were approximately 1. Interestingly, energy metabolism, pathogenicity/adaptation/chaperones and information transfer were “over-represented” in the transcriptome ranging from 1.57 to 2.23. As may be expected, transcriptionally silent prophage elements (ratio  $\sim 0.75$ ) are under represented, as are genes predicted to encode proteins required for degradation of both macromolecules and small molecules. Interestingly, pseudogenes represent 4.6% of the “coding sequences” yet a ratio of only 0.15 was observed for this gene class in the transcriptome.

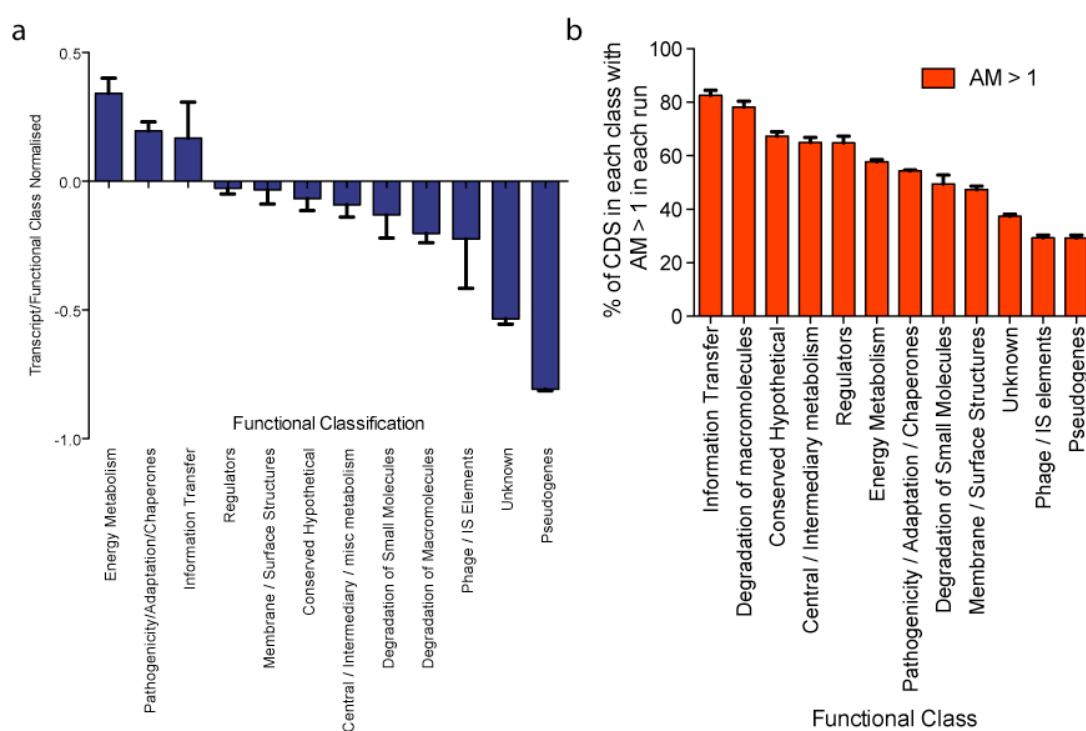


Figure 4.9. Functional classification of sequenced transcripts.

(a) The total number of reads/bp mapped to each CDS are assigned to functional class described previously [76]. These data were then normalised by the number of CDS for each function encoded within the entire genome. A ratio of 1 represents transcription of functional class on par with its genome content. A ratio of more than one represents a transcriptionally over-active class, and less than one, under-active. (b) Overview of *S. Typhi* Ty2 transcriptome assigned to functional class. The percentage of CDS in each functional class with an  $AM \geq 1$ .

The sequence coverage for the five most highly transcribed genes ranged from an AM of 996 to 596 reads/base on the sense strand. The reads mapping to large operons with

high levels of transcription generally tended to map more 5' than 3' but the previously published CDS annotation is generally supported by these data sets. As expected high coverage was observed for abundant proteins such as flagellin (*fliC*) [203] and outer membrane porin C (*ompC*) [138] as well as for genes in the TCA cycle, such as the succinate dehydrogenase operon (*sdhCDAB*, *sucABCD*) [204]. The resolution of the data is striking as even features such as transcriptional attenuators could be readily identified. For example, analysis of transcripts upstream of the threonine leader peptide, *thrL* (figure 4.10), provides a classic example of an attenuator [205]. The *thrL* peptide sequence contains eight threonine residues stimulating attenuation when threonine is abundant, as in the LB broth cultures employed in these studies.

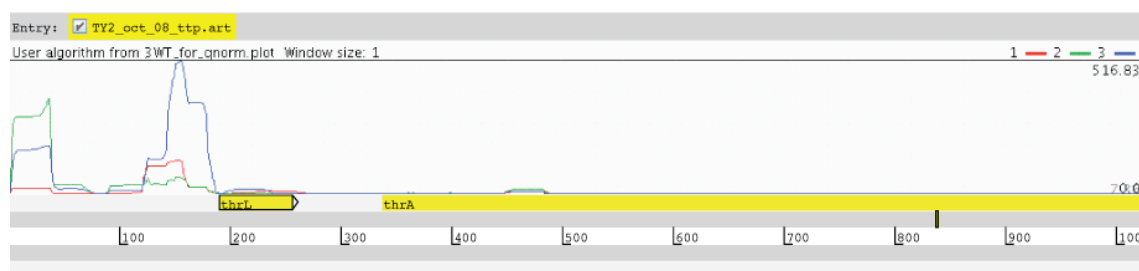


Figure 4.10. Attenuation.

RNA polymerase activity is inhibited after translation of the threonine rich leader peptide (*thrL*). LB is nutrient rich, thus reducing the requirement of threonine biosynthesis genes to be expressed. This feedback loop attenuates transcription. Run 1, red; run 2, green; run 3, blue. NB. Only data mapped to the forward strand is shown.

#### 4.3.2.2 Virulence Genes

Genes encoding pathogenicity/adaptation/chaperone, outer membrane and central metabolism proteins were generally highly expressed (figure 4.9). Under the growth conditions employed, the most significant region of highly transcribed classical virulence associated genes clustered with SPI-1. The regulation of SPI-1 expression is

complex and involves the interaction of several regulators including HilA, HilD, InvF, SprAB and OmpR [206]. Transcripts for each of these regulators were highly represented but not all transcripts of the SPI-1 needle complex translocon were equally represented (figure 4.11). For example, *spaN* (t2794) and *spaM* (t2795) had an AM of 40 nc/bp and 62 nc/bp respectively, while *spaI* (t2796) and *spaO* (t2793) that flank the latter, and are transcribed in the same operon had an AM of 12.9 nc/bp and 6.08 nc/bp respectively in one experiment. These data suggest that mRNA stability may play an important role in control of translocon expression. Furthermore, transcripts originating from genes encoding SPI-1 effector proteins encoded outside of SPI-1 were also highly expressed. For example, transcript mapping to the *sigE* (t1829, AM=32.7) and *sigD* (t1828, AM=53.5) of SPI-5 both had high sequence coverage. In sharp contrast to the relatively high coverage of SPI-1 transcripts, SPI-2 transcripts had an arithmetic-mean coverage of just 1.8/CDS (figure 4.12). This disparity was expected since it is known that SPI-2 encoded genes are up-regulated in response to environmental conditions in the *Salmonella*-containing vacuole, under the control of the *phoPQ* and *ssrAB* two component regulators (Deiwick et al, 1999). Perhaps surprisingly however, the transcripts from two genes, *sscB* (t1271, AM=16.3) and *sseF* (t1272, AM=12.5), have considerable sequence intensity.

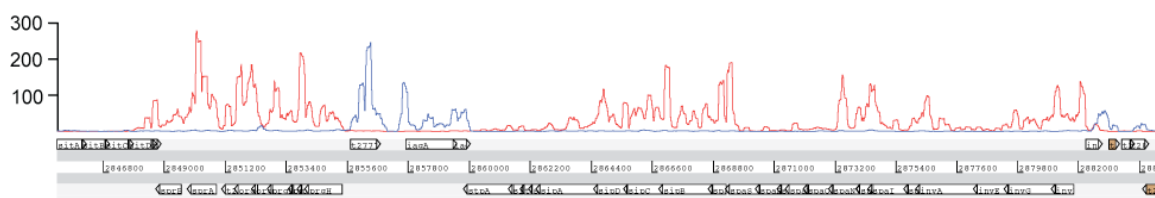


Figure 4.11 Coverage for SPI-1.

Plots represent fold-depth of sequence data mapped to each strand, forward strand, blue, red strand, reverse.



(*vexE*) to 179 reads/base (*tviA*), Ty2 is known to be highly Vi positive when grown in LB [116].

#### 4.3.2.3 Phage and putative cargo genes

*S. Typhi* harbours a number of distinct prophage, whose content can vary between the different evolutionary lineages [89,209]. Such prophages are regarded as being predominantly transcriptionally silent in the genome and can encode horizontally acquired ‘cargo’ genes potentially encoding factors that modify the virulence potential of the host bacteria. This analysis confirms that most of the resident prophage are indeed predominantly transcriptionally inactive (figure 4.13) but it is worth noting that the mapping was sufficiently sensitive to highlight low level transcription across phage regions involved in maintaining lysogeny. However, many of the prophage harbour transcriptionally active regions and some of these mapped over well known cargo or moron genes such as *sopE* (t4303, AM=186) encoded by the *sopE* phage (figure 4.13(a)). Similar analysis of this phage and others within the *S. Typhi* Ty2 genome highlights several transcriptionally active regions, which may encode novel cargo genes. Informatics analysis of these regions in some cases supports this hypothesis in that the genes do not encode known phage proteins but have functional protein similarities with genes in other pathogens such as *E. coli* 0157H7 (figure 4.13(b)), *Vibrio cholerae* (figure 4.13(c)) or the eukaryotic signalling enzymes phospholipase and putative threonine/serine kinases (figure 4.13(d)). Thus, these methodologies may provide a novel approach to identifying virulence genes expressed during the lysogenic phase.

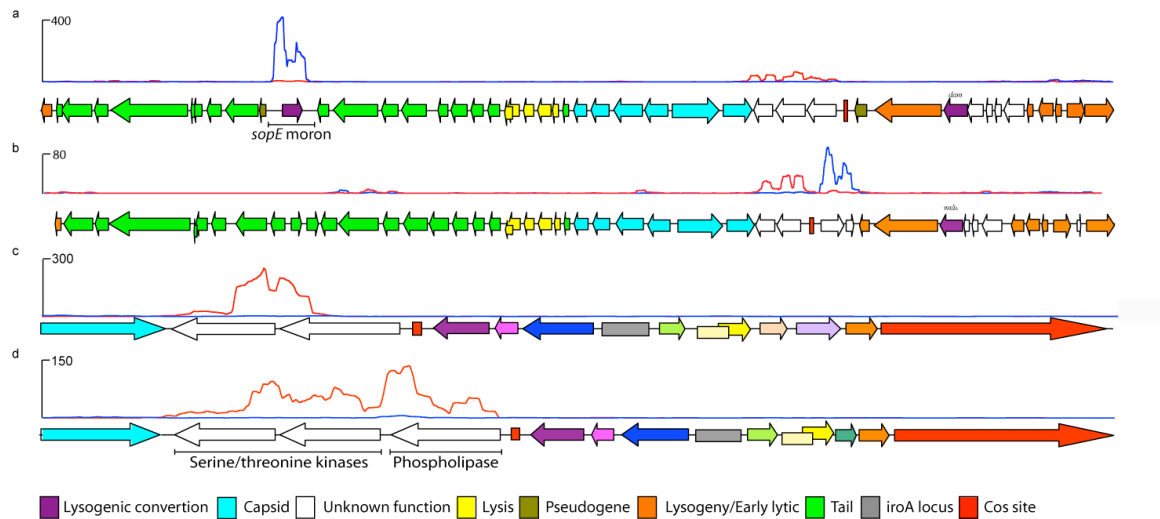


Figure 4.13. Sequenced transcripts identifying cargo genes within *Salmonella* prophage.

(a) Genetic organisation of the SopE prophage aligned with mapped sequence reads (blue forward strand, red reverse) illustrates “expression” of the *sopE* moron and a putative cargo region. (b) Genetic organisation of the ST35 prophage. The low-GC region maps significant sequence coverage compared with the prophage machinery putatively identifying it as cargo. (c) Putative ST2-27 Prophage Cargo. Low GC region maps significant sequence intensity. (d) Putative ST46 Prophage Cargo. Low GC region maps significant sequence intensity.

#### 4.3.2.4 Pseudogenes

*S. Typhi*, in common with other host-adapted pathogens, harbours a large number (~220) of putatively inactivated pseudogenes. Genome degradation may contribute to host restriction by inactivating pathways essential for infections in the non-permissive host. Theoretically, putative pseudogenes can still express a functional truncated protein domain, as for example has been demonstrated for the CTL gene encoding a toxin in *Chlamydia trachomatis* [210,211]. Based on the previous annotation there were nine pseudogenes in *S. Typhi* Ty2 that exhibited high levels of transcription, suggesting that they may be expressed as functional proteins. Interestingly, peptides that corresponded to the open reading frame upstream of the inactivating stop coding of one of the transcribed pseudogenes, *hdsM* (t4575) where detected using Mass

Spectrometric analysis of *S. Typhi* Ty2 extracts (our unpublished data). This represents the only evidence in this study of translated pseudogenes. This combined with the sheer lack of transcriptional abundance of other *S. Typhi* pseudogenes further supports the current interpretation that these genes are no longer active.

#### 4.3.2.5 Hypothetical Genes

Many genes in the *S. Typhi* genome were initially annotated as hypothetical coding sequences in the absence of any direct evidence for transcription or translation into a protein product. A considerable number of these hypothetical genes were also identified as orthologues in the genomes of *E. coli*, and other bacterial species. The annotation derived by Parkhill *et al* (2001) assigned each predicted coding sequence to a functional class according to *in silico* analyses. Thus, all genes that were annotated as putative, probable, predicted, hypothetical and possible were included in this analysis (appendix 9.5). This analysis encompassed ~2900 genes. 72 genes mapped no sequences for all three experiments and 677 genes had an average AM ranging from 0.01 to 0.10. 1751 genes had an average AM from 0.11 to 1.00 and 293 genes mapped an average AM ranging from 1.01 to 10.0. 75 genes ranged from an average AM 10.0 to 730.

#### 4.3.2.6 A *S. Typhi* Ty2 Specific Insertion

Annotation performed following the sequencing of the *S. Typhi* Ty2 genome failed to identify a variable region that mapped significant transcriptome sequence data [136]. This region, now referred to as ST20, shows variation between *S. Typhi* isolates and is composed of a mosaic of up to three known inserts, a, b and c [87]. *S. Typhi* strains CT18 and Ty2 encode forms a and b respectively. The synteny of this island is

specific to Ty2 and maps significant transcript sequence data (figure 4.14). This region, which encodes 27 putative CDSs in Ty2 mapped significant numbers of sequence reads. The highly transcribed genes are functionally homologous to two restriction enzymes, a negative regulator of *N*-acylhomoserine and protein with no significant sequence homology.

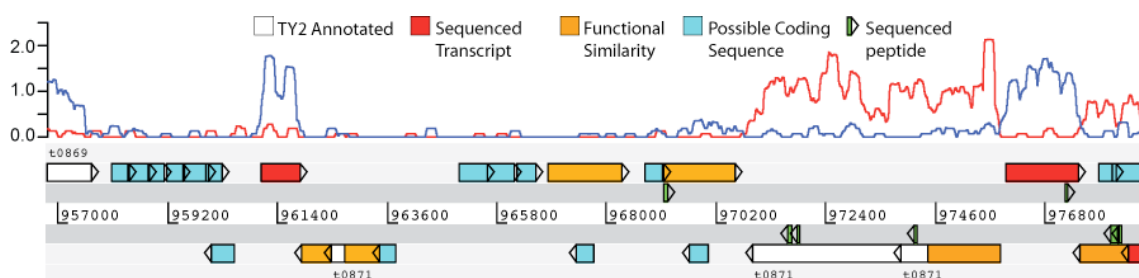


Figure 4.14. Third Party Annotation of the Ty2-specific Insertion.

Alignment of transcriptome sequence data to this previously poorly-annotated region of Ty2 represents further evidence for putative CDS. All possible CDS's are coloured blue, those with predicted functional protein homology are mustard, white are previously annotated and red annotated by transcriptome sequence data. Green features are mapped peptide sequences (FDR<0.076), which map to the same translation frame as the aligned

### 4.3.3 Non-Coding Sequences

The AM for each currently annotated sncRNA was determined. 67 of the known 151 have an average AM > 1, ranging from 1.17 to 1004 reads/base-pair (figure 4.15a). Furthermore, many of the transcripts identified by sequencing mapped back to regions of the *S. Typhi* Ty2 genome that were previously unannotated, predicting a further 40 regions as expressed non-coding sequences. Furthermore, many of these were not unique within the Ty2 genome and similar sequences were annotated as paralogues to include in this analysis. Furthermore, 127 CDS were identified that were preceded by putative 5'UTR transcripts, 31 of which were more than 150bps in length (appendix 9.6). There were also two novel putative 3'UTRs adjacent to *sprB* and *ramA*

respectively. Subsequently, we determined the AM for each prediction (figure 4.15b) and 85 of the 239 elements had an average AM  $> 1$ . Taken together, these sequence data suggests that there may be many previously unidentified functional non-coding RNAs present in *S. Typhi*, and potentially in other bacteria. Consequently, bioinformatics analyses were used to further interrogate these data [212].

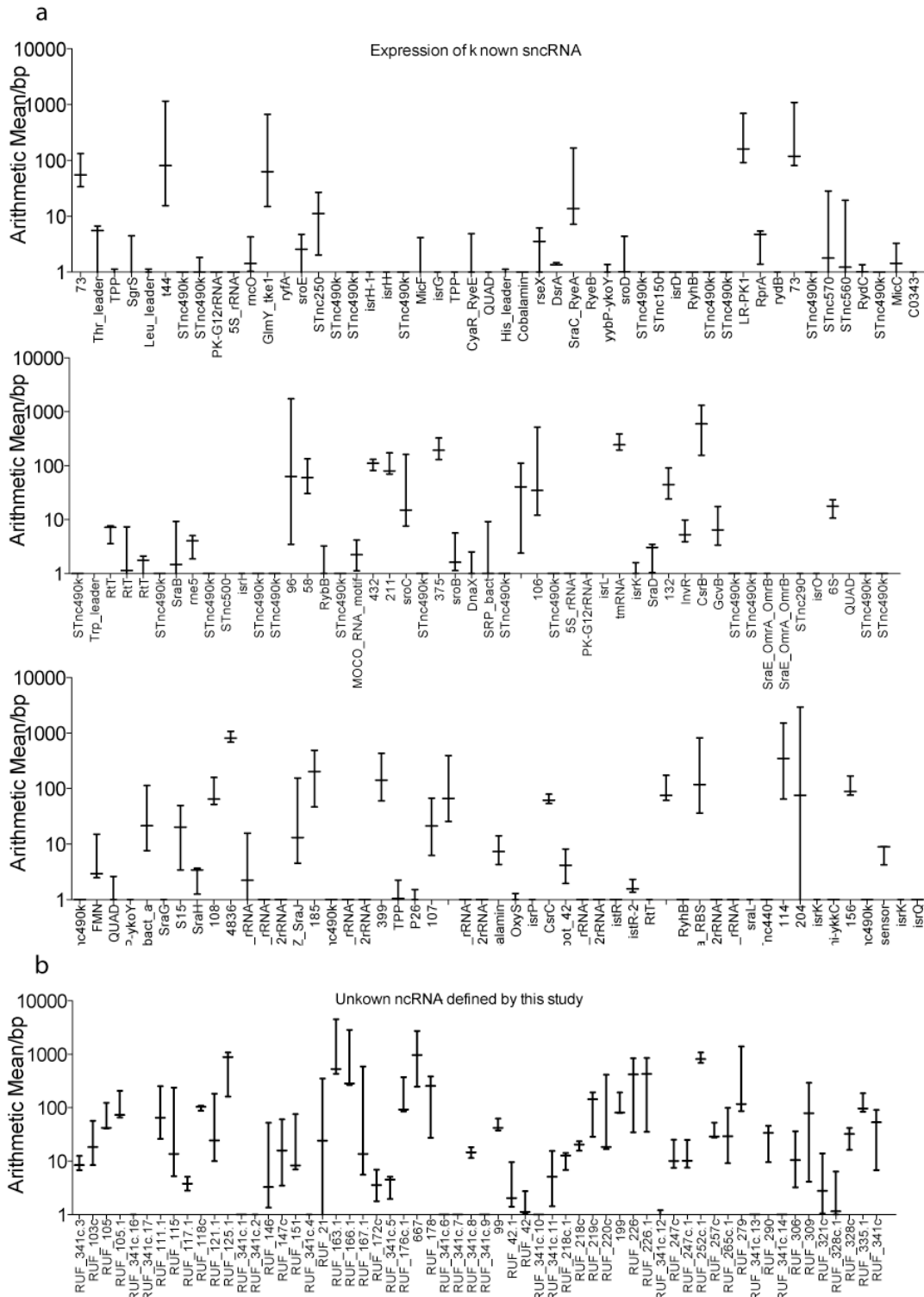


Figure 4.15 Coverage of putative non-coding RNAs.

(a) Sequence coverage (AM) of ncRNAs in each sequencing run, derived from previous annotation and publication. (b) Predicted ncRNA based on sequenced transcripts mapping to previously unannotated intergenic region or paralogous regions of ncRNA derived by this study.

Transcription and translation in prokaryotes is commonly regulated by changes to the conformational structure of *cis*-acting non-coding RNAs called riboswitches. These RNA's generally bind metabolites related to the function of each downstream gene [213,214] and have been identified bioinformatically based on sequence conservation of the 5' UTR. Several known riboswitches, such as *btuB* [200], *glmS* [202] and TPP [215] were highly represented in these data. Possibly one of the most interesting regions encoding novel non-coding RNAs was part of SPI-1 (figure. 4.16). Two of these SPI-1 associated transcripts were identified by the programme RNAz as candidate riboswitches, here designated SPIS1 and SPIS2 (RUF220c and RUF219c) (4.16b and c). SPIS1 and SPIS2 are located directly upstream of the AraC-like regulators *sprA* (t2988) and *sprB* (t2987), respectively. The third candidate element which is predicted to be a 3'UTR, named SPIS3 (RUF218c) (figure 4.16), is antisense to the *sitD* gene, an iron transport protein [216] and a hypothetical protein O30622 (t2767), which may have been acquired independently of the rest of SPI-1 [12]. The sequence of RUF218c is conserved across cyanobacteria, firmicutes and proteobacteria. The *sitA* gene maps sequenced transcripts (average AM=1.27), whereas *sitB*, *sitC* and *sitD* have slightly lower levels of expression (AM = 0.37, 0.38 and 0.61, respectively). It is possible that RUF218c is an antisense repressor of these proteins as it is predicted to form a moderately stable MFE secondary structure compared to a shuffled ensemble of sequences that have the same di-nucleotide composition (p=0.0090). The fourth candidate element, named SPIS4 (RUF221) maps to the 5' UTR of *iagA* (t2999) (figure 4.16), an invasion protein regulator. The structure of this RNA (34% G+C) is not predicted to be significant (RNAz p=0.0037 and shuffling p=0.2627) (figure 4.16c).

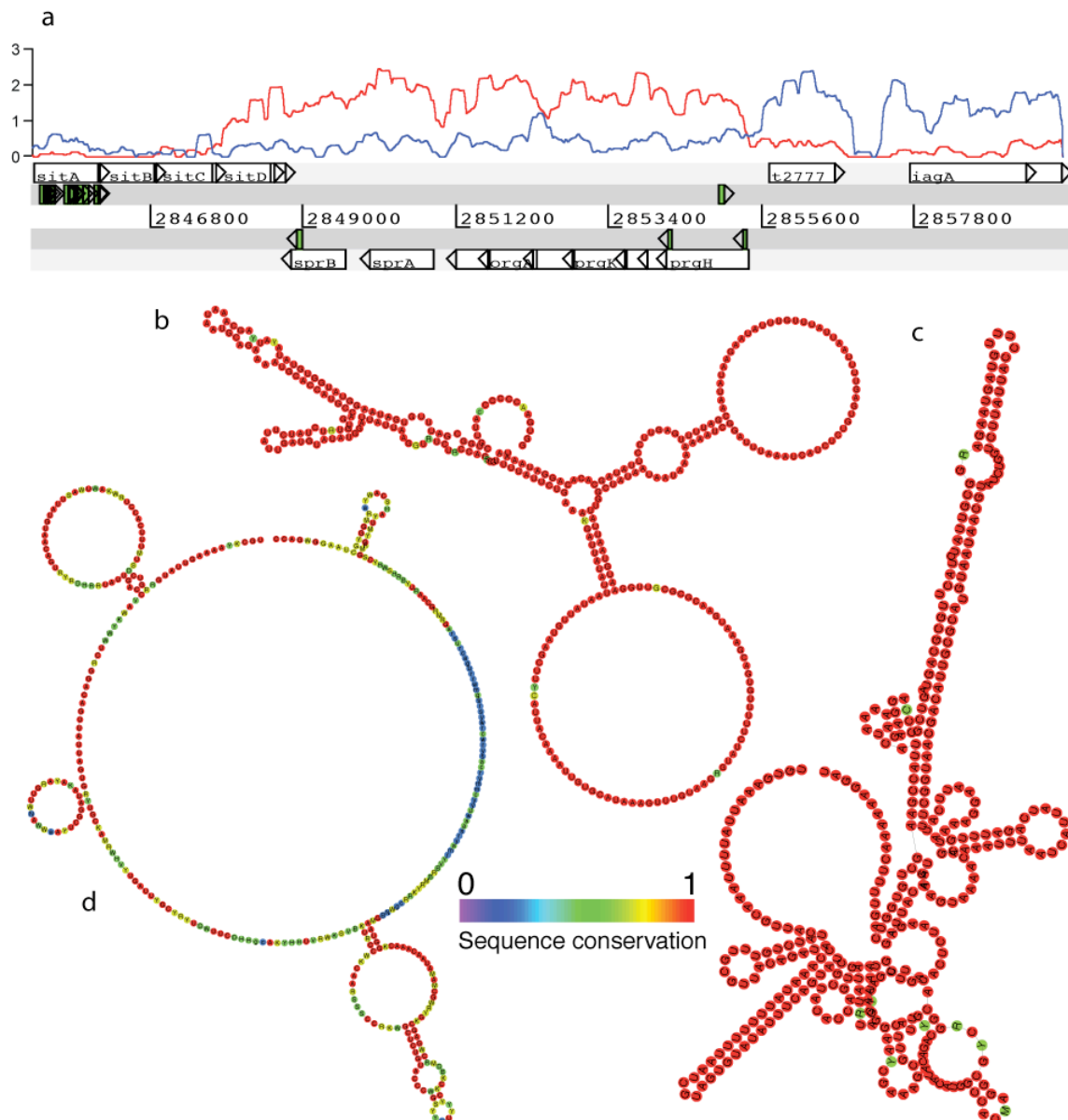


Figure 4.16. SPI-1 Sequence coverage

(a) Artemis representation of part sequenced transcript mapping to the highly transcribed region SPI-1. Log(10) +1 of Forward strand plot, blue; Log(10)+1 of reverse strand; red. SPIS1 is region upstream of *sprA*; SPIS2 is region upstream of *sprB*; SPIS3 is region downstream of *sprB*; SPIS4 is region upstream of *iagA*. Dark grey horizontal line represent translation of each CDS and green ORFs represent sequenced peptides mapping to those CDS. (b) Predicted secondary structure of data mapped to region 5' of *sprA* (p(RNAz)= 0.933655). (c) Predicted secondary structure of region 5' to *sprB* (p(RNAz)= 0.809987). (d) Predicted secondary structure of region 5' to *iagA* (p(RNAz)=0.003772).

A further non-coding feature of note is RUF107c, which is highly expressed in these *S. Typhi* samples. This element, predicted to be highly structured by RNAz

( $p=0.9396$ ), has approximately 115 paralogues in *Salmonella* (figure 4.17). Further, it is conserved across ~82 bacterial species but is chiefly restricted to Enterobacteriaceae. The genomic context of RUF107c is not consistent with a *cis*-regulatory or a transposable element, as the sequence does not consistently co-occur with either CDSs or near transposases, respectively.

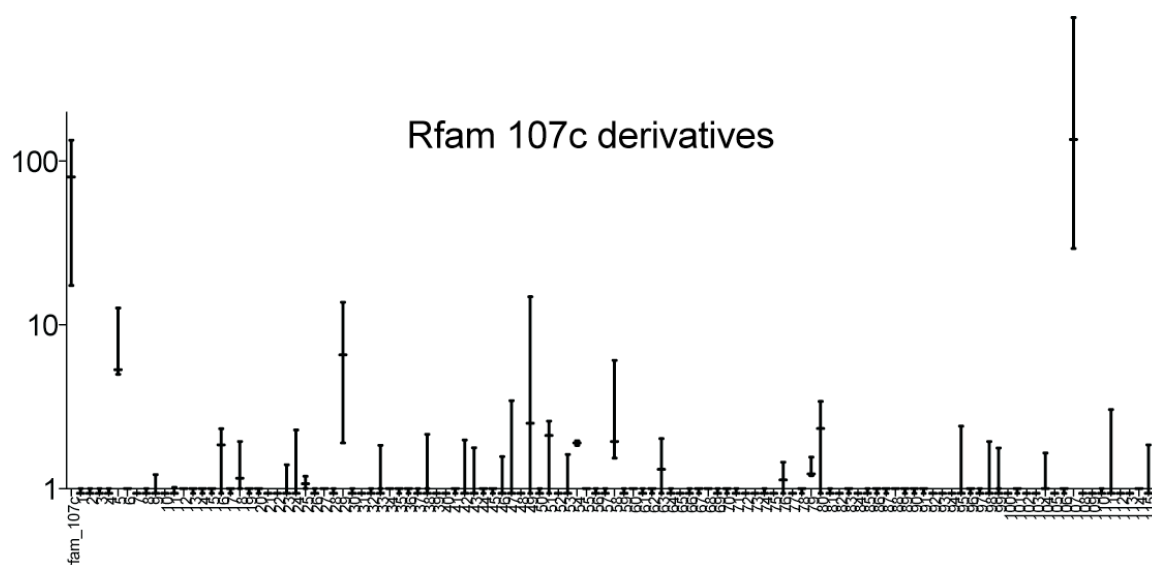


Figure 4.17. Paralogues of RUF107c.

AM of sequence coverage for predicted RUF107c paralogues in *S. Typhi*.

#### 4.3.4 Transcriptome comparison with *ompR*-null mutant

##### 4.3.4.1 Quality control

The validation of the first RNA-Seq experiments using *S. Typhi* Ty2 BRD948 prompted further experiments utilising similar RNA prepared from an *ompR* mutant of the same strain. The OmpR/EnvZ system is a key two component regulator controlling the expression of virulence phenotypes by *S. Typhi*. Hence, RNA-seq analysis could potentially reveal some novel aspects of OmpR mediated expression regulation.

The data described here are the product of a number of Illumina sequencing runs carried out during the development of protocols designed to deplete RNA preparations of 16s and 23s and provide high-throughput strand specific sequence information. Mapping statistics for each run and starting material are detailed in table 4.2.

Table 4.2. Sequence data mapped for each experiment

	Sequencing Data Table					
Flowcell/Lane	876/2	1104/2	1354/1	876/5	1354/2	1104/3
Strain	BRD948	BRD948	BRD948	BRD948DompR	BRD948DompR	BRD948DompR
rRNA Depletion	Oligo	Oligo	Oligo	Oligo	Oligo	Oligo
Mass of Total RNA	300	100	100	300	100	100
Read Length	36	36	36	36	36	36
No Reads	5608589	7183969	5848604	5375891	3098524	7399877
Total Mapped	5438270	6513814	5356994	5249193	2774513	6454861
Total Mapped (%)	96.9	90.7	91.5	97.6	89.5	87.2
Total Uniquely Mapped	1493759	2942477	2326287	1749240	610817	2119151
Total Mapped Uniquely (%)	0.266334189	0.409589323	0.397750814	0.325386062	0.197131602	0.286376517
Reads Mapped to CDS	1235932	2212650	1937241	1500449	491028	1617543
Reads Mapped to NC sequences	257827	729827	389046	248791	119789	501608
Reads mapped to pseudogenes	12403	45771	36678	14126	17185	50531
Reads mapped to hypothetical genes	131242	266139	264871	248791	78606	275976
GC content (ALL)	0.519742418	0.41840995	0.453535147	0.53065418	0.496200834	0.392659886
GC content (UNIQUE)	0.50318224	0.44267108	0.471412408	0.517455104	0.479285167	0.418839576
Coding Sense	1124620	1732827	1565156	1311837	295552	1056574
Antisense Coding	111461	480504	372642	188739	195794	561813

Genomic DNA contamination is evident in samples 876\_s\_1.wt and 876\_s\_3.mut when the data are plotted in Artemis (not shown). The histogram (figure 4.18(a)) or box-plot (figure 4.18(b)) also illustrates the differences between the runs with contaminating genomic DNA and cDNA. Further, samples were then prepared on a total of three occasions and the total RNA isolated ranged from 300µg to 100µg as the protocol was optimised. After mapping of read data with a quality score of 30 using MAQ, the AM for each CDS was determined and the data normalised by quantile normalization [187]. Due to inconsistencies in the box-plot and DNA contamination discovered during Illumina sequencing a subset was chosen for further analysis (figure 4.18(c)). Clustering of these sequence data intensities revealed each sample is

more closely related to the sample on the same flowcell rather than the strain it was isolated from. However, the comparative analysis was performed in order to assess the power of these experimental data and methods.

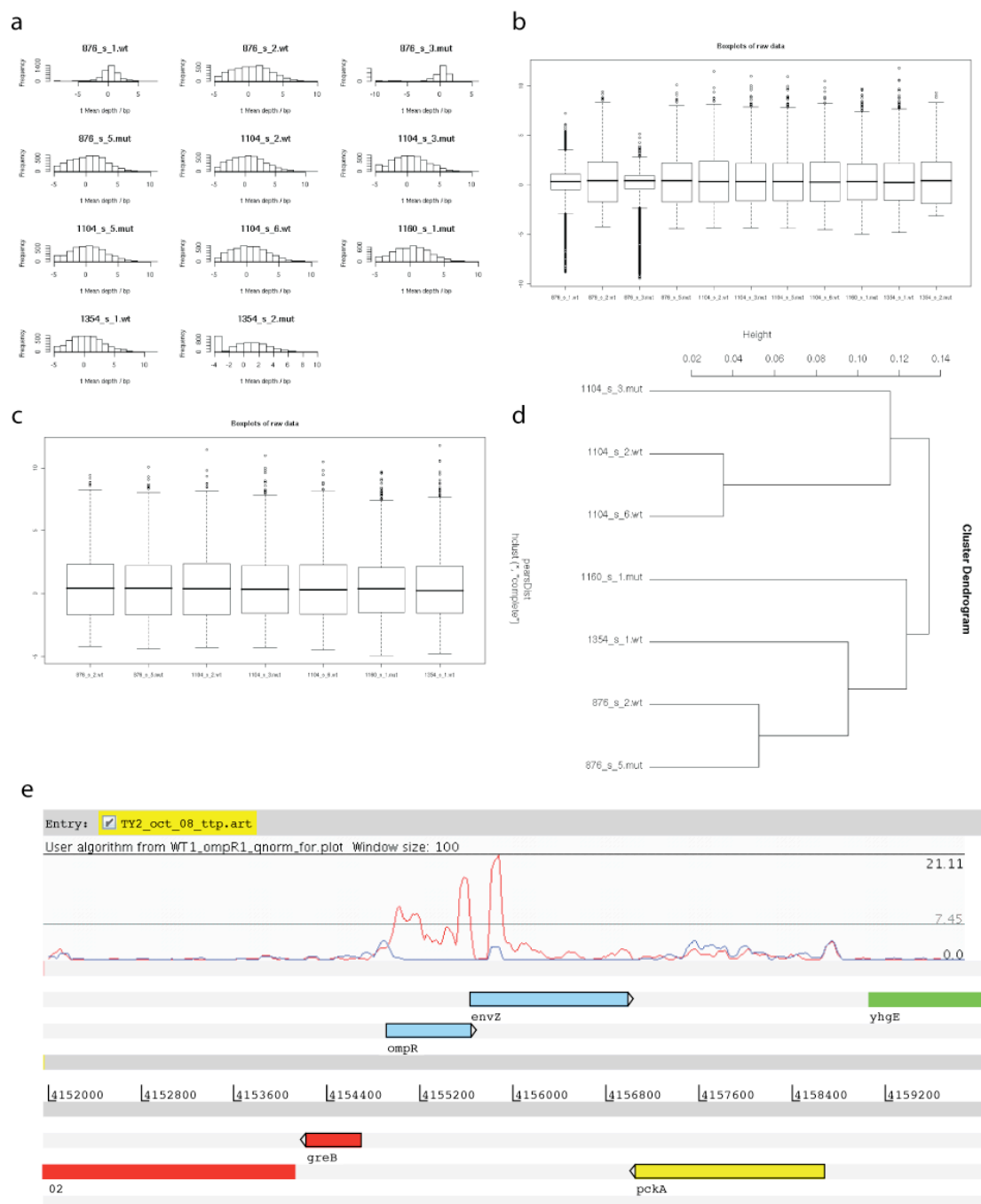


Figure 4.18 Quality control.

(a) Histogram of sequence data mapping to coding sequences (log2) (b) Box plots of depth over coding sequences (c) A subset chosen for further comparative analysis after quantile normalisation of the sequence data mapping to coding sequences (d) Pearson correlation clustering of subset. (e) OmpR deletion. No sequence data are mapped to the region deleted in *ompR* (blue), WT (red).

#### 4.3.4.2 General features of the BRD948 *ompR* transcriptome

Initially the data derived from the transcriptome of the BRD948 *ompR* mutant were mapped to each of the functional gene classes described by Parkhill *et al* (2001) (figure 4.19). Overall, as expected these data were similar between the wild-type and *ompR* mutant derivatives. OmpR is known to strongly activate expression of the *viaB* locus [118] and this region maps a very much reduced sequence data in the null-mutant (figure 4.20). Expression of *ompC* is also strongly activated by the presence of OmpR, and very few transcript reads mapped to this region in RNA seq of the *ompR* strain (figure 4.21).

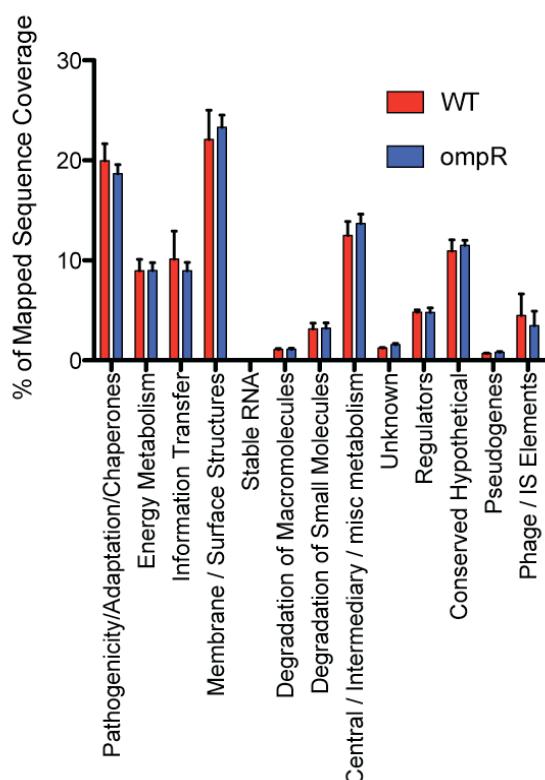


Figure 4.19 Assigning transcriptome to functional class.

Data mapped to the functional class derived by Parkhill *et al.* (2001) for CT18 to determine the overall AM/bp for each class.

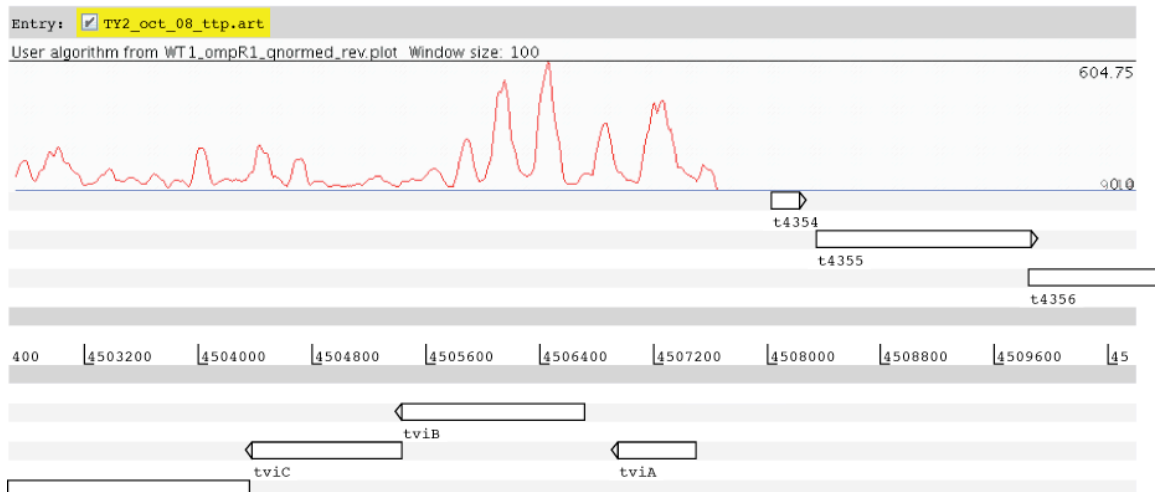


Figure 4.20. Start of the *viaB* locus.

Sequence data mapping to the coding strand of the *viaB* locus is considerably less in the *ompR* mutant (blue) compared with the WT (red).



Figure 4.21 Outer membrane porin C.

The transcript sequenced and mapped to the coding strand in the *ompR* mutant (blue) is clearly less abundance than WT (red).

#### 4.3.4.3 Quantified differences

The AM was determined for all annotated CDSs and these values were taken as intensity values much like values derived by microarray scanning. Using the AM and

the LIMMA package for microarray analysis the data were normalized by quantile normalization [187]. Quantile normalization attempts to match each percentile of each data set. This analysis identified 15 genes (adj  $p < 0.05$ ) that were differentially transcribed in the *ompR*-null mutant, none of which were significantly increased (table 4.3). These data were consistent with OmpR acting as a transcriptional activator [217].

Table 4.3 Statistically different genes in the *ompR* mutant

Gene Name	Start Co-ord	Stop Co-ord	logFC	t	P-value	adj P-value	B	Product
<i>tviC</i>	4504378	4505424	-8.17	-14.41	5.10E-06	4.24E-03	4.52	Vi polysaccharide biosynthesis protein, epimerase
<i>tviA</i>	4506949	4507488	-8.05	-8.72	1.02E-04	3.15E-02	2.01	Vi polysaccharide biosynthesis protein
<i>tviB</i>	4505427	4506704	-8.03	-13.21	8.63E-06	4.66E-03	4.13	Vi polysaccharide biosynthesis protein, UDP- glucose/GDP- mannose dehydrogenase
<i>tviD</i>	4501859	4504354	-7.09	-16.64	2.12E-06	4.24E-03	5.12	Vi polysaccharide biosynthesis protein
<i>vexB</i>	4498207	4499001	-6.98	-13.36	8.07E-06	4.66E-03	4.18	Vi polysaccharide export inner-membrane protein
<i>vexC</i>	4497486	4498181	-6.89	-9.52	6.11E-05	2.03E-02	2.49	Vi polysaccharide export ATP-binding protein
<i>vexA</i>	4499011	4500078	-5.73	-14.28	5.38E-06	4.24E-03	4.48	Vi polysaccharide export protein
<i>tviE</i>	4500123	4501859	-5.71	-14.07	5.89E-06	4.24E-03	4.41	Vi polysaccharide biosynthesis protein TviE
<i>vexE</i>	4494169	4496139	-5.62	-14.96	4.05E-06	4.24E-03	4.68	Vi polysaccharide export protein
<i>vexD</i>	4496159	4497463	-5.50	-11.28	2.23E-05	8.77E-03	3.36	Vi polysaccharide export inner-membrane protein
<i>hyaA</i>	1500712	1501815	-5.02	-14.36	5.20E-06	4.24E-03	4.50	uptake hydrogenase small subunit
<i>NA</i>	1502243	1503334	-4.82	-12.29	1.33E-05	6.39E-03	3.79	putative secreted hydrolase
<i>envZ</i>	4155629	4156981	-4.17	-11.09	2.47E-05	8.87E-03	3.28	osmolarity sensor protein
<i>slsA</i>	3873860	3874540	-3.63	-8.42	1.25E-04	3.60E-02	1.82	hypothetical protein
<i>NA</i>	1701305	1701808	-3.39	-11.50	1.98E-05	8.57E-03	3.46	hypothetical protein

The entire *viaB* locus is represented in these data, which agrees with Pickard *et al.* [116]. As expected, *envZ* is also significantly decreased. This is an important endogenous control as it is downstream of the defined mutation in *ompR*. The *ompR* gene is the first gene in the *ompB* locus and is not represented in these data as it was filtered out by the false discovery rate Benjamini-Hochberg correction. Benjamini-Hochberg correction is used as a post-analysis estimation to allow for false discovery

based on multiple analyses and variation between experimental testing [218]. However, this sort of correction also includes a level of false non-discovery rate (FNR) and the sheer lack of *ompR* reads in these data suggests such an estimation may be too stringent for this experiment.

Interestingly, there are genes in this dataset that were not previously described as *ompR* regulated. The *slsA* gene (12.3 fold down, adj  $p < 0.036$ ), encoded within SPI-3, is conserved throughout *Salmonella* and sequence identity suggests it is an inner membrane protein with homology to the isochorismatase hydrolase family of enzymes. Isochorismatase hydrolase has been characterised in the phenazine biosynthesis pathway in *Pseudomonas aeruginosa*. Phenazines are a group of 70 related compounds, some of which are antimicrobial and others induce neutrophil cell death [219]. The function of the *slsA* protein is not known so the consequence from its reduced expression remains to be elucidated.

The hydrogenase uptake gene, *hyaA2* is decreased (32.4 fold down, adj  $p < 0.004$ ). *Salmonella* encodes three predicted hydrogenase operons, two hydrogenase 1 operons (*hyaACDEFt1048* and *hyaA2B2C2D2E2F2t1454*) and a hydrogenase 2 operon (*hybOABCDEFG*) that are important factors in respiration. Interestingly, both *hyaA* and *hyaB2* are pseudogenes in *S. Typhi* Ty2 and CT18. The hydrogenase protein complex consists of Hya or Hyb subunits and enzymatically splits  $H_2$  to release electrons to reduce downstream components in the respiratory chain and all three of these operons contribute to virulence in the murine model [220]. Deletion in all three severely attenuates *S. Typhimurium*. Furthermore, the gene divergently transcribed from *hyaA2*, a putative secreted choloylglycine hydrolase, t1459 is also significantly decreased (28.0 fold down, adj  $p < 0.006$ ). The family of choloylglycine hydrolases

cleave carbon-nitrogen binds, exclusive of peptide bonds and includes conjugates bile acid hydrolase and penicillin acylase [221].

Expression of the conserved hypothetical gene, t1641, that exhibits no significant identity to any characterised gene but is conserved in *E. coli* K12, (annotated as yciF) is down (10.4 fold down, adj  $p < 0.009$ ).

#### **4.3.4.4 Quantified differences pre-Benjamini Hochberg correction.**

There were 305 genes with 2-fold differences ( $p < 0.05$ )(appendix 9.7) in these data prior to Benjamini-Hochberg false discovery rate estimation and correction. Ninety-nine of these genes were decreased in expression and 49 genes had one or more contiguous genes differentially regulated, suggesting they are encoded by an operon structure. These data set may provide useful information if corroborated by a different gene expression method as it contains genes reported previously to be *ompR* regulated in *Salmonella*, such as *ssrAB*, *ompC* and *ompS*. Furthermore, many of the genes predicted to be in an operon structure were differentially expressed with consistent fold-change direction.

## **4.4 Discussion**

### **4.4.1 General transcriptome results**

Overall, application of this Illumina sequencing technology has provided a detailed insight into the entire transcriptome of *S. Typhi* at one particular phase of growth. This technique has proven to be reproducible between experiments and has identified

many previously described features. However, the most interesting part of any study such as this, is the identification of novel expression features. This experiment has mapped sequence data to hypothetical open reading frames and has been able to identify putative errata in the existing annotation. Such information may be crucial for rationalising further work on *Salmonella* and possibly *E. coli*.

Identification of novel features specific to *Salmonella* has in the past permitted identification of virulence mediators. A global survey of the transcriptome on this scale *in vitro*, will unfortunately not identify such candidates without further *in vivo* experimentation, however, it will provide a rationale for further work. Combined with the comparison of *S. Typhi* with the *S. Typhimurium* and/or *E. coli* genomes, it is possible to further define variation and consequently highlight determinants potentially involved in pathogenicity.

Possibly the most interesting and exciting data are the transcripts that mapped to four intergenic regions of SPI-1. SPI-1 is predicted to be one of the oldest horizontally acquired regions of the *Salmonella* genome and this region does not exhibit any CDS degradation or intergenic region degradation within *Salmonella*. Such conservation suggests these intergenic regions are as functionally important as the coding sequences and the recent discovery of riboswitches underpins this logic. Investigation of possible ligands that bind to these 5'UTRs should further elucidate mechanisms of SPI-1 regulation. Currently, defined riboswitches are limited to non-coding regions and are generally predicted bioinformatically based on a significant gap between CDS and conservation. It is difficult to determine conservation for *Salmonella* specific islands, however, the identification of such transcribed regions in SPI-1 may permit

rational identification of such regions present in SPI-2, which also contains a significant number of intergenic regions.

Interesting work on riboswitches and the ligands that bind to alter their conformation has identified an antimicrobial compound pyrithiamine pyrophosphate, which inhibits binding of thiamine pyrophosphate and induces cellular dysfunction [222]. Identification of a natural ligand for *Salmonella* specific riboswitches may then facilitate the identification of synthetic ligands with microbicidal activity, thus permitting selective killing that would potentially not affect the overall microbiota. Indirect disruption of the normal flora through antimicrobial therapy has been shown, in many cases, to cause more harm than the targeted infection [223].

#### 4.4.2 OmpR comparison

Due to the lack of any precedent for the use of RNA-seq technology in bacteria, the methods and protocol were continuously modified during the course of these experiments in an attempt to optimize the approach. Cost was also a factor driving some of the changes.

Nagalakshmi *et al.* [188] and Wilhelm *et al.* [189] published the first method for transcriptional analysis by RNA-seq approximately three months after this experiment was designed. These techniques were used to sequence the more stable polyadenylated RNA of eukaryotes after enrichment using oligo(dT). Furthermore, Marioni *et al.* [224] found that during an analysis of technical reproducibility the cDNA concentration increased the variability of results. The method has undergone multiple variations and iteration throughout development and relies on a considerable

amount of RNA manipulation. This may limit the scope for definitive quantitative analysis.

The parent strain, BRD948, is deficient in part of the stress response (*htrA*) mechanism. The serine protease, HtrA, has been associated with virulence [72] and is homologous to a heat shock protein in *E. coli* involved in protein degradation and fidelity [225]. The aromatic amino acid biosynthesis pathway in *S. Typhi* BRD948 has also been disrupted promoting a need to scavenge *para*-aminobenzoic acid and dihydroxybenzoate [226], which is supplemented in all growth media. All comparisons are between isogenic strains (plus or minus the defined mutation in *ompR*), thus nullifying any effect these modifications may have on the global transcript.

This study was able to identify genes currently reported as being under the control of OmpR as well as extending current information on this regulon. Pickard *et al* (1994) demonstrated the *viaB* locus of *S. Typhi* is OmpR regulated and a deleting mutation arrested Vi expression. Expression of the Vi genes was found to be significantly decreased in the *ompR* mutant using RNA-seq analysis. Interestingly, *ompC* (figure 4.21) is highly activated by the presence of OmpR but is not found in the significantly different gene list. Under these conditions the OmpR regulon apparently includes the SPI-3 gene *slsA*, a putative hydrogenase gene, a putative choloylglycine hydrolases and a conserved hypothetical gene.

Improvements could be made to the experimental approach, based on the variation identified in these experiments and the observations published by Marioni *et al.* [224]. For example, ideally similar concentrations of cDNA should be prepared for a comparative experiment submitted to the same flowcell for sequencing.

## 4.5 Conclusion

The development of an RNA-seq approach to the analysis of the *S. Typhi* transcriptome provided an enormous amount of novel data. However, to some degree this technique should be regarded as ‘a work in progress’ and further developments of the experimental protocol and analysis tools will be beneficial. It has taken much iteration to get to this point and we are currently sequencing the transcriptome of *Clostridium difficile*, *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Campylobacter jejuni* and *Mycobacterium tuberculosis* using this refined protocol. Ultimately, this technique may provide an important quantitative analysis in both eukaryotes and prokaryotes.

## 5 Global transcriptome analysis of an *S. Typhi* *ompR*-null Mutant.

## 5.1 Aims of this chapter

The aim of the work described in this chapter was to characterise the *in vitro* regulon of the virulence mediating global transcriptional regulator, OmpR. This approach was facilitated by exploiting spotted PCR-based *pan-Salmonella* DNA microarrays and by developing a method for high-throughput chIP-seq.

## 5.2 Introduction

Bacterial species have evolved mechanisms through which to sense their environment and tightly regulate gene expression in response to specific signals. This dynamic response to changing stimulus has allowed bacteria to survive in their respective niches. Adaptive response studies have revealed many different strategies employed by bacteria to regulate gene transcription and translation. Transcriptional repressors and activators are the most widely studied regulatory systems and are essential for effective cellular function. However, more recently discovered *cis*-acting RNA elements and small RNA are believed to also play an important role. Approximately 227 genes in the *S. enterica* genome are thought to encode regulatory proteins [76,77,136]. Interestingly, many of these conserved regulators have been shown to tightly regulate recently horizontally acquired *Salmonella* DNA islands, not present in *E. coli*. The *S. Typhi* genome encodes 15 putative two-component transcriptional regulators. Each of these regulatory systems harbours both sensing and transcriptional activator domains either as a single, or as two individual, proteins. Transfer of the external signal is normally mediated via exchange of a phosphate group. The autophosphorylating sensing domains is an inner-membrane spanning sensor associated with a histidine kinase domain [217]. Changes in the environment, such as

reduced magnesium or low osmolarity, induce autophosphorylation and transfer of the phosphate group to the receiver domain of the transcriptional activator. Binding of the phosphate group alters the DNA-binding domain affinity to specific DNA motifs. Consensus sequences for most two-component regulators in *E. coli* have been described. The OmpR/EnvZ (*ompB* locus) two-component system is orthologous to the eponymous system in *E. coli* and senses changes in osmolarity [139]. The *S. Typhi* EnvZ protein has diverged from *E. coli* with substitutions in 13 amino acids, however, OmpR is completely conserved between these two species. While maintaining the key porin *ompC* within the regulon, the *S. Typhi* OmpR system has captured the Vi antigen locus [116] and co-regulates expression of SPI-1 [227] and SPI-2 [228]. *ompR*-null mutants are highly attenuated in the murine model [70].

DNA microarray technology has facilitated genome wide analysis of RNA expression in sequenced biological systems. Comparative genomic hybridisation (CGH) and comparative transcriptome analysis have been widely used to identify and quantify nucleic acid content of these bacterial cells. To further characterise the regulon of the *ompB* locus total RNA was isolated from *S. Typhi* under different growth and media conditions and directly compared to an isogenic strain with a defined deletion in the *ompR* gene on the WTSI generation III spotted DNA microarray. These data were confirmed using real time PCR and enzymatic assays to investigate differential protein abundance of the transcribed genes. To confirm OmpR directly regulates the differentially transcribed genes, a method for chIP-seq, to identify protein-binding sites, was developed using high-throughput Illumina sequencing technology and the data mapped back to the entire genome.

## 5.3 Results

### 5.3.1 Analysis of the *S. Typhi* OmpR regulon, preliminary work

Since this work was initiated at the WTSI before the CL3 containment facility was established, an attenuated and safe derivative of *S. Typhi* was selected for these experiments. BRD948 is a candidate vaccine strain derived from *S. Typhi* Ty2 that harbours mutations in *aroC*, *aroD*, and *htrA*. These mutations serve as attenuating lesions with *htrA* preventing the detection of bacteria in the blood of orally immunised volunteers. An *ompR* mutant of BRD948 was generated and both derivatives had similar and indistinguishable growth characteristics on LB agar and in LB broth. The DNA microarray used in this study was designed at the WTSI and was generated using specific unique PCR products (200-500bp) representative of the 4097 predicted coding sequences of CT18. Appropriate positive and negative controls were printed onto the arrays. The PCRs were carried out using specific primers (Sigma-Genosys) in a two-step protocol.

### 5.3.2 Microarray analysis

#### 5.3.2.1 Biological conditions, quality control and analysis

Total RNA was prepared from BRD948 and BRD948  $\Delta ompR::kan$  cells grown under three different conditions, mid-log ( $OD_{600}=0.6$ ), early stationary phase ( $OD_{600}=1.1$ ) and SPI-2 inducing conditions [157]. For each biological replicate, differentially

labelled cDNA was hybridised to four slides and the dye used to label each RNA was inverted for two of these. This was performed for three biological replicates in each condition. To identify systematic error and to assess if the microarray experiments were reproducible the samples were clustered (figure 5.1 (a)). Further analysis was carried out on these samples as the technical replicates generally clustered together. The relationship between the RNA populations isolated from each of the strains was calculated by average-linkage hierarchical clustering using the Pearson Correlation as part of the GeneSpring microarray analysis software V5.0 (Silicon Genetics). Lowess normalised intensities were plotted to determine possible RNA degradation or gDNA contamination (figure 5.1(b)). Differences of less than two-fold were discarded ( $p$ -values  $< 0.05$ ).

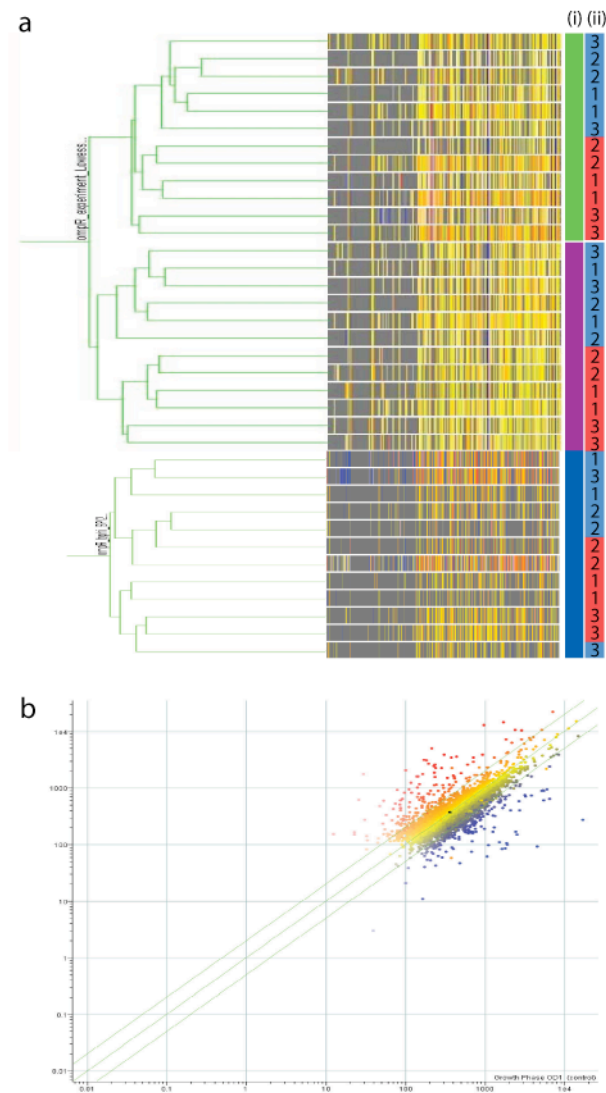


Figure 5.1 Quality control for expression microarrays.

(a) Clustering of microarray intensity data. The vertical colour bars (i) align with each experiment, green is exponential growth, purple is early stationary phase and dark blue represents SPI-2 inducing conditions. The (ii) colour bars represent dye labelling with red, Cy5 conjugated to BRD948 cDNA and Cy3, BRD948  $\Delta ompR$ . Blue was the inverse and the digits represent each biological replicate. (b) Fluorescence intensity scatter plot. Spots were assigned present or absent during the data extraction. Present spots were normalised using the LOWESS method and the intensity values for control (BRD948) were plotted against BRD948  $\Delta ompR$ . Data points outside the 2-fold range are clearly visible on this plot.

### 5.3.2.2 The OmpR regulon during exponential growth

(OD<sub>600</sub>=0.6)

A total of 219 genes had significantly different levels of fluorescence intensity directly related to labelled nucleic acid abundance and differential levels of transcription. Of the differentially transcribed genes, 153 were increased in BRD948  $\Delta ompR::kan$  (appendix 9.8). Differentially expressed genes were sorted (figure 5.2) according to functional classification as predicted in the *S. Typhi* CT18 sequence annotation. Briefly, for all differentially expressed genes, thirty-five were classified as structural, 82 respiratory, 12 regulatory, 25 transport associated, 29 enzymatic with the remaining 36 genes annotated as hypothetical proteins. The location of each of the differentially transcribed genes is illustrated in figure 5.3 using DNAPlot [229]. Genes previously shown to be under OmpR regulation, such as *ompC*, *ompS1* [143], *viaB* locus, *ompB* locus and *sprAB* were all significantly different.

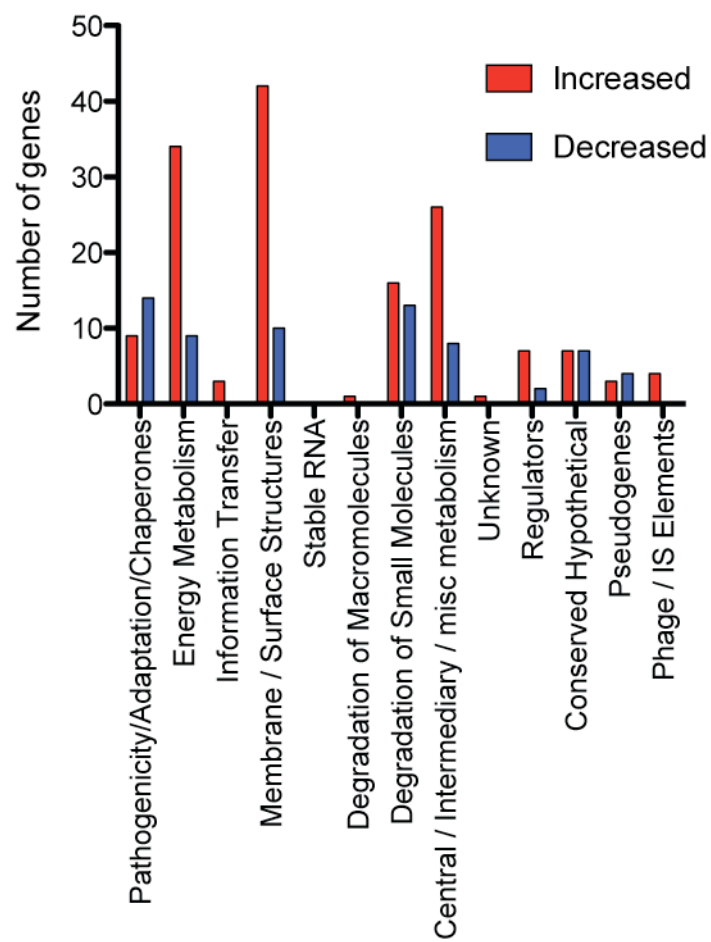


Figure 5.2 Numbers of genes differentially transcribed by functional class.

Genes were grouped according to functional classification. The number of genes in each class with increased transcription is represented by red bars and decreased, blue bars.

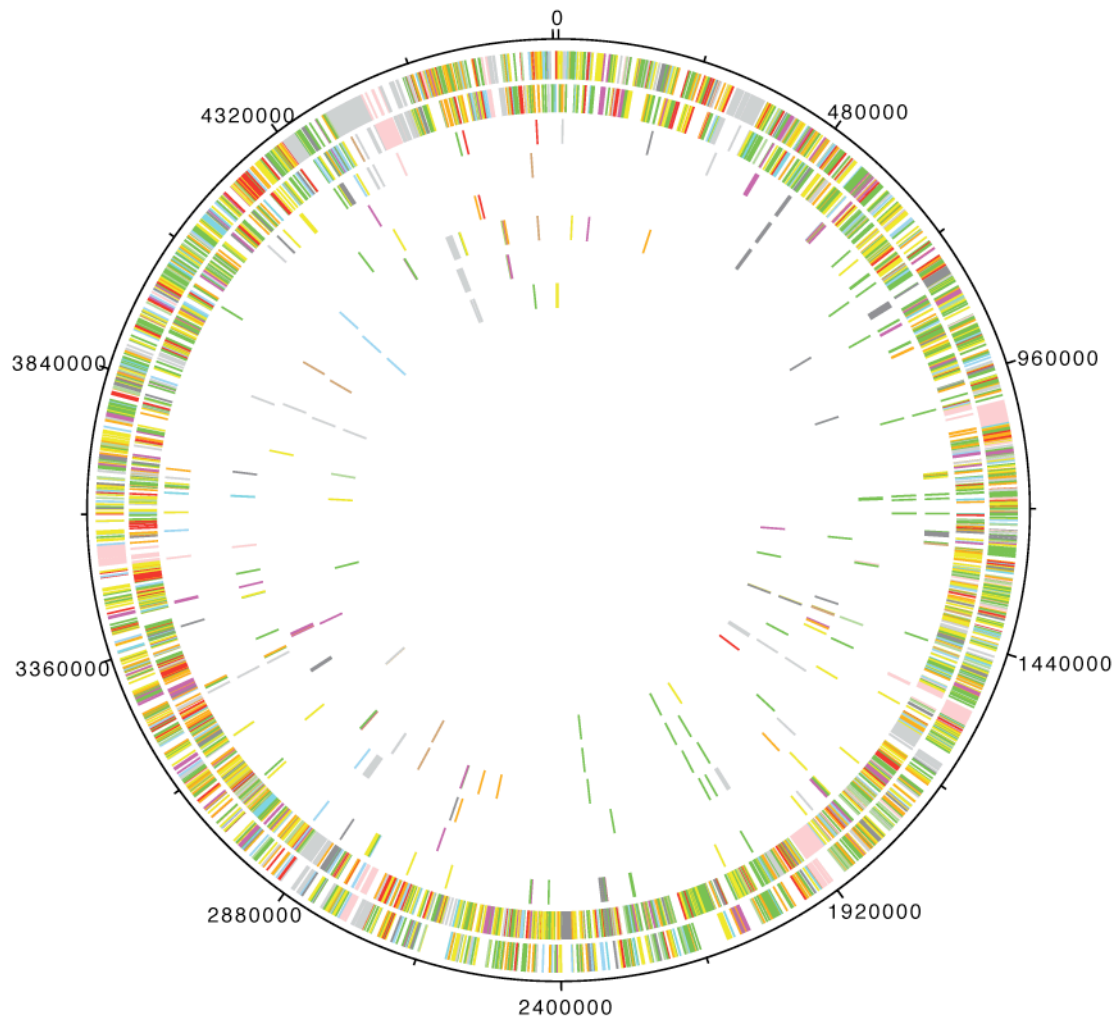


Figure 5.3 Circular plot of the *S. Typhi* genome and genes differentially transcribed in the *ompR* mutant.

Circles are numbered from 1(outermost) to 8(innermost). 1, forward strand annotated CDS; 2, reverse strand CDS; 3, midlog genes increased transcription; 4, stationary phase increased transcription; 5, SPI-2 genes increased transcription; 6, midlog genes decreased transcription; 7, stationary phase decreased transcription; 8 SPI-2 decreased transcription.

### 5.3.2.3 The OmpR regulon during early stationary phase

( $\text{OD}_{600}=1.1$ )

The global fluorescence intensity and quantity of fluorescent spots was reduced in these data, despite using the same mass of total RNA (16 $\mu\text{g}$ ). With fewer global mRNA transcripts, forty-three genes were identified as being differentially expressed

during early stationary phase in the BRD948  $\Delta ompR::kan$ , 16 of which were up regulated. Structural genes were the major functional class affected with 19 genes significantly different, including the *viaB* locus, *ompC* and *ompSI*. Other classes were as follows: 9 respiratory, 1 regulatory (*ompR*), 3 transport-associated, 5 enzymatic and 6 hypothetical proteins. Forty-one of the 43 genes were differentially expressed in both the exponential growth phase and early stationary phase (see appendix 9.8 for full gene list and comparison).

#### 5.3.2.4 The OmpR regulon grown under SPI-2 inducing conditions

SPI-2 expression is essential for the formation of and survival within the SCV. Expression of SPI-2 is influenced by OmpR and requires low pH and low  $Mg^{2+}$  conditions. Briefly, *S. Typhi* cells were grown to exponential phase, harvested, washed twice and resuspended in SPI-2 inducing media and incubated at 37°C for 20 hours. This was performed in parallel with BRD948 harbouring the plasmid p/c/1, a *lacZ* reporter construct with the promoter region for *ssaG* (from SPI-2) upstream of the *lacZ* gene. Expression of LacZ was confirmed by Miller  $\beta$ -galactosidase assays [181]. The *ssaG*-LacZ reporter was used to confirm that the growth conditions were optimal for SPI-2 activation.

The global fluorescence intensity and quantity of fluorescent spots was reduced in these data in a similar manner to the early stationary phase data. Of the 78 differentially transcribed genes, 54, were increased in expression in the BRD948  $\Delta ompR::kan$ . Forty-two of the genes were classified as structural, 3 transport-associated, 12 respiratory, 12 hypothetical, 8 enzymatic and 1 regulatory (*ompR*).

Interestingly, of the 80 genes differentially expressed in this experiment and the 218 identified during exponential growth, only 20 were identical in both data. Similarly, when compared with the early stationary phase set of 43 differentially expressed genes, only 12 were the same (see appendix 9.8 for full gene list and comparison).

### 5.3.3 Real time PCR confirmation of expression profile

To validate the expression profile, an experiment to test cDNA abundance was designed using real-time PCR [230]. Briefly, the technique requires quenched fluorescent probes (Applied Biosystems) which bind specifically to the target amplicon. During polymerase activity, the quencher is cleaved from the hybridised probe (by the polymerase) and fluoresces when excited by a LASER. An excess of probes ensures fluorescence increases every cycle and, if the primers are 100% efficient, this intensity should double every cycle. Requisite controls should ensure no contaminating gDNA is amplified (template mRNA prior to reverse transcription), inclusion of a no template control, and a gene, or set of genes, for normalisation. Eukaryote gene expression experiments, utilising real-time PCR, generally normalise to 18S,  $\beta$ -actin or GAPDH which are stably expressed under all conditions. The prokaryotic transcriptome is not thought to contain such stably expressed “housekeeping” genes, therefore, 10 different genes of varied, but essential biological function, were used to normalise to. 16S was not chosen due to the predicted difference in relative abundance to most mRNA transcripts.

A subset of genes that were of biological interest was chosen to interrogate further and potentially validate the array dataset (figure 5.4). From the real time data *sucA*, *sdhC*, *ompF*, *narK*, *hyaA*, *slsA*, *aceA* and *tviB* agree with the microarray data,

however, *fliC*, *cadA* do not. The *ompC* gene was too variable to determine any difference.

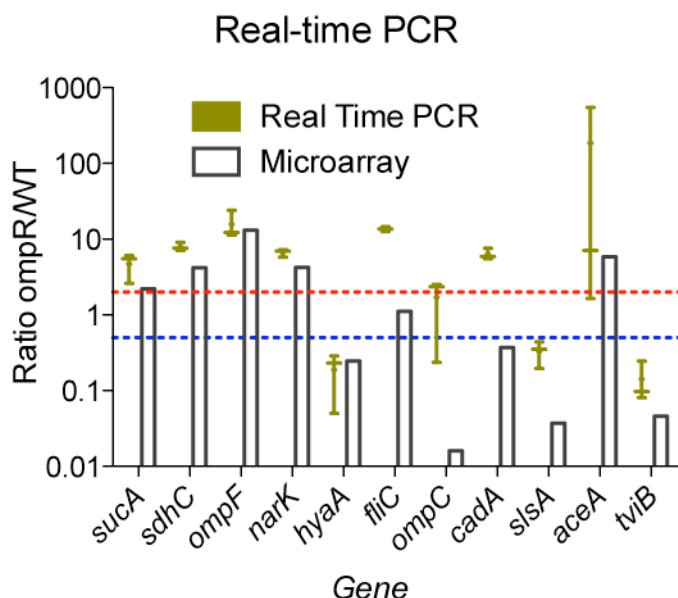


Figure 5.4 Comparison of expression values determined by real-time PCR and microarray.

Real time PCR intensity for each target gene in *BRD948ΔompR::kan* was compared to the average intensity determined for 10 control genes and then compared to intensity in *BRD948*.

### 5.3.4 Succinate dehydrogenase enzyme assay

The succinate dehydrogenase (*sdhCDABsucABCD*) operon, which encodes essential enzymatic components of the TCA cycle [204], has not previously been described as part of the OmpR regulon. Confirmation of the microarray data by real time PCR further confirmed a transcriptional difference between the levels of expression in *BRD948* compared to *BRD948 ΔompR::kan*. In order to determine if this increase in mRNA abundance is translated into functional proteins we performed a succinate dehydrogenase assay [231]. Briefly, *BRD948* and *BRD948 ΔompR::kan* were grown to the same density ( $OD_{600} = 0.6$ ) as the exponential growth microarray experiment

and harvested. The cells were lysed using a constant cell disrupter and spun (10 000 x g) to remove whole cell debris then centrifuged at 100 000 x g to collect the inner membrane. The pellet was then resuspended, the protein quantified and bound oxaloacetate removed. A colorimetric assay was used to determine the succinate dehydrogenase activity and subsequent kinetics.

The rate of change of absorbance during the colorimetric reaction reveals the enzyme activity, hence, the abundance of the target protein. Direct comparison using the same quantity of inner membrane proteins illustrates a significant difference ( $p=0.0094$ ) between BRD948 and BRD948  $\Delta ompR::kan$  (figure 5.5).

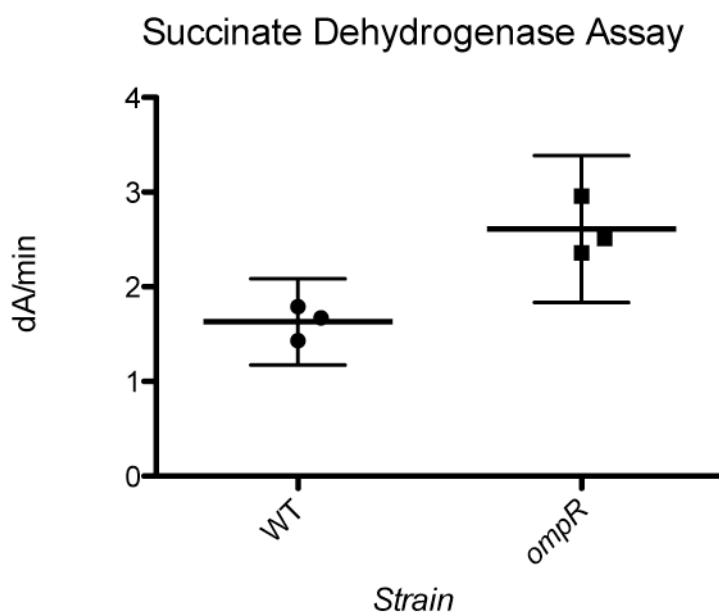


Figure 5.5 Succinate dehydrogenase assay.

Succinate dehydrogenase activity, measured by light absorption, of equal quantites of protein derived from both BRD948 and BRD948 $\Delta ompR$  grow under the same conditions as the microarray experiment.

## 5.3.5 chIP-seq to determine ompR-binding domains

### 5.3.5.1 Rabbit anti-OmpR

chIP-seq techniques are dependent on the ability to specifically precipitate the protein/DNA complexes under study. One method by which to achieve this is to use specific antibody to a particular DNA binding protein. Consequently, attempts were made to raise antibody to the *S. Typhi* OmpR protein. In order to express an OmpR antigen suitable for generation of anti-OmpR antibodies the commercially available vector, pBAD202 was used to clone and his-tag *ompR*. DNA encoding the *ompR* open reading frame was generated by PCR from *S. Typhi* BRD948 template DNA. The OmpR protein was expressed in *E. coli* K12 and the purified OmpR antigen was sent to CovalAb (UK) and an antibody raised in and purified from rabbits. Western blotting analysis after pre-absorption of sera in an *ompR*-null mutant fixed culture revealed polyclonal anti-OmpR was not specific enough to distinguish between OmpR and other proteins by western blot.

### 5.3.5.2 Construction of an OmpR:3xFLAG

An alternative approach to raising antibody is to tag a protein with a specific peptide sequence recognised by a commercially available generic antibody. Consequently, a BRD948 *ompR:3xFLAG* strain was constructed by exploiting a suicide vector to deliver a FLAG-tagged *ompR* open reading frame back *in situ* to the chromosome. Briefly, we cloned a PCR amplicon of the *ompR* open reading frame, designed to incorporate the 3xFLAG tag DNA sequence, by overlap extension PCR (figure 5.6). This sequence maintained translation of OmpR linked to the 3xFLAG peptide and

encoded a terminal stop codon. The sequence after the exogenous stop codon included part of the C-terminal sequence of *ompR* that was predicted to include the Shine-Dalgarno sequence requisite for *envZ* translation in the +1 frame. Homologous recombination was used to replace the wild type *ompR* gene with the FLAG-tagged version in situ on the BRD948 chromosome. To determine the functionality of the OmpR:3xFLAG protein *in vivo* Vi polysaccharide agglutinations were performed and the resulting strain TT53.8 was found to be still Vi positive. Vi expression is lost if the *ompR* gene of BRD948 is inactivated.

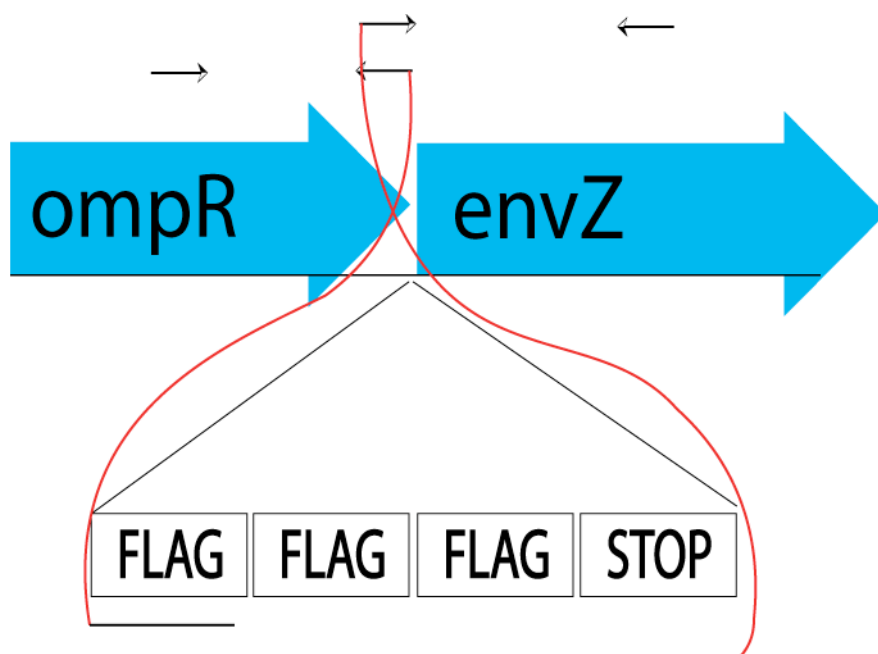


Figure 5.6 Schematic of overlap extension PCR to incorporate 3XFLAG peptide into the C-terminal region of OmpR.

Arrows represent primers homologous to Ty2 regions. Arrows are appended to sequences including 3XFLAG and 20bp overlapping regions. Two rounds of PCR are required to generate the cloning product.

### 5.3.5.3 chIP-Seq Quality Control

To determine if the monoclonal anti-FLAG antibody was efficacious in binding to OmpR:3xFLAG and able to enrich for the heterologous protein, a western blotting

and immunoprecipitation experiment was performed (figure 5.7). Further quality controls were performed to determine if it was possible to enrich for known OmpR-bound DNA sequences. This was undertaken by performing real time PCR comparison of enriched DNA sequences after immunoprecipitation of WT and the OmpR:3xFLAG. These data show enrichment for the heterologous protein and Illumina sequencing of co-purified DNA identified DNA binding sequences upstream of *tviA* and *ompC*.

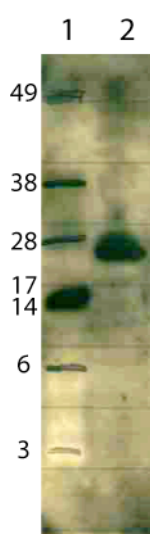


Figure 5.7 Western blot of immunoprecipitated OmpR:3XFLAG.

Lanes number from left to right. Lane 1, Seeblue protein ladder (Invitrogen), Lane 2, immunoprecipitated culture using anti-FLAG. No anti-FLAG antibodies bound to BRD948 cell extracts (data not shown). MW for the marker proteins shown in kDa.

#### 5.3.5.4 chIP-Seq using the Illumina sequencing platform

Paired-ended sequence data generated by sequencing DNA co-purified with OmpR:3xFLAG was mapped using the same parameters as previously performed for the transcriptome data without assigning the pileup of reads to each strand. Plots were z-score normalised (R, affy) and the differences between wild-type and ompR:3FLAG generated sequences determined. Plots were then read into Artemis and peakfinder

was used to determine enrichment for OmpR:3xFLAG bound sequences. The peakfinding programme identified 161 peaks. These were then filtered manually, based on the proximity to or location of the peak. Peaks in the middle of a CDS were removed unless there were multiple peaks nearby. 5 predicted DNA binding motifs (z-score > 10) were predicted and mapped back to 43 separate loci within the enriched sequences. Only 5 predicted motifs had a z-score greater than 10. Artemis assigns a score to each peak and this is represented in appendix 9.9 alongside information on the binding sites, number of sites and sequence motif.

Of the 5 putative OmpR-bound motifs, the TGTWACAW was the most common and it is found in all of the known OmpR regulated genes or operons. This motif was identified in 22 enriched sequences, 12 of which were found as a couplet. The couplet binding sites were upstream of *ompR*, *ompS*, *csgD*, *sdhC*, *mreB*, *dppA*, a probable phage integrase t4357 that is close to *tviA*, the conserved hypothetical t3244, putative outer membrane protein t3144, *galP*, *rfbB*, *fadL* and the conserved hypothetical *yabB*. Notably, upstream of *tviA*, *ompX* and *ompC*, only one of these sites TGTWACAW is present. The second commonest was AATATATA. This is found in 8 of these sequences and as a couplet upstream of *csgD* and as a singlet upstream of *ompS*, *stdA*, *pckA*, *glpF* and hypothetical proteins t4222 and t1022. The site GTGTSTAY is represented 7 times in these data and only as a singlet. These sites are upstream of *ssrA*, *cspE*, *ompR*, *glpF*, the probable phage integrase t4357 and genes t1320 and *yafK*. A further two sites were identified and these appear in 5 and 1 of the upstream sequences.

Peakfinder identified a series of three peaks upstream of the sigma factors *rpoS* and *rpoH*, fimbrial gene *stdA*, *hilC* and *glpF*. However, no motif was identified in these

sequences. Sequences upstream of subunits L10, L11, L20, L34 and L36 of the 50S ribosomal protein were also enriched in these data and again no motifs were identified by YMF [184,232]. Interestingly, there were a number of genes identified in the data with a common substrate or function such as L-arginine (*argG*, *argT*, *artJ*) and two cold shock proteins *cspE* and *cspA*.

### 5.3.5.6 Comparison with genes differentially transcribed in microarray data

Of the 127 upstream sequences that were enriched by immunoprecipitation 16 were associated with genes differentially transcribed in the microarray experiment (table 5.1) with 7 encoding the commonest motif (TGTWACAW).

Table 5.1 Genes differentially expressed in microarray experiments and with chIP-seq enrichment sites mapped upstream

Gene ID	Gene Name	Annotation	Expression Ratio	t-test P-value
STY0773	gltA	citrate synthase	3.24E+00	7.68E-11
STY0775	sdhC	succinate dehydrogenase cytochrome b-556 subunit	4.19E+00	1.85E-14
STY1167	-	hypothetical protein	2.03E+02	1.40E-03
STY1925	-	putative membrane protein	2.19E+00	2.64E-10
STY2203	ompS1	outer membrane protein S1	4.06E-01	4.38E-04
STY2424	mgIB	D-galactose-binding periplasmic protein precursor	2.11E+00	2.60E-07
STY2493	ompC	outer membrane protein C	1.60E-02	4.20E-06
STY2585	argT	lysine-arginine-ornithine-binding periplasmic protein precursor	3.51E+00	1.24E-06
STY2623	fadL	long-chain fatty acid transport protein precursor	2.11E+00	1.27E-03
STY3049	rpoS	RNA polymerase sigma subunit RpoS (sigma-38)	2.65E+00	1.44E-09
STY3176	stdB	probable outer membrane fimbrial usher protein	2.47E+00	2.92E-11
STY3395	air	aerotaxis receptor protein	2.68E+00	1.59E-08
STY3405	-	probable membrane transport protein	1.12E+01	2.92E-05
STY4168	dppA	periplasmic dipeptide transport protein precursor	2.17E+00	2.50E-07
STY4294	ompR	two-component response regulator	2.58E-01	1.26E-04
STY4662	tvIA	OmpR Vi polysaccharide biosynthesis protein	2.32E-02	5.92E-06

## 5.4 Discussion

### 5.4.1 Microarray Data

DNA microarrays are a high-throughput platform designed to represent a global snapshot of significant differences in a transcriptional profile. This experimental design offers only to highlight the differences between two conditions and precludes any information on the global transcriptome of the individual strains. Information is also limited to the double stranded DNA sequences present on the array, thus limiting the data to the sequenced strain global annotation, rather than the actual experimental organism transcriptome. It is not expected the genomic DNA would differ significantly in terms of genetic drift between the sequenced Ty2 and BRD948 Ty2. Also, a spotted array does not distinguish between the coding or non-coding strands, potentially increasing the noise, by hybridising any overlapping transcripts. Recently identified small non-coding RNA were not present on this array.

Previous literature on the OmpR regulon is summarised in table 5.2 and a comparison is made with our findings. These data generally agree with the literature. Individual genes such as the outer membrane porins, C and S1, and the *viaB* locus, which are strongly regulated by OmpR, and transcribed under these conditions, are important validations for these data. However, the *ssrA* gene and regulator of SPI-2 is co-regulated by OmpR. Under SPI-2 inducing conditions, there are 4 SPI-2 genes with reduced transcription, suggesting there is dysfunctional transcription of this region further validating the dataset.

Table 5.2 Experimentally verified and published genes in the OmpR regulon of *S. Typhi* or *S. Typhimurium* identified as being differentially regulated in the microarray experiments (2-fold,  $p < 0.05$ )

Gene	OmpR-regulated	Reference	Spotted Arrays		
			OD <sub>600</sub> = 0.6	OD <sub>600</sub> = 1.1	SPI-2 Induced
<i>ompC</i>	Yes	[233]	Yes	Yes	Yes
<i>ompF</i>	Yes	“”	Yes	Yes	No
<i>tviA</i>	Yes	[116]	Yes	Yes	Yes
<i>tviB</i>	Yes	“”	Yes	Yes	Yes
<i>tviC</i>	Yes	“”	Yes	Yes	No
<i>tviD</i>	Yes	“”	Yes	Yes	No
<i>tviE</i>	Yes	“”	Yes	Yes	Yes
<i>vexA</i>	Yes	“”	Yes	Yes	No
<i>vexB</i>	Yes	“”	Yes	Yes	No
<i>vexC</i>	Yes	“”	Yes	Yes	No
<i>vexD</i>	Yes	“”	Yes	Yes	No
<i>vexE</i>	Yes	“”	Yes	Yes	No
<i>ssrA</i>	Yes	[234]	No	Yes	No
<i>ompS1</i>	Yes	[144]	Yes	Yes	Yes
<i>ompS2</i> STY1649	Yes	[143]	No	Yes	No
<i>tppB</i> STY1670	Yes	[154]	No	Yes	No
<i>flhD</i>	No	[146]	No	No	No
<i>csgD</i>	Yes	[235]	No	No	No
<i>osmZ/hns</i> STY1299	No	[236]	No	No	No
<i>hilD/sprB</i>	Yes	[237]	Yes	No	No
<i>cadC/cadB</i>	Yes	[238]	Yes	No	No
<i>aas</i>	Yes	[155,156]	No	No	No

### 5.4.1.1 Regulators

There are 12 known regulators with differential expression patterns in the *S. Typhi* mid-log experiment. It appears the loss of OmpR could be impacting globally on these operons during exponential growth. Of these twelve genes, most are regulators of stress responses and of anaerobic metabolism. Anaerobic carnitine metabolism, regulated *caiF*, was decreased in expression (0.4,  $p = 2.7E-04$ ), allantoin regulator (*allR*) and the allantoin operon were significantly increased (range 2.0-4.0, all  $p < 1.1E-4$ ) indicating the OmpR deficient organism is exploiting allantoin as a anaerobic nitrogen source [239]. The expression of the two-component regulator for tricarboxylate transport is increased by approximately 4-fold ( $p = 2.6E-11$ ).

Tricarboxylates can be utilised as an anaerobic or aerobic carbon source [240]. Increased transcription of the *rpoS* gene has the potential to complicate this analysis as the RpoS regulon contains more than 50 genes [241]. Expression of the propionate catabolism regulator, *prpR*, is increased as is the propionate operon. This pathway is known as the 2-methylcitrate pathway and allows growth on propionate as the sole carbon and energy source [242]. Transcription of the two-component fumarate response regulator (*dcuR*) is also increased by approximately 4-fold. This regulator activates the anaerobic fumarate respiratory system, transport of fumarate and is activated by C4-dicarboxylates [243].

There are two putative regulatory genes also induced in the *S. Typhi* *ompR* mutant derivative, *rsD* and *ydeW*. In the *S. Typhi* exponential phase growth the SPI-1 AraC-like regulator transcripts, *sprA* (*hilC*) and *sprB* were repressed. These genes have been associated with the regulation of *hilA*, a regulator of SPI-1 mediated cell invasion. Over-expression of these regulatory genes also results in increased ability to enter host cells [94].

In these data, it is unclear whether OmpR is a direct activator or repressor of these regulators, or those indirect effects such as dysfunctional porin expression and changes in inner membrane transport induce anaerobiosis. Such major disruptions to the outer membrane, affecting homeostasis of the cell, may increase demand for less efficient systems such as anaerobic respiration thus indirectly increasing expression of anaerobic related transcripts.

#### 5.4.1.2 Membrane transport

Gram-negative bacteria require transport of nitrogen, carbon, water, electrons and ions across both the inner and outer membranes. The outer membrane acts as a general barrier to osmotic shock and maintains the ability to increase (OmpF) or decrease (OmpC) the porin size depending on the immediate environment [138]. OmpR is known to regulate this system [139]. There are also porins OmpS1 and OmpS2 that have been characterised in the *S. Typhi* OmpR regulon [143,144]. The genes encoding these proteins exhibit some form of deregulation in these data.

With many regulatory genes associated with stress response and anaerobic respiration it is no surprise to see an increase in inner membrane transport activity. During exponential growth of the *S. Typhi ompR* derivative there are 11 different transport systems significantly altered (figure 5.8). Allantoin permease increases the cellular availability of allantoin for anaerobic respiration [239]. Molybdenum uptake is induced to increase the amount of cofactor needed for reductases that catabolise nitrogen compounds [244,245]. Small peptide scavenging is also induced, suggesting limited availability of free peptides and recycling of peptides from the cell's outer wall [246,247]. Further evidence to suggest the cell is undergoing anaerobic respiration is the induction of the nitrite extrusion transport system (*narK*). Nitrite is toxic to the cell and a bi-product from anaerobic nitrate metabolism [248]. Glutamate and aspartate transport is also important for controlling cellular nitrogen concentration and is increased in the *S. Typhi ompR* derivative [249].

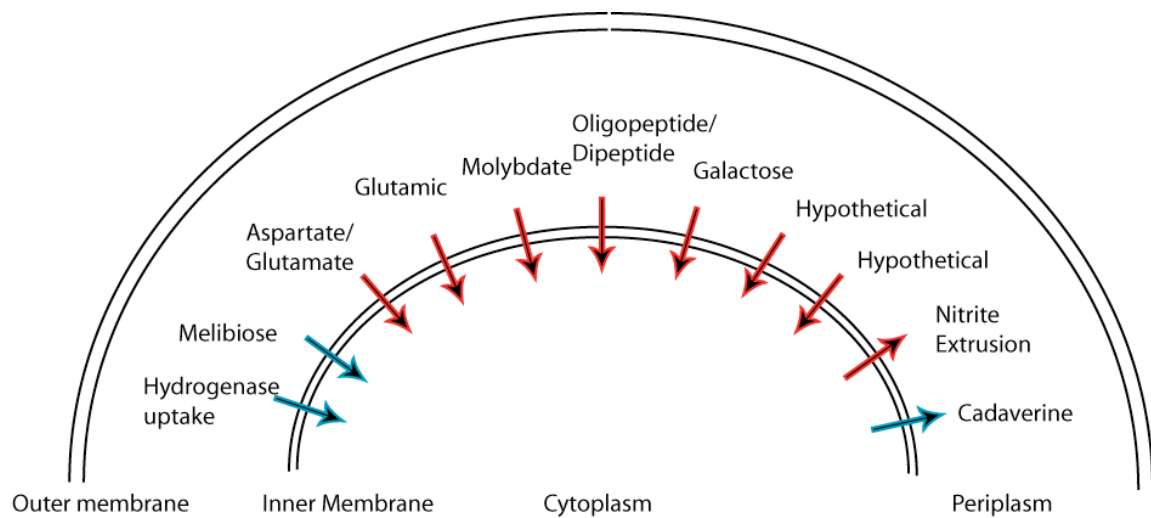


Figure 5.8 Transporter transcript changes.

Schematic diagram representing the shift in inner membrane transport in the *S. Typhi* *ompR*-null mutant. Arrows in red indicate significant increase (2-fold,  $p < 0.05$ ) and blue, decrease (2-fold,  $p < 0.05$ ). Arrow represents the predicted direction of substrate.

Cadaverine transport has been associated with blocking of outer membrane porins as well as playing a role in cellular response to low pH by neutralising the external media [250,251]. This may have important implications for the ability of *S. Typhi* to survive within the hostile environment of the stomach and host. The Vi polysaccharide is thought to inhibit the bacterium's ability to induce TTSS mediated endocytosis encoded within SPI-1, however, with cadaverine transport in the same regulon, this may confer resistance to low pH if Vi expression is reduced to invade the host. Unsurprisingly during early stationary phase, very little transport across the membrane is affected and this may be due in part to reduced mRNA transcripts during this period of attenuated growth.

Under SPI-2 inducing conditions there is an increase in transcript of genes required for the scavenging of  $Mg^{2+}$  ions (*mgtB*, *corA*) and an increase in amino acid permease (t3197 (STY3460)). The bacterium is starved of magnesium and essential amino acids in the host vacuole [252]. The two-component sensor, PhoP/PhoQ, which senses the

change from extracellular host to intracellular host environment, is activated by low magnesium, thus evidence of increase in the uptake of this integral cation is not unexpected.

### 5.4.1.3 Respiration

Respiration is the transfer, or donation of electrons by low redox potential donors, via a range of redox co-factors to the terminal electron acceptor, oxygen. Anaerobic respiration involves reduction of nitrogen oxy anions or nitrogen oxides and is less energy efficient. The energy released in this step-wise process is used to generate a *trans*-membrane electrochemical gradient to generate activation energy for the synthesis of ATP. The ability of different bacterial species to colonise particular hostile niches lies in part in the evolution of alternative strategies for this electron transport mechanism [253].

In the *S. Typhi* mid-log experiment, over half of the genes with significantly different transcription are annotated as genes associated with respiratory processes or are genes encoding enzymes. Figure 5.9 [253] represents the significantly different loci, both increased and decreased, associated with respiration. There is a significant increase in genes that respond to oxygen limitation, such as nitrogen metabolism (*narGHJdsbEccmFIEICII*, *nirDB* and *napCBHGAyojFnapF*) [254,255,256]. Glyoxylate induced genes, such as succinate dehydrogenase (*sdhCDAsucABC*) [204] and isocitrate lyase (*aceABK*) [257] are up-regulated as well as alternative cytochrome genes in the *nrf* locus (*nrfABCD,STY4480*) [258], which is known to be induced under glucose limitation. Interestingly, both aerobic (*fadAB*) [259] and anaerobic (*fadL*) systems for  $\beta$ -oxidation of acyl-CoA to acetyl-coA synthesis [260], a substrate that can be fed into the glyoxylate cycle, are increased in the *ompR* mutant.

This, combined with increases in propionate catabolism (*prpBCDE*) [261], suggests the organism's carbon metabolic programme has been rewired to catabolise poor carbon sources, possibly found in the mammalian gut. The significantly increased allantoin utilisation locus (*allARgipglxRallPybbWallB*) [239] expression appears to remain under regulatory control even though both *allA* and *ybbW* are pseudogenes.

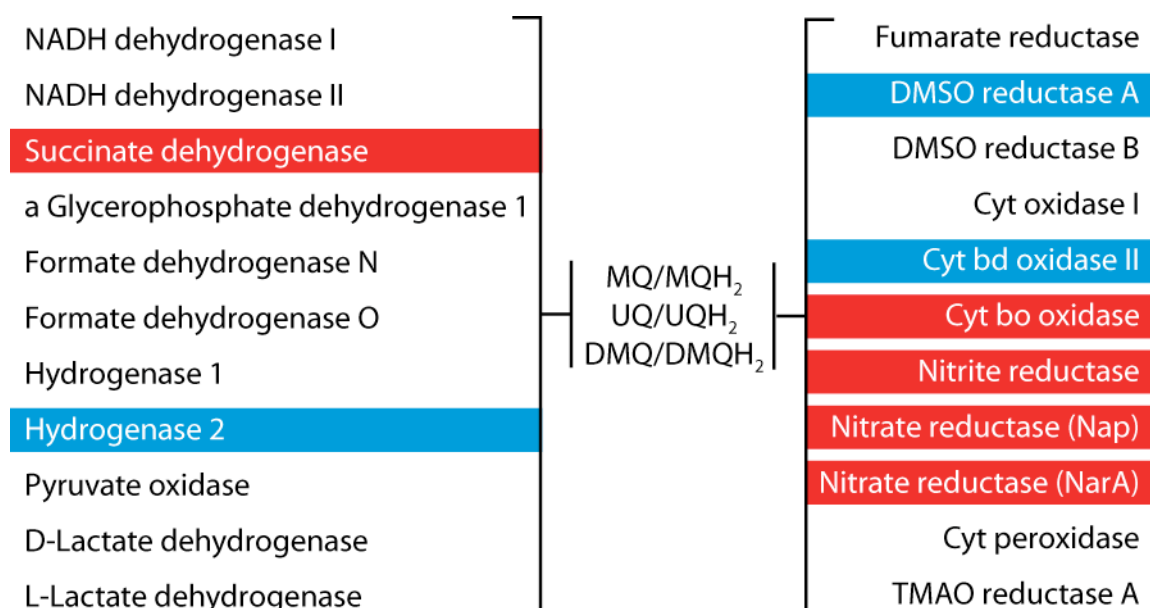


Figure 5.9 Respiration.

Significantly different genes involved in respiration. Red, increased (2-fold,  $p < 0.05$ ) and blue decreased (2-fold,  $p < 0.05$ ).

Respiratory operons with decreased expression also confer information about the cell's physiological state. Anaerobic dimethyl sulfoxide (DMSO) reduction (*dmsA*, *dmsA2CSTY1596*) and hydrogenase activity (*hybDCBAO*) also suggest the anaerobic metabolism switch is tightly regulated by OmpR.

The *S. Typhi* stationary phase experiment respiratory genes were not differentially expressed and these may not be expressed at high levels in either strain. However, the data indicate significant increase in transcription of the cytochrome-o biogenesis

pathway genes (*cyoABCDE*) [262]. This is also observed in the SPI-2 induced experiment.

#### **5.4.1.4 Chemotaxis and motility**

*S. Typhi* contains 10 chemotaxis genes involved in regulating motility two of which are putative pseudogenes. In this “stressed” exponentially growing *S. Typhi ompR* derivative, transcriptional increase of chemotaxis-associated genes is evident. Significantly different levels of expression in the aerotaxis gene, *air*, represent a natural response in order to translocate to a niche where, the more energy efficient aerobic respiration can be utilised [263]. For cells in early stationary phase, there is evidence for an increase in the flagellar regulation (*flgM*, sigma factor), structural (*flgK*, hook-associated) and biosynthesis genes (*flgN*) [264,265].

#### **5.4.1.5 Genes of unknown function**

In these data there are two putative operons encoding genes of unknown function that are differentially expressed (figure 5.10(a) and (b)) in the *ompR* mutant. The genes orthologous to STY1164-1170 are significantly increased in expression and the orthologous genes STY4801-4806 are significantly reduced. These genes may fall within the OmpR “virulence” or “respiratory” regulon or may be regulated due to indirect effects on the cell as discussed previously.

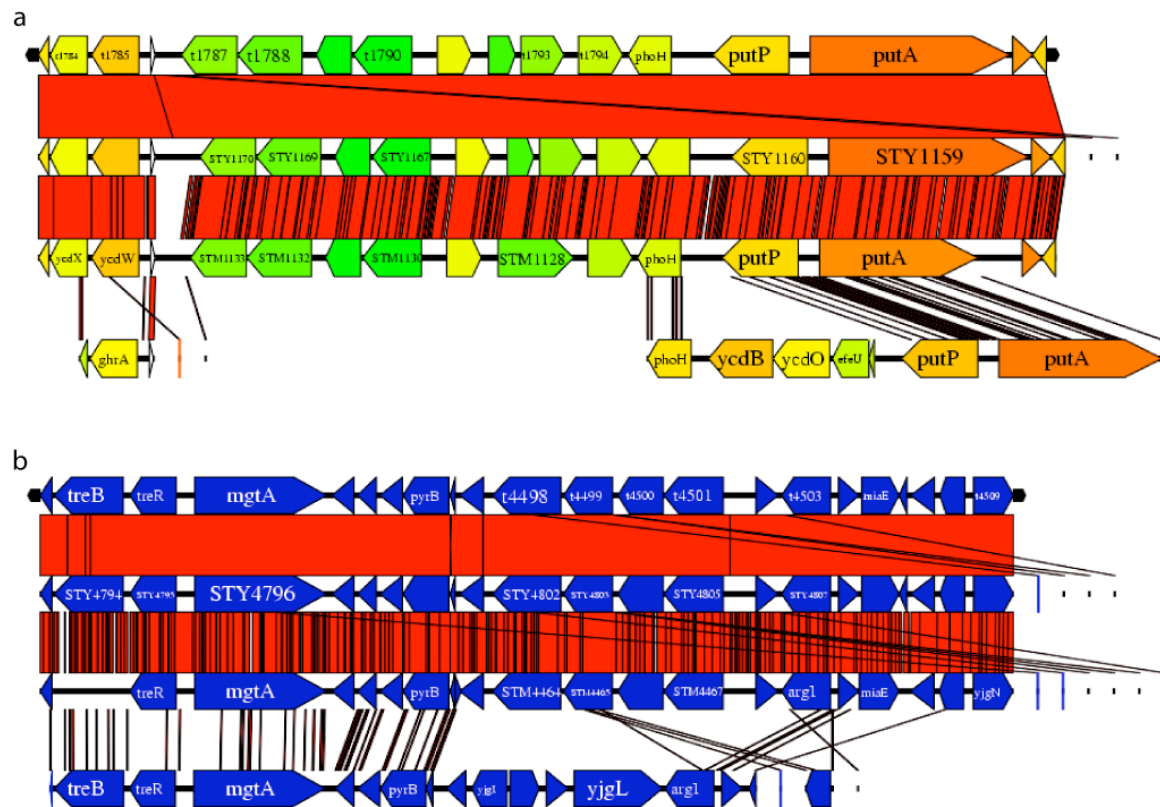


Figure 5.10 Genes of unknown function.

(a) STY1164-STY1170 genes compared with Ty2 (top), *E. coli* (bottom), and LT2 (second from bottom). (b) STY4801-4805 genes compared with Ty2 (top), *E. coli* (bottom), and LT2 (second from bottom). Both operons were consisted of hypothetical genes and were found to be differentially transcribed in the *ompR* mutant. Neither operon is present in *E. coli* as the above Colibase [266] comparison illustrates.

#### 5.4.1.6 Indirect comparison with RNA-seq data

The transcriptome sequencing data (chapter 4) identified 305 genes that were differentially transcribed (2-fold,  $p < 0.05$ ) and 39 of these genes were also differentially regulated in the microarray experiment. One of the genes, *yaeG*, was increased in the Illumina comparison and decreased in the microarray comparison and is considered an outlier (figure 5.11).

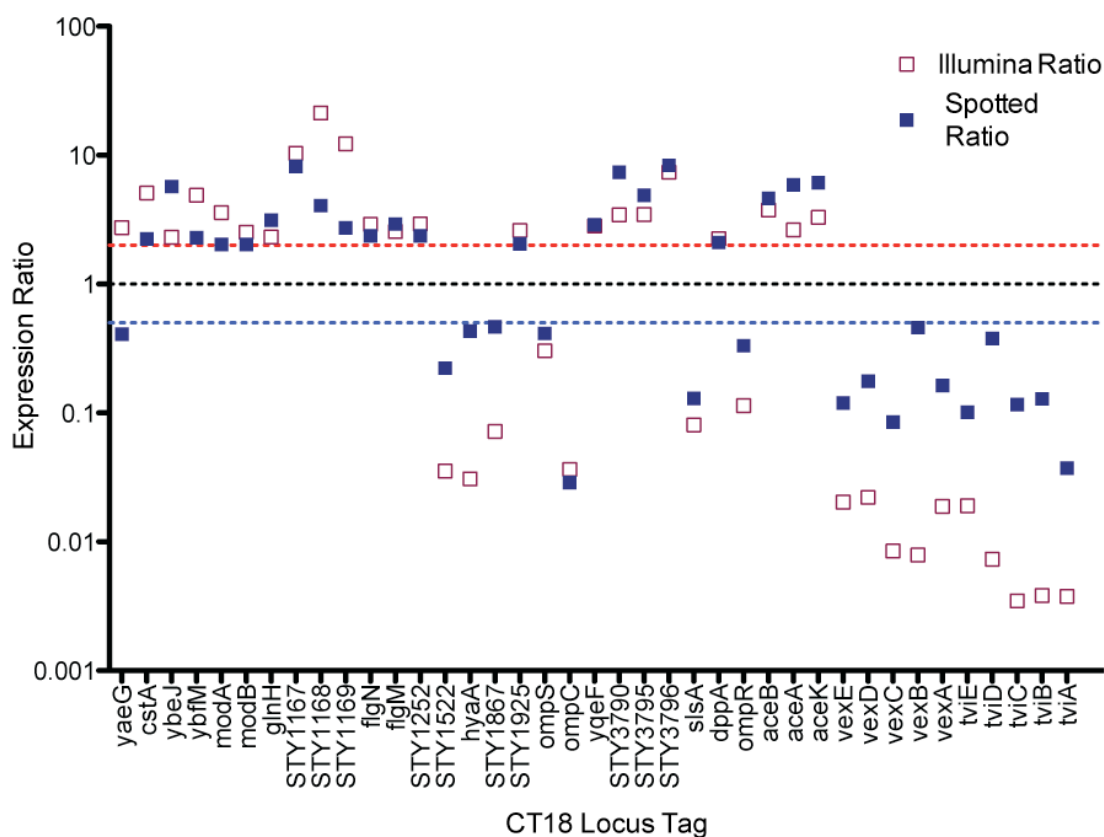


Figure 5.11 Expression ratio for genes found to be differentially expressed in microarray and deep-sequencing experiments.

Spotted array data are represented by filled blue square, Illumina data by red outline square. The red dotted line illustrates a two-fold increase and the blue dotted line, a two-fold decrease in nucleic acid content.

Of the 17 genes that were decreased in expression in both experiments, the Illumina data exhibited a consistently lower ratio of *ompR*/WT transcripts, suggesting background autofluorescence induced during microarray scanning may directly affect the results. Genes previously characterised as OmpR-regulated with decreased levels of expression in the mutant were *tviABCDE*, *vexABCDE*, *ompC*, *ompS*. Genes not previously described in the *ompR* regulon identified in these data include the *slsA* (t3757), *hyaA2* (t1458) and a putative secreted choloylglycine hydrolase (t1459). The function of these genes is discussed in section 4.3.4.3.

Intriguingly, 21 genes were increased in expression in response to the loss of *ompR*, which has previously described as a transcriptional activator. Two contiguous flagellin regulatory genes, *flgN* (t1749) and *flgM* (t1748) were increased in transcription in the *ompR* mutant and are both required for flagellin assembly. FlgM is a negative regulator of flagellar biosynthesis and a mutation is attenuating in *S. Typhimurium* [265]. FlgN is required for the efficient initiation of filament assembly [264]. The glyoxylate shunt genes (*aceBAK*) are also increased in expression; fatty acid catabolism by isocitrate lyase is crucial for macrophage persistence in *Mycobacterium tuberculosis* [267] and *Rhodococcus equi* [268]. Three genes *t3544*, *t3543* and *t3538* that are predicted components of a ribose/arabinose transport operon were also increased in expression. Furthermore, hypothetical genes *t1788-90* were greatly increased in expression in the *ompR* mutant. These genes are contiguous and encode proteins with sequence identity to sialic acid transport, a secreted protein and a sialic acid lyase respectively and are not present in *E. coli*. Molybdate transport is a crucial co-factor for anaerobic metabolism and the expression of two genes, *modAB*, required for transport are increased in the *ompR* mutant. The genes *dppA*, *cstA*, *ybeJ*, *ybfM*, *glnH*, *t1709* were also increased and these encode for proteins annotated as periplasmic dipeptide transporter, carbon starvation response, glutamate transport, putative outer membrane, glutamine transport and a hypothetical protein, respectively.

### 5.4.2 chIP-seq

The list of genes identified as being downstream of enriched peaks contains previously identified OmpR-regulated genes such as *ompS*, *ompC*, *tviA*, *csgD* and *ssrA*. Many of these genes are preceded by a preferred DNA-binding motif that is

commonly present as a doublet. This has increased the confidence in predicting genes not previously described as being OmpR regulated.

Combined with the microarray data, this technique has corroborated 16 genes, 7 of which encode the TGTWACAW motif. Three hypothetical genes are included in this dataset, t1790 (hypothetical protein), t1080 (putative membrane protein) and t3145 (probable membrane transport protein).

## 5.5 Conclusion

The OmpR protein of *S. enterica*, although not essential for *in vitro* cellular survival, appears to be an important regulator of the adaptive response in preparation for invasion of and survival within the host cells. It remains to be fully elucidated whether all the pleiotropic transcriptional effects observed with BRD948 harbouring the *ompR* null mutation are caused directly or indirectly by lack of OmpR binding to consensus sequences, directly activating or repressing transcription. Regardless of this, an interesting transcriptional signature has been identified. Genes of unknown function that are maintained within this regulon and are not present in *E. coli* are interesting inclusions to these regulon data, which are strongly bifurcated into control of anaerobic respiration or virulence determinants.

## **6 Characterisation of Hypothetical Genes in the OmpR Regulon**

## 6.1 Aims of this chapter

The OmpR protein is an essential regulator of virulence determinants. SPI-1, SPI-2 and SPI-7 are horizontally acquired regions of DNA that are OmpR regulated and essential to virulence. The OmpR regulon, partially defined in both chapters 4 and 5, reveals dysregulation of other horizontally acquired regions of DNA, previously annotated as hypothetical genes. This chapter seeks to characterise the virulence of isogenic mutants harbouring deletions in these hypothetical operons by exploiting the *in vivo* murine model for typhoid fever. *In silico* analysis and further *in vitro* characterisation seeks to determine their function.

## 6.2 Introduction

In the previous chapter, two loci were identified as differentially expressed in the *ompR* mutant and are specific to *Salmonella* compared with *E. coli*. This chapter is dedicated to further characterising these regions, one with increased expression and the other with reduced expression. By using the red recombinase mutagenesis method [180] in *S. Typhimurium*, it was possible to make defined deletions in these groups of hypothetical genes. Thus, assessment of differences in the virulence potential for each derived strain using the murine model was possible. Furthermore, by comparing sequence identity to previously characterised genes, the carbon source required for growth was determined for one of the derived mutants.

## 6.3 Results

Using expression DNA microarrays (chapter 5) we identified two regions whose transcriptional expression was significantly different in a recombinant strain in which the *ompR* gene was deleted. One of these regions that exhibited an increase in expression was t1787-t1793 and the second, with a decrease in transcription, was t4498-4501. They were both previously annotated as hypothetical genes and are not present in *E. coli*. We chose these groups of genes as they were similarly differentially expressed and were contiguous. The group of genes with increased expression are currently annotated as 7 genes and those with decreased expression are predicted to encode 4 genes (table 6.1).

Table 6.1 Differentially expressed hypothetical genes and the published annotation

Systematic ID/Gene Name	Base Range	Expression Values	<i>p</i> -value	GC content	Annotation	<i>S.Typhimurium</i> Orthologue
t1787	1846687..1847790	2.6	0.00583	0.468	Putative oxidoreductase	STM1133
t1788	1847804..1849084	2.5	0.00572	0.451	Putative transporter	STM1132
t1789	1849367..1850059	200	0.00140	0.363	Putative secreted protein	STM1131
t1790	1850105..1851259	7.4	4.93E-07	0.402	Conserved hypothetical protein	STM1130
t1791	1851766..1852446	3.9	0.00125	0.514	Conserved hypothetical protein	STM1129
t1792	1852783..1853310	5.3	8.27E-06	0.439	Putative membrane transporter	STM1128
t1793	1853423..1854280	1.3	0.00962	0.470	Putative membrane transporter	
t4498	4644775..4646178	0.26	2.07E-04	0.568	Arginine repressor	STM4464
t4499	4646234..4647238	0.17	5.08E-05	0.558	Ornithine carbamoyltransferase	STM4465
t4500	4647350..4648282	0.16	4.81E-05	0.592	Carbamate kinase	STM4466
t4501	4648293..4649513	0.16	4.60E-05	0.525	Arginine deiminase	STM4467

### 6.3.1 *In silico* analyses

In these analyses, we use the common genome annotation tools Artemis and ACT (<http://www.sanger.ac.uk/Software/>) to manipulate the information contained within the annotation. All applications of the BLAST (Basic Local Alignment Search Tool) were used to determine sequence identities of proteins and DNA sequences

(<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) [174,175]. For example, pBLAST queries characterised protein sequences of other species and kingdoms to determine any functional or conserved identities. The comparative information is represented as a ranked list indicating the sequence similarity (S, score) and E-value. E-value represents the probability due to chance that there is another alignment with a similarity greater than the given S value. During whole genome sequence and annotation of most species, many predicted genes have been annotated according to sequence identity predicted by BLAST alignment. Thus, with so many similar sequences available, it is often difficult to determine which protein has been fully characterised and which annotation is based on the sequence similarity. Therefore, the BLAST identities should be taken as broad predictor for functionality.

Pfam (<http://pfam.sanger.ac.uk>) is a database of protein families that aligns sequences in a similar manner to BLAST [269]. However, the algorithm is modified to identify conserved functional peptide domains contained within a protein sequence. This tool is useful for identifying conserved domains such as DNA binding motifs, phosphorylation sites and epimerase domains.

Current information on finished genomes and characterised DNA sequences has increase since the Ty2, LT2 and CT18 annotations were determined so functional classification should be re-derived using BLAST and Pfam.

#### **6.3.1.1 t1787-t1793 exhibited increased expression in the *ompR* mutant**

Figure 6.1 (a) represents the genetic arrangement of genes annotated in the *S. Typhi* Ty2 [136] genome and an alignment comparison with the published sequences of *S.*

Typhi, *S. Typhimurium* and *E. coli*. BLAST alignment and Pfam domain alignment shows re-annotation predictions for each gene (table 6.2)

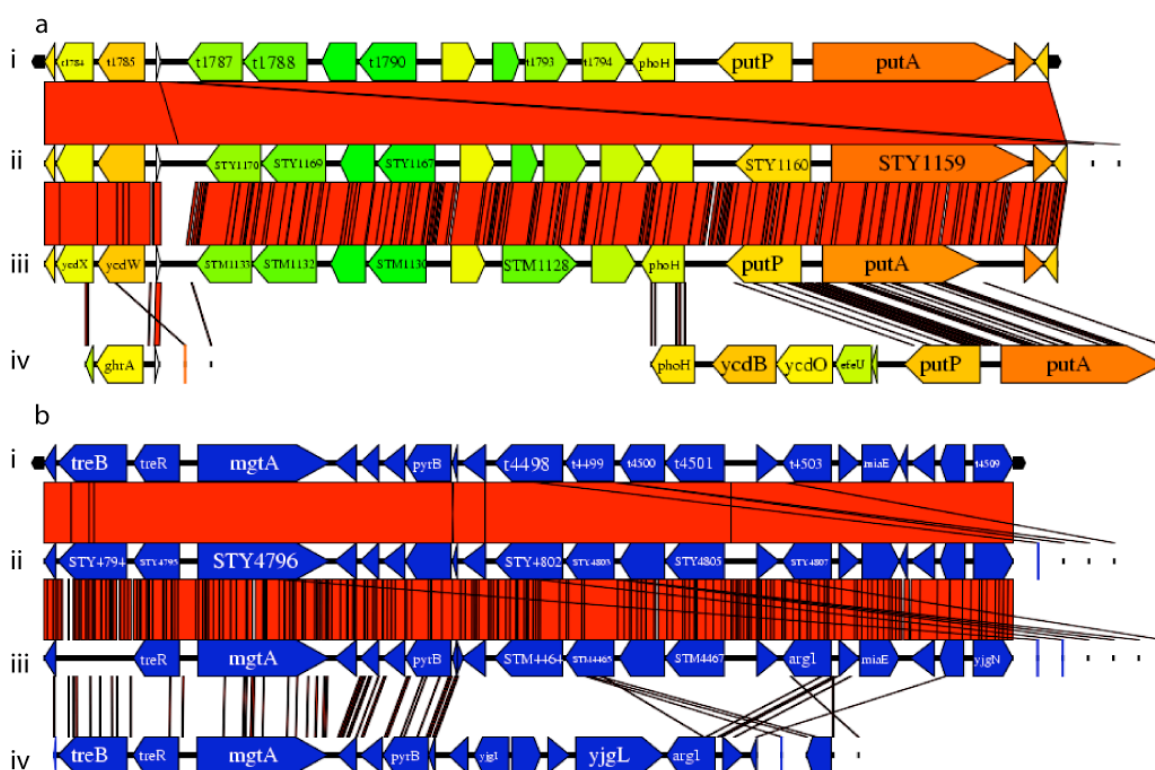


Figure 6.1. Comparative alignment of regions predicted to be OmpR regulated in *S. Typhi*.

(a) Ty2 (i) genes t1787-t1793 compared with CT18(ii), *S. Typhimurium* LT2 (iii) and *E. coli* K12 (iv). (b) Ty2 (i) genes t4498-4501 compared with CT18(ii), *S. Typhimurium* LT2 (iii) and *E. coli* K12 (iv).

To determine if this alignment is present in any other species we performed a nucleotide BLAST of the whole region (figure 6.1 (b)).

Sequence comparison suggested that t1787-t1793 may be involved in the uptake and metabolism of combined carbon and nitrogen sources such as amino sugars. Amino sugars are present in the bacterial cell wall in the form of *N*-acetylglucosamine (NAG) and *N*-acetylmuramic acid (NAM) [270] and are present in the eukaryotic cell in the form of sialic acid (*N*-acetylneuraminic acid) [271] (figure 6.2).

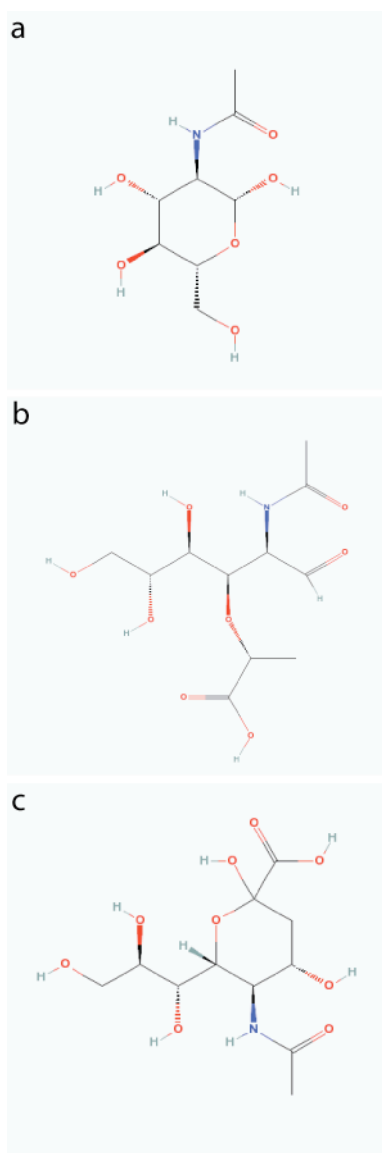


Figure 6.2 The chemical structure of related acetylated aminosugars.

Functional identity of each protein sequence suggests the t1787-t1793 group of contiguous genes were associated with amino sugar metabolism and transport. Derivatives and structures of biologically relevant amino sugars are (a) N-acetyl glucosamine (b) N-acetyl muramic acid (c) N-acetyl neuraminic acid

### 6.3.1.2 t4498-4501 exhibited decreased expression in the *ompR* mutant

Nucleotide BLAST alignment of the region t4498-4501 reveals syntenic conservation is only present in *S. enterica*. Serovars Newport, Paratyphi B, Enteritidis, Dublin,

Typhimurium, Paratyphi A, Cholerae, Gallinarum, Agona, Heidelberg, Schwarzengrund have sequence identity greater than 98% and Arizonae, 94%. No other sequenced species encodes four similar syntenic orthologous genes.

Interestingly, pBLAST and Pfam database searches of each individual predicted protein within this region, reveals each gene encodes functional identity to components of an arginine deiminase system (table 6.2). Arginine deiminase systems are important for maintaining pH homeostasis, can protect cells from acidification and provide ATP [272]. In *Streptococcus pyogenes* the ADS contributes to virulence [273]. A functioning ADS is required for anaerobic growth in some bacterial species [274].

Table 6.2 Differentially expressed hypothetical genes re-annotated according to functional identity derived using pBLAST and Pfam database searches.

Systematic ID/Gene Name	Base Range	Expression Values	p-value	GC content	Sequence Homologies	S.Typhimurium Orthologue
t1787	1846687..1847790	2.6	0.00583	0.468	Oxidoreductase	STM1133
t1788	1847804..1849084	2.5	0.00572	0.451	Sialic acid transporter	STM1132
t1789	1849367..1850059	200	0.00140	0.363	Secreted protein	STM1131
t1790	1850105..1851259	7.4	4.93E-07	0.402	Sialic acid lyase	STM1130
t1791	1851766..1852446	3.9	0.00125	0.514	N-acetylmannosamine-6-phosphate epimerase	STM1129
t1792	1852783..1853310	5.3	8.27E-06	0.439	Sodium/Glucose co-transporter	STM1128
t1793	1853423..1854280	1.3	0.00962	0.470	Transcriptional regulator	
t4498	4644775..4646178	0.26	2.07E-04	0.568	Arginine repressor	STM4464
t4499	4646234..4647238	0.17	5.08E-05	0.558	Ornithine carbamoyltransferase	STM4465
t4500	4647350..4648282	0.16	4.81E-05	0.592	Carbamate kinase	STM4466
t4501	4648293..4649513	0.16	4.60E-05	0.525	Arginine deiminase	STM4467

### 6.3.2 Identification of proteins using LC-MS

In order to identify proteins expressed in *S. Typhi*, LC-MS peptide sequencing technology was exploited to sequence trypsin-digested proteins and the sequence spectra were mapped back to an *in silico* generated 6-frame translation of the entire

Ty2 genome. The peptide sequence for each spectrum was predicted using MascotPercolator and assigned a q-value according to the similarity. Peptides with a score less than 0.01 were mapped to the genome resulting in an estimated FDR < 0.076.

Proteins were resolved using SDS-PAGE from a series of protein preparations derived from cellular fractionation [183], heat shock and by precipitation of the supernatant to increase the sensitivity of the possible number of peptides sequenced. Each gel was cut into small fragments then destained, trypsin digested and the peptides extracted for sequencing. This technique permitted identification of peptide data that mapped back to 1434 genes, 122 of which were not previously identified by Ansong *et al* [275].

Using this method we were able to identify peptides that mapped to the hypothetical genes t1789, t4499, t4500 and t4501 confirming these genes as being translated into proteins under normal conditions (figure 6.3).

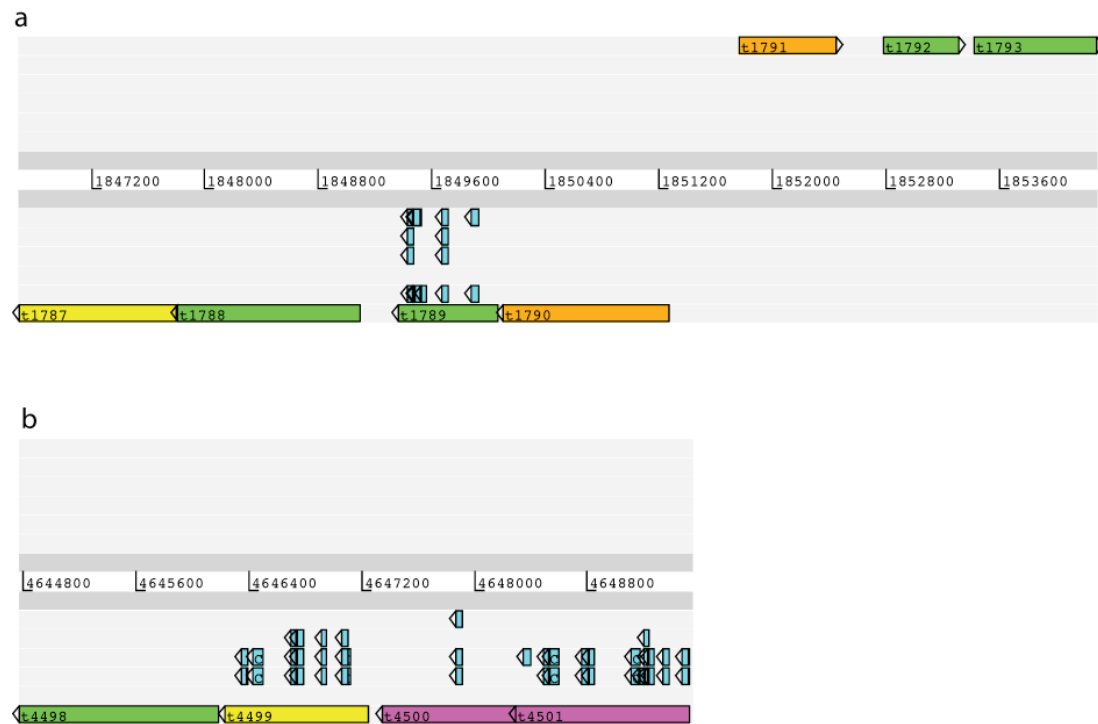


Figure 6.3 Sequenced peptides mapped to regions predicted to be OmpR-regulated.

Peptides are shown aligned to open reading frames in light blue. (a) Region with increased expression. (b) regions with decreased expression. Uniquely mapped peptides are shown (FDR<0.076) No peptides mapped to unannotated frame.

Although this data is not described further in detail here it has been used as a resource for this thesis and other workers to define peptides associated with *S. Typhi* and to complement transcriptome and other biological studies. The data are currently available on request.

### 6.3.3 Construction of recombinant *S. Typhimurium*

mutants in which homologues of t1787-1790, t1791-1793 and t4498-4501 are deleted

In order to try to determine the function of the hypothetical genes, we decided to use the broad host-range pathogen, *S. Typhimurium*, which can be readily employed in

murine infection studies and in vitro growth experiments. *S. Typhimurium* SL1344 requires L-Histidine for growth on minimal media. However, this amino acid cannot support SL1344 growth alone and must be used in combination with an alternative carbon and nitrogen source.

As a step towards identifying a phenotype for these genes using in vitro or in vivo assays, three isogenic mutants were constructed in *S. Typhimurium* SL1344, replacing three distinct groups of genes with a kanamycin cassette (figure 6.4), grouping the genes by the coding strand within each region. The mutant strains were RAK103, RAK105 and TT56.1.

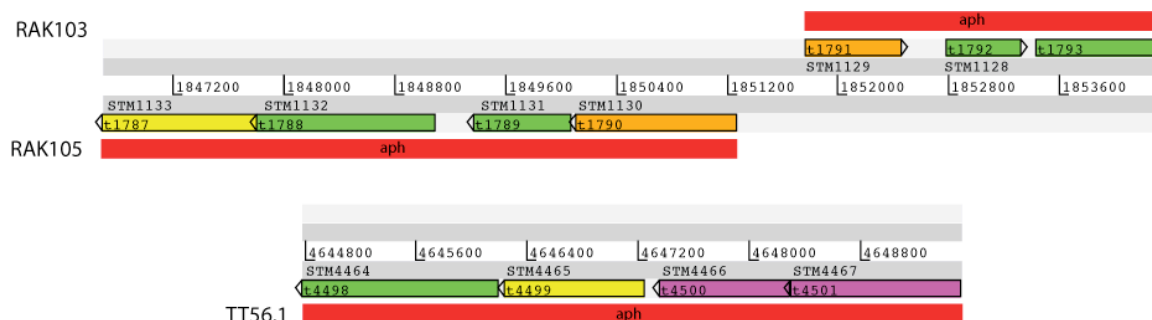


Figure 6.4 Details of mutagenesis for further phenotypic characterisation.

Red feature represents region replaced with kanamycin insert. Gene numbers for *S. Typhimurium* LT2 (STM) are shown. Mutant derivative names are shown on left in line with kanamycin insertion position.

### 6.3.4 *in vitro* Phenotyping of Mutants

#### 6.3.4.1 LPS silver stained gel

BLAST homology of the t1790 indicated that it had similarities to proteins involved in catabolism of amino sugars, which are present in the form of NAM and NAG as components of the bacterial cell wall. To determine if the lipopolysaccharide (LPS) was disrupted in any of the mutants we separated LPS prepared from the mutant strains and the SL1344 isogenic parent strain using SDS-PAGE and visualised the LPS by silver staining (figure 6.5). The pattern on the gel is consistent with the wild-type SL1344 LPS, thus the mutation does not detectably affect the endotoxin.

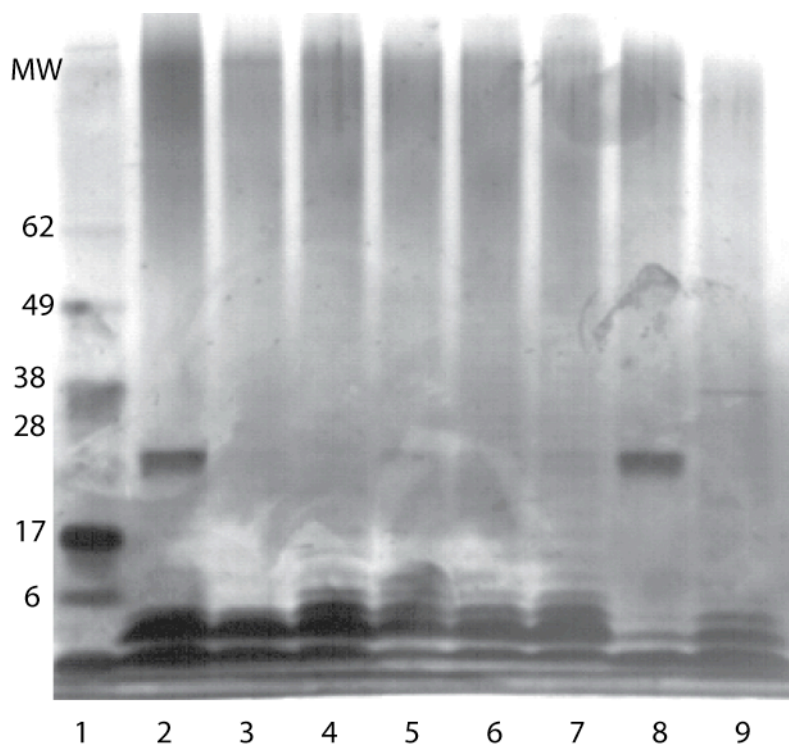


Figure 6.5 LPS silver stained gel.

MW, molecular weight, kDa; Lane 1, protein ladder; 2, BRD948; 3, BRD948  $\Delta ompR$ ; 4, SL1344; 5, SL1344  $\Delta ompR$ ; 6, RAK103; 7, RAK105; 8, LPS negative control (rough LPS); 9, LPS positive control (smooth LPS).

#### 6.4.4.2 Growth on different carbon sources

To determine if the growth of strains RAK103 and RAK105 were deficient using specific aminosugars as a sole carbon source, we grew them on minimal media containing a variety of amino linked carbon sources (table 6.3). To determine if nitrogen utilised from this sole carbon source we grew RAK105 and RAK103 on minimal media without nitrogen (table 6.3).

All wild type and mutant SL1344 derivatives were unable to grow on minimal media without a carbon source and where the sole carbon source was either pectin, galacturonic acid or muramic acid. All wild type and mutant SL1344 strains were able to grow on minimal media with glucose and sialic acid (*N-acetyl* neuraminic acid) as

the sole carbon source and attenuated growth was observed with cytidine sialic acid. SL1344 and RAK105 were competent for growth on minimal medium supplemented with *N-acetyl* muramic acid as the sole carbon source. However, RAK103 was not able to grow on this medium, suggesting the metabolic pathway for *N-acetyl* muramic acid catabolism was interrupted by the mutation of genes STM1128-29.

Table 6.3 Growth on specific sole carbon sources

Carbon Source	WT	RAK103	RAK105
No Carbon source	-	-	-
Glucose	+	+	+
Pectin	-	-	-
Galacturonic acid	-	-	-
Muramic acid	-	-	-
Cytidine sialic acid	-(‘+’)	-(‘+’)	-(‘+’)
<i>N</i> -acetyl muramic acid	+	-	+
Sialic acid	+	+	+

Legend. -, no growth; + significant growth; -(‘+’), very little growth but visible colonies.

## 6.4.5 *in vivo* phenotyping

### 6.4.5.1 Competitive infection assays

To determine if deletions in the t1787-1790 (STM1130-33), t1791-1793 (STM1128-29) and t4498-4501 (STM4464-67) hypothetical genes were attenuating, we performed a number of different assays using BALB/c mice. These mice are susceptible to *S. Typhimurium* SL1344, which causes a systemic infection. By day 5 post inoculation with a lethal dose of SL1344 mice are normally moribund and are humanely killed. Consequently, killing the cohort of 5 mice on day 5 terminated all *in vivo* assays, unless the mice were moribund prior to this day. Initial screening of

mutations was performed using a competitive infection assay [122]. Here mice were inoculated, by oral gavage, with equal CFU counts of both wild-type (SL1344 $\Delta$ *phoN*) and the mutant. The mice were killed on day 5 and the spleen, liver, mesenteric lymph nodes, caecum and terminal ileum were homogenised and serial 10-fold dilutions plated on selective media to determine viable counts. Selective plates contain kanamycin and Xphos (2-Dicyclohexylphosphino-2',4',6'-triisopropylbiphenyl) to distinguish between SL1344 $\Delta$ *phoN* and the mutants as colonies with functional PhoN are blue and without functional PhoN, are white. Ratios of the mutant: SL1344  $\Delta$ *phoN* and its consistency revealed if the mutant was attenuated and demanded further screening.

Initially we used this method to validate the use of SL1344  $\Delta$ *phoN* as the “wild-type” strain and compared it with SL1344 (figure 6.5). We directly compared SL1344  $\Delta$ *phoN* with SL1344  $\Delta$ *ompR*, as this strain is highly attenuated (figure 6.5) [70], to validate the system works for known attenuating mutations.

The comparison of our three mutations revealed no detectable attenuation in pathogenic phenotype (figure 6.7).

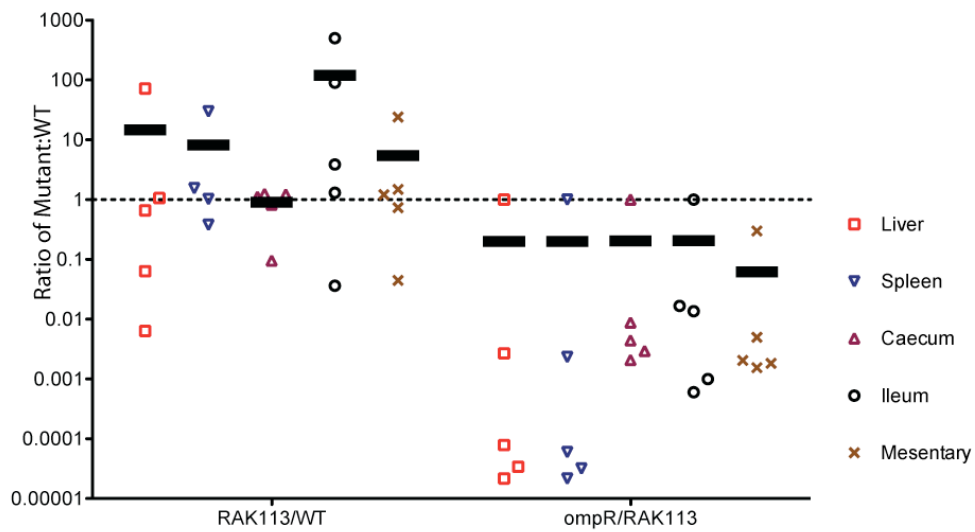


Figure 6.6 Control competitive infections.

Left group, comparison of RAK113 (SL1344  $\Delta phoN$ ) and SL1344 wild-type; right group, comparison of SL1344 $\Delta ompR$  and RAK113. Ratio below 1 indicates fewer bacteria survive compared with control.

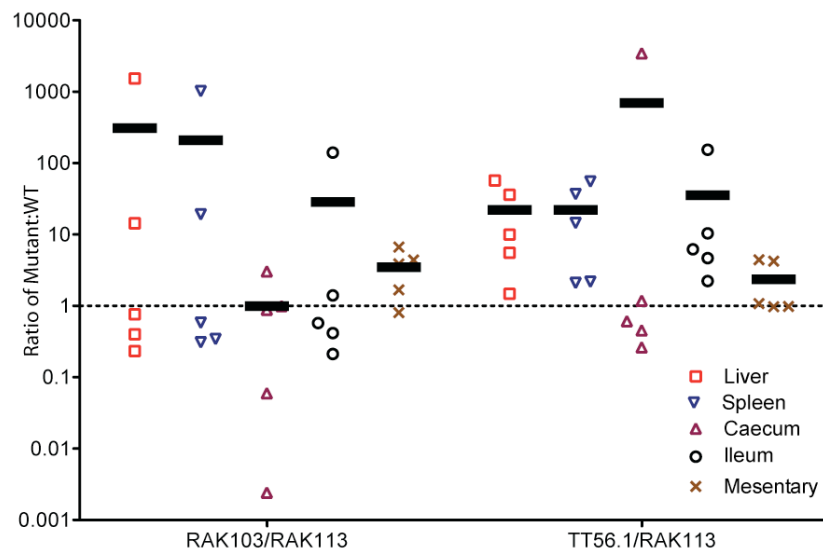


Figure 6.7 Comparison of mutants RAK103 and TT56.1 with RAK113 (SL1344  $\Delta phoN$ ).

Left group, comparison of RAK113 (SL1344  $\Delta phoN$ ) and RAK103; right group, comparison of RAK113 and TT56.1. Ratio below 1 indicates fewer bacteria survive compared with control.

## 6.5 Discussion

The aim of this chapter is to characterise the phenotypic difference caused by the deletion of hypothetical genes encoded within regions identified in Chapter 5. Experimental design encompasses *in vitro* and *in vivo* assays, rationally designed to determine the biological function and importance of OmpR-regulated *Salmonella* specific genes. Most importantly, this chapter is designed to further define *Salmonella* genes that are involved in pathogenesis, either directly or indirectly. Further knowledge of this “virulence regulon” could improve rational vaccine design and heterologous antigen delivery to the host immune system.

Limitations of this experiment lie in the murine co-infection model. The technique showed during the SL1344:SL1344  $\Delta phoN$ , wild-type validation comparison, the system can be quite variable. However, it was also shown by the SL1344:SL1344  $\Delta ompR$  comparison that the system can detect attenuation between co-infected bacterium. For phenotypes with limited pathogenic differences further work was abandoned due to ethical considerations.

### 6.5.1 RAK103 and RAK105 characterisation

It is important to note, *in vivo* assays were performed using the RAK105 strain in the early stages of these experiments. However, during the first round of murine infections the incorrect wild-type strain was used due to inexperience with P22 transduction and transferring the *phoN::kan* cassette from a different wild-type strain. This led to the use of a more virulent wild-type (ATC13028) as the control for this experiment. The results of this experiment indicated there was no attenuation of

RAK105 and in line with ethical guidelines, it was decided not to infect and kill another cohort of mice.

Functional homology of the contiguous genes deleted in RAK103 and RAK105 suggest the genes are involved in uptake and metabolism of an amino sugar-like substrate. The putative function assigns similarity to proteins involved in *N*-acetylneuraminic acid (sialic acid, NAN) transport, a *N*-acetylmannosamine-6-phosphate epimerase and a sialic acid lyase. Amino sugars are present in many biological systems, eukaryote and prokaryote, and alternating residues of NAM and NAG constitute the peptidoglycan layer of the bacterial cell wall [270]. Pectin, found in the primary cell wall of plants, is a polymer of  $\alpha$ -(1-4)-linked D-galacturonic acid [276] and sialic acid is abundant in many mammalian cell types [277] and is essential for immunological discrimination between host and pathogen [271]. In these data, the lack of NAM as a carbon source attenuates growth of RAK103. RAK103 harbours a mutation in the genes STM1128 and STM1129, with functional homologies to a *N*-acetylmannosamine-6-phosphate epimerase and a sodium/glucose co-transporter. Epimerase catalyses the inversion of an epimer, a well-characterised example is the intestinal enzyme lactase, which converts UDP-galactose to UDP-glucose via inversion of the epimeric C4-OH group. Muramic acid, the major compound in NAM, consists of glucosamine and lactate joined by an ether bond. This compound is widely available in biological systems and this operon may be required to scavenge bacterial cell wall remnants in the gut or environment. Other possible sources of NAM could be derived from nutrient rich media such as milk or eggs. Lysozyme is known to cleave glycosidic bonds that polymerise NAM and NAG and this pathway may be useful in maintaining bacterial growth in its presence [278]. The mutant strain RAK105, harbouring mutations in the 4 genes with functional identity to an

oxidoreductase, sialic acid transporter, a secreted protein and sialic acid lyase, was not attenuated in growth with NAM providing the sole carbon source. This suggests the predicted RAK105 operon may be involved in metabolising a closely related substrate possibly derived from cleaved peptidoglycan. However, this remains to be elucidated. It is entirely possible excess peptidoglycan is synthesised and the organism recycles it through this pathway or scavenged from an extracellular source, possibly derived from the death of other bacteria.

Analysis of *in vivo* competitive infection for both RAK103 and RAK105 reveals no significance between wild-type SL1344 and the mutant derivatives. Due to ethical considerations we chose not to further characterise the virulence of these mutants. However, the data suggest the operon is not directly involved in acute *Salmonella* murine infection. Lawley *et al* [127] found during a TraSH screen of mutants that the STM1131 gene was lost during a persistent infection. This gene was not further studied as it was a high-throughput screen and it was only negatively selected for in one round of infections.

In *S. Typhi*, the homologous region to the STM1128 gene is annotated as two separate genes in both Ty2 (t1792 and t1793) and CT18 (STY1164 and STY1165). Closer analysis suggests this gene may be a pseudogene, providing further evidence of genome degradation in this human-restricted organism. Pseudogenes may offer clues to genes redundant for non-host lifestyle and these data suggest functional deletion of this gene does not affect intracellular lifestyle.

### 6.5.2 TT56.1 phenotype

The arginine deiminase system (ADS) is not present in higher eukaryotes but is common among prokaryotes. Characterised systems are known to include three functional proteins, arginine deiminase, carbamoyltransferase and a carbamate kinase [272]. The conserved synteny and the predicted functional homology of the genes in *Salmonella* almost certainly suggest they are components of ADS. The ADS's function is to generate ATP and it does this by the catabolism of L-arginine to ammonia and carbon dioxide [272] and this respiratory process is tightly controlled. Function of ADS in *Pseudomonas aeruginosa* and *Bacillus licheniformis* permits growth under anaerobic conditions [279]. The ADS of *S. pyogenes* has been shown to be involved in cell adhesion and invasion of epithelial cells [273].

In these data we did not find these putative ADS in *Salmonella* to be directly involved in virulence or pathogenicity in a competitive infection assay. Successful colonisation of the host in this model relies on efficacious adhesion and TTSS-mediated endocytosis in the anaerobic gut. There appears to be no obvious attenuation in the mouse caused by the loss of this system.

## 6.6 Conclusion

Many of *Salmonella*'s virulence determinants have been identified through STM and TRASH screens with a great deal of success, effectively picking a great deal of low hanging fruit. However, the rationale for phenotyping such conserved *Salmonella* specific genes in this experiment, based on *ompR* regulation and virulence-associated orthologous systems, make these genes ideal candidates for further characterisation.

Its possible further murine work using a persistence model for Typhoid fever may elucidate a virulence potential of these genes.

## 7 General Discussion

New DNA sequencing technologies have greatly increased the throughput of genome sequencing and have facilitated novel approaches. A combination of Illumina and 454 pyrosequencing recently permitted the re-sequencing of genomes of ~21 *S. Typhi* isolates and the determination of a highly resolved phylogenetic tree [87]. Similar approaches have been applied to eukaryotic genomes to identify chromosome rearrangements, such as those in carcinoma cells (WTSI, unpublished data). This study has used Illumina sequencing technology in a novel manner to characterise the transcriptome of *S. Typhi* by exploiting the intrinsic secondary structure of RNA species. This secondary structure has enabled ligation of linkers and removed the need for a second strand synthesis step, which normally masks complete identification of the transcribed component. This facile method has allowed us to identify the template strand, something not possible using methods reported in recent related publications in Science [188] and Nature [189]. These RNA-seq articles illustrate the application of deep sequencing technology to further define the transcriptome, however, such techniques are biased to sequence enriched polyadenylated RNA. Prokaryotes do not adenylate mRNA so this method cannot readily be applied to *S. Typhi*. The 16s and 23s rRNA are the most abundant species in the bacterial transcriptome and it was predicted the presence of these RNA molecules would reduce sequencing capacity for less abundant transcripts. Sequencing of the total RNA population in Chapter 3 illustrates the importance of the rRNA depletion step in identifying transcription of less abundant genes. rRNA depletion increased the sequencing capacity available for less abundant transcripts, however, the extra steps involved also increased the manipulation of naturally labile bacterial RNA. Thus, these data are potentially

limited by the level of removal of rRNA, 16s and 23s, the ability of each target cDNA to form a sequencable secondary structure and the half-life of mRNA. Further work should include direct comparison with ds-cDNA Illumina sequencing and application to identifying directionality in eukaryotic transcriptomes.

Importantly, these data contain many endogenous controls such as the presence of high levels of transcription from well-characterised genes such as *fliC* [203], the *viaB* locus [118], *sopE* [103] and the SPI-1 locus [12]. The resolution of the technique is such that it is possible to identify RNA editing features such as attenuation of the threonine leader peptide region [205], recently identified RNA regulatory elements such as the *glmS* riboswitch [202] and the small non-coding RNA, *tkeI* [280]. Furthermore, comparison of transcriptomes with the *ompR* mutant *S. Typhi* supports the validity of the sequence data. There was very little sequence data mapping to the *ompC* [139] or *viaB* locus [116] as expected in this mutant. Subsequently, these endogenous controls permitted putative identification of non-coding elements, either as small non-coding RNAs or putative riboswitches. Many had not been previously identified and were supported by further *in silico* analyses. This technique has the capacity to identify the template strand and these data offer the scientific community an important database for further work on prokaryote transcriptomes and the virulence potential of such genes in *Salmonella*. Further work may include confirmation of these binding sites using gel shift assays. A high-throughput epitope tagging programme is currently being devised to study every gene with a predicted DNA binding domain in *Salmonella*.

DNA microarray technology was exploited in Chapter 5 to further define the OmpR regulon. These data indicate OmpR is involved in regulating outer membrane and

virulence associated genes and adjusting the respiratory programme. Many genes that were altered were associated with anaerobic metabolism, an essential component of survival in the anaerobic gut. The chIP-seq method using the Illumina platform was also developed and these data were used to identify binding sites for OmpR. Encoded within many of the chIP-enriched sites was a particular motif that was sometimes encoded as a couplet. This putative consensus sequence was upstream of many of the previously published genes in the OmpR regulon.

As OmpR is a conserved two-component regulator and is present in *E. coli*, why does it control so many horizontally acquired regions of DNA encoded within the *S. enterica* genome acquired since speciation? Some of the OmpR-regulated loci are essential for virulence and these data extend the OmpR regulon to include hypothetical genes, which are not present in *E. coli*. Two novel regions were identified and characterised based on sequence identity to previously described genes. The virulence potential of these genes was assessed using an acute infection murine infection model by competitive challenges but there was no significant difference detected. The gene knockouts were also characterised *in vitro* and growth of the mutant RAK103 was attenuated when the sole carbon source was *N*-acetyl muramic acid, a component of the peptidoglycan cell wall [270]. Persistence is also an important facet of *Salmonella* pathogenicity and mutants in these loci should be analysed in their ability to persist in the murine model for Typhoid fever.

This study encompasses methods developed for exploiting new technology in a novel manner to assess the transcriptome of bacteria. Interrogating a biologically conserved and virulence-essential regulator, the *ompR* gene, has supported the validity of these methods. The regulon was further defined by identifying OmpR bound sequences

using Illumina sequencing technology. Furthermore, newly defined OmpR regulated loci were characterised *in vivo* and *in vitro*. This body of work aims to increase the overall knowledge of the *S. Typhi* organism as a model and to develop methods suitable for interrogating both prokaryotes and eukaryotes and providing a more focussed approach to gene characterisation.

## 8 References

1. Ochman H, Wilson AC (1987) Evolution in Bacteria: Evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26: 74-86.
2. Kauffmann F (1973) [The classification and nomenclature of salmonella-species]. *Zentralbl Bakteriol [Orig A]* 223: 508-512.
3. Crosa JH, Brenner DJ, Ewing WH, Falkow S (1973) Molecular relationships among the Salmonelleae. *J Bacteriol* 115: 307-315.
4. Brenner FW, Villar RG, Angulo FJ, Tauxe R, Swaminathan B (2000) Salmonella nomenclature. *J Clin Microbiol* 38: 2465-2467.
5. Lee WS, Hafeez A, Hassan H, Raja NS, Puthucherry SD (2005) Focal non-typhoidal Salmonella infections from a single center in Malaysia. *Southeast Asian J Trop Med Public Health* 36: 678-682.
6. Gupta SK, Pandit A, White DG, Evans PD (2004) Salmonella osteomyelitis of the thoracic spine: an unusual presentation. *Postgrad Med J* 80: 110-111.
7. Langridge GC, Nair S, Wain J (2009) Nontyphoidal Salmonella serovars cause different degrees of invasive disease globally. *J Infect Dis* 199: 602-603.
8. Crump JA, Luby SP, Mintz ED (2004) The global burden of typhoid fever. *Bull World Health Organ* 82: 346-353.
9. Bitar R, Tarpley J (1985) Intestinal perforation and typhoid fever: a historical and state-of-the-art review. *Rev Infect Dis* 7: 257.
10. Kariuki S, Revathi G, Kariuki N, Kiiru J, Mwituria J, et al. (2006) Invasive multidrug-resistant non-typhoidal Salmonella infections in Africa: zoonotic or anthroponotic transmission? *J Med Microbiol* 55: 585-591.
11. Tsolis RM, Kingsley RA, Townsend SM, Ficht TA, Adams LG, et al. (1999) Of mice, calves, and men. Comparison of the mouse typhoid model with other Salmonella infections. *Adv Exp Med Biol* 473: 261-274.
12. Galán JE, Curtiss III R (1989) Cloning and molecular characterization of genes whose products allow *Salmonella typhimurium* to penetrate tissue culture cells. *Proc Natl Acad Sci USA* 86: 6383-6387.
13. Shea JE, Hensel M, Gleeson C, Holden DW (1996) Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. *Proc Natl Acad Sci USA* 93: 2593-2597.
14. Nnalue NA, Stocker BA (1987) Test of the virulence and live-vaccine efficacy of auxotrophic and galE derivatives of Salmonella choleraesuis. *Infect Immun* 55: 955-962.
15. Stocker BA (1988) Auxotrophic Salmonella typhi as live vaccine. *Vaccine* 6: 141-145.

16. Segall T, Lindberg AA (1991) *Salmonella dublin* experimental infection in calves: protection after oral immunization with an auxotrophic *aroA* live vaccine. *Zentralblatt Fur Veterinarmedizin Reihe B* 38: 142-160.
17. Drazek ES, Hough HS, Crawford RM, Hadfield TL, Hoover DL, et al. (1995) Deletion of *purE* attenuates *Brucella melitensis* 16M for growth in human monocyte-derived macrophages. *Infection & Immunity* 63: 3297-3301.
18. Kelly-Hope LA, Alonso WJ, Thiem VD, Anh DD, Canh do G, et al. (2007) Geographical distribution and risk factors associated with enteric diseases in Vietnam. *Am J Trop Med Hyg* 76: 706-712.
19. Garcia-del Portillo F, Foster JW, Finlay BB (1993) Role of acid tolerance response genes in *Salmonella typhimurium* virulence. *Infect Immun* 61: 4489-4492.
20. Selsted ME, Miller SI, Henschen AH, Ouellette AJ (1992) Enteric defensins: antibiotic peptide components in intestinal host defense. *J Cell Biol* 118: 929-936.
21. Michetti P, Mahan MJ, Slauch JM, Mekalanos JJ, Neutra MR (1992) Monoclonal secretory immunoglobulin A protects mice against oral challenge with the invasive pathogen *Salmonella typhimurium*. *Infect Immun* 60: 1786-1792.
22. Prouty AM, Brodsky IE, Falkow S, Gunn JS (2004) Bile-salt-mediated induction of antimicrobial and bile resistance in *Salmonella typhimurium*. *Microbiology* 150: 775-783.
23. Altmeyer RM, McNern JK, Bossio JC, Rosenshine I, Finlay BB, et al. (1993) Cloning and molecular characterization of a gene involved in *Salmonella* adherence and invasion of cultured epithelial cells. *Mol Microbiol* 7: 89-98.
24. Robertson JM, Grant G, Allen-Vercos E, Woodward MJ, Pusztai A, et al. (2000) Adhesion of *Salmonella enterica* var *Enteritidis* strains lacking fimbriae and flagella to rat ileal explants cultured at the air interface or submerged in tissue culture medium. *J Med Microbiol* 49: 691-696.
25. Edwards RA, Schifferli DM, Maloy SR (2000) A role for *Salmonella* fimbriae in intraperitoneal infections. *Proc Natl Acad Sci U S A* 97: 1258-1262.
26. Lindberg F, Lund B, Johansson L, Normark S (1987) Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus. *Nature* 328: 84-87.
27. Wilson RL, Elthon J, Clegg S, Jones BD (2000) *Salmonella enterica* serovars *gallinarum* and *pullorum* expressing *Salmonella enterica* serovar *typhimurium* type 1 fimbriae exhibit increased invasiveness for mammalian cells. *Infect Immun* 68: 4782-4785.
28. Bloch CA, Stocker BA, Orndorff PE (1992) A key role for type 1 pili in enterobacterial communicability. *Mol Microbiol* 6: 697-701.
29. Fujimura Y (1986) Functional morphology of microfold cells (M cells) in Payer's patches - phagocytosis and transport of BCG by M cells into rabbit Payer's patches. *Gastroenterol Jpn* 21: 325-335.

30. Vazquez-Torres A, Jones-Carson J, Baumler AJ, Falkow S, Valdivia R, et al. (1999) Extraintestinal dissemination of *Salmonella* by CD18-expressing phagocytes. *Nature* 401: 804-808.
31. Boyle EC, Brown NF, Finlay BB (2006) *Salmonella enterica* serovar Typhimurium effectors SopB, SopE, SopE2 and SipA disrupt tight junction structure and function. *Cell Microbiol* 8: 1946-1957.
32. Levine WC, Buehler JW, Bean NH, Tauxe RV (1991) Epidemiology of nontyphoidal *Salmonella* bacteremia during the human immunodeficiency virus epidemic. *J Infect Dis* 164: 81-87.
33. Bronzan RN, Taylor TE, Mwenechanya J, Tembo M, Kayira K, et al. (2007) Bacteremia in Malawian children with severe malaria: prevalence, etiology, HIV coinfection, and outcome. *J Infect Dis* 195: 895-904.
34. Sharma A, Qadri A (2004) Vi polysaccharide of *Salmonella typhi* targets the prohibitin family of molecules in intestinal epithelial cells and suppresses early inflammatory responses. *Proc Natl Acad Sci U S A* 101: 17492-17497.
35. Raffatellu M, Chessa D, Wilson RP, Dusold R, Rubino S, et al. (2005) The Vi capsular antigen of *Salmonella enterica* serotype Typhi reduces Toll-like receptor-dependent interleukin-8 expression in the intestinal mucosa. *Infect Immun* 73: 3367-3374.
36. Alpuche-Aranda CM, Berthiaume EP, Mock B, Swanson JA, Miller SI (1995) Spacious phagosome formation within mouse macrophages correlates with *Salmonella* serotype pathogenicity and host susceptibility. *Infec Immun* 63: 4456-4462.
37. Alpuche-Aranda CM, Racoosin EL, Swanson JA, Miller SI (1994) *Salmonella* stimulate macrophage macropinocytosis and persist within spacious phagosomes. *J Exp Med* 179: 601-608.
38. Chen LM, Hobbie S, Galan JE (1996) Requirement of cdc42 for salmonella-induced cytoskeletal and nuclear responses. *Science* 274: 2115-2118.
39. Frances CL, Ryan TA, Jones BD, Smith SJ, Falkow S (1993) Ruffles induced by *Salmonella* and other stimuli direct macropinocytosis of bacteria. *Nature* 364: 639-642.
40. Garcia-del Portillo F, Finlay BB (1995) Targeting of *Salmonella typhimurium* to vesicles containing lysosomal membrane glycoproteins bypasses compartments with mannose 6-phosphate receptors. *Journal of Cell Biology* 129: 81-97.
41. Harrison RE, Brumell JH, Khandani A, Bucci C, Scott CC, et al. (2004) *Salmonella* impairs RILP recruitment to Rab7 during maturation of invasion vacuoles. *Mol Biol Cell*.
42. Marsman M, Jordens I, Kuijl C, Janssen L, Neefjes J (2004) Dynein-mediated Vesicle Transport Controls Intracellular *Salmonella* Replication. *Mol Biol Cell* 15: 2954-2964.
43. Guignot J, Caron E, Beuzon C, Bucci C, Kagan J, et al. (2004) Microtubule motors control membrane dynamics of *Salmonella*-containing vacuoles. *J Cell Sci* 117: 1033-1045.

44. Steele-Mortimer O, Meresse S, Gorvel JP, Toh BH, Finlay BB (1999) Biogenesis of *Salmonella typhimurium*-containing vacuoles in epithelial cells involves interactions with the early endocytic pathway. *Cell Microbiol* 1: 33-49.
45. Garvis SG, Beuzon CR, Holden DW (2001) A role for the PhoP/Q regulon in inhibition of fusion between lysosomes and *Salmonella*-containing vacuoles in macrophages. *Cell Microbiol* 3: 731-744.
46. Abrahams GL, Muller P, Hensel M (2006) Functional dissection of SseF, a type III effector protein involved in positioning the salmonella-containing vacuole. *Traffic* 7: 950-965.
47. Yu XJ, Ruiz-Albert J, Unsworth KE, Garvis S, Liu M, et al. (2002) SpiC is required for secretion of *Salmonella* Pathogenicity Island 2 type III secretion system proteins. *Cell Microbiol* 4: 531-540.
48. Drecktrah D, Knodler LA, Howe D, Steele-Mortimer O (2007) *Salmonella* trafficking is defined by continuous dynamic interactions with the endolysosomal system. *Traffic* 8: 212-225.
49. Steele-Mortimer O, Brumell JH, Knodler LA, Meresse S, Lopez A, et al. (2002) The invasion-associated type III secretion system of *Salmonella enterica* serovar Typhimurium is necessary for intracellular proliferation and vacuole biogenesis in epithelial cells. *Cell Microbiol* 4: 43-54.
50. Cirillo DM, Valdivia RH, Monack DM, Falkow S (1998) Macrophage-dependent induction of the *Salmonella* pathogenicity island 2 type III secretion system and its role in intracellular survival. *Mol Microbiol* 30: 175-188.
51. Klein JR, Jones BD (2001) *Salmonella* pathogenicity island 2-encoded proteins SseC and SseD are essential for virulence and are substrates of the type III secretion system. *Infect Immun* 69: 737-743.
52. Brumell JH, Kujat-Choy S, Brown NF, Vallance BA, Knodler LA, et al. (2003) SopD2 is a novel type III secreted effector of *Salmonella typhimurium* that targets late endocytic compartments upon delivery into host cells. *Traffic* 4: 36-48.
53. Brumell JH, Rosenberger CM, Gotto GT, Marcus SL, Finlay BB (2001) SifA permits survival and replication of *Salmonella typhimurium* in murine macrophages. *Cell Microbiol* 3: 75-84.
54. Gallois A, Klein JR, Allen LA, Jones BD, Nauseef WM (2001) *Salmonella* pathogenicity island 2-encoded type III secretion system mediates exclusion of NADPH oxidase assembly from the phagosomal membrane. *J Immunol* 166: 5741-5748.
55. Lindberg AA, Segall T, Weintraub A, Stocker BA (1993) Antibody response and protection against challenge in mice vaccinated intraperitoneally with a live aroA O<sub>4</sub>-O<sub>9</sub> hybrid *Salmonella dublin* strain. *Infection & Immunity* 61: 1211-1221.
56. Segall T, Jacobsson SO, Karlsson K, Lindberg AA (1994) Mucosal immune responses in calves orally vaccinated with a live auxotrophic aroA *Salmonella dublin* strain. *Zentralblatt Fur Veterinarmedizin Reihe B* 41: 305-312.

57. Formica S, Roach TI, Blackwell JM (1994) Interaction with extracellular matrix proteins influences Lsh/Ity/Bcg (candidate Nramp) gene regulation of macrophage priming/activation for tumour necrosis factor- $\alpha$  and nitrite release. *Immunology* 82: 42-50.
58. Jabado N, Cuellar-Mata P, Grinstein S, Gros P (2003) Iron chelators modulate the fusogenic properties of Salmonella-containing phagosomes. *Proc Natl Acad Sci U S A* 100: 6127-6132.
59. Martin-Orozco N, Touret N, Zaharik ML, Park E, Kopelman R, et al. (2006) Visualization of vacuolar acidification-induced transcription of genes of pathogens inside macrophages. *Mol Biol Cell* 17: 498-510.
60. Engels EA, Falagas ME, Lau J, Bennish ML (1998) Typhoid fever vaccines: a meta-analysis of studies on efficacy and toxicity. *Bmj* 316: 110-116.
61. Szu SC, Taylor DN, Trofa AC, Clements JD, Shiloach J, et al. (1994) Laboratory and preliminary clinical characterization of Vi capsular polysaccharide-protein conjugate vaccines. *Infect Immun* 62: 4440-4444.
62. Germanier R, Fürer E (1975) Isolation and Characterization of *galE* mutant Ty21a of *Salmonella typhi*: a candidate strain for a live, oral typhoid vaccine. *J Infect Dis* 131: 553-558.
63. Wahdan MH, Serie C, Germanier R, Lackany A, Cerisier Y, et al. (1980) A controlled field trial of liver oral typhoid vaccine Ty21a. *Bull World Health Organ* 58: 469-474.
64. Levine MM, Ferreccio C, Black RE, Germanier R (1987) Large-scale field trial of Ty21a live oral typhoid vaccine in enteric-coated capsule formulation. *Lancet* i: 1049-1052.
65. Silva B, Gonzalez, C., Mora, GC, and Cabello, F (1987) Genetic characteristics of the *Salmonella typhi* Ty21a vaccine. *J Infect Dis* 155: 1077-1078.
66. Fraser A, Paul M, Goldberg E, Acosta CJ, Leibovici L (2007) Typhoid fever vaccines: systematic review and meta-analysis of randomised controlled trials. *Vaccine* 25: 7848-7857.
67. Smith BP, Reina GM, Hoiseth SK, Stocker BA, Habasha F, et al. (1984) Aromatic-dependent *Salmonella typhimurium* as modified live vaccines for calves. *American Journal of Veterinary Research* 45: 59-66.
68. Stocker BA, Hoiseth SK, Smith BP (1983) Aromatic-dependent "*Salmonella* sp." as live vaccine in mice and calves. *Developments in Biological Standardization* 53: 47-54.
69. Hindle Z, Chatfield SN, Phillimore J, Bentley M, Johnson J, et al. (2002) Characterization of *Salmonella enterica* derivatives harboring defined *aroC* and *Salmonella* pathogenicity island 2 type III secretion system (*ssaV*) mutations by immunization of healthy volunteers. *Infect Immun* 70: 3457-3467.
70. Dorman CJ, Chatfield S, Higgins CF, Hayward C, Dougan G (1989) Characterization of porin and *ompR* mutants of a virulent strain of *Salmonella typhimurium*: *ompR* mutants are attenuated in vivo. *Infect Immun* 57: 2136-2140.

71. Hohmann EL, Oletta CA, Killeen KP, Miller SI (1996) *phoP/phoQ*-deleted *Salmonella typhi* (Ty800) is a safe and immunogenic single-dose typhoid fever vaccine in volunteers. *J Infect Dis* 173: 1408-1414.
72. Johnson K, Charles I, Dougan G, Pickard D, O'Gaora P, et al. (1991) The role of a stress-response protein in *Salmonella typhimurium* virulence. *Mol Microbiol* 5: 401-407.
73. Covone MG, Brocchi M, Palla E, Dias da Silveira W, Rappuoli R, et al. (1998) Levels of expression and immunogenicity of attenuated *Salmonella enterica* serovar typhimurium strains expressing *Escherichia coli* mutant heat-labile enterotoxin. *Infect Immun* 66: 224-231.
74. Kramer U, Rizos K, Apfel H, Autenrieth IB, Lattemann CT (2003) Autodisplay: development of an efficacious system for surface display of antigenic determinants in *Salmonella* vaccine strains. *Infect Immun* 71: 1944-1952.
75. Roland KL, Tinge SA, Killeen KP, Kochi SK (2005) Recent advances in the development of live, attenuated bacterial vectors. *Curr Opin Mol Ther* 7: 62-72.
76. Parkhill J, Dougan G, James KD, Thomson NR, Pickard D, et al. (2001) Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413: 848-852.
77. 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.
78. Bäumlér AJ (1997) The record of horizontal gene transfer in *Salmonella*. *Trends in Microbiol* 5: 318-322.
79. Zhou D, Galán J (2001) *Salmonella* entry into host cells: the work in concert of type III secreted effector proteins. *Microbes Infect* 3: 1293-1298.
80. Bäumlér AJ, Gilde AJ, Tsolis RM, van der Velden AWM, Ahmer BMM, et al. (1997) Contribution of horizontal gene transfer and deletion events to the development of distinctive patterns of fimbrial operons during evolution of *Salmonella* serotypes. *J Bacteriol* 179: 317-322.
81. Bäumlér AJ, Heffron F (1995) Identification and sequence analysis of *lpfABCDE*, a putative fimbrial operon of *Salmonella typhimurium*. *J Bacteriol* 177: 2087-2097.
82. 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. *Nat Genet* 36: 1268-1274.
83. 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 Res* 18: 1624-1637.

84. 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. *J Bacteriol* 185: 5055-5065.
85. Kolyva S, Waxin H, Popoff MY (1992) The Vi antigen of *Salmonella typhi*: molecular analysis of the *viaB* locus. *J Gen Microbiol* 138: 297-304.
86. Wilson RP, Raffatellu M, Chessa D, Winter SE, Tukel C, et al. (2008) The Vi-capsule prevents Toll-like receptor 4 recognition of *Salmonella*. *Cell Microbiol* 10: 876-890.
87. 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*. *Nat Genet* 40: 987-993.
88. Klugman KP, Gilbertson IT, Koornhof HJ, Robbins JB, Schneerson R, et al. (1987) Protective activity of Vi capsular polysaccharide vaccine against typhoid fever. *Lancet* 2: 1165-1169.
89. 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. *J Mol Biol* 339: 279-300.
90. Herrington DA, Hall RH, Losonsky G, Mekalanos JJ, Taylor RK, et al. (1988) Toxin, toxin coregulated pili and *toxR* regulon are essential for *Vibrio cholerae* pathogenesis in humans. *J Exp Med* 168: 1487-1492.
91. Knothe H, Schmidt D (1952) [Diphtheria toxin formation in incubated egg.]. *Arztl Wochensh* 7: 275-277.
92. Desiere F, McShan WM, van Sinderen D, Ferretti JJ, Brussow H (2001) Comparative genomics reveals close genetic relationships between phages from dairy bacteria and pathogenic *Streptococci*: evolutionary implications for prophage-host interactions. *Virology* 288: 325-341.
93. Miold S, Rabsch W, Tschape H, Hardt WD (2001) Transfer of the *Salmonella* type III effector *sopE* between unrelated phage families. *J Mol Biol* 312: 7-16.
94. Altier C (2005) Genetic and environmental control of salmonella invasion. *J Microbiol* 43 Spec No: 85-92.
95. Sukhan A, Kubori T, Wilson J, Galan JE (2001) Genetic analysis of assembly of the *Salmonella enterica* serovar Typhimurium type III secretion-associated needle complex. *J Bacteriol* 183: 1159-1167.
96. Galan JE, Wolf-Watz H (2006) Protein delivery into eukaryotic cells by type III secretion machines. *Nature* 444: 567-573.
97. Raffatellu M, Wilson RP, Chessa D, Andrews-Polymenis H, Tran QT, et al. (2005) SipA, SopA, SopB, SopD, and SopE2 contribute to *Salmonella enterica* serotype typhimurium invasion of epithelial cells. *Infect Immun* 73: 146-154.
98. Jepson MA, Kenny B, Leard AD (2001) Role of sipA in the early stages of *Salmonella typhimurium* entry into epithelial cells. *Cell Microbiol* 3: 417-426.

99. Bakshi CS, Singh VP, Wood MW, Jones PW, Wallis TS, et al. (2000) Identification of SopE2, a *Salmonella* secreted protein which is highly homologous to SopE and involved in bacterial invasion of epithelial cells. *J Bacteriol* 182: 2341-2344.
100. Norris FA, Wilson MP, Wallis TS, Galyov EE, Majerus PW (1998) SopB, a protein required for virulence of *Salmonella dublin*, is an inositol phosphate phosphatase [see comments]. *Proc Natl Acad Sci U S A* 95: 14057-14059.
101. Jones MA, Wood MW, Mullan PB, Watson PR, Wallis TS, et al. (1998) Secreted effector proteins of *Salmonella dublin* act in concert to induce enteritis. *Infect Immun* 66: 5799-5804.
102. Hardt WD, Urlaub H, Galan JE (1998) A substrate of the centisome 63 type III protein secretion system of *Salmonella typhimurium* is encoded by a cryptic bacteriophage. *Proc Natl Acad Sci U S A* 95: 2574-2579.
103. Hardt WD, Chen LM, Schuebel KE, Bustelo XR, 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.
104. Galyov EE, Wood MW, Rosqvist R, Mullan PB, Watson PR, et al. (1997) A secreted effector protein of *Salmonella dublin* is translocated into eukaryotic cells and mediates inflammation and fluid secretion in infected ileal mucosa. *Molecular Microbiology* 25: 903-912.
105. Collazo CM, Galan JE (1997) The invasion-associated type III system of *Salmonella typhimurium* directs the translocation of Sip proteins into the host cell. *Mol Microbiol* 24: 747-756.
106. Wood MW, Rosqvist R, Mullan PB, Edwards MH, Galyov EE (1996) SopE, a secreted protein of *Salmonella dublin*, is translocated into the target eukaryotic cell via a *sip*-dependent mechanism and promotes bacterial entry. *Mol Microbiol* 22: 327-338.
107. McGhie EJ, Hayward RD, Koronakis V (2001) Cooperation between actin-binding proteins of invasive *Salmonella*: SipA potentiates SipC nucleation and bundling of actin. *Embo J* 20: 2131-2139.
108. Haraga A, Ohlson MB, Miller SI (2008) *Salmonellae* interplay with host cells. *Nat Rev Microbiol* 6: 53-66.
109. Worley MJ, Ching KH, Heffron F (2000) *Salmonella* SsrB activates a global regulon of horizontally acquired genes. *Mol Microbiol* 36: 749-761.
110. 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. *Mol Microbiol* 30: 163-174.
111. Nikolaus T, Deiwick J, Rappl C, Freeman JA, Schroder W, et al. (2001) SseBCD proteins are secreted by the type III secretion system of *Salmonella* pathogenicity island 2 and function as a translocon. *J Bacteriol* 183: 6036-6045.
112. Waterman SR, Holden DW (2003) Functions and effectors of the *Salmonella* pathogenicity island 2 type III secretion system. *Cell Microbiol* 5: 501-511.

113. Kuhle V, Hensel M (2004) Cellular microbiology of intracellular *Salmonella enterica*: functions of the type III secretion system encoded by *Salmonella* pathogenicity island 2. *Cell Mol Life Sci* 61: 2812-2826.
114. Felix A, Pitt R (1934) Virulence of *B. typhosus* and Resistance to O Antibody. *J Path & Bact* 38: 409-420.
115. Hashimoto Y, Khan AQ (1997) Comparison of *ViaB* regions of Vi-positive organisms. *FEMS Microbiol Lett* 157: 55-57.
116. Pickard D, Li J, Roberts M, Maskell D, Hone D, et al. (1994) Characterization of defined *ompR* mutants of *Salmonella typhi*: *ompR* is involved in the regulation of Vi polysaccharide expression. *Infect Immun* 62: 3984-3993.
117. Virlogeux I, Waxin H, Ecobichon C, Lee JO, Popoff MY (1996) Characterization of the *rcsA* and *rcsB* genes from *Salmonella typhi*: *rcsB* through *tviA* is involved in regulation of Vi antigen synthesis. *J Bacteriol* 178: 1691-1698.
118. Virlogeux I, Waxin H, Ecobichon C, Popoff MY (1995) Role of the *viaB* locus in synthesis, transport and expression of *Salmonella typhi* Vi antigen. *Microbiology* 141 ( Pt 12): 3039-3047.
119. Blanc-Potard AB, Solomon F, Kayser J, Groisman EA (1999) The SPI-3 pathogenicity island of *salmonella enterica* [In Process Citation]. *J Bacteriol* 181: 998-1004.
120. Kiss T, Morgan E, Nagy G (2007) Contribution of SPI-4 genes to the virulence of *Salmonella enterica*. *FEMS Microbiology Letters* 275: 153-159.
121. Wood MW, Jones MA, Watson PR, Hedges S, Wallis TS, et al. (1998) Identification of a pathogenicity island required for *Salmonella enteropathogenicity*. *Mol Microbiol* 29: 883-891.
122. 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.
123. Bäumler AJ, Tsois RM, Bowe F, Kusters JG, Hoffmann S, et al. (1996) The *pef* fimbrial operon mediates adhesion to murine small intestine and is necessary for fluid accumulation in infant mice. *Infect Immun* 64: 61-68.
124. Bäumler AJ, Tsois RM, Bowe F, Kusters JG, Hoffmann S, et al. (1996) The *pef* fimbrial operon of *Salmonella typhimurium* mediates adhesion to murine small intestine and is necessary for fluid accumulation in the infant mouse. *Infect Immun* 64: 61-68.
125. Bäumler AJ, Tsois RM, Heffron F (1996) The *lpf* fimbrial operon mediates adhesion to murine Peyer's patches. *Proc Natl Acad Sci USA* 93: 279-283.
126. Bäumler AJ, Tsois RM, Heffron F (1996) Contribution of fimbrial operons to attachment to and invasion of epithelial cell lines by *Salmonella typhimurium*. *Infect Immun* 64: 1862-1865.

127. Lawley TD, Chan K, Thompson LJ, Kim CC, Govoni GR, et al. (2006) Genome-Wide Screen for Salmonella Genes Required for Long-Term Systemic Infection of the Mouse. *PLoS Pathogens* 2: e11.
128. Tsui IS, Yip CM, Hackett J, Morris C (2003) The type IVB pili of *Salmonella enterica* serovar Typhi bind to the cystic fibrosis transmembrane conductance regulator. *Infect Immun* 71: 6049-6050.
129. 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.
130. Sun Y-H, Rolan HG, Tsois RM (2007) Injection of Flagellin into the Host Cell Cytosol by *Salmonella enterica* Serotype Typhimurium. *J Biol Chem* 282: 33897-33901.
131. Soncini FC, Garcia Vescovi E, Solomon F, Groisman EA (1996) Molecular basis of the magnesium deprivation response in *Salmonella typhimurium*: identification of PhoP-regulated genes. *J Bacteriol* 178: 5092-5099.
132. Bijlsma JJ, Groisman EA (2005) The PhoP/PhoQ system controls the intramacrophage type three secretion system of *Salmonella enterica*. *Mol Microbiol* 57: 85-96.
133. Alpuche Aranda CM, Swanson JA, Loomis WP, Miller SI (1992) *Salmonella typhimurium* activates virulence gene transcription within acidified macrophage phagosomes. *Proc Natl Acad Sci U S A* 89: 10079-10083.
134. van Velkinburgh JC, Gunn JS (1999) PhoP-PhoQ-regulated loci are required for enhanced bile resistance in *Salmonella* spp. *Infect Immun* 67: 1614-1622.
135. Gunn JS, Miller SI (1996) PhoP-PhoQ activates transcription of pmrAB, encoding a two-component regulatory system involved in salmonella typhimurium antimicrobial peptide resistance. *J Bacteriol* 178: 6857-6864.
136. Deng W, Liou SR, Plunkett G, 3rd, Mayhew GF, Rose DJ, et al. (2003) Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18. *J Bacteriol* 185: 2330-2337.
137. Blattner FR, Plunkett G, Bloch CA, Perna NT, Burland V, et al. (1997) The complete genome sequence of *Escherichia coli* K-12 [Review]. *Science* 277.
138. Nakae T, Ishii JN, Tokunaga H, Kobayashi Y, Nakae R (1982) The solute selectivity of porin pores of *Escherichia coli* and *Salmonella typhimurium*. *Tokai J Exp Clin Med* 7 Suppl: 141-148.
139. Slauch JM, Silhavy TJ (1989) Genetic analysis of the switch that controls porin gene expression in *Escherichia coli* K-12. *J Mol Biol* 210: 281-292.
140. Thomas AD, Booth IR (1992) The regulation of expression of the porin gene ompC by acid pH. *J Gen Microbiol* 138: 1829-1835.
141. Heyde M, Portalier R (1987) Regulation of major outer membrane porin proteins of *Escherichia coli* K 12 by pH. *Mol Gen Genet* 208: 511-517.
142. Contreras I, Munoz L, Toro CS, Mora GC (1995) Heterologous expression of *Escherichia coli* porin genes in *Salmonella typhi* Ty2: regulation by

medium osmolarity, temperature and oxygen availability. FEMS Microbiol Lett 133: 105-111.

143. Fernandez-Mora M, Puente JL, Calva E (2004) OmpR and LeuO positively regulate the *Salmonella enterica* serovar Typhi ompS2 porin gene. J Bacteriol 186: 2909-2920.
144. Oropeza R, Sampieri CL, Puente JL, Calva E (1999) Negative and positive regulation of the non-osmoregulated ompS1 porin gene in *Salmonella typhi*: a novel regulatory mechanism that involves OmpR. Mol Microbiol 32: 243-252.
145. Bang IS, Kim BH, Foster JW, Park YK (2000) OmpR regulates the stationary-phase acid tolerance response of *Salmonella enterica* serovar typhimurium. J Bacteriol 182: 2245-2252.
146. Kutsukake K (1997) Autogenous and global control of the flagellar master operon, *flhD*, in *Salmonella typhimurium*. Mol Gen Genet 254: 440-448.
147. Gerstel U, Park C, Romling U (2003) Complex regulation of *csgD* promoter activity by global regulatory proteins. Mol Microbiol 49: 639-654.
148. Romling U, Sierralta WD, Eriksson K, Normark S (1998) Multicellular and aggregative behaviour of *Salmonella typhimurium* strains is controlled by mutations in the *agfD* promoter. Mol Microbiol 28: 249-264.
149. Bajaj V, Lucas RL, Hwang C, Lee CA (1996) Co-ordinate regulation of *Salmonella typhimurium* invasion genes by environmental factors is mediated by control of *hilA* expression. Mol Microbiol 22: 703-714.
150. Robbins JD, Robbins JB (1984) Reexamination of the protective role of the capsular polysaccharide (Vi antigen) of *Salmonella typhi*. J Infect Dis 150: 436-449.
151. Looney RJ, Steigbigel RT (1986) Role of the Vi antigen of *Salmonella typhi* in resistance to host defense in vitro. J Lab Clin Med 108: 506-516.
152. Bishop A, House D, Perkins T, Baker S, Kingsley RA, et al. (2008) Interaction of *Salmonella enterica* serovar Typhi with cultured epithelial cells: roles of surface structures in adhesion and invasion. Microbiology 154: 1914-1926.
153. Brumell JH, Grinstein S (2004) *Salmonella* redirects phagosomal maturation. Curr Opin Microbiol 7: 78-84.
154. Gibson MM, Ellis EM, Graeme-Cook KA, Higgins CF (1987) OmpR and EnvZ are pleiotropic regulatory proteins: positive regulation of the tripeptide permease (*tppB*) of *Salmonella typhimurium*. Mol Gen Genet 207: 120-129.
155. Valdivia RH, Falkow S (1997) Fluorescence-based isolation of bacterial genes expressed within host cells. Science 277: 2007-2011.
156. Valdivia RH, Falkow S (1996) Bacterial genetics by flow cytometry: rapid isolation of *Salmonella typhimurium* acid-inducible promoters by differential fluorescence induction. Mol Microbiol 22: 367-378.

157. Garmendia J, Beuzon CR, Ruiz-Albert J, Holden DW (2003) The roles of SsrA-SsrB and OmpR-EnvZ in the regulation of genes encoding the *Salmonella typhimurium* SPI-2 type III secretion system. *Microbiology* 149: 2385-2396.
158. Sugiura M, Okamoto T, Takanami M (1970) RNA polymerase sigma-factor and the selection of initiation site. *Nature* 225: 598-600.
159. Popham DL, Szeto D, Keener J, Kustu S (1989) Function of a bacterial activator protein that binds to transcriptional enhancers. *Science* 243: 629-635.
160. Fang FC, Libby SJ, Buchmeier NA, Loewen PC, Switala J, et al. (1992) The alternative sigma factor katF (rpoS) regulates *Salmonella* virulence. *Proc Natl Acad Sci U S A* 89: 11978-11982.
161. Testerman TL, Vazquez-Torres A, Xu Y, Jones-Carson J, Libby SJ, et al. (2002) The alternative sigma factor sigmaE controls antioxidant defences required for *Salmonella* virulence and stationary-phase survival. *Mol Microbiol* 43: 771-782.
162. Conlin CA, Miller CG (2000) opdA, a *Salmonella enterica* serovar Typhimurium gene encoding a protease, is part of an operon regulated by heat shock. *J Bacteriol* 182: 518-521.
163. Baumberg S (1999) *Prokaryotic Gene Expression*. Oxford.
164. Dorman CJ (1994) *Genetics of Bacterial Virulence*: Blackwell Science. 390 p.
165. Bourret RB, Hess JF, Borkovich KA, Pakula AA, Simon MI (1989) Protein phosphorylation in chemotaxis and two-component regulatory systems of bacteria. *J Biol Chem* 264: 7085-7088.
166. Wadhams GH, Armitage JP (2004) Making sense of it all: bacterial chemotaxis. *Nat Rev Mol Cell Biol* 5: 1024-1037.
167. Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270: 467-470.
168. 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.
169. Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. *Genomics* 92: 255-264.
170. Sittka A, Lucchini S, Papenfort K, Sharma CM, Rolle K, et al. (2008) Deep sequencing analysis of small noncoding RNA and mRNA targets of the global post-transcriptional regulator, Hfq. *PLoS Genet* 4: e1000163.
171. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, et al. (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16: 944-945.
172. Abbott JC, Aanensen DM, Rutherford K, Butcher S, Spratt BG (2005) WebACT--an online companion for the Artemis Comparison Tool. *Bioinformatics* 21: 3665-3666.

173. Carver TJ, Rutherford KM, Berriman M, Rajandream MA, Barrell BG, et al. (2005) ACT: the Artemis Comparison Tool. *Bioinformatics* 21: 3422-3423.
174. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
175. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25: 3389-3402.
176. Sambrook J, Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning - a lab manual*: Cold Spring Harbour Laboratory Press, USA.
177. Levine MM, Galen J, Barry E, Noriega F, Chatfield S, et al. (1996) Attenuated *Salmonella* as live oral vaccines against typhoid fever and as live vectors. *J Biotechnol* 44: 193-196.
178. Hoiseth SK, Stocker BA (1985) Genes *aroA* and *serC* of *Salmonella typhimurium* constitute an operon. *J Bacteriol* 163: 355-361.
179. Herrero M, de Lorenzo V, Timmis KN (1990) Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *J Bacteriol* 172: 6557-6567.
180. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc Natl Acad Sci U S A* 97: 6640-6645.
181. McKelvie ND, Stratford R, Wu T, Bellaby T, Aldred E, et al. (2004) Expression of heterologous antigens in *Salmonella Typhimurium* vaccine vectors using the in vivo-inducible, SPI-2 promoter, *ssaG*. *Vaccine* 22: 3243-3255.
182. Doyle M, Fookes M, Ivens A, Mangan MW, Wain J, et al. (2007) An H-NS-like stealth protein aids horizontal DNA transmission in bacteria. *Science* 315: 251-252.
183. Hantke K (1981) Regulation of the ferric iron transport in *Escherichia coli* K12: Isolation of a constitutive mutant. *Mol Gen Genet* 182: 288-292.
184. Sinha S, Tompa M (2003) YMF: A program for discovery of novel transcription factor binding sites by statistical overrepresentation. *Nucleic Acids Res* 31: 3586-3588.
185. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18: 1851-1858.
186. Team RDC (2005) R: A language and environment for statistical computing.
187. Wettenhall JM, Smyth GK (2004) limmaGUI: a graphical user interface for linear modeling of microarray data. *Bioinformatics* 20: 3705-3706.
188. Nagalakshmi U, Wang Z, Waern K, Shou C, Raha D, et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320: 1344-1349.

189. Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, et al. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453: 1239-1243.
190. Graveley BR (2008) Molecular biology: power sequencing. *Nature* 453: 1197-1198.
191. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5: 621-628.
192. Sultan M, Schulz MH, Richard H, Magen A, Klingenhoff A, et al. (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. *Science* 321: 956-960.
193. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, et al. (2008) Highly integrated single-base resolution maps of the epigenome in *Arabidopsis*. *Cell* 133: 523-536.
194. Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, et al. (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5: 613-619.
195. Kuhn H, Frank-Kamenetskii MD (2005) Template-independent ligation of single-stranded DNA by T4 DNA ligase. *Febs J* 272: 5991-6000.
196. Carver T, Berriman M, Tivey A, Patel C, Bohme U, et al. (2008) Artemis and ACT: viewing, annotating and comparing sequences stored in a relational database. *Bioinformatics* 24: 2672-2676.
197. Pak J, Fire A (2007) Distinct populations of primary and secondary effectors during RNAi in *C. elegans*. *Science* 315: 241-244.
198. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, et al. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496-512.
199. Bentley DR, Deloukas P, Dunham A, French L, Gregory SG, et al. (2001) The physical maps for sequencing human chromosomes 1, 6, 9, 10, 13, 20 and X. *Nature* 409: 942-943.
200. Gallo S, Oberhuber M, Sigel RK, Krautler B (2008) The corrin moiety of coenzyme B12 is the determinant for switching the *btuB* riboswitch of *E. coli*. *Chembiochem* 9: 1408-1414.
201. Nahvi A, Barrick JE, Breaker RR (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* 32: 143-150.
202. McCarthy TJ, Plog MA, Floy SA, Jansen JA, Soukup JK, et al. (2005) Ligand requirements for *glmS* ribozyme self-cleavage. *Chem Biol* 12: 1221-1226.
203. Kutsukake K, Ohya Y, Iino T (1990) Transcriptional analysis of the flagellar regulon of *Salmonella typhimurium*. *J Bacteriol* 172: 741-747.
204. Cunningham L, Guest JR (1998) Transcription and transcript processing in the *sdhCDAB-sucABCD* operon of *Escherichia coli*. *Microbiology* 144 (Pt 8): 2113-2123.

205. Bennett GN, Schweingruber ME, Brown KD, Squires C, Yanofsky C (1976) Nucleotide sequence of region preceding trp mRNA initiation site and its role in promoter and operator function. *Proc Natl Acad Sci U S A* 73: 2351-2355.
206. Bajaj V, Lucas RL, Hwang C, Lee CA (1996) Co-ordinate regulation of *Salmonella typhimurium* invasion genes by environmental and regulatory factors is mediated by control of hilA expression. *Mol Microbiol* 22: 703-714.
207. Bäumlér AJ, Tsoilis RM, Heffron F (1997) Fimbrial adhesins of *Salmonella typhimurium* - role in bacterial interactions with epithelial cells. In: Paul PS, Francis DH, Benfield D, editors. *Mechanisms in the pathogenesis of enteric diseases*. New York: Plenum Press. pp. 149-158.
208. van der Velden AWM, Bäumlér AJ, Tsoilis RM, Heffron F (1998) Multiple fimbrial adhesins are required for full virulence of *Salmonella typhimurium* in mice. *Infection & Immunity* 66: 2803-2808.
209. Cooke FJ, Wain J, Fookes M, Ivens A, Thomson N, et al. (2007) Prophage sequences defining hot spots of genome variation in *Salmonella enterica* serovar Typhimurium can be used to discriminate between field isolates. *J Clin Microbiol* 45: 2590-2598.
210. Carlson JH, Hughes S, Hogan D, Cieplak G, Sturdevant DE, et al. (2004) Polymorphisms in the *Chlamydia trachomatis* cytotoxin locus associated with ocular and genital isolates. *Infect Immun* 72: 7063-7072.
211. Carlson JH, Porcella SF, McClarty G, Caldwell HD (2005) Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains. *Infect Immun* 73: 6407-6418.
212. Washietl S, Hofacker IL, Stadler PF (2005) Fast and reliable prediction of noncoding RNAs. *Proc Natl Acad Sci U S A* 102: 2454-2459.
213. Weinberg Z, Regulski EE, Hammond MC, Barrick JE, Yao Z, et al. (2008) The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *Rna* 14: 822-828.
214. Barrick JE, Breaker RR (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* 8: R239.
215. Serganov A, Polonskaia A, Phan AT, Breaker RR, Patel DJ (2006) Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* 441: 1167-1171.
216. Zhou D, Hardt WD, Galan JE (1999) *Salmonella typhimurium* encodes a putative iron transport system within the centisome 63 pathogenicity island. *Infect Immun* 67: 1974-1981.
217. Nixon BT, Ronson CW, Ausubel FM (1986) Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes ntrB and ntrC. *Proc Natl Acad Sci U S A* 83: 7850-7854.
218. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. *Stat Med* 9: 811-818.

219. Muller PK, Krohn K, Muhlradt PF (1989) Effects of pyocyanine, a phenazine dye from *Pseudomonas aeruginosa*, on oxidative burst and bacterial killing in human neutrophils. *Infect Immun* 57: 2591-2596.
220. Maier RJ, Olczak A, Maier S, Soni S, Gunn J (2004) Respiratory hydrogen use by *Salmonella enterica* serovar Typhimurium is essential for virulence. *Infect Immun* 72: 6294-6299.
221. Lambert JM, Bongers RS, de Vos WM, Kleerebezem M (2008) Functional analysis of four bile salt hydrolase and penicillin acylase family members in *Lactobacillus plantarum* WCFS1. *Appl Environ Microbiol* 74: 4719-4726.
222. Sudarsan N, Cohen-Chalamish S, Nakamura S, Emilsson GM, Breaker RR (2005) Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem Biol* 12: 1325-1335.
223. Young GP, Bayley N, Ward P, St John DJ, McDonald MI (1986) Antibiotic-associated colitis caused by *Clostridium difficile*: relapse and risk factors. *Med J Aust* 144: 303-306.
224. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. *Genome Res* 18: 1509-1517.
225. Pallen MJ, Wren BW (1997) The HtrA family of serine proteases. *Mol Microbiol* 26: 209-221.
226. Hoiseth SK, Stocker BAD (1981) Aromatic-dependent *Salmonella typhimurium* are non-virulent and effective as live oral vaccines. *Nature* 291: 238-239.
227. Song M, Kim HJ, Kim EY, Shin M, Lee HC, et al. (2004) ppGpp-dependent stationary phase induction of genes on salmonella pathogenicity island 1 (SPI1). *J Biol Chem*.
228. Kim CC, Falkow S (2004) Delineation of upstream signaling events in the salmonella pathogenicity island 2 transcriptional activation pathway. *J Bacteriol* 186: 4694-4704.
229. Carver T, Thomson N, Bleasby A, Berriman M, Parkhill J (2008) DNAPlotter: Circular and linear interactive genome visualisation. *Bioinformatics*.
230. Fey A, Eichler S, Flavier S, Christen R, Hofle MG, et al. (2004) Establishment of a real-time PCR-based approach for accurate quantification of bacterial RNA targets in water, using *Salmonella* as a model organism. *Appl Environ Microbiol* 70: 3618-3623.
231. Yang X, Yu L, Yu CA (1997) Resolution and reconstitution of succinate-ubiquinone reductase from *Escherichia coli*. *J Biol Chem* 272: 9683-9689.
232. Sinha S, Tompa M (2000) A statistical method for finding transcription factor binding sites. *Proc Int Conf Intell Syst Mol Biol* 8: 344-354.
233. Liljestrom P, Luokkamaki M, Palva ET (1987) Isolation and characterization of a substitution mutation in the *ompR* gene of *Salmonella typhimurium* LT2. *J Bacteriol* 169: 438-441.

234. Lee AK, Detweiler CS, Falkow S (2000) OmpR regulates the two-component system SsrA-ssrB in *Salmonella* pathogenicity island 2. *J Bacteriol* 182: 771-781.
235. Romling U, Bian Z, Hammar M, Sierralta WD, Normark S (1998) Curli fibers are highly conserved between *Salmonella typhimurium* and *Escherichia coli* with respect to operon structure and regulation. *J Bacteriol* 180: 722-731.
236. Mills SD, Ruschkowski SR, Stein MA, Finlay BB (1998) Trafficking of porin-deficient *Salmonella typhimurium* mutants inside HeLa cells: ompR and envZ mutants are defective for the formation of *Salmonella*-induced filaments. *Infect Immun* 66: 1806-1811.
237. Lucas RL, Lostroh CP, DiRusso CC, Spector MP, Wanner BL, et al. (2000) Multiple factors independently regulate hilA and invasion gene expression in *Salmonella enterica* serovar typhimurium. *J Bacteriol* 182: 1872-1882.
238. Lee YH, Kim BH, Kim JH, Yoon WS, Bang SH, et al. (2007) CadC has a global translational effect during acid adaptation in *Salmonella enterica* serovar Typhimurium. *J Bacteriol* 189: 2417-2425.
239. Rintoul MR, Cusa E, Baldoma L, Badia J, Reitzer L, et al. (2002) Regulation of the *Escherichia coli* allantoin regulon: coordinated function of the repressor AllR and the activator AllS. *J Mol Biol* 324: 599-610.
240. Krom BP, Warner JB, Konings WN, Lolkema JS (2003) Transporters involved in uptake of di- and tricarboxylates in *Bacillus subtilis*. *Antonie Van Leeuwenhoek* 84: 69-80.
241. Ibanez-Ruiz M, Robbe-Saule V, Hermant D, Labrude S, Norel F (2000) Identification of RpoS (sigma(S))-regulated genes in *Salmonella enterica* serovar typhimurium. *J Bacteriol* 182: 5749-5756.
242. Palacios S, Escalante-Semerena JC (2004) 2-Methylcitrate-dependent activation of the propionate catabolic operon (prpBCDE) of *Salmonella enterica* by the PrpR protein. *Microbiology* 150: 3877-3887.
243. Abo-Amer AE, Munn J, Jackson K, Aktas M, Golby P, et al. (2004) DNA interaction and phosphotransfer of the C<sub>4</sub>-dicarboxylate-responsive DcuS-DcuR two-component regulatory system from *Escherichia coli*. *J Bacteriol* 186: 1879-1889.
244. Mejean V, Iobbi-Nivol C, Lepelletier M, Giordano G, Chippaux M, et al. (1994) TMAO anaerobic respiration in *Escherichia coli*: involvement of the tor operon. *Mol Microbiol* 11: 1169-1179.
245. Siddiqui RA, Warnecke-Eberz U, Hengsberger A, Schneider B, Kostka S, et al. (1993) Structure and function of a periplasmic nitrate reductase in *Alcaligenes eutrophus* H16. *J Bacteriol* 175: 5867-5876.
246. Abouhamad WN, Manson MD (1994) The dipeptide permease of *Escherichia coli* closely resembles other bacterial transport systems and shows growth-phase-dependent expression. *Mol Microbiol* 14: 1077-1092.
247. Abouhamad WN, Manson M, Gibson MM, Higgins CF (1991) Peptide transport and chemotaxis in *Escherichia coli* and *Salmonella typhimurium*:

- characterization of the dipeptide permease (Dpp) and the dipeptide-binding protein. *Mol Microbiol* 5: 1035-1047.
248. Rowe JJ, Ubbink-Kok T, Molenaar D, Konings WN, Driessen AJ (1994) NarK is a nitrite-extrusion system involved in anaerobic nitrate respiration by *Escherichia coli*. *Mol Microbiol* 12: 579-586.
  249. Kustu SG, McFarland NC, Hui SP, Esmon B, Ames GF (1979) Nitrogen control of *Salmonella typhimurium*: co-regulation of synthesis of glutamine synthetase and amino acid transport systems. *J Bacteriol* 138: 218-234.
  250. Dela Vega AL, Delcour AH (1996) Polyamines decrease *Escherichia coli* outer membrane permeability. *J Bacteriol* 178: 3715-3721.
  251. delaVega AL, Delcour AH (1995) Cadaverine induces closing of *E. coli* porins. *Embo J* 14: 6058-6065.
  252. Soncini FC, Vescovi EG, Solomon F, Groisman EA (1996) Molecular basis of the magnesium deprivation response in *Salmonella typhimurium*: identification of PhoP-regulated genes. *J Bacteriol* 178: 5092-5099.
  253. Richardson DJ (2000) Bacterial respiration: a flexible process for a changing environment. *Microbiology* 146 (Pt 3): 551-571.
  254. Cheneby D, Hallet S, Mondon M, Martin-Laurent F, Germon JC, et al. (2003) Genetic characterization of the nitrate reducing community based on narG nucleotide sequence analysis. *Microb Ecol* 46: 113-121.
  255. Harborne NR, Griffiths L, Busby SJ, Cole JA (1992) Transcriptional control, translation and function of the products of the five open reading frames of the *Escherichia coli* nir operon. *Mol Microbiol* 6: 2805-2813.
  256. Brondijk TH, Nilavongse A, Filenko N, Richardson DJ, Cole JA (2004) NapGH components of the periplasmic nitrate reductase of *Escherichia coli* K-12: location, topology and physiological roles in quinol oxidation and redox balancing. *Biochem J* 379: 47-55.
  257. Chung T, Klumpp DJ, LaPorte DC (1988) Glyoxylate bypass operon of *Escherichia coli*: cloning and determination of the functional map. *J Bacteriol* 170: 386-392.
  258. Clarke TA, Mills PC, Poock SR, Butt JN, Cheesman MR, et al. (2008) *Escherichia coli* cytochrome c nitrite reductase NrfA. *Methods Enzymol* 437: 63-77.
  259. Campbell JW, Morgan-Kiss RM, Cronan JE, Jr. (2003) A new *Escherichia coli* metabolic competency: growth on fatty acids by a novel anaerobic beta-oxidation pathway. *Mol Microbiol* 47: 793-805.
  260. Higashitani A, Nishimura Y, Hara H, Aiba H, Mizuno T, et al. (1993) Osmoregulation of the fatty acid receptor gene fadL in *Escherichia coli*. *Mol Gen Genet* 240: 339-347.
  261. Horswill AR, Escalante-Semerena JC (1999) *Salmonella typhimurium* LT2 catabolizes propionate via the 2-methylcitric acid cycle. *J Bacteriol* 181: 5615-5623.

262. Nakamura H, Yamato I, Anraku Y, Lemieux L, Gennis RB (1990) Expression of *cyoA* and *cyoB* demonstrates that the CO-binding heme component of the *Escherichia coli* cytochrome *o* complex is in subunit I. *J Biol Chem* 265: 11193-11197.
263. Shioi J, Dang CV, Taylor BL (1987) Oxygen as attractant and repellent in bacterial chemotaxis. *J Bacteriol* 169: 3118-3123.
264. Aldridge P, Karlinsey J, Hughes KT (2003) The type III secretion chaperone FlgN regulates flagellar assembly via a negative feedback loop containing its chaperone substrates FlgK and FlgL. *Mol Microbiol* 49: 1333-1345.
265. Karlinsey JE, Lonner J, Brown KL, Hughes KT (2000) Translation/secretion coupling by type III secretion systems. *Cell* 102: 487-497.
266. Chaudhuri RR, Khan AM, Pallen MJ (2004) coliBASE: an online database for *Escherichia coli*, *Shigella* and *Salmonella* comparative genomics. *Nucleic Acids Res* 32 Database issue: D296-299.
267. McKinney JD, Honer zu Bentrup K, Munoz-Elias EJ, Miczak A, Chen B, et al. (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* 406: 735-738.
268. Wall DM, Duffy PS, Dupont C, Prescott JF, Meijer WG (2005) Isocitrate lyase activity is required for virulence of the intracellular pathogen *Rhodococcus equi*. *Infect Immun* 73: 6736-6741.
269. Finn RD, Tate J, Mistry J, Coghill PC, Sammut SJ, et al. (2008) The Pfam protein families database. *Nucleic Acids Res* 36: D281-288.
270. Meroueh SO, Bencze KZ, Hesek D, Lee M, Fisher JF, et al. (2006) Three-dimensional structure of the bacterial cell wall peptidoglycan. *Proc Natl Acad Sci U S A* 103: 4404-4409.
271. Crocker PR, Paulson JC, Varki A (2007) Siglecs and their roles in the immune system. *Nat Rev Immunol* 7: 255-266.
272. Burne RA, Parsons DT, Marquis RE (1989) Cloning and expression in *Escherichia coli* of the genes of the arginine deiminase system of *Streptococcus sanguis* NCTC 10904. *Infect Immun* 57: 3540-3548.
273. Degnan BA, Palmer JM, Robson T, Jones CE, Fischer M, et al. (1998) Inhibition of human peripheral blood mononuclear cell proliferation by *Streptococcus pyogenes* cell extract is associated with arginine deiminase activity. *Infect Immun* 66: 3050-3058.
274. Makhlin J, Kofman T, Borovok I, Kohler C, Engelmann S, et al. (2007) *Staphylococcus aureus* ArcR controls expression of the arginine deiminase operon. *J Bacteriol* 189: 5976-5986.
275. Ansong C, Yoon H, Norbeck AD, Gustin JK, McDermott JE, et al. (2008) Proteomics analysis of the causative agent of typhoid fever. *J Proteome Res* 7: 546-557.
276. Sun-Waterhouse D, Melton LD, O'Connor CJ, Kilmartin PA, Smith BG (2008) Effect of apple cell walls and their extracts on the activity of dietary antioxidants. *J Agric Food Chem* 56: 289-295.

277. Keppler OT, Horstkorte R, Pawlita M, Schmidt C, Reutter W (2001) Biochemical engineering of the N-acyl side chain of sialic acid: biological implications. *Glycobiology* 11: 11R-18R.
278. Meng F, Ma G, Kong J (1998) [Part of biological activity of peptidoglycan from *Streptococcus lactis* SB900]. *Wei Sheng Wu Xue Bao* 38: 376-380.
279. Gruening P, Fulde M, Valentin-Weigand P, Goethe R (2006) Structure, regulation, and putative function of the arginine deiminase system of *Streptococcus suis*. *J Bacteriol* 188: 361-369.
280. Rivas E, Klein RJ, Jones TA, Eddy SR (2001) Computational identification of noncoding RNAs in *E. coli* by comparative genomics. *Curr Biol* 11: 1369-1373.

## 9 Appendices

### 9.1 Maqdepth2pileup.pl

=head1 SYNOPSIS

Takes a maq pileup file as input, get depth information out of it.

The output can be loaded as user plot in Artemis, output contains read depths for different strands.

=head1 AUTHORS

Miao He <mh10(at)sanger.ac.uk>

=head1 COPYRIGHT

=head1 BUGS

if you witness any bug please contact the authors

=head1 LICENSE

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.

=cut

use strict;

sub show\_help {

    die <<EOF;

    \$0: take maq pileup file as input, get depth information out of it.

    The output can be loaded as user plot in artemis.

Output contains depths for different strands.

Usage: \$0 maq\_pileup\_file > output

EOF

}

my \$infile = shift;

my \$prefix = shift;

unless (\$infile) {

    &show\_help;

    exit;

}

warn "\n\nreading input file...\n\n";

open(IN, \$infile) or die;

while (<IN>) {

    chomp;

    my @fields = split /\t/, \$\_;

    my \$pos = \$fields[1];

    my \$depth = \$fields[3];

    my \$read\_bases = substr(\$fields[4],1);

    my @read\_bases\_array = split //, \$read\_bases;

my \$depth\_fwd = "";

my \$depth\_rev = "";

foreach my \$character (@read\_bases\_array) {

    if (\$character =~ /[ACGT\,]/) {

        \$depth\_fwd++;

```

    }

    elsif ($character =~ /[acgt\.\.]/) {

        $depth_rev++;

    }

}

if (!$depth_rev) {

    $depth_rev = 0;

}

if (!$depth_fwd) {

    $depth_fwd = 0;

}

print "$depth_fwd $depth_rev\n";

}

close IN;

exit;

```

## 9.2 extractLines.pl

```
#!/usr/local/bin/perl
```

```
=head1 NAME
```

```
extractLines.pl
```

```
=head1 SYNOPSIS
```

Read in a file and output only the lines in a specified region.

```
=head1 AUTHORS
```

Kathryn Holt <kh2(at)sanger.ac.uk>

```
=head1 COPYRIGHT
```

```
=head1 BUGS
```

if you witness any bug please contact the authors

=head1 LICENSE

This program is free software; you can redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2 of the License, or (at your option) any later version.

This program is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details. You should have received a copy of the GNU General Public License along with this program; if not, write to the Free Software Foundation, Inc., 59 Temple Place - Suite 330, Boston, MA 02111-1307, USA.

=cut

use Getopt::Long;

sub show\_help {

    print STDERR <<EOF;

\$0: Read in a file and output only the lines in a specified region.

Usage: \$0 -i input\_file -start linenum -stop linenum > output\n

EOF

}

    &GetOptions( "i=s"=>\\$inputfile,

                "start=i"=>\\$start,

                "stop=i"=>\\$stop,

                "h"=>\\$help,

                "help"=>\\$help);

if( \$help ) {

    &show\_help();

    exit(1);

}

```

$line=0;

open (IN,$inputfile) or die ("Couldn't find input file $inputfile\n");

while ($x=<IN>) {

    $line++;

    if ($line >= $start && $line <=$stop) {

        print $x;

    }

}

close IN;

END:

```

## 9.3 tram.pl

```

#!/usr/local/bin/perl -w

use strict;

use warnings;

#use lib '/nfs/team81/tk2/snp_analysis/latest/SNP_analysis/';

#use SNP_analysis;

use lib '/nfs/team81/sa4/3D7/transcriptome/scripts';

use TrAM;

#####

#This script will find various statistical measures for a given region on a chromosome

#This is a transcriptome analysis module

#USAGE:    tram.pl          <gff_file>    <coverage_file>    <uniqueness_file>
           <uniqueness_cutoff>

#  <gff_file> : Gff annotation file for a single chromosome

#  <coverage_file> : coverage file for a single chromosome

#  <uniqueness_file>: a single column file generated by a k-mer analysis

```

```

# <uniqueness_cutoff>: integer value that determines which bases to consider for
further analysis\n";

#

#

#

#####

use POSIX qw(ceil floor);

if (@ARGV != 6)

{

    print "USAGE:  tram.pl    <gff_file>  <coverage_file>  <uniqueness_file>
<uniqueness_cutoff> <coverage> <genome si
ze>\n";

    exit;

}

TrAM::ComputeGmeansGff($ARGV[0],$ARGV[1],          $ARGV[2],$ARGV[3],
$ARGV[4], $ARGV[5]);

```

## 9.4 LIMMA Analysis for Illumina data

```

library(limma)

library(vsn)

readExprs <- function(fileNames) {

    rd <- function (fn) {

        read.table(fn, sep = "\t", header = TRUE, row.names = 1,

            colClasses = c("character", "numeric", "numeric",

                "character", "numeric", "numeric", "numeric",

                "numeric", "numeric"))
    }
}

```

```

}

rc <- function(x) {
  x$DepthAMean
}

tables <- lapply(fileNames, rd)
expr <- cbind(sapply(tables, rc))
rownames(expr) <- rownames(tables[[1]])

expr
}

plotStdDevMedian <- function(matrix, targets, sample) {
  cols <- targets$sample == sample
  subset <- cbind(matrix[TRUE, cols])
  sd <- apply(subset, 1, sd)
  median <- apply(subset, 1, median)

  plot(sd, median,
       main = paste("Std. dev vs. median for replicates of", sample))
}

plotRawDepth <- function(matrix, targets) {
  ## par(mfrow = c(nrow(targets), nrow(targets)))

  par(ask = TRUE)
  sample <- targets$sample
  lane <- targets$lane

  for (i in 1:nrow(targets)) {
    for (j in 1:nrow(targets)) {
      plot(matrix[TRUE, i], matrix[TRUE, j],
           main = paste(sample[i], "vs", sample[j]),

```

```

        xlab = lane[i],
        ylab = lane[j])
    }
}
}

targets <- read.table("s_enterica_typhi_ty2_ompr_targets.txt", sep = "\t",
                      header = TRUE)

fileNames <- as.character(targets$fileName)

expr <- readExprs(fileNames)

colnames(expr) <- paste(targets$lane, targets$sample, sep = ".")

png(filename = "mean_depth_hist.png", width = 800, height = 600)

par(mfrow = c(4, 3))

for (colname in colnames(expr)) {
  hist(expr[TRUE, colname],
        freq = FALSE,
        main = colname,
        xlab = "Mean depth / bp",
        ylab = "Frequency")
}

dev.off()

texpr <- justvsn(expr)

png(filename = "mean_depth_t_hist.png", width = 800, height = 600)

par(mfrow = c(4, 3))

for (colname in colnames(texpr)) {
  x <- texpr[ TRUE, colname]

  hist(x,
        main = colname,

```

```

      xlab = "t Mean depth / bp",
      ylab = "Frequency")
}
dev.off()

png(filename = "mean_depth_t_box.png", width = 1000, height = 600)
par(cex = 0.7)
boxplot(texpr ~ col(texpr),
        main = "Boxplots of raw data",
        names = colnames(texpr))
dev.off()

texpr <- texpr[TRUE, c(2, 4, 5, 6, 8, 9, 10)]
colnames(texpr) <- (colnames(expr))[c(2, 4, 5, 6, 8, 9, 10)]
png(filename = "mean_depth_t_subset_box.png", width = 1000, height = 600)
par(cex = 0.7)
boxplot(texpr ~ col(texpr),
        main = "Boxplots of raw data",
        names = colnames(texpr))
dev.off()

png(filename = "mean_depth_t_subset_ma.png", width = 800, height = 800)
mva.pairs(texpr,
          log.it = FALSE,
          main = "M (t value ratio) vs A (average log2 values)",
          labels = colnames(texpr),
          cex = 0.7)
dev.off()

norm <- normalizeQuantiles(texpr)
png(filename = "mean_depth_t_norm_box.png", width = 800, height = 600)

```

```

par(cex = 0.7)

boxplot(norm ~ col(norm),

        main = "Boxplots of quantile normalized data",

        names = colnames(norm))

dev.off()

png(filename = "mean_depth_t_norm_ma.png", width = 800, height = 800)

mva.pairs(norm,

          log.it = FALSE,

          main = "M (log2 value ratio) vs A (average log2 values)",

          labels = colnames(norm),

          cex = 0.7)

dev.off()

sampleGroups <- targets$sample[c(2, 4, 5, 6, 8, 9, 10)]

exptDesign <- model.matrix(~ 0 + sampleGroups)

colnames(exptDesign) <- levels(sampleGroups)

contrasts <- makeContrasts('mut-wt', levels = exptDesign)

fit <- lmFit(norm, exptDesign)

efit <- eBayes(contrasts.fit(fit, contrasts))

annot <- read.table("s_enterica_typhi_ty2.table.1.txt", sep = "\t",

                  header = TRUE,

                  colClasses = c("character", "character",

                                "character", "numeric", "numeric", "numeric",

                                "character", "character", "character"))

table <- topTable(efit, n = 1000)

table2 <- merge(table, annot, by.x = "ID", by.y = "ID.0", sort = FALSE)

write.table(table2, "s_enterica_typhi_ty2.diff.expr.txt",

```

```

sep = "\t", quote = FALSE, col.names = NA)

eucDist <- dist(t(norm))

eucClust <- hclust(eucDist)

png(filename = "cluster_norm_euclid_dist.png", width = 800, height = 800)

plot(eucClust, hang = -1)

dev.off()

pearsDist <- as.dist(1- cor(norm))

pearsClust <- hclust(pearsDist)

png(filename = "cluster_norm_pearson_dist.png", width = 800, height = 800)

plot(pearsClust, hang = -1)

dev.off()

batches <- targets$batch[c(2, 4, 5, 6, 8, 9, 10)]

batchNorm <- cbind(normalizeQuantiles(expr[TRUE, batches == 1]),
                    normalizeQuantiles(expr[TRUE, batches == 2]),
                    expr[TRUE, batches == 3],
                    expr[TRUE, batches == 4])

colnames(batchNorm) <- colnames(expr)

png(filename = "mean_depth_t_batch_norm_box.png", width = 800, height = 600)

par(cex = 0.7)

boxplot(batchNorm ~ col(batchNorm),
        main = "Boxplots of batch quantile normalized data",
        names = colnames(batchNorm))

dev.off()

exptDesign <- model.matrix(~ 0 + sampleGroups + batches)

colnames(exptDesign) <- c("mut", "wt", "batch")

contrasts <- makeContrasts('mut-wt', levels = exptDesign)

fit <- lmFit(batchNorm, exptDesign)

```

```
efit <- eBayes(contrasts.fit(fit, contrasts))

table <- topTable(efit, n = 1000)

table2 <- merge(table, annot, by.x = "ID", by.y = "ID.0", sort = FALSE)

write.table(table2, "s_enterica_typhi_ty2.batch.diff.expr.txt",
            sep = "\t", quote = FALSE, col.names = NA)
```

## 9.5 Hypothetical gene sequence coverage

<b>CDS</b>	<b>WT1</b>	<b>WT2</b>	<b>WT3</b>	<b>Average</b>
t0005	0.17	0.32	0.32	0.27
t0006	0.21	0.2	0.11	0.17
t0009	0.45	0.38	0.27	0.37
t0010	0.1	0.05	0.08	0.08
t0011	0	0	0	0
t0014	0.05	0.12	0.06	0.08
t0015	0	0.06	0	0.02
t0016	0	0	0	0
t0017	0	0.05	0.22	0.09
t0018	0.05	0.03	0.1	0.06
t0019	0	0.04	0.11	0.05
t0020	0.01	0.07	0.19	0.09
t0021	0.03	0.02	0	0.02
t0029	0.1	0.13	0.25	0.16
t0030	0	0.17	0.29	0.15
t0031	0.42	1.16	2.83	1.47
t0032	0	0.05	0.24	0.1
t0033	0	0.25	0.02	0.09
t0034	0.12	0.21	0.21	0.18
t0035	0.1	0.19	0.25	0.18
t0036	0.05	0.14	0.13	0.11
t0037	0.1	0.06	0.25	0.14
t0038	0.21	0.08	0.35	0.21
t0039	0	0.04	0.14	0.06
t0040	0.43	0.21	0.56	0.4
t0043	0.06	0.09	0.15	0.1
t0044	0.09	0.16	0.12	0.12
t0046	0.3	1.58	0.16	0.68
t0047	2.31	4.03	3.41	3.25
t0050	0.05	0.16	0.1	0.1
t0052	0.12	0.53	0.39	0.35
t0053	0.36	0.32	0.38	0.35
t0072	0.09	0.21	0.16	0.15
t0074	0.02	0.06	0.06	0.05
t0075	0.06	0.08	0.27	0.14
t0080	0.18	0.11	0.66	0.32
t0081	0	0.1	0.49	0.2
t0082	0.27	0.68	0.38	0.44
t0083	0	0	0.44	0.15

t0084	0.12	0.29	0.26	0.22
t0085	0	0	0.27	0.09
t0087	0.08	0	0.1	0.06
t0088	0.12	0.14	0.24	0.17
t0099	0.12	0.16	0.24	0.17
t0101	0.77	0.83	0.17	0.59
t0101	0.77	0.83	0.17	0.59
t0102	0.54	0.33	0.62	0.5
t0103	4.62	4.51	8.82	5.98
t0103	4.62	4.51	8.82	5.98
t0104	0	0	0.07	0.02
t0110	0.09	0.05	0.45	0.2
t0111	0.15	0.14	0.24	0.18
t0113	0.2	0.1	0.26	0.19
t0119	0	0.05	0.26	0.1
t0122	4.19	2.36	4.36	3.64
t0123	0.3	0.65	0.74	0.56
t0124	0.16	0.6	0.97	0.58
t0139	0.09	0.32	0.43	0.28
t0142	0.13	0.14	0.16	0.14
t0143	0.27	0.28	0.39	0.31
t0144	0.33	0.06	0.72	0.37
t0145	0.3	0.12	0.14	0.19
t0146	0.09	0.24	0.24	0.19
t0155	0.09	0	0.04	0.04
t0157	1.32	1.01	0.62	0.98
t0158	5.78	16.49	14.93	12.4
t0159	3.16	2.36	3.51	3.01
t0160	56.12	22.29	149.83	76.08
t0161	0.23	0.25	0.16	0.21
t0162	0.03	0.05	0.41	0.16
t0163	0.15	0.14	0.25	0.18
t0165	0.4	0.31	0.51	0.41
t0167	0.01	0.04	0.07	0.04
t0169	0.22	0.09	0.12	0.14
t0172	0.05	0.12	0.07	0.08
t0173	0.13	0.15	0.44	0.24
t0178	0.14	0.13	0.1	0.12
t0179	1.02	0.69	0.42	0.71
t0180	1.39	1.14	1.71	1.41
t0184	0.7	0.47	0.93	0.7
t0185	0.12	0.36	0.5	0.33
t0186	0.43	1.06	1.54	1.01
t0187	0.32	0.08	0.23	0.21
t0190	0.08	0	0.4	0.16
t0204	0.09	0.05	0.31	0.15
t0205	0.59	0.56	1.29	0.81
t0206	2.86	1.41	2.68	2.32
t0211	2.33	1.13	3.69	2.38
t0212	1.62	2.43	1.55	1.87
t0213	5.67	1.82	9.52	5.67
t0224	0.12	0.26	0.25	0.21
t0236	3.16	2.54	5.68	3.79

t0237	0.06	0.12	0.02	0.07
t0240	0.21	0.31	0.65	0.39
t0241	0.2	0.29	0.12	0.2
t0242	0.06	0.21	0.15	0.14
t0246	0.19	0.22	0.34	0.25
t0248	0.61	1.97	1.8	1.46
t0249	0.17	0.31	0.32	0.27
t0250	0.17	0.16	0.32	0.22
t0251	0.42	0.48	2.06	0.99
t0255	0	0.2	0	0.07
t0257	0.51	0.5	0.17	0.39
t0259	0.24	0.15	0.27	0.22
t0260	1.45	0.73	0.53	0.9
t0262	0.39	0.54	0.49	0.47
t0264	0.06	0.28	0	0.11
t0265	0	0.04	0.19	0.08
t0266	0.06	0.21	0.26	0.18
t0268	0.21	0.18	0.17	0.19
t0279	0.1	0.16	0.24	0.17
t0281	1.81	2.09	4.57	2.82
t0281	10.01	2.25	1.11	4.46
t0282	0.09	0.36	0.75	0.4
t0283	0.02	0.1	0.03	0.05
t0284	0.15	0.12	0.38	0.22
t0285	0.33	0.19	0.35	0.29
t0286	0.21	0.1	0.14	0.15
t0287	0.07	0.06	0.16	0.1
t0288	0.08	0.12	0.47	0.22
t0289	0.12	0	0.11	0.08
t0290	0.11	0.1	0.06	0.09
t0292	0.25	0.08	0.21	0.18
t0293	0.05	0.25	0.21	0.17
t0294	0.09	0.48	0.41	0.33
t0296	0.77	0.39	1	0.72
t0298	0.46	0.26	0.53	0.42
t0302	0.42	0.13	0.68	0.41
t0304	0.04	0.04	0.25	0.11
t0305	0.1	0.09	0.37	0.19
t0306	0.23	0.39	0.7	0.44
t0307	0.17	0.35	0.52	0.35
t0308	0.32	0.28	0.75	0.45
t0309	0.6	0.36	1.21	0.72
t0311	1.19	0.47	1.38	1.01
t0312	0.48	0.25	2	0.91
t0313	0.46	1.49	2.32	1.42
t0315	0.42	0.66	2.12	1.07
t0318	24.21	20.01	77.3	40.51
t0319	0.99	2.99	2.04	2.01
t0322	0.72	0.25	0.61	0.53
t0323	0.2	0.28	0.47	0.32
t0324	0.14	0.11	0.19	0.15
t0326	0.03	0.06	0.16	0.08
t0327	0.28	0.11	0.2	0.2

t0328	0.03	0.03	0.07	0.04
t0329	0.14	0.03	0.34	0.17
t0331	0.22	0.54	0.22	0.33
t0332	0.25	0.44	0.44	0.38
t0335	0.47	0.72	0.84	0.68
t0336	0.02	0.2	0.24	0.15
t0337	0.07	0.2	0.08	0.12
t0338	0.11	0	1.04	0.38
t0339	0.02	0.1	0.14	0.09
t0340	0	0.18	0.16	0.11
t0341	0.07	0.11	0.17	0.12
t0342	0.05	0.06	0.15	0.09
t0343	0.11	0.11	0.11	0.11
t0344	0.06	0.06	0.18	0.1
t0348	13.47	7.64	48.29	23.13
t0349	0.67	4.14	4.05	2.95
t0350	0.22	0.2	0.43	0.28
t0351	0.13	0.09	0.31	0.18
t0351	0.13	0.09	0.31	0.18
t0352	5.93	4.53	1.72	4.06
t0353	0	0.31	0	0.1
t0354	0.14	0.29	0.49	0.31
t0361	0.68	1.29	1.1	1.02
t0363	0.44	0.72	0.56	0.57
t0364	0.11	0.11	0.16	0.13
t0365	0.07	0.16	0.22	0.15
t0366	0.15	0.11	0.17	0.14
t0367	0.93	0.28	1.27	0.83
t0371	0.35	0.65	0.71	0.57
t0373	0.24	0.2	0.26	0.23
t0374	0.41	0.18	0.35	0.31
t0375	0.87	0.24	0.53	0.55
t0377	0.24	0.58	0.86	0.56
t0378	0.25	0.12	0.66	0.34
t0380	0.02	0.06	0.1	0.06
t0381	0.36	0.38	0.58	0.44
t0382	0.52	0.25	0.21	0.33
t0383	0.09	0.13	0.18	0.13
t0384	1.81	0.38	1.8	1.33
t0389	0	0.07	0	0.02
t0390	0.1	0	0.06	0.05
t0391	0.03	0.05	0	0.03
t0392	0.06	0.05	0.16	0.09
t0393	0.05	0.04	0.15	0.08
t0394	0.03	0.17	0.13	0.11
t0395	0	0.03	0.09	0.04
t0396	0.1	0	0.07	0.06
t0397	0.02	0.04	0.11	0.06
t0398	0.09	0.06	0.28	0.14
t0399	0	0.03	0.21	0.08
t0405	0.5	0.29	0.56	0.45
t0406	0.07	0.13	0.05	0.08
t0408	0.19	0.42	0.3	0.3

t0409	0.3	0.46	0.86	0.54
t0410	0.23	0.11	0.07	0.14
t0411	0.06	0.11	0.05	0.07
t0412	0.18	0.59	0.76	0.51
t0413	0.9	1.46	1.94	1.43
t0419	0.22	0.13	0.21	0.19
t0420	0.23	0.16	0.15	0.18
t0421	0.06	0.14	0.05	0.08
t0422	0.1	0.18	0.4	0.23
t0425	0.47	1.4	0.81	0.89
t0425	0.47	1.4	0.81	0.89
t0428	0.13	0.27	0.13	0.18
t0430	0.35	0.37	0.29	0.34
t0432	0.84	0.14	0.16	0.38
t0433	0.08	0.08	0.39	0.18
t0434	0	0.03	0.25	0.09
t0443	0.37	0.27	0.36	0.33
t0444	0.32	0.27	0.4	0.33
t0447	0.08	0.12	0.3	0.17
t0450	0.23	0.27	0.09	0.2
t0451	0.18	0.1	0.06	0.11
t0452	0.16	0.04	0.17	0.12
t0453	0.23	0	0.29	0.17
t0455	0.13	0.17	0.36	0.22
t0457	0.08	0.09	0.13	0.1
t0458	0.15	0	0.34	0.16
t0464	0.2	0.18	0.12	0.17
t0465	0.05	0.26	0.27	0.19
t0466	0.01	0.57	0.23	0.27
t0469	0.11	0	0	0.04
t0471	0.03	0.07	0.23	0.11
t0474	0.05	0.32	0.37	0.25
t0475	0.15	0.1	0.77	0.34
t0476	0.22	0.15	0.5	0.29
t0478	0.53	0.1	0.5	0.38
t0479	0.14	0.67	0.63	0.48
t0482	0.3	0.18	0.23	0.24
t0483	0.04	0.06	0.56	0.22
t0484	0	0.07	0	0.02
t0485	0.16	0.13	0.17	0.15
t0487	0.15	0.13	0.15	0.14
t0488	0.05	0.34	0.06	0.15
t0489	0	0	0	0
t0490	0	0	0.12	0.04
t0490	0	0	0.12	0.04
t0491	0.06	0.06	0.17	0.1
t0492	0.17	0.1	0.08	0.12
t0495	0.3	0.28	0.35	0.31
t0503	0.12	0.49	0.39	0.33
t0504	0	0.08	0.1	0.06
t0505	0.51	0.28	0.36	0.38
t0506	0.13	0.12	0.22	0.16
t0507	0.07	0.07	0.21	0.12

t0508	0.17	0.08	0.32	0.19
t0514	0.31	0.16	0.47	0.31
t0515	0.39	0.56	0.32	0.42
t0516	0.17	0.25	0.23	0.22
t0517	0.24	0.69	0.44	0.46
t0518	0.48	0.41	0.34	0.41
t0519	0.75	0.32	0.2	0.42
t0520	0.38	0.44	0.76	0.53
t0521	0.23	0.09	0.73	0.35
t0522	0.11	0.14	0.55	0.27
t0523	0.12	0.19	1.56	0.62
t0524	0.12	0.19	0.73	0.35
t0525	0.95	0.49	1.33	0.92
t0528	0	0.19	0.21	0.13
t0529	0.23	1.1	0.63	0.65
t0530	0.09	0.02	0.07	0.06
t0531	0.13	0.09	0.14	0.12
t0532	0.37	0.45	0.05	0.29
t0533	0.26	0.56	0.97	0.6
t0534	0.82	0.74	0.74	0.77
t0535	0.16	0.42	0.2	0.26
t0536	15.06	24.82	14.86	18.25
t0537	32.29	32.91	23.43	29.54
t0539	60.06	63.58	80.48	68.04
t0540	17.33	13.99	25.98	19.1
t0541	16.28	8.22	11.54	12.01
t0542	6.89	6.77	6.64	6.77
t0543	15.61	14.27	17.92	15.93
t0544	10.49	5.73	7.99	8.07
t0545	10.28	10.08	7.2	9.19
t0546	9.7	6.49	8.91	8.37
t0547	9.26	11.52	9.25	10.01
t0548	4.8	1.86	2.39	3.02
t0549	1.92	6.43	5.99	4.78
t0550	0.16	0.19	0.09	0.15
t0551	0.15	0.29	0.39	0.28
t0552	0	0.59	0.42	0.34
t0555	0.1	0.19	0.17	0.15
t0560	12.94	16.73	36.81	22.16
t0561	1.77	0.67	0.76	1.07
t0563	0.22	0.23	0.29	0.25
t0564	0.35	0.54	0.46	0.45
t0565	0.61	0.91	0.62	0.71
t0566	0.56	1.21	0.73	0.83
t0568	0.38	0.36	0.59	0.44
t0569	0.6	0.5	0	0.37
t0570	0.11	0.15	0.35	0.2
t0571	0	0.11	0.08	0.06
t0572	0.17	0.13	0.94	0.41
t0573	0.12	0.07	0.33	0.17
t0574	0.06	0.24	0.28	0.19
t0575	0	0.44	0	0.15
t0576	0.06	0.22	0.16	0.15

t0577	0	0.42	0	0.14
t0578	18.19	9.49	31.6	19.76
t0579	16.44	11.51	52.71	26.89
t0580	16.01	18.3	42.97	25.76
t0583	0.5	0.71	0.96	0.72
t0584	0.08	0.11	0.21	0.13
t0585	0	0.21	0	0.07
t0585	0.88	3.48	1.7	2.02
t0589	0	0.05	0.1	0.05
t0589	0	0.05	0.1	0.05
t0590	0	0.07	0.05	0.04
t0591	0	0.06	0.21	0.09
t0596	0.18	0.34	0.25	0.26
t0601	0.1	0.19	0.32	0.2
t0603	0	0.04	0	0.01
t0604	0.54	0.86	0.94	0.78
t0605	0.51	0.91	1.39	0.94
t0606	0.44	0.16	0.8	0.47
t0607	0.2	0.11	0.21	0.17
t0608	1.03	0.64	3.73	1.8
t0609	2.18	2.6	5.68	3.49
t0614	0	0	0	0
t0615	0	0	0	0
t0617	0	0	0	0
t0618	5.68	9.14	7.93	7.58
t0619	0	0	0	0
t0621	0	0.06	0.26	0.11
t0626	0.13	0.22	0.13	0.16
t0627	0.02	1.37	0.45	0.61
t0629	0.38	0.54	0.57	0.5
t0631	0.08	0.12	0.14	0.11
t0634	5.94	12.8	41.1	19.95
t0635	0.19	0.06	0.63	0.29
t0636	0.46	0.19	0.35	0.33
t0637	0.06	0.2	0.27	0.18
t0638	0.17	0.27	0.44	0.29
t0639	0.03	0.18	0.12	0.11
t0641	1.5	3.77	0.66	1.98
t0642	0.41	1.06	1.76	1.08
t0643	0.21	0.14	0.24	0.2
t0645	0.36	1.74	0.58	0.89
t0646	0	0.2	0.24	0.15
t0647	0	0.07	0.25	0.11
t0649	1.07	0.45	2.21	1.24
t0653	0.96	0.42	1.55	0.98
t0654	0.23	0.17	0.04	0.15
t0657	0.04	0.09	0.33	0.15
t0657	0.04	0.09	0.33	0.15
t0658	0.06	0.16	0.03	0.08
t0659	0.17	0.08	0.16	0.14
t0660	2.71	1.17	1.28	1.72
t0661	1.93	2.22	2.63	2.26
t0663	0.1	0.08	0.08	0.09

t0668	0.28	0.23	0.37	0.29
t0668	0.28	0.23	0.37	0.29
t0669	0.15	0.13	0.47	0.25
t0670	0.55	1.05	0.58	0.73
t0674	0	0	0.08	0.03
t0675	0.15	0.17	0.03	0.12
t0676	0.02	0.05	0.04	0.04
t0677	0.05	0.11	0.21	0.12
t0680	0.08	0.04	0.12	0.08
t0681	0.05	0.05	0.22	0.11
t0682	0.46	0	0	0.15
t0683	0.1	0.03	0.05	0.06
t0684	0.03	0.06	0.22	0.1
t0685	0.15	0.04	0.05	0.08
t0686	0.19	0.12	0.23	0.18
t0687	0.25	0.05	0.17	0.16
t0688	9.41	9.65	25	14.69
t0690	0	0	0.14	0.05
t0692	0.12	0.06	0.31	0.16
t0693	1.26	0.38	1.01	0.88
t0694	0	0.4	0.28	0.23
t0695	0.22	0.03	0.24	0.16
t0696	0.35	0.2	0.15	0.23
t0697	0	0.13	0.22	0.12
t0698	0.21	0.22	0.41	0.28
t0699	0.44	0.5	0.79	0.58
t0700	0.05	0.53	1.12	0.57
t0702	0.21	1.15	0.39	0.58
t0703	0.01	0.76	0.48	0.42
t0704	0	0.17	0.06	0.08
t0705	0	0.04	0.04	0.03
t0707	0	0.13	0	0.04
t0708	0.88	0.9	0.51	0.76
t0711	0.22	0.06	0.12	0.13
t0712	0.14	0.1	0.38	0.21
t0713	0.07	0.1	0	0.06
t0714	0.09	0.09	0.12	0.1
t0715	0.42	0.31	0.42	0.38
t0716	0.03	0.05	0	0.03
t0716	0.03	0.05	0	0.03
t0717	0	0.36	0	0.12
t0718	0	0.2	0.59	0.26
t0719	0.16	0.44	0.14	0.25
t0720	0.02	0.16	0.2	0.13
t0721	0.11	0.08	0.16	0.12
t0722	0	0	0	0
t0723	0	0.03	0	0.01
t0724	0.06	0	0.3	0.12
t0725	0	0.15	0	0.05
t0727	0	0	0.48	0.16
t0728	0	0	0.13	0.04
t0729	0	0.06	0	0.02
t0730	0	0.09	0.11	0.07

t0731	0	0.18	0.1	0.09
t0732	0	0	0	0
t0733	0	0	0.07	0.02
t0734	0	0.05	0.11	0.05
t0735	0	0.09	0.13	0.07
t0736	0.44	0.16	0.3	0.3
t0737	0.17	0.07	0.13	0.12
t0738	0.13	0.09	0.31	0.18
t0739	0.05	0.05	0.55	0.22
t0740	0	0.1	0.03	0.04
t0741	0.03	0.09	0.04	0.05
t0742	0.35	0.13	0.28	0.25
t0743	0.08	0.1	0.04	0.07
t0744	0.07	0.07	0.26	0.13
t0745	0.11	0.07	0.15	0.11
t0746	0.02	0.18	0.12	0.11
t0747	0	0.11	0.16	0.09
t0749	0.1	0.09	0.19	0.13
t0752	0.13	0.17	0.26	0.19
t0753	0.42	0.23	0.2	0.28
t0754	0.04	0	0.08	0.04
t0755	0.05	0.05	0	0.03
t0756	0.02	0.03	0.1	0.05
t0757	0.02	0.14	0.05	0.07
t0758	0.05	0.05	0.1	0.07
t0759	0	0.09	0.1	0.06
t0760	0	0.02	0.07	0.03
t0761	0	0.03	0.04	0.02
t0762	0	0.04	0.16	0.07
t0764	0	0	0.06	0.02
t0766	0.02	0.04	0.02	0.03
t0769	0.02	0.06	0.14	0.07
t0770	0.15	0.24	0.37	0.25
t0771	0.01	0.04	0.16	0.07
t0772	0	0.06	0.05	0.04
t0773	0.02	0.06	0.09	0.06
t0779	0.29	1.54	0.62	0.82
t0782	0.51	1.67	1.42	1.2
t0785	0.06	1.1	0.26	0.47
t0786	0.17	1.89	0.22	0.76
t0787	0.15	0.88	0.21	0.41
t0788	0.21	1.57	0.06	0.61
t0792	14.84	50.84	14.39	26.69
t0804	0.09	0.15	0.86	0.37
t0805	0.1	0.1	0.23	0.14
t0806	0.04	0.31	0.16	0.17
t0809	13.57	9.02	53.08	25.22
t0810	7.82	3.64	25.53	12.33
t0811	3.75	2.76	7.92	4.81
t0813	0.77	0.75	0.74	0.75
t0814	0.17	0.33	0.11	0.2
t0815	0.44	1.83	0.9	1.06
t0816	0.6	0.29	1.02	0.64

t0817	0.02	0.04	0.14	0.07
t0818	0	0.05	0.13	0.06
t0819	0.07	0	0	0.02
t0820	0	0	0.12	0.04
t0821	0	0.04	0.09	0.04
t0821	0.05	0.02	0.05	0.04
t0822	0.1	0.04	0.17	0.1
t0823	0	0.09	0.18	0.09
t0824	0.02	0.07	0.03	0.04
t0825	0	0.09	0	0.03
t0825	0	0.09	0	0.03
t0826	0.05	0	0.13	0.06
t0827	0	0.08	0	0.03
t0828	0.17	0.1	0	0.09
t0829	0	0.09	0.11	0.07
t0835	0	0.03	0.09	0.04
t0836	0	0	0	0
t0849	0	0	0.04	0.01
t0852	0.3	0	0.11	0.14
t0853	0.11	0.1	0	0.07
t0854	0.03	0.09	0.12	0.08
t0855	0.07	0.05	0.16	0.09
t0859	0.44	0.15	0.29	0.29
t0860	2.31	0.78	0.01	1.03
t0862	0.07	0.06	0.12	0.08
t0864	0	1.08	0	0.36
t0865	0.49	0.41	0.16	0.35
t0866	0.28	0.8	0.59	0.56
t0867	1	1.81	1.33	1.38
t0874	0	0	0	0
t0875	0	0.24	0.33	0.19
t0877	1.05	0.7	0.87	0.87
t0879	0.27	0.3	0.13	0.23
t0884	0.29	0.08	0.1	0.16
t0885	0.14	0.1	0.28	0.17
t0888	0.1	0.1	0.2	0.13
t0889	0.1	0.05	0.1	0.08
t0890	0	0.1	0.12	0.07
t0891	0.05	0.03	0.16	0.08
t0892	0.17	0.09	0.04	0.1
t0893	0	0.26	0.17	0.14
t0910	5.91	6.25	21.53	11.23
t0911	0.92	1.26	1.55	1.24
t0912	0.11	0.14	0.19	0.15
t0913	0	0	0.08	0.03
t0923	0.08	0.17	0.29	0.18
t0924	0.17	0.23	0.44	0.28
t0925	0.19	0.22	0.31	0.24
t0927	1.79	1.62	1.57	1.66
t0928	3.21	4.43	1.99	3.21
t0935	1.05	0.86	2.11	1.34
t0936	0	0.44	0.56	0.33
t0937	0.07	0.25	0.32	0.21

t0938	0.17	0.13	0.16	0.15
t0939	0.55	0.21	0.26	0.34
t0940	0.04	0.12	0.5	0.22
t0942	46.24	2.12	4.23	17.53
t0944	0.72	0.95	1.08	0.92
t0946	0.04	0.1	0.16	0.1
t0950	0.12	0.12	0.9	0.38
t0951	0	0.81	1.7	0.84
t0966	0.02	0.1	0.16	0.09
t0969	0	0.13	0.05	0.06
t0970	0.19	0.17	0.12	0.16
t0971	0.18	0.26	0.36	0.27
t0972	0.16	0.35	0.34	0.28
t0973	0.2	0.17	1.05	0.47
t0974	0.09	0.11	0.12	0.11
t0975	0.04	0.17	0.28	0.16
t0978	0.49	0.58	0.72	0.6
t0980	0.14	0.47	0.48	0.36
t0981	0	0.07	0.09	0.05
t0987	0.51	0.53	0.21	0.42
t0989	7.93	3.38	17.1	9.47
t0990	0.56	0.3	0.42	0.43
t0991	13.64	14.2	25.09	17.64
t0993	0.21	0.31	0.3	0.27
t0995	0.3	0.38	0.31	0.33
t0996	0.07	0.14	0.11	0.11
t0997	0.43	0.23	0.45	0.37
t1000	0.04	0.28	0.09	0.14
t1002	0.47	0.35	1.15	0.66
t1003	0.21	0.43	0.52	0.39
t1004	0.39	0.38	0.23	0.33
t1005	24.54	9.49	317.28	117.1
t1007	0.18	0.07	0.17	0.14
t1007	0.18	0.07	0.17	0.14
t1008	0	0	0	0
t1009	0	0	0	0
t1010	0.19	0.31	0.02	0.17
t1011	0.01	0.08	0.3	0.13
t1012	0.04	0.11	0.33	0.16
t1013	0	0.06	0	0.02
t1014	0.08	0.22	0.25	0.18
t1015	0	0.3	0.94	0.41
t1016	0	0	0.1	0.03
t1018	0.14	0	0.74	0.29
t1019	0.18	0.09	0.25	0.17
t1020	0.98	4.66	1.19	2.28
t1021	1.06	0.53	1.52	1.04
t1022	0	0.29	0.29	0.19
t1023	0.15	0.76	0.22	0.38
t1024	0	0	0.33	0.11
t1026	0.34	0.47	0	0.27
t1027	0	0.18	0	0.06
t1028	0.02	0.03	0.06	0.04

t1029	0.06	0.08	0.25	0.13
t1030	0.19	0.06	0.26	0.17
t1031	0.33	0.21	0.35	0.3
t1035	0.18	0.08	0.18	0.15
t1035	0.18	0.08	0.18	0.15
t1037	0.8	0.1	0.2	0.37
t1038	1.26	0.91	1.73	1.3
t1039	3.1	2.43	3.52	3.02
t1040	5.86	23.91	2.65	10.81
t1044	0	0.16	0.1	0.09
t1045	0.1	0.29	0.07	0.15
t1049	0.2	0.24	0.14	0.19
t1050	0.06	0.06	0.17	0.1
t1052	0.16	0.14	0.16	0.15
t1054	0.85	0.71	1.02	0.86
t1055	1.68	1.11	1.68	1.49
t1056	0.06	0.12	0.12	0.1
t1057	0.2	0.14	0.11	0.15
t1058	0.04	0.04	0.22	0.1
t1064	0.61	0.95	0.63	0.73
t1065	0.11	0.21	0.19	0.17
t1066	0.26	0.63	1.26	0.72
t1067	0.61	0.96	1.67	1.08
t1068	2.9	0.56	1.99	1.82
t1069	1.52	0.35	2.42	1.43
t1073	0.2	0.07	0.24	0.17
t1076	0.31	0.15	0.41	0.29
t1077	0.08	0.05	0.27	0.13
t1079	0.21	0.31	0.34	0.29
t1080	0	0.09	0	0.03
t1083	1.25	0.36	0.72	0.78
t1083	0.14	0.1	0.11	0.12
t1085	0.03	0.03	0.04	0.03
t1086	0	0.06	0.07	0.04
t1087	0	0.04	0.05	0.03
t1088	0.24	0.18	0.32	0.25
t1088	0.03	0	0.08	0.04
t1091	0.11	0.29	0.07	0.16
t1092	0.24	0.82	0.32	0.46
t1094	0.21	0.77	1.2	0.73
t1095	0.56	0.22	0.83	0.54
t1102	0.06	0.73	0.25	0.35
t1103	0.03	0.85	0.41	0.43
t1106	0.04	0.72	0.15	0.3
t1107	0	0.22	0.35	0.19
t1108	0	0.05	0.18	0.08
t1109	0.09	0.18	0.11	0.13
t1110	0.03	0.15	0.02	0.07
t1111	0.11	0.03	0.16	0.1
t1112	0	0.12	0	0.04
t1113	0.65	0.91	1.49	1.02
t1114	0.09	0.5	0.06	0.22
t1115	0	0	0.31	0.1

t1116	0	0.11	0	0.04
t1117	0	0.23	0.18	0.14
t1118	0	0.08	0	0.03
t1120	0	0.14	0	0.05
t1122	0.52	0.28	2.26	1.02
t1123	0.11	0.7	0.86	0.56
t1124	4.66	13.6	11.03	9.76
t1125	0.71	1.43	3.03	1.72
t1126	0	0	0	0
t1127	0.21	0.39	0	0.2
t1128	0.15	0.25	1.15	0.52
t1129	0	0.12	0.64	0.25
t1130	0.12	0.09	0.39	0.2
t1131	3.01	2.43	4.82	3.42
t1132	0.19	6.78	0.2	2.39
t1133	0.08	0.14	0.08	0.1
t1134	0.11	0.11	0.04	0.09
t1135	0.17	0.01	0.06	0.08
t1136	0.43	0.09	0.4	0.31
t1137	0.09	0.15	0.29	0.18
t1138	0.61	0.46	0.66	0.58
t1139	0	0.15	0.09	0.08
t1141	0.04	0.34	0.24	0.21
t1143	0	0	0	0
t1144	0.06	0.18	0.59	0.28
t1145	0	0.17	0	0.06
t1146	0	0.15	0.19	0.11
t1147	0.04	0.12	0.33	0.16
t1148	1.15	0.95	1.24	1.11
t1149	0.07	0.27	0.03	0.12
t1150	0	0.13	0.1	0.08
t1151	0	0	0.09	0.03
t1152	0.35	0.35	0	0.23
t1153	0	0	0	0
t1155	0.65	1	1.07	0.91
t1156	0.26	0.39	0.84	0.5
t1157	4.2	0.45	1.72	2.12
t1157	4.2	0.45	1.72	2.12
t1158	0.06	0.08	0.36	0.17
t1159	0.41	0.25	0.3	0.32
t1160	2.3	1.14	1.88	1.77
t1161	0.69	0.43	0.88	0.67
t1162	0.16	0.25	0.1	0.17
t1162	0.16	0.25	0.1	0.17
t1163	0.04	0.24	0.61	0.3
t1164	0.02	0.08	0.13	0.08
t1165	1.2	2.9	4.95	3.02
t1167	0.19	0.23	0.2	0.21
t1168	0.55	0.07	0.23	0.28
t1168	0.55	0.07	0.23	0.28
t1169	90.95	164.05	211.64	155.55
t1170	0.76	4.34	3.28	2.79
t1171	1.05	2.43	1.8	1.76

t1175	0.13	0.34	0.47	0.31
t1179	2.64	3.27	10.13	5.35
t1180	6.82	2.47	9.14	6.14
t1187	0.82	2.02	1.13	1.32
t1188	0.08	0.23	0.25	0.19
t1194	0.08	0	0.53	0.2
t1196	0.09	0.37	0.41	0.29
t1200	0.12	0.15	0.09	0.12
t1201	0.65	0.33	0.32	0.43
t1202	0.16	0.27	0.33	0.25
t1203	6.47	6.25	26.81	13.18
t1204	0.56	0.74	1.81	1.04
t1205	0.08	0.12	0.21	0.14
t1206	9.83	2.42	7.57	6.61
t1207	0.54	0.33	0.59	0.49
t1208	0.19	1.49	0.23	0.64
t1209	3.34	3.44	1.07	2.62
t1210	0.03	0.21	0.11	0.12
t1211	0	0.15	0.19	0.11
t1222	0.13	0.11	0.12	0.12
t1224	0.19	0.29	0.19	0.22
t1225	0.03	0.07	0.12	0.07
t1226	0.13	0.11	0.29	0.18
t1227	3.86	1.15	4.86	3.29
t1229	1.03	0.79	1.09	0.97
t1232	0.71	0.61	1.57	0.96
t1233	0.12	0.22	0.24	0.19
t1234	0.94	0.48	2.01	1.14
t1235	0.3	0.5	0.58	0.46
t1236	0.02	0.27	0.31	0.2
t1237	0	0.07	0.31	0.13
t1238	0.02	0.17	0.12	0.1
t1239	0.09	0.13	0.15	0.12
t1240	0.12	0.15	0.26	0.18
t1241	0.02	0.02	0.1	0.05
t1242	0.06	0.22	0.62	0.3
t1243	0.14	0.02	0.32	0.16
t1246	6.32	7.06	7.27	6.88
t1247	1.21	0.79	1.64	1.21
t1248	0.19	0	0.17	0.12
t1248	0.19	0	0.17	0.12
t1254	0.28	0.12	0.98	0.46
t1255	0	0.28	0.31	0.2
t1256	0.05	0.32	0.47	0.28
t1257	1.2	0.47	1.05	0.91
t1258	0.02	0.07	0.13	0.07
t1259	0	0.61	0.05	0.22
t1260	0.02	0.32	0.1	0.15
t1261	0.06	0	0	0.02
t1262	0.04	0.02	0.1	0.05
t1263	0.04	0.05	0.16	0.08
t1264	0.1	0.28	0.05	0.14
t1265	0	0.29	0	0.1

t1266	0.13	0.19	0.17	0.16
t1267	0.07	0.05	0.28	0.13
t1268	0.05	0.18	0.41	0.21
t1269	0.15	0.45	0.85	0.48
t1270	0.2	1.8	0.28	0.76
t1271	0.31	0.94	0.21	0.49
t1272	0.11	0.08	0.04	0.08
t1273	0.07	0.03	0.21	0.1
t1274	0	0.11	0	0.04
t1275	0	0.11	0.09	0.07
t1276	0	0	0	0
t1277	0.12	0.13	0.13	0.13
t1278	0	0.08	0.12	0.07
t1279	0	0.29	0.36	0.22
t1280	0	0.13	0.44	0.19
t1281	0	0	0.08	0.03
t1282	0.01	0.04	0.21	0.09
t1283	0.09	0.2	0.31	0.2
t1284	0.13	0.13	0.17	0.14
t1285	0.23	0.13	0.26	0.21
t1286	0.22	0.06	0.31	0.2
t1287	0.04	0.21	0.09	0.11
t1288	0	0	0.11	0.04
t1289	0	0.08	0.08	0.05
t1290	0.02	0.11	0.14	0.09
t1293	0.46	0.17	0.46	0.36
t1296	0.08	0.04	0.19	0.1
t1297	0.03	0.15	0.16	0.11
t1300	1.94	6.7	0.88	3.17
t1301	1.36	2.75	0.63	1.58
t1305	0.11	0.35	0.34	0.27
t1306	0.1	0	0	0.03
t1307	0.08	0.4	0.21	0.23
t1309	0.22	0.15	0.25	0.21
t1310	0.22	0.53	0.6	0.45
t1311	1.46	0.93	2.09	1.49
t1314	0.02	0.08	0.31	0.14
t1315	0.17	0.63	0.24	0.35
t1320	2.9	2.15	4.01	3.02
t1322	0.06	0.18	0.19	0.14
t1323	0.04	0.35	0.03	0.14
t1324	0.19	0.17	0.19	0.18
t1325	0.09	0.26	0.13	0.16
t1326	0	0.01	0.05	0.02
t1327	0.04	0.11	0	0.05
t1328	0.22	0.4	0.1	0.24
t1329	0.25	6.4	1.04	2.56
t1330	1.04	1.36	4.21	2.2
t1332	0.4	0	0	0.13
t1332	0.4	0	0	0.13
t1334	0.37	0.95	0.43	0.58
t1336	86.08	37.73	127.57	83.79
t1337	22.95	13.68	58.9	31.84

t1340	0.01	0.11	0.06	0.06
t1342	0.23	0.53	0.25	0.34
t1343	0.29	0.41	0.17	0.29
t1344	0.14	0.69	0.48	0.44
t1345	0.21	0.03	0.32	0.19
t1349	0.58	3.74	0.88	1.73
t1352	0	0.16	0.03	0.06
t1352	0	0.16	0.03	0.06
t1353	0.02	0	0.14	0.05
t1353	0.02	0	0.14	0.05
t1354	0	0.07	0.09	0.05
t1354	0	0.07	0.09	0.05
t1355	0	0.05	0.06	0.04
t1355	0	0.05	0.06	0.04
t1356	0	0.04	0.08	0.04
t1356	0	0.04	0.08	0.04
t1357	0	0.11	0.14	0.08
t1357	0	0.11	0.14	0.08
t1358	0.11	0.07	0.04	0.07
t1358	0.11	0.07	0.04	0.07
t1359	0.02	0.04	0.04	0.03
t1359	0.02	0.04	0.04	0.03
t1360	0	0.07	0.29	0.12
t1360	0	0.07	0.29	0.12
t1361	0	0.09	0.22	0.1
t1361	0	0.09	0.22	0.1
t1364	0	0	0.26	0.09
t1364	0	0	0.26	0.09
t1365	0	0	0	0
t1365	0	0	0	0
t1366	0	0	0.07	0.02
t1366	0	0	0.07	0.02
t1367	0	0.07	0.09	0.05
t1367	0	0.07	0.09	0.05
t1368	0	0	0.03	0.01
t1368	0	0	0.03	0.01
t1369	0.02	0.02	0.08	0.04
t1369	0.02	0.02	0.08	0.04
t1370	0.05	0	0.21	0.09
t1370	0.05	0	0.21	0.09
t1371	0.06	0.22	0.04	0.11
t1371	0.06	0.22	0.04	0.11
t1372	0.02	0.03	0.06	0.04
t1372	0.02	0.03	0.06	0.04
t1373	0.07	0.08	0.1	0.08
t1373	0.07	0.08	0.1	0.08
t1374	0.04	0	0.26	0.1
t1374	0.04	0	0.26	0.1
t1375	0	0.08	0.06	0.05
t1375	0	0.08	0.06	0.05
t1376	0.26	0.03	0	0.1
t1376	0.26	0.03	0	0.1
t1377	0	0.16	0.09	0.08

t1377	0	0.16	0.09	0.08
t1378	0	0	0.22	0.07
t1378	0	0	0.22	0.07
t1379	0	0.07	0.12	0.06
t1379	0	0.07	0.12	0.06
t1380	0	0	0.09	0.03
t1380	0	0	0.09	0.03
t1381	0	0.05	0	0.02
t1381	0	0.05	0	0.02
t1382	0	0	0.14	0.05
t1383	0.15	0.15	0	0.1
t1384	0.04	0.04	0.05	0.04
t1385	0.01	0.05	0.25	0.1
t1386	0	0	0	0
t1388	0	0.12	0.15	0.09
t1389	0	0.04	0	0.01
t1390	0	0.03	0.1	0.04
t1391	0	0.06	0.16	0.07
t1392	0.11	0.22	0.29	0.21
t1393	0.04	0.16	0.19	0.13
t1394	0	0	0.7	0.23
t1395	0.39	0.11	0.71	0.4
t1396	0	0	0.07	0.02
t1397	0.45	0.16	0.2	0.27
t1398	0.2	0.94	0.69	0.61
t1401	0.21	0.21	0.4	0.27
t1402	0.07	0	0.19	0.09
t1403	0.52	0.26	1.84	0.87
t1404	0.47	0.12	0.81	0.47
t1405	0.32	0.4	0.77	0.5
t1407	0.13	0.14	0.31	0.19
t1408	0.03	0	0.1	0.04
t1409	0	0.25	0.24	0.16
t1410	0.12	0.09	0.48	0.23
t1411	0.41	0.25	0.12	0.26
t1412	0.08	0.16	0.37	0.2
t1413	0.24	0.14	0.46	0.28
t1414	0.1	0.05	0.25	0.13
t1415	0.07	0.02	0.08	0.06
t1416	0.71	0.5	0.1	0.44
t1417	0.03	0.05	0.43	0.17
t1418	0	0	0	0
t1419	0.1	0.06	0.13	0.1
t1420	0.18	0.22	0.31	0.24
t1421	0.57	0.41	1.93	0.97
t1421	0.57	0.41	1.93	0.97
t1422	6.86	1.56	1.13	3.18
t1424	0.61	1.56	1.59	1.25
t1425	0.07	0	0	0.02
t1428	0.08	0.09	0.47	0.21
t1430	0	0	0.15	0.05
t1431	0.05	0.39	0.42	0.29
t1432	0.51	0.47	0.19	0.39

t1434	0	0.12	0.62	0.25
t1435	0	0	0	0
t1436	0	0.2	0.08	0.09
t1437	0.53	0.29	0.42	0.41
t1438	0.13	0.11	0.21	0.15
t1442	0.2	0.29	0.09	0.19
t1443	0.47	0.34	0.58	0.46
t1444	0.04	0.08	0.31	0.14
t1445	0.06	0.1	0.06	0.07
t1447	0.3	0.12	0.26	0.23
t1448	0.34	0.21	0.93	0.49
t1449	0.04	0.49	0.21	0.25
t1450	0.79	1.07	0.28	0.71
t1451	4.27	3.19	0.55	2.67
t1452	2.78	5.01	2.24	3.34
t1453	1.48	5.75	2.02	3.08
t1454	2.44	2.6	1.64	2.23
t1455	1.21	1.14	0.83	1.06
t1456	0.65	0.81	0.44	0.63
t1458	13.29	14.65	0.73	9.56
t1459	0.64	1.86	0.51	1
t1460	0.24	0.43	0.6	0.42
t1461	0.75	0.61	1.36	0.91
t1463	0.07	0.29	0.15	0.17
t1464	0.3	0.2	0.29	0.26
t1465	0.12	0.19	0.15	0.15
t1465	0.12	0.19	0.15	0.15
t1466	0.33	0.39	0.27	0.33
t1467	0.04	0.13	0.12	0.1
t1468	0.03	0.14	0.39	0.19
t1470	0	0.04	0.37	0.14
t1471	0.13	0.15	0.23	0.17
t1472	0.11	0.1	0.13	0.11
t1473	0.08	0.06	0.11	0.08
t1474	0.03	0.06	0.19	0.09
t1475	0.27	1.04	1.17	0.83
t1476	0	0.19	0.12	0.1
t1479	0	0.11	0.53	0.21
t1482	4.26	6.93	1.19	4.13
t1484	1	0.17	0.86	0.68
t1486	0.2	0.14	0.2	0.18
t1488	0.59	0.62	0.47	0.56
t1489	1.09	0.58	0.24	0.64
t1491	0.03	0.21	0.26	0.17
t1493	1.14	2.34	1.41	1.63
t1495	0	0.37	0	0.12
t1496	0.11	0.24	0.3	0.22
t1497	0.02	0.07	0.32	0.14
t1498	0.07	0.17	0.23	0.16
t1499	0.16	0.41	0.5	0.36
t1500	0.23	0.3	0.39	0.31
t1501	0.8	0.28	3.68	1.59
t1502	0.14	0.59	0.47	0.4

t1503	0.28	0.47	0.81	0.52
t1504	0.21	0.41	0.47	0.36
t1505	0.26	0.28	0.32	0.29
t1506	0.15	0.07	0.44	0.22
t1507	0.15	0.1	0.36	0.2
t1509	0.01	0.1	0.13	0.08
t1510	0	0	0.16	0.05
t1511	0	0	0.09	0.03
t1512	0.1	0.29	0.05	0.15
t1513	0.01	0.04	0.09	0.05
t1514	0	0	0.34	0.11
t1515	0.43	0.08	0.44	0.32
t1516	1.34	1.42	1.33	1.36
t1519	0.19	0.39	0.63	0.4
t1521	0.23	0.25	0.22	0.23
t1522	0.4	0.32	0.84	0.52
t1523	0.02	0.09	0.23	0.11
t1524	0.17	0	0.19	0.12
t1525	0.13	0.07	0.1	0.1
t1526	0.04	0.08	0.05	0.06
t1527	0.11	0.12	0.12	0.12
t1529	0.09	0.14	0.24	0.16
t1530	0.11	0	0.17	0.09
t1531	0.17	0.35	0.19	0.24
t1532	0.18	0.04	0.21	0.14
t1533	0	0.04	0.06	0.03
t1533	0	0.04	0.06	0.03
t1535	0.07	0.13	0.49	0.23
t1536	0.03	0.26	0.08	0.12
t1537	0	0	0.05	0.02
t1539	0.07	0.11	0.31	0.16
t1540	0.09	0.23	0.15	0.16
t1541	0	0.25	0.17	0.14
t1542	0.28	0.42	0.28	0.33
t1542	1.33	2.67	5.14	3.05
t1543	0.21	0.76	0.81	0.59
t1546	0.11	0.26	0.42	0.26
t1547	0.12	0.15	0.54	0.27
t1548	0.13	0.58	0.77	0.49
t1550	6.06	4.67	6.51	5.75
t1552	0.52	0.3	0	0.27
t1553	0.06	0.13	0.14	0.11
t1555	2.29	5.34	3.05	3.56
t1556	5.01	4.3	18.77	9.36
t1557	0.06	0.11	0.12	0.1
t1558	0.14	0.42	0.08	0.21
t1560	0.53	0.24	0.87	0.55
t1561	0.29	0.33	0.51	0.38
t1562	0.16	0.05	0.16	0.12
t1565	5.26	3.6	3.01	3.96
t1566	3.35	8.04	13.78	8.39
t1567	0.12	0.33	1	0.48
t1568	0	0.15	0.2	0.12

t1569	0.09	0.2	0.09	0.13
t1570	0.19	0.34	0.37	0.3
t1571	0.22	0.29	0.27	0.26
t1572	0.03	0.11	0.04	0.06
t1574	0	0.05	0	0.02
t1575	0	0.06	0.11	0.06
t1576	0.85	1.91	1.35	1.37
t1577	0.91	0.62	3.07	1.53
t1578	0.12	0.16	0.4	0.23
t1579	0.26	0.54	0.45	0.42
t1580	0.85	0.6	1.92	1.12
t1581	0.13	0.11	0.22	0.15
t1582	0.9	0.38	1.03	0.77
t1584	3.34	1.51	2.1	2.32
t1585	0.24	0.35	0.43	0.34
t1589	0.09	0.09	0.25	0.14
t1590	0.03	0.1	0.08	0.07
t1599	0	0.41	0.78	0.4
t1601	0.14	0.18	0.64	0.32
t1601	0.14	0.18	0.64	0.32
t1602	0.06	0.1	0.23	0.13
t1604	0	0	0	0
t1604	0	0	0	0
t1605	0.05	0.05	0	0.03
t1606	0.47	0.24	0.37	0.36
t1611	0.52	0.53	0.59	0.55
t1612	0	0.13	0.07	0.07
t1614	0	0.03	0.1	0.04
t1616	0.07	0.01	0.19	0.09
t1617	0.32	0.71	0.57	0.53
t1619	0.15	0.47	0.2	0.27
t1621	0.18	0.11	0.22	0.17
t1622	0.08	0.49	0	0.19
t1625	10.47	5.59	22.56	12.87
t1628	0.89	0.24	1.12	0.75
t1629	0.28	0.68	0.65	0.54
t1630	0.67	0.89	0.24	0.6
t1632	0.13	0.2	0.11	0.15
t1633	0.1	0.49	0.43	0.34
t1634	0.14	0.17	0.22	0.18
t1640	0.13	0	0	0.04
t1641	0.14	0.19	0.06	0.13
t1642	0	0.62	0	0.21
t1643	0.26	0.23	0.45	0.31
t1644	0.56	0.65	2.88	1.36
t1645	0.06	0.22	0.83	0.37
t1646	0.08	0.2	0.34	0.21
t1647	0.33	0.17	0.52	0.34
t1648	0.33	0.49	0.34	0.39
t1650	0.72	1.35	3.83	1.97
t1652	0.97	0.08	0.49	0.51
t1653	1.29	0.76	1.24	1.1
t1659	0.33	0.1	0.57	0.33

t1664	0.2	0.36	0.23	0.26
t1665	0.52	0.32	0.48	0.44
t1666	3.64	2.27	3.27	3.06
t1668	10.92	6.4	3.27	6.86
t1669	0.02	0.16	0.39	0.19
t1670	0.24	0.13	0.4	0.26
t1671	0.2	0.53	0.44	0.39
t1672	0.4	0.4	0.28	0.36
t1673	0.31	0.51	0.58	0.47
t1675	5.97	5.21	8.45	6.54
t1676	2.92	2.4	1.84	2.39
t1677	0.98	0.43	1.34	0.92
t1678	0.25	0.39	0.23	0.29
t1680	0.11	0.04	0.23	0.13
t1682	105.76	141.03	616.59	287.79
t1683	0.05	0.04	0.46	0.18
t1684	0	0.21	0.09	0.1
t1685	0.49	0.28	0.11	0.29
t1687	0.11	0.25	0.09	0.15
t1691	0.14	0.12	0.31	0.19
t1692	0.34	0.3	0.62	0.42
t1696	0.02	0.07	0.15	0.08
t1699	0.44	0.23	0.31	0.33
t1700	0.04	0.09	0.17	0.1
t1705	0.11	0.18	0.11	0.13
t1706	0.09	0	0	0.03
t1707	0.35	0.54	0.93	0.61
t1708	0.76	0.95	1.13	0.95
t1709	0.88	0.61	1.25	0.91
t1710	0.18	0.69	0.38	0.42
t1711	0.16	0.16	0.55	0.29
t1712	0.2	0.2	0.15	0.18
t1713	0.53	0.41	0.3	0.41
t1714	0.07	0.29	0.07	0.14
t1715	0.2	0.51	0.78	0.5
t1718	0.39	0.08	0.23	0.23
t1721	0.17	0.19	0.42	0.26
t1730	1.76	4.31	0.46	2.18
t1731	0.31	0.34	0.48	0.38
t1733	2.61	4.03	0.97	2.54
t1742	1.36	2.54	4.06	2.65
t1745	1.37	2.53	2.03	1.98
t1746	1.61	3.38	2.9	2.63
t1751	0.03	0	0.24	0.09
t1752	0.16	0.14	1.22	0.51
t1754	0.15	0.14	0.38	0.22
t1756	0.14	0.22	0.14	0.17
t1759	0.16	0.15	4.21	1.51
t1760	0.2	0.13	0.19	0.17
t1761	0.21	1.14	0.75	0.7
t1763	0.08	0.18	0.41	0.22
t1764	0.02	0.22	0.09	0.11
t1766	0.39	0.22	0.45	0.35

t1768	0.05	0.19	0.34	0.19
t1772	0.21	0.19	0.34	0.25
t1773	0.76	0.63	2.1	1.16
t1774	0.41	0.65	3.7	1.59
t1775	0	0.3	0.46	0.25
t1778	0.04	0.07	0.02	0.04
t1782	0.04	0.12	0.06	0.07
t1783	0.16	0.2	0.41	0.26
t1784	0.25	0.21	0.41	0.29
t1785	0.29	0.58	0.32	0.4
t1787	3.14	0.53	60.99	21.55
t1788	0.7	0.17	41.41	14.09
t1789	1.24	0.3	11.41	4.32
t1790	3.34	0.54	43.7	15.86
t1791	0.33	0.18	7.73	2.75
t1792	0.74	0.21	1.9	0.95
t1793	0.33	0.25	0.68	0.42
t1794	0.54	0.45	1.04	0.68
t1798	0.84	0.76	0.17	0.59
t1799	3.88	0.86	1.35	2.03
t1800	0.53	0.64	0.07	0.41
t1802	0.4	0.1	0.42	0.31
t1809	0.03	0.17	1.33	0.51
t1810	0.06	0.03	0.28	0.12
t1812	0.19	0.08	0.15	0.14
t1822	0.43	0.43	1.7	0.85
t1825	0.43	0.63	1.6	0.89
t1826	0.21	7.88	1.42	3.17
t1827	5.01	1.1	8.38	4.83
t1830	0	0.32	0.18	0.17
t1831	0.11	0.36	0.3	0.26
t1834	1.28	4.49	1.54	2.44
t1835	0.26	0.29	0.09	0.21
t1836	0.11	0.05	0.83	0.33
t1839	0.04	0.09	0.13	0.09
t1839	0.04	0.09	0.13	0.09
t1840	0.31	0.39	0.28	0.33
t1841	0.08	0.15	0.39	0.21
t1842	1.36	1.07	3.5	1.98
t1843	0.08	0.07	0.14	0.1
t1846	0.2	0.69	0.35	0.41
t1847	0.25	0.22	0.82	0.43
t1848	0.18	0.15	0.2	0.18
t1851	0.76	0.62	0.94	0.77
t1852	0.15	0.41	0.13	0.23
t1855	0.09	0.27	0.14	0.17
t1856	0.06	0.25	0.15	0.15
t1857	0.1	0.23	0.35	0.23
t1859	0.18	0.53	0.32	0.34
t1860	0.82	0.28	3.24	1.45
t1861	0.19	0.15	0.81	0.38
t1864	0.49	0.45	0.45	0.46
t1865	0.55	7.03	0.89	2.82

t1867	0	0.02	0.12	0.05
t1867	0	0.02	0.12	0.05
t1868	0.12	0.06	0.14	0.11
t1868	0.12	0.06	0.14	0.11
t1869	0	0	0.12	0.04
t1869	0	0	0.12	0.04
t1870	0.04	0.12	0.33	0.16
t1870	0.04	0.12	0.33	0.16
t1871	0	0.07	0.48	0.18
t1871	0	0.07	0.48	0.18
t1872	0	0.16	0.59	0.25
t1872	0	0.16	0.59	0.25
t1873	0.06	0.09	0.04	0.06
t1873	0.06	0.09	0.04	0.06
t1874	0.06	1.14	0.21	0.47
t1874	0.06	1.14	0.21	0.47
t1875	0.12	1.95	0.53	0.87
t1875	0.12	1.95	0.53	0.87
t1876	0.36	1.09	0.43	0.63
t1876	0.36	1.09	0.43	0.63
t1877	0.31	0.17	0.12	0.2
t1877	0.31	0.17	0.12	0.2
t1878	0.22	0.24	0.21	0.22
t1878	0.22	0.24	0.21	0.22
t1879	0.04	0.09	0.09	0.07
t1879	0.04	0.09	0.09	0.07
t1880	0.05	0.11	0.16	0.11
t1880	0.05	0.11	0.16	0.11
t1881	0	0.05	0.13	0.06
t1881	0	0.05	0.13	0.06
t1882	0	0.05	0.22	0.09
t1882	0	0.05	0.22	0.09
t1883	0.02	0.04	0.03	0.03
t1883	0.02	0.04	0.03	0.03
t1884	0.21	0.14	0.12	0.16
t1884	0.21	0.14	0.12	0.16
t1885	0	0.05	0	0.02
t1885	0	0.05	0	0.02
t1886	0.04	0.04	0.11	0.06
t1886	0.04	0.04	0.11	0.06
t1887	0	0.16	0.3	0.15
t1887	0	0.16	0.3	0.15
t1888	0	0	0.08	0.03
t1888	0	0	0.08	0.03
t1889	0.07	0.14	0.28	0.16
t1889	0.07	0.14	0.28	0.16
t1890	0	0.05	0.15	0.07
t1890	0	0.05	0.15	0.07
t1891	0	0.08	0.17	0.08
t1891	0	0.08	0.17	0.08
t1892	0.05	0.12	0.32	0.16
t1892	0.05	0.12	0.32	0.16
t1893	0.14	0.04	0.1	0.09

t1893	0.14	0.04	0.1	0.09
t1894	0	0.08	0.03	0.04
t1894	0	0.08	0.03	0.04
t1895	0.06	0.13	0.12	0.1
t1895	0.06	0.13	0.12	0.1
t1896	1.37	0.58	0.11	0.69
t1896	1.37	0.58	0.11	0.69
t1897	0.06	0.25	0.16	0.16
t1897	0.06	0.25	0.16	0.16
t1898	0.09	0.13	0.12	0.11
t1898	0.09	0.13	0.12	0.11
t1899	0.04	0	0.05	0.03
t1899	0.04	0	0.05	0.03
t1900	0.21	0.09	0	0.1
t1900	0.21	0.09	0	0.1
t1901	0.02	0.34	0.15	0.17
t1901	0.02	0.34	0.15	0.17
t1902	0.05	0.4	1.11	0.52
t1903	1.82	1.42	0.75	1.33
t1903	1.82	1.42	0.75	1.33
t1904	0.03	0.02	0.13	0.06
t1904	0.03	0.02	0.13	0.06
t1905	0.04	0.04	0.11	0.06
t1905	0.04	0.04	0.11	0.06
t1906	0	0	0.05	0.02
t1906	0	0	0.05	0.02
t1907	0	0.08	0.47	0.18
t1907	0	0.08	0.47	0.18
t1908	0.06	0.1	0.26	0.14
t1908	0.06	0.1	0.26	0.14
t1909	0.15	0.73	0.17	0.35
t1909	0.15	0.73	0.17	0.35
t1910	0.53	0.76	0.25	0.51
t1910	0.53	0.76	0.25	0.51
t1911	0.39	0.2	0.65	0.41
t1911	0.39	0.2	0.65	0.41
t1912	0	0.09	0.24	0.11
t1912	0	0.09	0.24	0.11
t1914	0.05	0.14	0.21	0.13
t1914	0.05	0.14	0.21	0.13
t1915	0.05	0.1	0.21	0.12
t1915	0.05	0.1	0.21	0.12
t1916	0	0.19	0.31	0.17
t1916	0	0.19	0.31	0.17
t1917	0.03	0.12	0.16	0.1
t1917	0.03	0.12	0.16	0.1
t1918	0.22	0.23	0.72	0.39
t1918	0.22	0.23	0.72	0.39
t1919	0.07	0	0.08	0.05
t1919	0.07	0	0.08	0.05
t1920	0.21	0.51	0.95	0.56
t1925	0	0.1	0.28	0.13
t1926	0.05	0.23	0.16	0.15

t1927	0.1	0.17	0.25	0.17
t1931	0.03	0.11	0.08	0.07
t1932	0	0.01	0	0
t1932	0	0.01	0	0
t1933	0.29	0.05	0.63	0.32
t1937	0.35	0.49	0.68	0.51
t1938	1.16	2.6	0.83	1.53
t1939	0.16	0.06	0.15	0.12
t1944	0.32	0.16	0.15	0.21
t1945	0.1	0.15	0.03	0.09
t1947	1.56	0.92	2.5	1.66
t1948	0.06	0.02	0.22	0.1
t1948	0.06	0.02	0.22	0.1
t1950	0.08	0.29	0.16	0.18
t1951	0.09	0.07	0.21	0.12
t1955	0.08	0.18	0.16	0.14
t1958	0	0.03	0	0.01
t1959	0.05	0.04	0.08	0.06
t1960	0.25	0.93	0.83	0.67
t1961	20.15	17.62	37.8	25.19
t1962	0.03	0.19	0.17	0.13
t1962	0.03	0.19	0.17	0.13
t1963	0.7	0.18	0.15	0.34
t1964	5.89	6.85	7.72	6.82
t1965	0.09	0.17	0.2	0.15
t1966	0.13	0.28	0.68	0.36
t1967	0	0.22	0	0.07
t1968	0.03	0.12	0.04	0.06
t1969	0.13	0.03	0.96	0.37
t1970	0.15	0.27	0.46	0.29
t1972	0.14	0.2	0.16	0.17
t1981	0.36	0.25	0.43	0.35
t1982	0	0.02	0.04	0.02
t1983	0.04	0	0.15	0.06
t1985	0	0	1.4	0.47
t1986	0.93	1.17	7.04	3.05
t1987	0	0	0.91	0.3
t1989	0.42	1.16	2.22	1.27
t1991	0.36	0.24	0.52	0.37
t1992	0.37	0.47	0.26	0.37
t1993	13.47	13.03	14.91	13.8
t1994	1.28	1.6	1.45	1.44
t1995	0.11	0.12	0.33	0.19
t1997	0.06	0.07	0.1	0.08
t2000	0.11	0.05	0.09	0.08
t2001	1.24	0.55	0.35	0.71
t2002	0.18	0.24	0.19	0.2
t2003	0.17	0.42	0.54	0.38
t2004	1.39	1.48	0.8	1.22
t2010	0.1	0.07	0.14	0.1
t2011	0.21	0	0.21	0.14
t2012	0.04	0.01	0.09	0.05
t2013	0	0.11	0	0.04

t2014	0.3	0.2	0.18	0.23
t2015	0.11	0.19	0.37	0.22
t2020	0	0	0.13	0.04
t2022	1.12	1.76	1.5	1.46
t2023	0.25	0.39	1.16	0.6
t2025	4.81	1.89	3.82	3.51
t2026	0.22	0.18	0.19	0.2
t2027	0.01	0.13	0.11	0.08
t2028	0.03	0.14	0.11	0.09
t2029	0.36	0.1	0.16	0.21
t2031	0.22	0.3	0	0.17
t2035	0.31	0.14	0.38	0.28
t2036	0.22	0.12	0.08	0.14
t2037	0.15	0.34	0.19	0.23
t2038	0.07	0.13	0.32	0.17
t2039	0.05	0	0.2	0.08
t2040	0.17	0.16	0.99	0.44
t2041	0.08	0.1	0.26	0.15
t2042	0.03	0.13	0.16	0.11
t2045	0.03	0.05	0.05	0.04
t2046	0.21	0.07	0.39	0.22
t2047	0.24	0.37	0.27	0.29
t2048	0.09	0.09	0.18	0.12
t2050	0.11	0.65	0.16	0.31
t2051	0.15	0.32	0.54	0.34
t2053	0.18	0.18	0.21	0.19
t2056	0.11	0.05	0.35	0.17
t2061	0.27	0.27	0.39	0.31
t2062	0.15	0.19	0.16	0.17
t2062	0.15	0.19	0.16	0.17
t2063	0.08	0.34	1.28	0.57
t2064	0.05	0.02	0.76	0.28
t2065	0.09	0.09	0.03	0.07
t2066	0.07	0.02	0.16	0.08
t2066	0.07	0.02	0.16	0.08
t2067	0.06	0.03	0.14	0.08
t2067	0.06	0.03	0.14	0.08
t2068	0.02	0.08	0.04	0.05
t2069	0.21	0.18	0.34	0.24
t2071	0.06	0.06	0.17	0.1
t2072	0.12	0.18	0.29	0.2
t2073	0.61	0.23	0.9	0.58
t2074	0.43	0.34	0.56	0.44
t2075	0.03	0	0.04	0.02
t2076	0	0.04	0.23	0.09
t2077	0.05	0.09	0.16	0.1
t2078	0	0.1	0.9	0.33
t2078	0	0.1	0.9	0.33
t2079	0	0.06	0.2	0.09
t2080	0.27	0.38	0.44	0.36
t2080	0.27	0.38	0.44	0.36
t2086	0.14	0.07	0.16	0.12
t2088	0.11	1.02	0.03	0.39

t2095	0.29	0.25	0.44	0.33
t2101	0.29	0.89	0.52	0.57
t2102	0.34	0.45	0.55	0.45
t2103	1.33	0.4	3.05	1.59
t2107	0	0.37	0.2	0.19
t2108	0.14	0.17	0.9	0.4
t2110	0	0.13	0.17	0.1
t2115	50.3	22.28	54.21	42.26
t2119	0.53	0.22	0.08	0.28
t2120	0.46	0.39	1.07	0.64
t2126	0.88	1.69	0.84	1.14
t2132	0.06	0.39	0.05	0.17
t2133	0.09	0.33	0.24	0.22
t2134	0.21	1.11	0	0.44
t2135	10.16	8.9	11.84	10.3
t2136	6.01	14.6	10.69	10.43
t2137	141.41	93.37	166.12	133.63
t2138	110.86	125.18	309.79	181.94
t2140	77.8	111.72	265.46	151.66
t2141	136.54	211.48	480.84	276.29
t2142	71.64	73.42	248.52	131.19
t2143	46.46	455.43	847.97	449.95
t2144	140.48	372.25	804.1	438.94
t2145	187.85	296.55	159.03	214.48
t2146	172.83	228.4	1188.75	529.99
t2147	0.19	0.13	0.25	0.19
t2149	0.43	0	0.12	0.18
t2150	0.02	0.06	0.08	0.05
t2150	0.02	0.06	0.08	0.05
t2151	0	0.2	0.11	0.1
t2152	0.04	0.18	0.19	0.14
t2152	0.04	0.18	0.19	0.14
t2153	0	0.16	0.01	0.06
t2154	0	0.17	0.24	0.14
t2154	0	0.17	0.24	0.14
t2156	0	0.17	0.03	0.07
t2158	0.06	0.15	0.28	0.16
t2159	0	0.14	0.42	0.19
t2160	0.04	0.2	0.27	0.17
t2161	0.4	0.34	0.89	0.54
t2162	0.18	0.03	0.13	0.11
t2163	0.2	0.17	1	0.46
t2164	0.12	0.1	0.76	0.33
t2165	0.38	0.3	2.7	1.13
t2168	0	0.07	0	0.02
t2176	0	0.1	0	0.03
t2179	0.22	0.16	0.18	0.19
t2180	0.81	0.32	0.47	0.53
t2181	25.5	73.88	70.58	56.65
t2184	0.07	0.13	0.25	0.15
t2187	0.41	0.07	0.46	0.31
t2188	0.09	0.03	0.42	0.18
t2202	0.21	0.08	0.45	0.25

t2205	0.05	0.25	0.48	0.26
t2212	0.34	0.36	0.57	0.42
t2213	4.85	1.82	4.63	3.77
t2214	0.68	0.56	1.22	0.82
t2215	0.39	0.43	0.94	0.59
t2216	0.09	0.03	0.24	0.12
t2217	0	0.08	0.17	0.08
t2218	0.09	0	0	0.03
t2222	1.42	0.57	0.9	0.96
t2223	0.27	0.08	0.55	0.3
t2225	0.58	0.4	0.27	0.42
t2226	0	0.31	0.28	0.2
t2231	0.37	1.36	0.68	0.8
t2236	0.11	0.06	0.08	0.08
t2236	0.11	0.06	0.08	0.08
t2237	0.06	0.18	0.16	0.13
t2247	0.06	0.06	0.23	0.12
t2252	0.06	0.06	0.12	0.08
t2253	0.87	2.07	1.17	1.37
t2257	0.1	0.14	0.2	0.15
t2262	0	0.4	0.17	0.19
t2263	0.04	0.15	0.07	0.09
t2264	0.18	0.04	0.42	0.21
t2265	0.06	0.06	0.13	0.08
t2266	0.05	0.14	1.15	0.45
t2267	0	0.4	0.48	0.29
t2269	0	0.06	0	0.02
t2275	0.04	0.04	0.23	0.1
t2281	0	0	0	0
t2285	0.3	0.34	0.51	0.38
t2286	11.38	11.69	19.63	14.23
t2288	0.18	0.4	1.02	0.53
t2289	0.36	0.62	0.76	0.58
t2292	0.15	0.25	0.39	0.26
t2294	0.15	0.17	0.29	0.2
t2295	0.05	0.15	0.21	0.14
t2296	0.3	0.28	0.68	0.42
t2297	0	0.21	0.13	0.11
t2298	0.26	0.13	0.42	0.27
t2299	0.64	0.73	1	0.79
t2300	2.38	1.47	2.65	2.17
t2301	0.17	1.29	3.46	1.64
t2301	0.17	1.29	3.46	1.64
t2302	0	0.19	0.24	0.14
t2304	0.03	1.65	0.23	0.64
t2306	0.04	0.31	0.43	0.26
t2307	0.13	0.06	0	0.06
t2308	0.04	0	1.44	0.49
t2312	0	0.08	0.19	0.09
t2319	0.2	0.12	0.12	0.15
t2319	0.2	0.12	0.12	0.15
t2322	0	0.2	0.28	0.16
t2323	1.58	0.43	0.31	0.77

t2326	0.16	0.52	0.14	0.27
t2330	0.01	0.06	0.12	0.06
t2331	0.04	0.09	0.2	0.11
t2335	0.06	0.07	0.49	0.21
t2337	0.08	0.11	0.19	0.13
t2338	0.05	0.11	0.11	0.09
t2339	0	0.05	0.14	0.06
t2339	0	0.05	0.14	0.06
t2340	0.11	0.02	0.15	0.09
t2345	0.05	0.1	0.3	0.15
t2345	0.05	0.1	0.3	0.15
t2346	0.34	0.24	0.54	0.37
t2347	0.17	0.1	0.36	0.21
t2352	0.06	0.02	0.28	0.12
t2354	0.01	0.31	0.19	0.17
t2355	0.09	0.35	0.43	0.29
t2357	1.08	0.68	1.14	0.97
t2358	0.13	0.18	0.14	0.15
t2359	0.21	0.4	0.79	0.47
t2360	5.89	2.39	4.47	4.25
t2363	0.06	0.06	0.2	0.11
t2364	0.46	0.53	0.29	0.43
t2365	2.51	0.09	0.88	1.16
t2368	0.15	0.07	0.27	0.16
t2375	0.2	0.07	0.34	0.2
t2378	0.24	0.06	0.16	0.15
t2380	0	0.06	0.07	0.04
t2381	0.19	0.52	0.57	0.43
t2386	0.04	0.49	0.14	0.22
t2389	0.45	0.9	0.92	0.76
t2390	0.93	2.07	1.42	1.47
t2391	0.14	0.25	0.03	0.14
t2393	0	0.14	0.15	0.1
t2394	0.56	0.3	0.42	0.43
t2396	0.07	0.11	0.15	0.11
t2398	0.08	0.14	0.12	0.11
t2399	0.01	0.07	0.03	0.04
t2400	0.32	0.05	0.43	0.27
t2401	0.35	0.21	0.55	0.37
t2403	0.22	0.14	0.58	0.31
t2404	0.28	0.13	0.46	0.29
t2405	0.16	0.43	0.13	0.24
t2406	2.71	2.24	1.94	2.3
t2407	0	0.26	0.38	0.21
t2415	0.66	2.97	0.74	1.46
t2417	103.3	258.19	276.83	212.77
t2418	48.74	95.52	68.81	71.02
t2419	105.99	347.1	133.64	195.58
t2420	45.16	178.77	98.53	107.49
t2421	52.92	141.67	88.19	94.26
t2422	0.62	1.13	0.99	0.91
t2423	0	0.09	0	0.03
t2424	0.08	0.17	0.96	0.4

t2425	0.09	0.12	0.49	0.23
t2426	0	0.24	0.31	0.18
t2427	0.15	0.12	0.13	0.13
t2428	1.41	0.91	0.24	0.85
t2433	0.19	0.03	0.09	0.1
t2434	0.21	0.11	0.15	0.16
t2435	0.17	0	0.25	0.14
t2436	0.03	0.03	0.1	0.05
t2437	0	0.03	0.13	0.05
t2442	0.09	0.5	0.39	0.33
t2448	0.04	0.39	0.66	0.36
t2448	0.04	0.39	0.66	0.36
t2449	1.16	0.2	1.08	0.81
t2451	1.8	0.07	0.42	0.76
t2452	0.39	0.63	0.57	0.53
t2453	0	0.06	0.14	0.07
t2454	0.25	0	0.08	0.11
t2457	0.83	1.16	0.83	0.94
t2460	0.04	0.27	0.28	0.2
t2461	3.52	9.11	2.98	5.2
t2470	0.92	0.71	1.02	0.88
t2471	0.1	0.08	0.2	0.13
t2473	1.3	0.59	1.28	1.06
t2475	0	0.47	1.38	0.62
t2477	0	0.13	0.07	0.07
t2481	10.97	16.25	40.15	22.46
t2482	0.27	0.21	0.31	0.26
t2483	0.07	0.1	0.07	0.08
t2485	0.93	0.51	0.45	0.63
t2486	0.49	0.28	0.4	0.39
t2487	0.08	0.11	0.16	0.12
t2488	0.07	0.12	0.35	0.18
t2490	0.08	0.14	0.05	0.09
t2496	0.05	0.23	0.14	0.14
t2498	0.75	0.88	0	0.54
t2502	2.39	0.63	1.29	1.44
t2503	0.34	0.09	0.35	0.26
t2503	0.12	0.14	0.45	0.24
t2504	0.05	0.06	0.33	0.15
t2504	0.11	0.07	0.15	0.11
t2505	0	0.24	0.29	0.18
t2508	0.44	0.17	0.29	0.3
t2509	0.03	1.32	0.01	0.45
t2510	0	0.05	0.06	0.04
t2511	0	0.14	0.18	0.11
t2512	0	0.12	0	0.04
t2513	0	0.11	0	0.04
t2514	0	0.08	0.3	0.13
t2515	0	0.05	0.06	0.04
t2516	0.12	0.12	0.16	0.13
t2517	0	0.19	0.16	0.12
t2518	0.06	0.13	0.52	0.24
t2519	0	0	0	0

t2520	0.03	0.43	0.37	0.28
t2521	0.04	0.13	0.3	0.16
t2524	0.1	0.16	0.14	0.13
t2526	0	0.37	0	0.12
t2532	0.14	0.46	0.91	0.5
t2535	1.05	0.39	1.15	0.86
t2536	0.05	0.06	0.16	0.09
t2538	0.52	0.96	0.81	0.76
t2539	0.86	1.35	1.24	1.15
t2541	0.08	0.15	0.46	0.23
t2542	0.11	0.27	0.27	0.22
t2543	1.12	1.27	1.41	1.27
t2544	0.48	1.88	1.79	1.38
t2545	0.04	0.07	0.33	0.15
t2547	0.47	0.48	1.18	0.71
t2549	0.04	0.07	0.05	0.05
t2550	0.26	0.31	0.27	0.28
t2551	0	0.17	0.16	0.11
t2552	0.09	0	1.22	0.44
t2552	0.09	0	1.22	0.44
t2553	0	0.09	0	0.03
t2555	0	0.62	0.76	0.46
t2555	0	0.62	0.76	0.46
t2556	0.06	0.09	0.24	0.13
t2557	0.05	0.15	0.34	0.18
t2560	0.05	0.19	0.55	0.26
t2560	0.05	0.19	0.55	0.26
t2561	0.05	0.22	0.13	0.13
t2563	0	0.08	0.11	0.06
t2564	0.19	0.33	0.05	0.19
t2565	0	0.21	0.59	0.27
t2567	0	0.21	0.1	0.1
t2568	0.07	1.09	0.04	0.4
t2570	0.06	0.05	0	0.04
t2572	0.61	0	0.97	0.53
t2573	0	0.05	0.14	0.06
t2574	0.02	0.08	0.38	0.16
t2575	0	0	0.13	0.04
t2576	0	0.14	0.13	0.09
t2577	0.03	0.12	0.08	0.08
t2578	0.03	0.04	0.05	0.04
t2578	0.03	0.04	0.05	0.04
t2579	0	0.49	0.12	0.2
t2580	0.15	0.13	0.18	0.15
t2581	0.32	0.11	0.45	0.29
t2582	0.05	0.07	0.17	0.1
t2583	0	0	0.06	0.02
t2584	0.05	0.09	0	0.05
t2585	0	0.06	0	0.02
t2586	0.12	0.16	0.31	0.2
t2587	0.06	0	0.17	0.08
t2587	0.06	0	0.17	0.08
t2588	0.13	0.24	0.17	0.18

t2589	0	0	0.33	0.11
t2590	0	0.24	0.19	0.14
t2592	0	0	0.1	0.03
t2593	0.07	0.07	0.12	0.09
t2594	0.19	0.16	0.24	0.2
t2595	0.11	0	0.25	0.12
t2596	0	0.02	0.06	0.03
t2597	0.07	0.08	0.07	0.07
t2598	0	0.06	0.07	0.04
t2599	0.32	0.06	0.17	0.18
t2603	0.54	1.15	1.2	0.96
t2604	0.72	1.27	0.68	0.89
t2606	0.16	0.4	0.17	0.24
t2607	0.27	0.22	0.96	0.48
t2608	0.02	0.08	0.08	0.06
t2609	0.03	0.05	0.22	0.1
t2610	0.46	0.34	0.46	0.42
t2617	0.31	0.32	0.18	0.27
t2619	0.46	0.64	0.93	0.68
t2620	14.26	6.09	44.42	21.59
t2623	0.15	0.1	0.14	0.13
t2626	0.05	0.2	0.22	0.16
t2627	0.08	0.12	0.15	0.12
t2633	0.31	0.25	0.76	0.44
t2634	0.17	0.12	0.15	0.15
t2635	0.36	0.39	0.23	0.33
t2637	0.3	0.41	1.81	0.84
t2640	0	0.09	0.11	0.07
t2641	0.3	0.69	0.37	0.45
t2644	0.19	0.1	0.29	0.19
t2645	0.01	0.07	0.18	0.09
t2646	0.31	0.18	0.3	0.26
t2657	0.95	0.75	0.91	0.87
t2657	0.95	0.75	0.91	0.87
t2658	2.26	0.7	1.21	1.39
t2658	2.26	0.7	1.21	1.39
t2659	0	0.51	2.75	1.09
t2659	0	0.51	2.75	1.09
t2660	0	0	0	0
t2660	0	0	0	0
t2661	0	0	0	0
t2661	0	0	0	0
t2662	0	0	0	0
t2662	0	0	0	0
t2663	0	0	0	0
t2663	0	0	0	0
t2664	0	0	0	0
t2664	0	0	0	0
t2665	0	0	0	0
t2665	0	0	0	0
t2666	0	0	0	0
t2666	0	0	0	0
t2667	0	0	0	0

t2667	0	0	0	0
t2668	0.02	0.06	0.23	0.1
t2669	0.11	0.03	0.18	0.11
t2670	0.48	0.06	0.17	0.24
t2671	0.05	0.13	0.26	0.15
t2673	0	0.2	0	0.07
t2674	0	0.19	0.29	0.16
t2674	0	0.19	0.29	0.16
t2676	0.57	6.08	1.01	2.55
t2677	0.78	0.59	1.1	0.82
t2677	0.78	0.59	1.1	0.82
t2678	0.23	0.21	0.42	0.29
t2680	0	0	0.24	0.08
t2681	0.03	0.08	0.24	0.12
t2682	0.03	0.17	0.15	0.12
t2683	0.31	0.14	0.2	0.22
t2685	0.09	0.18	0.57	0.28
t2689	0.04	0.04	0.23	0.1
t2690	0	0.15	0.47	0.21
t2690	0	0.15	0.47	0.21
t2691	0.3	0.42	0.77	0.5
t2692	0.11	0.19	1.33	0.54
t2693	0.12	0.33	0.44	0.3
t2694	6.91	7.93	10.97	8.6
t2696	0.22	0.71	1.37	0.77
t2697	2.92	1.76	4.24	2.97
t2698	0.8	0.17	0.76	0.58
t2699	0.09	0.16	0.34	0.2
t2699	0.09	0.16	0.34	0.2
t2700	0.11	0.06	0.22	0.13
t2701	0	0	0.17	0.06
t2702	0	0	0.27	0.09
t2709	0.52	0.18	0.42	0.37
t2710	0.36	2.1	0.88	1.11
t2713	0.01	0.06	0.02	0.03
t2716	0.12	0.31	0.67	0.37
t2717	0.2	0.04	0.37	0.2
t2729	0.1	0.06	0	0.05
t2731	0.08	0.1	0.23	0.14
t2738	0	0.93	2.14	1.02
t2740	0.04	0.14	0.33	0.17
t2741	0.23	0.05	0.17	0.15
t2742	0.13	0.06	0.12	0.1
t2743	0.02	0.06	0.18	0.09
t2744	1.44	0.73	1.39	1.19
t2745	0	0.04	0.06	0.03
t2746	0.12	0.06	0.2	0.13
t2748	0.25	0.09	0.94	0.43
t2749	0.06	0.17	0.38	0.2
t2750	0.04	0.27	0.42	0.24
t2751	0.11	0.06	0.16	0.11
t2752	0.09	0.05	0.05	0.06
t2753	0.1	0.05	0.15	0.1

t2754	0.08	0.08	0.15	0.1
t2755	0	0.05	0.54	0.2
t2756	1.45	0.86	0.57	0.96
t2757	2.78	2.37	4.2	3.12
t2758	2.85	2.13	5.58	3.52
t2759	1.93	1.13	1.99	1.68
t2760	0.64	0.22	0.7	0.52
t2762	0	0	0.09	0.03
t2767	47.35	34.56	30.59	37.5
t2768	0.51	3.39	0.72	1.54
t2769	1.66	5.69	2.32	3.22
t2770	1.67	8.36	2.56	4.2
t2783	0.1	0.5	0.46	0.35
t2798	0.18	0.98	1.13	0.76
t2801	1.07	3.68	5.74	3.5
t2803	0.51	0	0.63	0.38
t2804	0.6	1.36	0.88	0.95
t2805	0.09	0.25	0.43	0.26
t2806	0.38	0.11	0.58	0.36
t2806	0.38	0.11	0.58	0.36
t2807	0	0.04	0.09	0.04
t2808	0.62	0.24	0.14	0.33
t2809	2.7	5.1	0.62	2.81
t2811	0	0.26	0.1	0.12
t2812	0.04	0.06	0.15	0.08
t2813	0.19	0.27	0.25	0.24
t2814	0.06	0.09	0.22	0.12
t2815	0.01	0	0.11	0.04
t2816	0	0.15	0.32	0.16
t2817	0.08	0.04	0.24	0.12
t2818	0.04	0.03	0.23	0.1
t2818	0.04	0.03	0.23	0.1
t2819	0	0.06	0.12	0.06
t2819	0	0.06	0.12	0.06
t2821	0.12	0	0.07	0.06
t2822	0	0	0.05	0.02
t2823	2.12	0.57	11.22	4.64
t2824	8.81	1.33	21.99	10.71
t2829	0	0.38	0.36	0.25
t2832	0.28	0.49	0.46	0.41
t2833	0.07	0.42	0.08	0.19
t2838	0.22	0.08	0	0.1
t2839	0.11	0.03	0.17	0.1
t2840	0.07	0.23	0.3	0.2
t2841	0.07	0.16	0.13	0.12
t2842	0.05	0.05	0.18	0.09
t2843	0	0.14	0.05	0.06
t2844	0.13	0.06	0.2	0.13
t2845	0.03	0.06	0.15	0.08
t2846	0.09	1.97	0.16	0.74
t2850	0.63	0.62	1.49	0.91
t2851	0.23	0.24	0.58	0.35
t2852	0.16	0.23	0.78	0.39

t2853	17.37	28.87	23.7	23.31
t2855	0.03	0.15	0.08	0.09
t2862	0.04	0.11	0.14	0.1
t2863	0	0.06	0.28	0.11
t2864	0	0.8	1.06	0.62
t2866	0.04	0.23	0.2	0.16
t2868	0.17	0.08	2.84	1.03
t2869	0.17	0.13	3.04	1.11
t2870	0.09	0.07	6.77	2.31
t2871	0.57	0.6	2.4	1.19
t2872	0.11	0.32	0.32	0.25
t2873	0.11	0.2	0.19	0.17
t2874	0.07	0.07	0	0.05
t2876	0.16	0.36	0.59	0.37
t2877	1.53	0.78	1.08	1.13
t2878	0.61	1.82	0.73	1.05
t2888	0.24	0.44	0.32	0.33
t2889	0.12	0.04	0.09	0.08
t2891	0.85	0.37	0.23	0.48
t2892	0.1	0.04	0.13	0.09
t2893	0.21	0.05	0	0.09
t2894	0.63	0.72	0.43	0.59
t2895	0.42	0.16	0.16	0.25
t2906	0	0.06	0	0.02
t2913	0.63	0.82	1.13	0.86
t2915	0.47	0.43	1.1	0.67
t2916	0.6	0.33	0.87	0.6
t2917	0.74	0.47	0.67	0.63
t2918	0.25	0.13	0.37	0.25
t2922	0.29	0.19	0.63	0.37
t2925	0.27	0.26	0.97	0.5
t2929	0.28	0.4	0.61	0.43
t2930	0.1	0	0.19	0.1
t2931	0.37	0.58	0.25	0.4
t2932	0.58	0.31	0.63	0.51
t2933	0	0.09	0.43	0.17
t2934	0.37	0.27	0.15	0.26
t2935	0.06	0	0.02	0.03
t2936	0.05	0.05	0.12	0.07
t2937	0.07	0.17	0.17	0.14
t2938	0.06	0.03	0.16	0.08
t2939	0.07	0.02	0.2	0.1
t2940	0.04	0.29	0.15	0.16
t2941	0.53	0.3	1.2	0.68
t2946	0.38	0.37	0.22	0.32
t2956	0.24	0.3	0.22	0.25
t2957	0.2	0.24	0.58	0.34
t2963	1.84	3.73	2.35	2.64
t2964	0.63	0.26	0.76	0.55
t2965	0.75	1.42	2.58	1.58
t2966	0.19	0.23	1.02	0.48
t2967	0.43	0.44	0.71	0.53
t2968	0	0.18	0.19	0.12

t2970	0.03	0.14	0.16	0.11
t2974	0.05	0.11	0.27	0.14
t2977	0.12	0.5	0.61	0.41
t2978	0.77	2.16	0.47	1.13
t2979	0.04	0.04	0.18	0.09
t2983	0.06	0.57	1.04	0.56
t2984	0.11	0.14	0.28	0.18
t2985	0.12	0.23	0.26	0.2
t2987	58.9	93.5	206.37	119.59
t2988	25.65	36.66	151.14	71.15
t2990	0	0.11	0.07	0.06
t2991	0.23	0.11	0.1	0.15
t2992	0	0.11	0.18	0.1
t2993	0	0.04	0.18	0.07
t2994	0.17	0	0.34	0.17
t2997	0.24	0.67	0.24	0.38
t3000	0.18	0	0.04	0.07
t3001	0	0.35	0.01	0.12
t3004	0.17	0.1	0	0.09
t3006	0.27	0.32	0.23	0.27
t3008	0.08	0.46	0.33	0.29
t3009	0.17	0.12	0.22	0.17
t3010	0.07	0.06	0.31	0.15
t3012	0.13	0.47	0.43	0.34
t3013	0.13	0.16	0.13	0.14
t3014	0.7	0.76	2.02	1.16
t3015	0.56	0.84	1.4	0.93
t3016	0.08	0.14	0.29	0.17
t3017	0.08	0.14	0.09	0.1
t3020	0.16	0.61	0.4	0.39
t3021	0.14	2.22	0.25	0.87
t3022	0.32	0.53	0.2	0.35
t3024	0.42	1.57	1.25	1.08
t3028	0.01	0.03	0.17	0.07
t3031	0	0	0.07	0.02
t3032	0.2	0.2	0	0.13
t3033	0	0	0	0
t3034	0	0	0.12	0.04
t3039	0	0.26	0	0.09
t3040	0	0.19	0.77	0.32
t3041	0	0	0.06	0.02
t3042	0	0.04	0	0.01
t3043	0.14	0	0	0.05
t3044	0.03	0.22	0.06	0.1
t3045	0	0	0.13	0.04
t3047	0.43	0.1	0.38	0.3
t3048	0.11	0.19	0.25	0.18
t3049	0.34	0.09	0	0.14
t3050	0.12	0.09	0.38	0.2
t3051	0.06	0.17	0.12	0.12
t3052	0.14	0	0.22	0.12
t3053	0.03	0.08	0.24	0.12
t3054	0	0.11	0.12	0.08

t3059	0.64	2.67	2.4	1.9
t3061	0.23	0.12	0.1	0.15
t3062	0.09	0.03	0.36	0.16
t3063	0	0.02	0.06	0.03
t3064	0.5	0.27	0.43	0.4
t3065	0.31	0.58	0.24	0.38
t3066	0.4	1.34	0.62	0.79
t3067	3.94	2.66	4.08	3.56
t3068	2.18	2.43	4.86	3.16
t3069	0.05	0.19	0.23	0.16
t3069	0.34	0.56	0.51	0.47
t3070	0.99	1.11	1.83	1.31
t3071	2.93	2.6	2.29	2.61
t3072	0.08	0.63	0.3	0.34
t3074	0.1	0.05	0.11	0.09
t3075	0.86	3.07	4.12	2.68
t3076	0.31	1.66	0.6	0.86
t3077	0.08	0.14	0.2	0.14
t3080	0.44	0.18	0	0.21
t3083	0.17	0.33	0.38	0.29
t3084	0.39	0.85	1.11	0.78
t3085	0.22	0.44	0.27	0.31
t3086	0.05	0.06	0.3	0.14
t3087	0.21	0.4	0.65	0.42
t3088	0.16	0.14	0.17	0.16
t3089	0.05	0.15	0.21	0.14
t3090	0.31	0.2	0.26	0.26
t3091	0.06	0.16	0.17	0.13
t3092	0.17	0.23	0.9	0.43
t3096	0	0.03	0.46	0.16
t3097	0.2	0.66	1.12	0.66
t3098	0.1	0.11	0.19	0.13
t3099	0.15	0.11	0.16	0.14
t3100	0.15	0.36	0.19	0.23
t3101	0.97	0.89	0.92	0.93
t3103	0	0.39	0.1	0.16
t3104	0.13	0.69	0.17	0.33
t3105	0.62	0.41	0.26	0.43
t3106	0.33	0.67	0.52	0.51
t3108	0.49	1.01	0.85	0.78
t3109	0.26	0.6	0.61	0.49
t3110	0.31	0.43	0.9	0.55
t3111	0.23	0.12	0.17	0.17
t3112	0.08	0.19	0.23	0.17
t3116	0.2	0.38	0.57	0.38
t3118	0.2	0.36	0.63	0.4
t3119	0.18	0.09	0.33	0.2
t3122	0.12	0.15	0.26	0.18
t3123	0.08	1.22	1.63	0.98
t3127	0.49	0.57	0.6	0.55
t3128	0.19	0.27	0.12	0.19
t3134	0.06	0.09	0.33	0.16
t3135	2.08	1.95	3.97	2.67

t3138	0.02	0.04	0.12	0.06
t3139	0.09	0.15	0.22	0.15
t3140	0.13	0.1	0.13	0.12
t3141	0.18	0.1	0.19	0.16
t3142	0.07	0.08	0.52	0.22
t3143	0.28	0.26	0.4	0.31
t3144	0.08	0.24	0.34	0.22
t3145	0.16	0.2	0.76	0.37
t3146	0.52	0.35	0.35	0.41
t3147	0.52	0.17	0	0.23
t3148	1.06	6	7.41	4.82
t3149	0.42	0.48	0.94	0.61
t3150	0.54	0.45	0.08	0.36
t3151	1.04	0.62	0.3	0.65
t3152	0	0.17	0.06	0.08
t3153	0.16	0.11	0.43	0.23
t3154	3.35	1.26	0.79	1.8
t3155	0.12	0.16	0.38	0.22
t3156	0.03	0.16	0.14	0.11
t3157	0.27	0	0.09	0.12
t3158	0.26	0.1	0.34	0.23
t3159	0.6	0.35	0.86	0.6
t3159	0.6	0.35	0.86	0.6
t3161	0.08	0.08	0.19	0.12
t3161	0.39	0.28	0.49	0.39
t3167	0.24	0.16	3.37	1.26
t3171	0.03	0.19	0.38	0.2
t3174	0	0.05	0.14	0.06
t3177	0.19	0.11	0.24	0.18
t3183	0.3	0.45	0.14	0.3
t3184	0.33	2.16	0.61	1.03
t3185	0.12	0.26	0.39	0.26
t3186	0.65	0.29	1.25	0.73
t3187	0.45	0.81	1.13	0.8
t3188	1.14	0.5	0.62	0.75
t3189	0.72	0.05	0.46	0.41
t3190	0.18	0.05	0.29	0.17
t3191	1.07	0.23	0.7	0.67
t3192	0.3	0.22	0.27	0.26
t3193	0.53	0.74	0.89	0.72
t3194	0.06	0.25	0.12	0.14
t3195	0.05	0	0.28	0.11
t3196	0.31	0.13	0.27	0.24
t3197	0.13	0.12	0.35	0.2
t3199	0.54	1.19	0.8	0.84
t3206	0.22	1.98	0.38	0.86
t3215	19.68	8.45	6.87	11.67
t3218	0.15	0.08	0.16	0.13
t3219	0.06	0.07	0.16	0.1
t3225	0.56	1.19	0.6	0.78
t3226	0.08	0.36	0	0.15
t3227	0.32	0.86	0.82	0.67
t3228	0.16	0.49	0.28	0.31

t3229	0	0.2	0.15	0.12
t3230	0.08	0.16	0.17	0.14
t3231	0.26	0.26	0.29	0.27
t3232	0.41	0.45	0.41	0.42
t3233	0.16	0.21	0.25	0.21
t3234	0.45	0.99	1.19	0.88
t3235	0.33	2.56	0.7	1.2
t3236	0.13	0.52	0.49	0.38
t3238	2.53	3.59	6.38	4.17
t3240	0.61	0.91	0.88	0.8
t3242	0.56	0.26	0.39	0.4
t3244	0.92	0.42	1.55	0.96
t3246	0.14	0.03	0.06	0.08
t3249	0.02	0.14	0.14	0.1
t3252	0.29	0.09	1.14	0.51
t3253	0.22	0.03	1.06	0.44
t3254	0.26	0.1	1.13	0.5
t3254	0.26	0.1	1.13	0.5
t3255	0.25	0.07	0.84	0.39
t3257	0.08	0.2	0.26	0.18
t3262	0.12	0.15	0.19	0.15
t3263	0.77	0.69	1	0.82
t3269	0.11	0.22	0.32	0.22
t3270	0.1	0.47	0.03	0.2
t3271	0.13	0.13	0.36	0.21
t3272	0.23	0.12	0.38	0.24
t3273	0.06	0.18	0.1	0.11
t3274	309.22	508.66	1372.63	730.17
t3276	0.07	0.31	0.11	0.16
t3277	0.27	0	0.22	0.16
t3278	0.09	0	0	0.03
t3279	0.25	0.44	0.66	0.45
t3280	0.37	0.38	0.33	0.36
t3281	2.08	4.62	5.57	4.09
t3282	0.03	0.15	0.15	0.11
t3284	0.08	0.04	0.11	0.08
t3286	0.13	0.12	0.4	0.22
t3290	0.07	0.41	0.5	0.33
t3291	0.31	0.29	0.35	0.32
t3292	0.39	1.88	0.64	0.97
t3293	0.33	0.47	0.74	0.51
t3296	0.1	0.36	0.67	0.38
t3299	0.18	0.61	0.07	0.29
t3301	0.2	0.11	0.44	0.25
t3302	0.11	0.1	0.31	0.17
t3305	0.77	0.8	1.63	1.07
t3313	0.08	0.26	0.1	0.15
t3317	0.31	0.1	0.47	0.29
t3318	5.62	6.12	6.24	5.99
t3319	0.28	0.56	0.69	0.51
t3321	0.29	0.55	0.53	0.46
t3326	0.12	0.3	0.28	0.23
t3328	0.13	0.17	0.48	0.26

t3330	0.34	0.47	0.45	0.42
t3331	0.57	0.42	0.31	0.43
t3334	0.05	0.05	0.25	0.12
t3335	0.03	0.05	0.04	0.04
t3341	0.04	0.1	0.13	0.09
t3343	0.1	0.28	0.07	0.15
t3344	0	0.64	0.08	0.24
t3347	0.03	0.07	0.21	0.1
t3349	0.2	0.11	0.08	0.13
t3351	0.26	0.45	0.33	0.35
t3352	0.13	0.02	0.39	0.18
t3353	0.24	0.26	0.27	0.26
t3355	0.13	0.26	1.11	0.5
t3356	0.04	0.09	0.11	0.08
t3357	0	0	0.19	0.06
t3367	0.52	0.4	2.01	0.98
t3368	0.1	0.1	0.11	0.1
t3369	0.13	0.2	0.74	0.36
t3370	0.08	0.13	0.16	0.12
t3371	0.09	0.09	0.24	0.14
t3379	0.14	0.23	0.14	0.17
t3382	0.28	0.6	0.43	0.44
t3385	0.03	0.02	0.17	0.07
t3386	0	0.23	0.17	0.13
t3387	0.06	0.31	0.51	0.29
t3391	0.23	0.55	0.11	0.3
t3392	0.18	0.3	0.31	0.26
t3399	1.11	1.89	2.96	1.99
t3400	0	0.07	0.21	0.09
t3400	0	0.07	0.21	0.09
t3401	0.16	0.27	0	0.14
t3401	0.16	0.27	0	0.14
t3403	0.11	0.13	0.39	0.21
t3405	0	0.11	0.13	0.08
t3406	0.09	0.24	0.19	0.17
t3407	0	0.31	0.24	0.18
t3408	0	0	0.03	0.01
t3410	0.02	0.1	0.13	0.08
t3411	0.11	0.38	0.31	0.27
t3412	0.8	1.18	1.29	1.09
t3412	0.8	1.18	1.29	1.09
t3413	0.82	2.09	2.6	1.84
t3413	0.82	2.09	2.6	1.84
t3414	0.75	1.67	1.11	1.18
t3415	0.09	0.91	0.61	0.54
t3416	0.2	0.25	0.37	0.27
t3416	0.2	0.25	0.37	0.27
t3418	0.08	0.25	0.2	0.18
t3421	0.05	0.05	0.06	0.05
t3422	0	0	0.6	0.2
t3423	0	0	0	0
t3424	0	0.1	0.19	0.1
t3425	0.06	0	0.49	0.18

t3426	0	0	0.14	0.05
t3427	0.01	0.1	0.31	0.14
t3428	0.06	0.06	0.16	0.09
t3429	0	0.12	0.08	0.07
t3432	0	0	0	0
t3433	0.18	0.32	0.4	0.3
t3434	0	0.08	0.4	0.16
t3435	0.2	0.16	0.2	0.19
t3435	0.2	0.16	0.2	0.19
t3438	0	0	0	0
t3439	0	0	0	0
t3440	0	0	0	0
t3441	0	0	0	0
t3442	0	0	0	0
t3443	0	0	0	0
t3444	0	0	0	0
t3445	0	0	0	0
t3448	1.05	2.23	1.79	1.69
t3449	0.27	0.27	1.22	0.59
t3457	1.05	0.79	1.23	1.02
t3458	2.27	2.47	1.8	2.18
t3459	0.54	0.69	1.87	1.03
t3460	0.06	0.35	0.47	0.29
t3462	0.04	0.34	0.34	0.24
t3463	0.14	0.04	0.31	0.16
t3465	0.08	0.14	0.08	0.1
t3466	2.36	2.35	2.76	2.49
t3486	0	2.89	0.17	1.02
t3497	0.41	0.5	0.29	0.4
t3498	0.14	0.57	0.45	0.39
t3499	1.49	4.54	3.7	3.24
t3505	4.43	4.56	8.87	5.95
t3506	0.1	0.44	0.29	0.28
t3507	0.22	0.18	0.25	0.22
t3508	0.25	0.11	0.19	0.18
t3512	0.1	0.14	0.17	0.14
t3514	0.02	0.08	0.29	0.13
t3515	0.08	0.03	0.04	0.05
t3516	0.31	0.47	0.21	0.33
t3520	0.01	0.08	0.36	0.15
t3521	0.01	0.1	0.32	0.14
t3530	2.65	8.59	18.07	9.77
t3533	0.09	0.22	0.27	0.19
t3535	0.25	0.41	0.07	0.24
t3536	0.04	0.15	0.4	0.2
t3537	27.14	63.38	34.43	41.65
t3538	1.56	1.6	7.12	3.43
t3539	0.74	1.49	0.24	0.82
t3540	1.61	6.52	5.25	4.46
t3541	2.14	1.89	6.49	3.51
t3542	0.25	0.36	0.65	0.42
t3543	0.14	0.16	0.29	0.2
t3544	0.85	0.35	1.32	0.84

t3545	0.7	0.19	0.86	0.58
t3546	0.43	0.24	0.88	0.52
t3547	0.26	0.75	0.84	0.62
t3549	1.5	1.21	0.99	1.23
t3549	1.5	1.21	0.99	1.23
t3550	0.39	0.28	0.26	0.31
t3551	0.14	0.13	0.07	0.11
t3552	0.03	0.1	0.13	0.09
t3553	0	0	0.14	0.05
t3554	0.08	0.11	0.06	0.08
t3557	6.96	4.69	9.5	7.05
t3558	0.07	0.15	0.34	0.19
t3562	0.15	0.16	0.3	0.2
t3563	0.19	0.49	0.3	0.33
t3565	0.02	0.03	0.07	0.04
t3566	0.13	0.19	0.23	0.18
t3567	0.14	0.21	0.63	0.33
t3568	0	0.3	0.09	0.13
t3576	0.91	0.85	2.17	1.31
t3577	0.32	0.35	0.43	0.37
t3578	0.1	0.11	0.28	0.16
t3579	0	0.08	0.12	0.07
t3580	0.02	0.09	0.12	0.08
t3581	0.02	0.05	0.06	0.04
t3582	0.96	0.89	0.91	0.92
t3583	44.66	18.86	129.92	64.48
t3584	42.02	14.8	67	41.27
t3585	12.38	14.92	19.2	15.5
t3586	5.18	7.64	5.91	6.24
t3587	0.28	1.21	1.2	0.9
t3588	0.34	2.42	1.34	1.37
t3589	0.21	0.22	0.35	0.26
t3590	0.6	1.59	1.14	1.11
t3591	0.94	1.33	1.71	1.33
t3592	0.27	0.73	1.1	0.7
t3593	0.18	0.16	0.14	0.16
t3595	0.18	1.03	0.32	0.51
t3596	0.18	0.18	0.34	0.23
t3597	0.13	0.1	0.33	0.19
t3598	0.1	0.03	0.1	0.08
t3599	0.07	0.03	0.11	0.07
t3600	0.14	0.11	0.32	0.19
t3601	0.3	0.09	0.24	0.21
t3602	0.06	0.1	0.3	0.15
t3603	0.25	0.24	0.44	0.31
t3604	0.02	0.12	0.08	0.07
t3605	0	0.14	0.04	0.06
t3606	0.06	0.18	0	0.08
t3607	0.08	0.15	0.11	0.11
t3608	0.08	0.12	0.1	0.1
t3609	0.06	0.16	0.08	0.1
t3610	0.05	0.2	0.07	0.11
t3611	0.16	0.18	0.29	0.21

t3612	0.16	0.41	0.55	0.37
t3618	0.54	0.74	1.19	0.82
t3619	11.76	14.35	47.33	24.48
t3622	0.19	0.15	0.43	0.26
t3624	0.86	1.02	1.22	1.03
t3625	0.98	4.88	2.2	2.69
t3631	0.03	0.23	0.38	0.21
t3632	1.1	0.84	1.42	1.12
t3640	0.18	0.13	0.27	0.19
t3641	0.28	0.15	0.42	0.28
t3656	0.15	0.35	0.27	0.26
t3659	0.04	0.08	0.05	0.06
t3662	0.11	0.13	0.26	0.17
t3663	0.21	0.04	0.27	0.17
t3664	0.06	0.12	0.11	0.1
t3665	1.02	0.5	1.55	1.02
t3671	0.78	0.19	0.46	0.48
t3672	0.23	0.39	1.24	0.62
t3673	0.31	0.36	0.41	0.36
t3674	0.06	0.24	0.17	0.16
t3675	0.39	0.21	0.7	0.43
t3676	0.18	0.11	0.4	0.23
t3678	0.27	2.55	0.33	1.05
t3685	0.26	0.13	0.35	0.25
t3686	0.02	0.08	0.31	0.14
t3687	0.07	0.29	0.2	0.19
t3688	0.44	0.53	1.19	0.72
t3689	0.3	0.21	0.21	0.24
t3690	0.1	0.08	0.08	0.09
t3691	0.07	0.02	0.16	0.08
t3692	0.19	0.13	0.15	0.16
t3693	0.36	0.56	0.5	0.47
t3695	0.18	0.13	0.34	0.22
t3696	0.23	0.18	0.23	0.21
t3697	0	0.02	0.09	0.04
t3698	0.05	0.11	0.04	0.07
t3703	0	0	0	0
t3704	0	0	0	0
t3706	0	0	0	0
t3707	0.13	0.04	0.24	0.14
t3708	0.86	0.31	0.24	0.47
t3711	0.06	0.19	0.12	0.12
t3712	0	0	0.19	0.06
t3713	0.34	0.52	0.48	0.45
t3714	0.03	0.2	0.59	0.27
t3718	0.35	0.26	0.25	0.29
t3720	0.33	0.13	0.2	0.22
t3720	0.33	0.13	0.2	0.22
t3724	0.59	0.86	0.66	0.7
t3725	0.09	0.15	0.44	0.23
t3726	0.1	0.25	0.35	0.23
t3727	0.23	0.11	0.21	0.18
t3732	0.05	0.1	0.51	0.22

t3733	0.2	1.01	0.31	0.51
t3734	0	0.29	0.38	0.22
t3735	0.08	0.44	0.46	0.33
t3736	0.33	0.47	0.71	0.5
t3737	0.17	0.08	0.22	0.16
t3738	0.03	0.14	0	0.06
t3739	0.09	0.09	0.36	0.18
t3740	0	0.09	0.34	0.14
t3741	0	0.34	0.1	0.15
t3741	0	0.34	0.1	0.15
t3742	0.06	0.12	0.3	0.16
t3743	0.1	0.18	0.31	0.2
t3744	0.3	0	0.1	0.13
t3745	0.12	0.07	0.15	0.11
t3746	0	0.07	0.37	0.15
t3747	0	0.16	0.22	0.13
t3748	0.16	0.08	0.24	0.16
t3749	0.11	0.1	0.08	0.1
t3750	0.06	0.04	0.05	0.05
t3751	0.03	0.03	0.11	0.06
t3752	0.08	0.19	0.71	0.33
t3753	2.35	0.51	2.13	1.66
t3756	0.4	0.1	1.96	0.82
t3757	1.2	1.15	0.77	1.04
t3758	0.11	0.14	0.14	0.13
t3759	0.05	0.09	0.37	0.17
t3760	0.18	0.08	0.15	0.14
t3760	0.18	0.08	0.15	0.14
t3761	0	0.08	0.36	0.15
t3762	0	0	0	0
t3762	0	0	0	0
t3763	0	0	0	0
t3764	0.15	0.16	0.28	0.2
t3765	0.03	0.01	0.11	0.05
t3765	0.03	0.01	0.11	0.05
t3766	0.12	0.08	0.16	0.12
t3767	0.07	0.28	0.17	0.17
t3770	0.13	0.16	0.23	0.17
t3771	0.1	0.23	0.13	0.15
t3772	0.38	0.23	0.43	0.35
t3779	0.13	0.16	0.11	0.13
t3780	0.19	0.17	0.2	0.19
t3781	0.43	0.12	0.94	0.5
t3782	0.62	0.5	0.56	0.56
t3783	0.25	0.25	0.52	0.34
t3786	0.1	0.96	0.22	0.43
t3788	0.46	1.31	0.65	0.81
t3789	0.07	0.13	0.05	0.08
t3798	0	0.92	0.08	0.33
t3799	0.04	1.15	0.1	0.43
t3812	0.21	0.12	0.06	0.13
t3813	0.16	0.15	0.21	0.17
t3814	0.22	0.23	0.3	0.25

t3816	0.53	0.82	0.21	0.52
t3819	7.2	2.4	2.77	4.12
t3821	2.22	4.22	6.29	4.24
t3821	2.22	4.22	6.29	4.24
t3822	0.24	0.28	1.33	0.62
t3823	0.06	0.09	0.15	0.1
t3824	0.29	0.38	0.28	0.32
t3825	0.22	0.1	0.29	0.2
t3825	6.19	0.36	5.47	4.01
t3826	0.09	0.17	0.32	0.19
t3826	1.52	0.27	1.93	1.24
t3828	0.04	0.1	0.24	0.13
t3829	0	0.06	0.04	0.03
t3830	0.23	1.95	0.08	0.75
t3831	0	0.35	0.27	0.21
t3835	0.12	0.31	0.31	0.25
t3838	0.08	0.12	0.18	0.13
t3840	0.11	0.18	0.35	0.21
t3841	0	0.09	0.15	0.08
t3842	0.07	0	0.53	0.2
t3843	0.06	0.06	0.09	0.07
t3844	0	0.02	0.18	0.07
t3845	0.05	0.07	0.02	0.05
t3846	0	0.09	0.23	0.11
t3846	0	0.09	0.23	0.11
t3847	0.12	0	0.21	0.11
t3848	0	0	0.2	0.07
t3849	0.17	0.09	0.25	0.17
t3850	0.05	0.28	0.25	0.19
t3851	0.02	0.05	0.04	0.04
t3852	0.03	0.14	0.23	0.13
t3853	0.77	0.45	0.9	0.71
t3856	0.22	0.82	0.37	0.47
t3860	0	0	0	0
t3861	0.99	0.57	0.58	0.71
t3861	0.99	0.57	0.58	0.71
t3862	0	0.08	0	0.03
t3865	0	0.15	0.07	0.07
t3866	0.05	0.19	0.14	0.13
t3867	0.23	0.4	0.21	0.28
t3868	0.19	0.76	0.54	0.5
t3869	0.62	1.78	0.4	0.93
t3871	0	0.08	0.07	0.05
t3872	0.4	0.33	0.31	0.35
t3873	0.32	0.2	0.33	0.28
t3874	0.62	1.93	1.26	1.27
t3876	0.44	0.54	1.48	0.82
t3878	1.17	0.52	0.76	0.82
t3879	0.07	0.08	0.15	0.1
t3881	0.07	0.39	0	0.15
t3882	0.07	0.19	0.2	0.15
t3883	0.07	0.16	0.2	0.14
t3884	0.71	0.27	0.6	0.53

t3890	0.33	0.19	0.7	0.41
t3891	0.68	0.65	0.5	0.61
t3892	0.47	0.12	0.36	0.32
t3892	0.47	0.12	0.36	0.32
t3893	0.22	0.15	0.23	0.2
t3894	0.25	0.25	0.32	0.27
t3895	0.32	0.26	0.39	0.32
t3896	0.12	0.18	0.57	0.29
t3897	0.12	0.03	0.08	0.08
t3898	0.09	0.06	0.2	0.12
t3899	0.12	0.31	0.45	0.29
t3900	0.28	0.04	0.38	0.23
t3901	0.06	0.05	0.2	0.1
t3902	0.15	0.09	0.21	0.15
t3904	0.21	0.2	0.21	0.21
t3906	0.24	0.89	1.27	0.8
t3907	0.09	0.11	0.16	0.12
t3908	0.19	0.1	0.24	0.18
t3909	1.25	1.13	2.16	1.51
t3910	0.03	0.1	0.29	0.14
t3911	0.08	0.08	0.11	0.09
t3913	0.24	0.4	0.29	0.31
t3916	0.54	0.62	0.75	0.64
t3919	0.14	0.49	0.42	0.35
t3920	0	0.09	0.06	0.05
t3921	0	0.14	0.53	0.22
t3923	2.24	1.04	4.35	2.54
t3924	0.39	0.76	0.83	0.66
t3927	0.3	0.5	0.46	0.42
t3928	0.04	0.12	0.13	0.1
t3929	0	0.59	0	0.2
t3930	0.41	0.25	0.8	0.49
t3930	0.41	0.25	0.8	0.49
t3931	0.1	0.36	0.24	0.23
t3932	0.04	0.13	0.99	0.39
t3933	0.02	0.47	0.36	0.28
t3934	0.29	0.17	0.6	0.35
t3935	0.12	0.07	0.17	0.12
t3936	0.12	0.08	0.25	0.15
t3937	0.23	0.14	0.28	0.22
t3939	0.6	0.33	0.5	0.48
t3940	0	0.09	0.31	0.13
t3941	0.36	0.43	0.35	0.38
t3942	0.59	0.65	0.62	0.62
t3943	0.01	0.07	0.15	0.08
t3944	5.48	3.34	8.13	5.65
t3947	0.22	0.32	0.15	0.23
t3948	0.26	0.47	0.16	0.3
t3949	0.82	1.2	2.18	1.4
t3950	0.2	0.44	0.43	0.36
t3955	0.02	0.09	0.14	0.08
t3956	0	0.18	0	0.06
t3957	0.01	0.27	0.08	0.12

t3963	0.11	0.14	0.89	0.38
t3969	0.48	0.31	0.46	0.42
t3972	0.06	0.12	0.42	0.2
t3973	0	0.15	0.16	0.1
t3974	0	0.19	0.18	0.12
t3975	0.21	0.05	0.27	0.18
t3976	0.1	0.23	0.24	0.19
t3977	0.09	0.03	0.18	0.1
t3980	0.16	0.09	0.45	0.23
t3987	0.89	1.65	1.99	1.51
t3988	0.11	0.23	0.26	0.2
t3989	0.03	0.17	0.04	0.08
t3990	5.6	5.45	3.63	4.89
t3995	0.82	2.6	2.2	1.87
t3996	0.85	0.23	0.4	0.49
t3997	0.12	0.18	0.7	0.33
t3998	0.25	0.17	0.3	0.24
t3999	0.04	0	0	0.01
t4001	0.45	0.41	0.39	0.42
t4002	0.2	0.62	0.55	0.46
t4007	0.18	0.13	0.14	0.15
t4010	0.23	0.38	0.43	0.35
t4011	0.23	0.07	1	0.43
t4012	0.11	0.15	0.12	0.13
t4014	0.03	0.03	0.48	0.18
t4015	0.05	0.05	0.19	0.1
t4016	0	0.16	0.18	0.11
t4017	0.06	0	0.22	0.09
t4026	0	0.14	0.18	0.11
t4027	0.11	0.11	0.16	0.13
t4027	0.11	0.11	0.16	0.13
t4029	0.31	0.17	0.03	0.17
t4030	0	0	0.46	0.15
t4031	0.17	0.25	0.21	0.21
t4032	0.02	0.08	0.34	0.15
t4034	0	0	0.24	0.08
t4038	0.16	0.13	0.19	0.16
t4038	0.16	0.13	0.19	0.16
t4040	0.96	1.01	1.5	1.16
t4042	0.11	0.54	0.44	0.36
t4043	0.1	0.21	0.2	0.17
t4043	0.1	0.21	0.2	0.17
t4044	0.58	0.44	1.17	0.73
t4045	0	0.35	0	0.12
t4046	0.08	0.16	0.14	0.13
t4047	0.09	0.13	0.24	0.15
t4049	0.13	0.12	1.47	0.57
t4051	0.05	0.91	0.22	0.39
t4053	0.44	0.51	0.51	0.49
t4054	0.15	0.14	0.53	0.27
t4055	0	0.07	0.16	0.08
t4056	0.88	0.82	0	0.57
t4092	0.4	0.25	0.08	0.24

t4093	0.11	0.37	0.33	0.27
t4099	0.06	0.14	0.27	0.16
t4100	0.33	1.05	0.84	0.74
t4101	0.04	0.05	0.31	0.13
t4102	0.19	0.23	0.18	0.2
t4104	0.35	0.39	0.5	0.41
t4105	0.61	0.68	1.44	0.91
t4109	0.05	0.4	0.85	0.43
t4116	0.04	0.11	0.1	0.08
t4118	0.24	0.41	0.26	0.3
t4119	0.18	1.76	0.52	0.82
t4120	0.17	1.11	0.4	0.56
t4121	0.5	0.24	0.22	0.32
t4123	0.1	0.43	0.49	0.34
t4124	0.15	0.03	0.22	0.13
t4125	0	0.05	0.14	0.06
t4127	10.43	13.59	18.63	14.22
t4128	0	0.17	0.02	0.06
t4129	0	0.03	0.09	0.04
t4130	0	0.05	0.08	0.04
t4131	0.03	0.07	0.16	0.09
t4132	0	0.12	0.22	0.11
t4144	0.09	0.12	0.09	0.1
t4145	0.1	0.5	1.29	0.63
t4146	0.31	0.14	1.97	0.81
t4148	0	0.19	0.12	0.1
t4149	0	0.15	0.06	0.07
t4150	0.47	0.2	0.52	0.4
t4151	2.49	2.01	3.46	2.65
t4156	0.87	1.05	3.2	1.71
t4157	0.24	0.38	1.24	0.62
t4158	2.6	1.36	3.85	2.6
t4159	1.15	0.67	1.4	1.07
t4162	0	0.05	0.15	0.07
t4163	0.02	0.1	0.08	0.07
t4164	0.04	0.14	0.07	0.08
t4165	0	0.09	0.05	0.05
t4168	0.02	0.23	0.21	0.15
t4169	0	0.3	0.21	0.17
t4170	0.09	0.17	0.39	0.22
t4173	0.04	0.09	0.26	0.13
t4174	0.02	0.11	0.22	0.12
t4175	0.03	0.06	0.44	0.18
t4177	0.81	0.7	0.93	0.81
t4178	0.29	0.13	0.41	0.28
t4180	0	0.08	0.29	0.12
t4182	0.1	0.42	0.35	0.29
t4183	0.63	0.37	0.82	0.61
t4184	0.33	0.19	0.33	0.28
t4185	0.34	0.29	0.41	0.35
t4186	0.05	0.05	0.09	0.06
t4187	0.15	0.04	0.11	0.1
t4188	0.09	0.14	0.39	0.21

t4191	0.07	0.29	0.14	0.17
t4192	0.22	0.14	0.36	0.24
t4193	0.4	0.51	0.31	0.41
t4194	0.1	0.19	0.56	0.28
t4195	0.05	0.07	0.07	0.06
t4196	0	0.21	0.26	0.16
t4200	0.56	1.2	1.13	0.96
t4201	0.29	0.18	0.51	0.33
t4202	1.21	3.11	5.94	3.42
t4207	0.6	0.57	0.58	0.58
t4209	0	0.09	0	0.03
t4214	0.07	0.06	0.22	0.12
t4215	0.03	0.1	0.13	0.09
t4216	0.29	0.11	0.41	0.27
t4217	0.15	0.3	0.29	0.25
t4218	0	0.06	0.08	0.05
t4219	0.1	0.26	0.76	0.37
t4220	0.58	1.57	0.75	0.97
t4222	5.43	35.57	3.14	14.71
t4223	0.41	1.2	1.85	1.15
t4224	0.37	0.71	0.4	0.49
t4226	0.08	0.04	0.16	0.09
t4228	0.06	0.13	0.11	0.1
t4229	0.29	0.14	0.1	0.18
t4230	0.04	0.11	0.11	0.09
t4231	0.04	0.27	0.28	0.2
t4232	0.06	0.09	0.21	0.12
t4233	0.08	0.18	0.2	0.15
t4235	0	0	0.24	0.08
t4236	0	0.11	0.14	0.08
t4238	0.12	0.03	0.23	0.13
t4239	0.11	0.09	0.27	0.16
t4240	0.5	0.48	1.23	0.74
t4243	0.04	0.06	0.18	0.09
t4246	0.1	0.02	0.34	0.15
t4248	0.9	0.13	0.19	0.41
t4253	0.1	0.2	0	0.1
t4254	0.08	0.12	0.21	0.14
t4255	0.08	0.09	0.14	0.1
t4256	0	0.08	0.25	0.11
t4257	0	0.21	0	0.07
t4258	0.17	0.05	0.18	0.13
t4259	0	0.07	0.12	0.06
t4260	0.03	0.14	0.27	0.15
t4261	0.06	0.12	0.25	0.14
t4262	0.26	0.15	0	0.14
t4263	0.1	0.22	0.51	0.28
t4264	0.13	0.14	0.78	0.35
t4265	0.04	0.07	0.78	0.3
t4266	0.46	0.6	2.23	1.1
t4267	0.2	0	0.15	0.12
t4268	0.03	0.24	0.05	0.11
t4269	0.06	0.17	0.01	0.08

t4270	0.01	0.13	0.18	0.11
t4271	0	0.12	0.09	0.07
t4272	0	0.06	0	0.02
t4273	0.07	0.19	0.17	0.14
t4274	0.02	0.12	0.17	0.1
t4275	0	0	0.43	0.14
t4276	0.14	0.13	0.14	0.14
t4277	0.33	0.9	1.08	0.77
t4278	0.16	0.33	1.12	0.54
t4279	0	0.12	0.37	0.16
t4280	0	0.05	0.27	0.11
t4281	0	0	0.06	0.02
t4282	0.04	0.26	0	0.1
t4283	0	0	0.05	0.02
t4284	0.22	0	0.22	0.15
t4285	0	0.16	0.07	0.08
t4286	0.03	0	0.17	0.07
t4287	0.06	0.03	0.1	0.06
t4288	0.03	0.07	0.41	0.17
t4289	0	0.12	0.08	0.07
t4290	0.09	0.33	0.11	0.18
t4293	0.17	0.25	0.16	0.19
t4294	0.22	0.94	0.19	0.45
t4294	0.22	0.94	0.19	0.45
t4295	0.12	0.1	0.34	0.19
t4295	0.12	0.1	0.34	0.19
t4296	0	0.14	0.35	0.16
t4296	0	0.14	0.35	0.16
t4297	0.05	0.19	0.34	0.19
t4297	0.05	0.19	0.34	0.19
t4298	0	0	0	0
t4298	0	0	0	0
t4299	0.08	0.08	0	0.05
t4299	0.08	0.08	0	0.05
t4300	0.21	0.17	1.56	0.65
t4300	0.21	0.17	1.56	0.65
t4301	0.04	0.06	0.05	0.05
t4301	0.04	0.06	0.05	0.05
t4302	0.17	0.23	0.67	0.36
t4302	0.17	0.23	0.67	0.36
t4304	0.34	0.61	1.59	0.85
t4304	0.34	0.61	1.59	0.85
t4305	0.15	0.09	0.27	0.17
t4305	0.15	0.09	0.27	0.17
t4306	0	0	0	0
t4306	0	0	0	0
t4307	0.05	0	0.38	0.14
t4307	0.05	0	0.38	0.14
t4309	0	0.12	0.16	0.09
t4309	0	0.12	0.16	0.09
t4310	0	0.06	0.16	0.07
t4310	0	0.06	0.16	0.07
t4312	0	0	0.18	0.06

t4312	0	0	0.18	0.06
t4313	0	0	0.03	0.01
t4313	0	0	0.03	0.01
t4314	0	0.1	0.31	0.14
t4314	0	0.1	0.31	0.14
t4315	0.25	0	0	0.08
t4315	0.25	0	0	0.08
t4316	0	0	0	0
t4316	0	0	0	0
t4317	0	0.21	0.07	0.09
t4317	0	0.21	0.07	0.09
t4318	0.04	0	0	0.01
t4318	0.04	0	0	0.01
t4319	0.02	0.12	0.22	0.12
t4319	0.02	0.12	0.22	0.12
t4320	0.09	0.21	0	0.1
t4320	0.09	0.21	0	0.1
t4321	0.42	0.53	0.22	0.39
t4322	0.05	0.05	0.16	0.09
t4323	0.13	1.02	0.61	0.59
t4323	0.13	1.02	0.61	0.59
t4324	0.81	4.28	4.02	3.04
t4324	0.81	4.28	4.02	3.04
t4325	0.75	8.74	2.38	3.96
t4325	0.75	8.74	2.38	3.96
t4326	1.3	0.26	4.15	1.9
t4326	1.3	0.26	4.15	1.9
t4328	0.03	0.04	0.06	0.04
t4328	0.03	0.04	0.06	0.04
t4330	0	0.22	0.19	0.14
t4330	0	0.22	0.19	0.14
t4331	0	0.53	0.19	0.24
t4331	0	0.53	0.19	0.24
t4332	0.14	0.81	0.01	0.32
t4332	0.14	0.81	0.01	0.32
t4333	0.01	0.03	0.27	0.1
t4333	0.01	0.03	0.27	0.1
t4334	0	0	0	0
t4334	0	0	0	0
t4340	0	0	0	0
t4341	0	0.71	0.18	0.3
t4343	1.46	0.83	0.58	0.96
t4354	0.21	0.36	0.32	0.3
t4355	0.02	0.01	0.02	0.02
t4356	0.03	0.1	0.38	0.17
t4357	0.07	0.27	0.18	0.17
t4358	3.08	4.42	2.45	3.32
t4359	0.02	0.96	0.6	0.53
t4360	1.32	7.46	1.5	3.43
t4361	0.01	0.1	0.14	0.08
t4362	1.55	2.05	3.18	2.26
t4363	0.05	0.34	0.33	0.24
t4365	0.77	0.42	0.21	0.47

t4366	0	0	0	0
t4367	0.03	0.03	0.11	0.06
t4368	0.3	0	0	0.1
t4373	0.04	0.75	0.13	0.31
t4378	1.5	4.24	1.24	2.33
t4380	0.49	0.32	0.44	0.42
t4383	0.87	1.18	1.51	1.19
t4384	0.47	0.36	0.62	0.48
t4385	0.73	0.29	0.35	0.46
t4391	0.29	0.23	0.17	0.23
t4394	9.82	5.68	10.91	8.8
t4395	15.71	8.43	36.66	20.27
t4397	0.06	0.07	0.16	0.1
t4398	0.41	0.43	0.11	0.32
t4399	1.03	0.1	0.2	0.44
t4401	0.17	0.23	0.61	0.34
t4403	0.24	0.31	1.48	0.68
t4406	0.08	0.1	0.2	0.13
t4407	0.31	0.55	0.3	0.39
t4408	0	0.05	0	0.02
t4413	0.04	0.2	0.6	0.28
t4416	0	0.12	0.33	0.15
t4418	0.06	0.69	0.85	0.53
t4420	0.03	0.1	0.36	0.16
t4421	0	0.06	0	0.02
t4422	0	0.03	0.18	0.07
t4423	0.1	0.19	0.59	0.29
t4424	0.04	0.08	0.14	0.09
t4425	0	0.41	0.17	0.19
t4426	0.04	0.28	0.29	0.2
t4427	0.04	0.08	0.24	0.12
t4428	1.31	0.43	0.95	0.9
t4429	0.09	0.21	0.52	0.27
t4430	0.33	0.32	0.28	0.31
t4431	0.59	0.24	0.39	0.41
t4432	0.57	0.22	0.41	0.4
t4433	0.16	0.09	0.68	0.31
t4434	0.03	0.08	0.26	0.12
t4435	0.13	0.04	0.06	0.08
t4436	0	0	0.19	0.06
t4437	0.07	0.1	0.09	0.09
t4438	0.06	0.03	0.42	0.17
t4439	0.07	0.2	0.34	0.2
t4440	0	0.26	0.08	0.11
t4441	0.17	0.47	0.47	0.37
t4446	0.39	0.09	0.57	0.35
t4447	0.18	0.24	0.64	0.35
t4448	0.26	0.48	1.08	0.61
t4450	0.16	0.06	0.19	0.14
t4451	0.25	0.09	0.28	0.21
t4452	0.25	0.21	0.36	0.27
t4453	0.07	0.07	0.06	0.07
t4454	0	0.05	0.11	0.05

t4455	0.41	0.13	0.68	0.41
t4458	0.03	0.3	0.84	0.39
t4459	0.85	1.13	0	0.66
t4461	0	0.13	0.18	0.1
t4463	0.03	0.1	0.16	0.1
t4464	0.13	0.06	0.25	0.15
t4465	0.14	0.53	0.32	0.33
t4466	0.88	0.3	2.75	1.31
t4467	0.06	0.2	0.31	0.19
t4471	0.09	0.67	0.68	0.48
t4472	0.35	0.45	0.39	0.4
t4473	25.13	143.9	138.54	102.52
t4474	0.07	0.09	0.34	0.17
t4475	0	0	0.37	0.12
t4476	0	0.03	0.44	0.16
t4477	0.17	0.06	0.25	0.16
t4478	0.04	0	0.09	0.04
t4479	0.1	0.15	0.03	0.09
t4480	0.06	0.08	0.05	0.06
t4481	0.11	0.02	0.14	0.09
t4482	0.07	0.04	0.29	0.13
t4484	0.08	0.15	0.11	0.11
t4487	1.54	1.58	4.94	2.69
t4492	1.22	3.86	2.82	2.63
t4493	1.24	2.25	2.04	1.84
t4497	0.05	0.16	0.25	0.15
t4498	0.22	0.18	0.89	0.43
t4502	0.21	0.03	0.34	0.19
t4504	0.54	1.1	1.5	1.05
t4506	0.15	0.27	0	0.14
t4507	0.15	0.32	0	0.16
t4508	0.11	0.09	0.34	0.18
t4509	0.26	0.26	0.05	0.19
t4513	0.07	0.13	0.05	0.08
t4514	0.08	0.04	0.37	0.16
t4516	0.04	0.09	0.29	0.14
t4516	0.04	0.09	0.29	0.14
t4521	1.07	6.52	2.07	3.22
t4521	1.07	6.52	2.07	3.22
t4524	0.07	0.13	0.32	0.17
t4524	0.07	0.13	0.32	0.17
t4527	0	0.15	0.13	0.09
t4528	0.07	0.15	0.09	0.1
t4530	0	0.96	0.21	0.39
t4531	2.13	1.88	2.93	2.31
t4532	0.02	0.01	0.19	0.07
t4532	0.02	0.01	0.19	0.07
t4535	0.05	0.11	0	0.05
t4537	0.02	0.2	0.2	0.14
t4537	0.02	0.2	0.2	0.14
t4538	0	0.2	0.7	0.3
t4539	0.17	0	0.03	0.07
t4541	0.06	0.11	0.04	0.07

t4542	0.03	0.17	0.17	0.12
t4545	0.2	0.53	0.66	0.46
t4546	2.49	0.19	1.34	1.34
t4547	0.05	0.05	0.11	0.07
t4548	0.09	0.26	0	0.12
t4549	0.03	2.26	0.72	1
t4550	0.08	0.23	0.38	0.23
t4551	0.16	0.34	0	0.17
t4552	0.04	0.06	0.05	0.05
t4553	0.07	0.15	0.54	0.25
t4554	0.37	0.38	0.06	0.27
t4555	0.97	0.8	2.21	1.33
t4557	0.1	0.28	0.34	0.24
t4558	0	0.27	0	0.09
t4559	0.2	0.21	0.43	0.28
t4560	0.04	0.19	0.46	0.23
t4561	0.63	0.34	0.5	0.49
t4562	0.08	0.14	0.05	0.09
t4563	0.33	0.05	0.31	0.23
t4564	0.17	0.04	0.6	0.27
t4565	2.35	0.58	0.71	1.21
t4566	0.77	0.51	0.69	0.66
t4567	0.2	0.06	0.3	0.19
t4568	0.09	0.07	0.02	0.06
t4569	0.2	0.07	0.48	0.25
t4571	0.46	0	0	0.15
t4572	0.12	2.08	0.09	0.76
t4573	0.16	0.12	0.15	0.14
t4578	0.2	0.25	0.5	0.32
t4579	0.38	0.51	0.26	0.38
t4580	0.05	0.48	0.36	0.3
t4581	1.34	2.89	12.57	5.6
t4582	0.46	0.74	1.09	0.76
t4584	0.25	0.25	0.82	0.44
t4585	0.38	0	1.1	0.49
t4586	0.13	0.67	0.68	0.49
t4588	0.49	0.1	0.51	0.37
t4589	0.16	0.1	0.18	0.15
t4590	0	0.06	0.08	0.05
t4591	3.64	2.59	4.92	3.72
t4592	0.36	0.43	0.55	0.45
t4594	0.45	0.3	1.02	0.59
t4595	1.16	0.18	0.64	0.66
t4599	0	0.13	0.06	0.06
t4602	0.02	0.18	0.37	0.19
t4605	0.13	0	0.29	0.14
t4606	0.11	0.01	0.2	0.11
t4607	0.21	0.23	0.51	0.32
t4608	0.13	0.23	0.17	0.18
t4609	0.09	0.02	0.24	0.12
t4614	0.49	0.31	0.31	0.37
t4616	0.07	0.49	0.3	0.29
t4617	0.07	0.23	0.22	0.17

t4618	0	0.13	0.11	0.08
t4619	0.1	0.26	0.3	0.22
t4620	0.58	0.49	0.63	0.57
t4622	0.27	0.43	0.83	0.51
t4623	1.23	2.01	0.32	1.19
t4623	1.23	2.01	0.32	1.19
t4624	15.94	26.19	37.23	26.45
t4626	1.23	1.08	2.23	1.51
t4627	0.12	0.1	0.18	0.13
t4628	0.04	0.1	0.22	0.12
t4630	0.29	0.12	0.42	0.28
t4630	0.29	0.12	0.42	0.28
t4631	0.03	0.09	0.32	0.15
t4633	0.17	0.19	0.62	0.33
t4634	0.04	0.16	0.05	0.08
t4635	0.22	0.5	0.19	0.3
t4636	0.08	0.09	0.25	0.14
t4638	33.18	124.04	76.03	77.75
t4639	0.4	0.39	0.53	0.44

## 9.6 Non-coding RNA details

Gene	Start	Stop	Arithmetic Mean			Average	
			WT1	WT2	WT3		
	73	1	189	33.72	54.98	132.54	73.74666667
Thr_leader		191	311	5.57	0.4	6.68	4.216666667
rfam_107c.105		5014	5123	0.54	0.05	0.25	0.28
rfam_107c.114		81427	81537	0	0	0	0
rfam_175c.63		95014	95134	0	0	1.05	0.35
rfam_107c.1		101116	101223	0.22	0.78	0	0.333333333
rfam_107c.2		109017	109116	0	0.05	0.66	0.236666667
rfam_107c.3		119385	119481	0.25	0	0.97	0.406666667
TPP		128108	128207	0.04	0.74	1.13	0.636666667
SgrS		129980	130218	4.48	0.28	0.67	1.81
Leu_leader		135184	135332	1.13	0.32	0	0.483333333
rfam_107c.6		139451	139563	0	0	0	0
rfam_341c.3		152481	152561	6.78	12.57	8.45	9.266666667
rfam_107c.11		209243	209353	1.02	0.1	0.51	0.543333333
rfam_107c.19		227845	227966	0	0	0.84	0.28
rfam_107c.20		232065	232178	0.79	0	0	0.263333333
rfam_107c.21		234106	234213	0.74	0.71	0.08	0.51
rfam_107c.22		237505	237612	0.3	0.27	0.22	0.263333333
rfam_107c.25		244215	244323	0	1.19	1.07	0.753333333
rfam_107c.29		249818	249932	1.9	6.56	13.75	7.403333333
rfam_175c.21		250821	250945	3.75	1.49	0.56	1.933333333
t44		251051	251144	80.8	1155.27	15.5	417.19
rfam_175c.22		251911	252034	0	0.19	0	0.063333333
rfam_175c.23		256051	256173	0	0.36	0	0.12
STnc490k		270919	271066	0	0	0	0
rfam_107c.32		277526	277635	0	0.07	0.43	0.166666667
STnc490k		280538	280686	0	0	1.83	0.61
PK-G12rRNA		291750	291857	0	0	0	0

5S_rRNA	292474	292589	0	0	0	0
rncO	315098	315309	1.42	4.27	1.03	2.24
rfam_107c.41	318688	318796	0.07	0	0	0.023333333
GlmY_tke1	332914	333061	62.8	673.15	14.98	250.31
rfam_107c.54	357149	357251	1.96	1.9	1.84	1.9
rfam_175c.42	360064	360185	1.42	0.37	0.41	0.733333333
ryfA	365400	365695	0.18	0.53	0.44	0.383333333
sroE	384343	384434	2.56	4.74	0.4	2.566666667
rfam_103c	417252	417501	18.42	8.41	56.79	27.873333333
rfam_105	447517	447787	40.88	42.1	123.37	68.783333333
rfam_105.1	447531	447708	65.6	73.4	206.99	115.33
STnc250	447726	447819	11.2	2.02	26.91	13.376666667
STnc490k	464568	464716	0.17	0.5	0.21	0.293333333
rfam_107c.103	486861	486963	0.19	0.1	0.64	0.31
rfam_107c.104	491872	491980	0	1.65	0	0.55
rfam_107c.106	511700	511807	0	0.24	0	0.08
rfam_107c	526043	526232	79.98	17.42	134.14	77.18
rfam_107c.107	526122	526232	135.51	29.22	226.98	130.57
STnc490k	528129	528277	0	0	0	0
rfam_175c.59	530324	530408	0.2	0	0.35	0.183333333
rfam_175c.60	531065	531186	1.87	0.58	1.12	1.19
rfam_107c.108	585236	585337	0.34	0.63	0.53	0.5
rfam_107c.109	598213	598328	0.61	0.21	0.17	0.33
rfam_107c.110	605698	605801	0.13	0.06	0	0.063333333
isrH-1	652339	652789	0	0.36	0.13	0.163333333
isrH	652510	652789	0	0.22	0.22	0.146666667
rfam_341c.16	652676	652756	0	0.14	0.17	0.103333333
rfam_107c.111	663264	663372	0.76	3.04	0.83	1.543333333
rfam_341c.17	670425	670507	0.78	0.7	0.36	0.613333333
STnc490k	673377	673525	0	0	0	0
rfam_107c.112	677005	677113	0.5	0.29	0.9	0.563333333
MicF	680757	680850	4.15	0.36	0.42	1.643333333
rfam_175c.61	692867	692990	0	0	0.79	0.263333333
isrG	702748	703029	0	0.09	0.21	0.1
rfam_107c.113	723906	724007	0.24	0.19	0.11	0.18
rfam_111.1	750299	750449	65.25	252.32	26.15	114.573333333
TPP	796380	796476	0.8	0.9	0	0.566666667
CyaR_RyeE	806251	806335	4.88	0.65	0	1.843333333
QUAD	831957	832102	0	0.16	0.21	0.123333333
rfam_115.1	879438	879534	13.66	237.5	5.23	85.463333333
rfam_107c.115	896523	896640	0.03	1.85	0.95	0.943333333
His_leader	897580	897703	0.85	1.12	0.24	0.736666667
rfam_175c.62	926317	926432	0	0	1.34	0.446666667
rfam_117.1	932965	933267	5.1	3.79	2.82	3.903333333
Cobalamin	933281	933456	0	0.39	0	0.13
rfam_118c	975648	975783	103.85	109.08	87.81	100.2466667
rseX	985610	985704	3.54	6.18	0	3.24
DsrA	994134	994217	1.35	1.36	1.47	1.393333333
rfam_121.1	1007471	1007596	10.03	182.98	24.4	72.47
rfam_125.1	1046823	1046862	162.07	885.15	1080.4	709.2066667
rfam_175c.1	1060576	1060698	0	0	0.49	0.163333333
rfam_175c.2	1066422	1066530	0.09	0.15	0	0.08
rfam_175c.3	1086839	1086963	1	0	0.58	0.526666667

SraC_RyeA	1092651	1092807	13.78	7.22	166.99	62.66333333
RyeB	1092692	1092791	0.49	0.31	0.79	0.53
yybP-ykoY	1127473	1127595	0.03	1.37	0	0.466666667
sroD	1144268	1144356	0.49	1.01	4.37	1.956666667
rfam_175c.4	1181921	1182043	1.3	2.09	1.82	1.736666667
STnc490k	1189168	1189316	0.67	0.67	0	0.446666667
STnc150	1189846	1190112	0	0	0.4	0.133333333
isrD	1212905	1212955	0	0.69	0	0.23
rfam_341c.1	1216232	1216306	0	0	0	0
rfam_175c.5	1216336	1216457	0.05	0	0	0.016666667
RyhB	1220046	1220113	0.97	0	0	0.323333333
STnc490k	1220604	1220752	0	0	0	0
rfam_107c.4	1229615	1229720	0.46	0.59	0	0.35
rfam_175c.6	1236390	1236509	2.64	1.74	1.65	2.01
rfam_341c.2	1236462	1236543	0	0.37	0.37	0.246666667
STnc490k	1264979	1265127	0.67	0	0	0.223333333
LR-PK1	1282880	1283221	92.05	698.74	160.82	317.2033333
rfam_107c.5	1298589	1298691	5.32	5	12.67	7.663333333
RprA	1299513	1299620	4.74	5.42	1.39	3.85
rydB	1305097	1305163	0.2	0.31	0	0.17
73	1375243	1375288	119.09	1097.91	81.59	432.8633333
rfam_175c.7	1446388	1446511	0	0	0	0
STnc490k	1450411	1450559	0.11	0	0	0.036666667
STnc570	1482097	1482247	28.36	0.74	1.78	10.29333333
STnc560	1482115	1482334	19.46	0.55	1.22	7.076666667
rfam_146.1	1482145	1482226	52.22	1.35	3.28	18.95
rfam_147c	1482289	1482424	15.79	60.33	3.47	26.53
rfam_151.1	1512962	1513010	8.32	75.75	7.01	30.36
rfam_175c.8	1566390	1566514	0	0	0	0
RydC	1597161	1597226	0.19	1.01	1.36	0.853333333
STnc490k	1613762	1613910	0	0	0	0
MicC	1613895	1614016	3.28	1.43	0.34	1.683333333
rfam_175c.9	1615277	1615400	0.3	0.03	0.14	0.156666667
C0343	1620403	1620477	0.49	0	0.42	0.303333333
STnc490k	1646853	1647001	0	0.67	0	0.223333333
rfam_175c.10	1685013	1685130	0.1	0	0	0.033333333
Trp_leader	1693907	1694002	0.07	0.69	0.16	0.306666667
rfam_175c.11	1722224	1722347	0.74	0.53	0.04	0.436666667
rfam_341c.4	1722300	1722385	0	0.14	0.43	0.19
RtT	1728063	1728204	7.68	7.21	3.58	6.156666667
RtT	1728244	1728384	7.34	1.14	0.39	2.956666667
RtT	1728435	1728566	1.76	0.64	2.1	1.5
STnc490k	1745868	1746016	0	0	0	0
SraB	1796277	1796444	1.47	9.24	0.23	3.646666667
rfam_107c.7	1797013	1797124	0.17	0.45	0.13	0.25
rne5	1798276	1798613	5.03	4.05	1.87	3.65
STnc490k	1802590	1802738	0.67	0	0	0.223333333
rfam_107c.8	1814278	1814376	0	0.15	0.67	0.273333333
rfam_21	1849205	1849367	24.17	0.22	350.62	125.0033333
STnc500	1854281	1854565	0.57	0.34	0.49	0.466666667
rfam_107c.9	1878383	1878511	0.3	0.34	1.22	0.62
isrI	1892814	1892912	0	0.29	0.82	0.37
STnc490k	1895477	1895625	0	0	0	0

STnc490k		1898689	1898837	0	0	0	0
rfam_163.1		1909619	1909754	431.53	4502.98	527.71	1820.74
	96	1914137	1914203	62.75	1743.6	3.48	603.2766667
rfam_166.1		1979573	1979690	284.02	2851.54	265.84	1133.8
rfam_107c.10		1985793	1985884	0.19	0.25	0.15	0.196666667
rfam_167.1		2011998	2012142	13.67	589.52	5.55	202.9133333
	58	2033798	2034068	30.32	133.8	59.94	74.68666667
rfam_172c		2083929	2084072	6.91	1.78	3.56	4.083333333
RybB		2084035	2084113	3.24	0.28	0	1.173333333
STnc490k		2117037	2117185	0	0	0	0
rfam_175c.12		2119800	2119910	0	0.38	0.18	0.186666667
rfam_107c.12		2130211	2130321	0	0	0	0
MOCO_RNA_motif		2145861	2146005	4.19	1.12	2.23	2.513333333
rfam_107c.13		2180122	2180226	0	0	0.27	0.09
rfam_107c.14		2186940	2187049	0	0.36	0	0.12
rfam_175c.13		2193776	2193899	0.05	0.21	0.24	0.166666667
	432	2204563	2204890	81.32	111.08	130.86	107.7533333
rfam_175c.14		2204596	2204720	62.53	36.47	46.28	48.42666667
rfam_341c.5		2204673	2204754	1.95	4.52	5.09	3.853333333
rfam_175c.15		2204829	2204950	0.95	0.41	2.74	1.366666667
rfam_107c.15		2204981	2205089	0	0.44	0	0.146666667
rfam_176c.1		2208270	2208547	92.22	85.31	371.53	183.02
	211	2208319	2208547	79.93	69.9	172.91	107.58
rfam_177.1		2208716	2209067	247.41	967.56	2734.25	1316.406667
rfam_178.1		2239201	2239294	384.53	27.45	255.98	222.6533333
rfam_107c.16		2264196	2264298	2.32	0.72	1.85	1.63
rfam_107c.17		2267003	2267104	0	0	0.12	0.04
rfam_107c.18		2271552	2271659	1.16	1.94	0	1.033333333
sroC		2275477	2275637	14.97	7.6	161.66	61.41
STnc490k		2323856	2324004	0	0	0	0
rfam_175c.16		2325847	2325970	0.87	0.11	0.46	0.48
rfam_341c.6		2325923	2326006	0	0	0	0
rfam_341c.7		2340724	2340803	0	0	0	0
rfam_175c.17		2340756	2340878	0	0	0	0
	375	2357968	2358067	193.41	327.02	130.45	216.96
rfam_175c.18		2422015	2422129	52.42	11.26	7.96	23.88
rfam_341c.8		2422082	2422161	14.62	11.42	18.33	14.79
rfam_107c.23		2423858	2423959	1.4	0.65	0.82	0.956666667
rfam_175c.19		2429940	2430060	0	0	0	0
rfam_341c.9		2430091	2430173	0.75	0	0	0.25
sroB		2431991	2432070	5.65	1.13	1.61	2.796666667
rfam_107c.24		2432673	2432783	2.28	0.79	0.33	1.133333333
rfam_107c.26		2442165	2442277	0.14	0.05	0	0.063333333
DnaX		2446019	2446083	2.52	0.32	0	0.946666667
SRP_bact		2463711	2463810	9.22	0	0	3.073333333
rfam_107c.27		2465974	2466066	0	0	0	0
STnc490k		2473717	2473865	0	0	0.43	0.143333333
rfam_107c.28		2477272	2477392	0.5	0.37	0	0.29
		2480850	2480970	111.06	40.44	2.39	51.29666667
rfam_175c.20		2484971	2485094	3.77	2.93	2.98	3.226666667
	106	2487291	2487546	34.66	520.27	12.04	188.99
rfam_107c.30		2502662	2502764	0.1	0.16	0	0.086666667
rfam_191.1		2511029	2511230	37.67	62.48	42.15	47.43333333

rfam_107c.31	2515289	2515406	0.89	0.12	0.66	0.556666667
STnc490k	2534836	2534984	0.67	0	0	0.223333333
5S_rRNA	2687483	2687598	0	0	0	0
PK-G12rRNA	2688215	2688322	0	0	0	0
isrL	2729595	2729939	0.49	0.33	0.09	0.303333333
tmRNA	2734487	2734848	244.91	194.71	387.59	275.7366667
rfam_42.1	2734714	2734850	9.57	2.03	1.41	4.336666667
isrK	2740919	2740996	1.59	0.61	0	0.733333333
rfam_42	2745336	2745590	1.12	1.07	2.74	1.643333333
rfam_341c.10	2766907	2766988	0	0.35	0	0.116666667
rfam_175c.24	2766941	2767064	0	0.16	0	0.053333333
rfam_107c.33	2781633	2781742	0.22	0	1.84	0.686666667
SraD	2805604	2805679	3.04	3.48	1.04	2.52
132	2809895	2810078	44.49	90.32	24.18	52.99666667
rfam_175c.25	2812777	2812881	1.14	5.71	5.85	4.233333333
rfam_175c.26	2820233	2820356	0	0.9	1.99	0.963333333
rfam_341c.11	2820309	2820390	5.11	1.43	15.5	7.346666667
rfam_107c.34	2826380	2826493	0	0	0	0
rfam_107c.35	2828741	2828849	0	0.31	0	0.103333333
rfam_218c.1	2848000	2848552	14.23	6.83	12.7	11.25333333
rfam_218c	2848000	2848783	23.52	15.7	20.49	19.90333333
rfam_219c	2849589	2849964	28.62	143.68	193.34	121.88
rfam_220c	2850855	2851204	16.8	414.03	18.43	149.7533333
rfam_221.1	2857461	2857722	79.59	80.8	191.84	117.41
InvR	2882780	2882870	5.25	3.89	9.81	6.316666667
rfam_175c.27	2927268	2927396	0.84	0.68	0.98	0.833333333
CsrB	2958127	2958482	596.82	156.13	1316.25	689.7333333
rfam_226	2958340	2958475	421.01	34.83	840.67	432.17
rfam_226.1	2958340	2958473	426.96	35.31	852.28	438.1833333
rfam_107c.36	2965835	2965946	0.05	0	0	0.016666667
rfam_107c.37	2971421	2971519	0.24	0.09	0	0.11
GcvB	2976750	2976955	6.44	17.42	3.37	9.076666667
rfam_107c.38	2985603	2985710	0	0.09	2.15	0.746666667
STnc490k	2999969	3000117	0	0	0.83	0.276666667
STnc490k	3012831	3012979	0	0	0	0
SraE_OmrA_OmrB	3012974	3013060	0.29	0.07	0	0.12
SraE_OmrA_OmrB	3013175	3013259	0.24	0.45	0.35	0.346666667
rfam_175c.28	3036726	3036846	0	0.71	0.23	0.313333333
rfam_341c.12	3036877	3036959	0.39	1.21	0.5	0.7
STnc290	3037052	3037134	0.18	0.31	0	0.163333333
isrO	3046397	3046597	0.54	0.09	0	0.21
rfam_107c.39	3055923	3056032	0.93	0.15	0.24	0.44
6S	3070118	3070301	10.62	23.33	17.69	17.21333333
QUAD	3070933	3071085	0.71	0.63	0.92	0.753333333
STnc490k	3078053	3078201	0	0	0	0
STnc490k	3085644	3085792	0	0	0	0
rfam_107c.40	3088852	3088963	0.06	0.13	0.23	0.14
STnc490k	3107876	3108024	0	0	0	0
rfam_175c.29	3122587	3122710	0	0.21	0.86	0.356666667
rfam_175c.30	3141805	3141929	1.19	1.21	2.53	1.643333333
FMN	3201608	3201806	15.06	2.92	2.48	6.82
QUAD	3205321	3205470	2.62	0.2	0	0.94
rfam_107c.42	3209886	3209993	0.51	0.61	1.98	1.033333333

rfam_107c.43		3213467	3213566	1.78	0.5	0.44	0.906666667
rfam_175c.31		3217366	3217488	0.72	0.62	0.24	0.526666667
yybP-ykoY		3235058	3235197	0.16	0.2	0	0.12
rfam_107c.44		3244895	3244993	0	0.24	0	0.08
RNaseP_bact_a		3257454	3257829	113.59	7.66	21.36	47.536666667
rfam_107c.45		3284802	3284908	0.17	0	0	0.056666667
rfam_107c.46		3286000	3286102	0.6	0.47	1.57	0.88
SraG		3292745	3292912	0	0.14	0.03	0.056666667
S15		3293143	3293258	20.2	49.77	3.44	24.47
rfam_175c.32		3294792	3294915	0.77	0.65	0.68	0.7
rfam_247c		3307380	3307692	25.04	10.1	7.51	14.216666667
rfam_247c.1		3307384	3307692	24.94	10.12	7.53	14.196666667
rfam_107c.47		3316573	3316679	0.19	3.45	0.27	1.303333333
rfam_107c.48		3328810	3328914	0.23	0.23	0.44	0.3
SraH		3330400	3330504	3.43	3.67	1.26	2.786666667
rfam_107c.49		3348024	3348126	2.51	0.01	14.87	5.796666667
	108	3351196	3351300	64.65	52.13	158.43	91.736666667
rfam_252c.1		3366708	3366887	1078.07	689.9	811.55	859.84
	4836	3366708	3366888	1072.11	686.09	807.06	855.0866667
rfam_107c.50		3373366	3373480	0.16	0.37	0	0.176666667
rfam_107c.51		3374897	3375001	2.11	2.59	0.19	1.63
rfam_175c.33		3403128	3403248	0.6	0	0	0.2
5S_rRNA		3403435	3403550	15.7	2.24	0.6	6.18
5S_rRNA		3403816	3403931	0	0	0	0
PK-G12rRNA		3404548	3404655	0	0	0	0
rfam_175c.34		3417036	3417159	0.04	0.19	0.24	0.156666667
rfam_175c.35		3424702	3424826	0.5	0	0.81	0.436666667
rfam_175c.36		3435426	3435548	0.03	0.2	0.29	0.173333333
rfam_107c.52		3447335	3447443	0.3	0.22	0.02	0.18
rfam_257c		3460420	3460582	52.76	28.8	27.52	36.36
GlmZ_SraJ		3464776	3464983	13.1	154.43	4.53	57.35333333
	185	3481711	3481948	46.72	202.41	487.48	245.5366667
rfam_175c.37		3483874	3483979	0.52	0.4	3.75	1.556666667
rfam_107c.53		3490286	3490396	0.31	0.3	1.62	0.743333333
rfam_175c.38		3497276	3497394	1.39	0.54	0.66	0.863333333
STnc490k		3536068	3536216	0	0	0	0
5S_rRNA		3538393	3538508	0	0	0	0
PK-G12rRNA		3539125	3539232	0	0	0	0
	399	3551487	3551589	141.58	431.31	60.12	211.0033333
TPP		3555495	3555720	2.25	0.3	1.05	1.2
rfam_175c.39		3555518	3555641	0.39	0.38	0.42	0.396666667
P26		3571216	3571277	1.52	0	0.03	0.516666667
rfam_265c.1		3572219	3572432	29.28	99.24	9.12	45.88
	107	3572219	3572546	21.23	67.06	6.25	31.51333333
rfam_175c.40		3572434	3572556	0.24	0.49	0.78	0.503333333
		3576263	3576387	66.23	393.03	25.64	161.6333333
5S_rRNA		3580447	3580562	0	0	0	0
PK-G12rRNA		3581180	3581287	0	0	0	0
Cobalamin		3588817	3589010	14.16	4.29	7.32	8.59
OxyS		3593876	3593987	1.29	0.97	0	0.753333333
rfam_107c.55		3595460	3595565	0	0	0	0
rfam_175c.41		3598816	3598924	2.03	1.23	2.7	1.986666667
rfam_107c.56		3603539	3603650	0	0.39	0.32	0.236666667

isrP	3624207	3624354	0	0.45	0.2	0.216666667
rfam_175c.43	3633616	3633739	11.65	4.16	15.52	10.44333333
rfam_107c.57	3638704	3638813	0.32	0	0.67	0.33
rfam_107c.58	3640825	3640929	1.94	6.08	1.54	3.186666667
rfam_107c.59	3657700	3657808	0.12	0	0.35	0.156666667
rfam_107c.60	3671047	3671151	0	0	0	0
rfam_107c.61	3675922	3676033	0	0.31	0.31	0.206666667
rfam_107c.62	3677502	3677609	0	0	0.23	0.076666667
rfam_175c.44	3715402	3715526	0	0.27	0.24	0.17
rfam_107c.63	3718349	3718451	2.02	1.31	0.44	1.256666667
CsrC	3720580	3720834	53.73	79.22	61.78	64.91
Spot_42	3721814	3721933	8.15	4.16	1.96	4.756666667
5S_rRNA	3729441	3729556	0	0	0	0
PK-G12rRNA	3730173	3730280	0	0	0	0
rfam_175c.45	3745139	3745262	0.27	0.36	0.53	0.386666667
rfam_279.1	3753696	3754101	86.66	1394.16	117.02	532.6133333
rfam_107c.64	3776368	3776481	0	0.37	0.78	0.383333333
rfam_107c.65	3779959	3780064	0.6	0.27	0.21	0.36
rfam_107c.66	3785966	3786069	0.05	0	0.63	0.226666667
istR	3836071	3836200	0	0.26	0	0.086666667
rfam_107c.67	3843022	3843131	0.22	0	0	0.073333333
rfam_107c.68	3861819	3861926	0	0	0	0
rfam_175c.46	3862890	3863013	0	0	0	0
rfam_107c.69	3890041	3890149	0.07	0	0.11	0.06
istR-2	3911672	3911803	1.36	2.32	1.56	1.746666667
rfam_175c.47	3931761	3931881	1.83	0.67	0	0.833333333
rfam_341c.13	3931912	3931996	0	0	0	0
rfam_290.1	3940547	3940691	9.55	33.73	45.71	29.66333333
rfam_175c.48	3956280	3956403	0	0	0	0
rfam_175c.49	3994068	3994191	0	0	0	0
rfam_341c.14	3994144	3994225	0	0	0	0
RtT	4007496	4007614	0.11	0.47	0	0.193333333
rfam_107c.70	4017516	4017619	0.23	0	0	0.076666667
rfam_107c.71	4061631	4061739	0.22	0	0.21	0.143333333
	4063962	4064100	61.17	75.22	173.59	103.3266667
rfam_107c.72	4078794	4078903	0.08	0.41	0.64	0.376666667
rfam_306.1	4085930	4085969	10.52	36.12	3.23	16.62333333
rfam_175c.50	4096143	4096268	0	0	0	0
rfam_107c.73	4103978	4104080	0.08	0.27	0.83	0.393333333
rfam_309.1	4114109	4114157	79.05	292.36	4.14	125.1833333
RyhB	4114856	4114920	0	0	0	0
rfam_107c.74	4120397	4120505	0	0.26	0.44	0.233333333
rfam_107c.75	4131661	4131763	0.51	0.22	0.19	0.306666667
rfam_107c.76	4135516	4135626	1.13	0.46	1.45	1.013333333
rfam_107c.77	4160755	4160863	0.26	0.3	0.06	0.206666667
rfam_107c.78	4190890	4190995	0	0.21	0	0.07
rfam_107c.79	4199308	4199413	1.23	1.2	1.56	1.33
rfam_175c.51	4204421	4204555	0.29	0.35	7.51	2.716666667
rfam_175c.52	4207837	4207961	1.29	3.63	0.29	1.736666667
Alpha_RBS	4228339	4228450	118.1	818.59	36.1	324.2633333
PK-G12rRNA	4246422	4246529	0	0	0	0
5S_rRNA	4247146	4247261	0	0	0	0
rfam_107c.80	4252316	4252412	3.41	0.72	2.33	2.153333333

rfam_107c.81	4255159	4255278	0	0.23	0.55	0.26
rfam_107c.82	4271566	4271676	0.26	0.94	0.2	0.466666667
rfam_107c.83	4278999	4279102	0	0.13	0	0.043333333
rfam_107c.84	4289660	4289784	0.03	0.22	0.31	0.186666667
rfam_107c.85	4298889	4299000	0.67	0.59	0	0.42
sraL	4334645	4334785	0.28	0.17	0	0.15
rfam_107c.86	4339061	4339158	0	0	0.01	0.003333333
rfam_175c.53	4343263	4343386	0	0	2.59	0.863333333
rfam_107c.87	4347310	4347414	0	0	0.29	0.096666667
rfam_107c.88	4351334	4351444	0	0	0	0
rfam_107c.89	4354103	4354211	0	0.51	0	0.17
rfam_107c.90	4355036	4355147	0	0	0.27	0.09
rfam_321c	4381431	4381470	2.77	13.95	1.05	5.923333333
STnc440	4388957	4389041	0.76	0	0	0.253333333
114	4390880	4390991	64.8	1509.64	348.07	640.8366667
rfam_328c.1	4391654	4391696	0	1.15	6.41	2.52
204	4507489	4507654	75.71	2937.07	0.18	1004.32
rfam_328c	4513001	4513043	41.69	32.25	16.19	30.043333333
isrK	4519423	4519501	0.15	0	0.38	0.176666667
rfam_175c.54	4536161	4536274	5.67	3.52	2.24	3.81
mini-ykkC	4540829	4540873	0	0	0.04	0.013333333
rfam_107c.91	4548601	4548713	0	0.04	0	0.013333333
rfam_175c.55	4556010	4556133	0	0.19	0	0.063333333
156	4565560	4565703	76.12	168.8	88.59	111.17
rfam_335.1	4565574	4565703	84.11	186.7	96.62	122.4766667
rfam_107c.92	4600744	4600840	0.08	0.82	0.47	0.456666667
STnc490k	4605467	4605615	0	0.67	0	0.223333333
rfam_107c.93	4620031	4620142	0.45	0.06	0.34	0.283333333
Mg_sensor	4638446	4638558	9.1	8.95	4.24	7.43
rfam_107c.94	4647252	4647358	0.7	0.64	0.62	0.653333333
rfam_107c.95	4655873	4655996	2.41	0.19	0.53	1.043333333
rfam_341c	4660949	4661030	53.27	90.44	6.77	50.16
rfam_175c.56	4660983	4661107	2.8	1.81	3.3	2.636666667
rfam_107c.96	4663515	4663619	0.16	0	0	0.053333333
isrK	4674069	4674146	0.15	0.54	0	0.23
isrQ	4709935	4710096	0	0.15	0	0.05
rfam_175c.57	4715805	4715925	0	0	0.67	0.223333333
rfam_107c.97	4727945	4728047	0	0.08	0.3	0.126666667
rfam_107c.98	4739069	4739187	0.08	0.33	1.94	0.783333333
rfam_107c.99	4761696	4761800	0.3	0	1.77	0.69
rfam_175c.58	4764420	4764544	2.46	3.05	7.74	4.416666667
rfam_107c.100	4765346	4765468	0.1	0.02	0.4	0.173333333
rfam_107c.101	4766034	4766139	0	0.2	0	0.066666667
rfam_107c.102	4770391	4770507	0.21	0.3	0.21	0.24

## 9.7 Differential expression pre-Benjamini-Hochberg

correction

Ratio	P.Value	adj.P.Val	Ty2	Gene	CT18	ProductString
2.914	1.66E-02	4.65E-01	t0011	yaaI	STY0011	hypothetical protein

0.454	3.41E-02	5.08E-01	t0030	NA	STY0034	hypothetical protein
2.383	1.85E-02	4.74E-01	t0065	citG	STY0072	CitG protein
0.433	3.33E-02	5.08E-01	t0067	carA	STY0076	carbamoyl phosphate synthase s
0.434	4.63E-03	3.25E-01	t0068	carB	STY0077	carbamoyl phosphate synthase la
2.568	2.50E-02	4.86E-01	t0070	caiE	STY0079	carnitine operon protein CaiE
2.044	7.59E-03	3.66E-01	t0083	NA	STY0095	probable secreted protein
0.302	8.82E-03	3.66E-01	t0090	folA	STY0102	dihydrofolate reductase
2.728	7.48E-03	3.66E-01	t0121	ilvH	STY0136	acetolactate synthase 3 regulator
0.145	1.48E-03	1.99E-01	t0210	htrA	STY0231	serine endoprotease
2.742	2.64E-02	4.86E-01	t0211	yaeG	STY0232	carbohydrate diacid transcription
2.413	4.58E-02	5.45E-01	t0248	metQ	STY0272	DL-methionine transporter substi
2.618	3.96E-02	5.23E-01	t2598	NA	STY0287	hypothetical protein
0.461	1.44E-02	4.49E-01	t2567	NA	STY0323	hypothetical protein
0.120	5.96E-04	1.14E-01	t2543	NA	STY0352	probable secreted protein
2.616	1.87E-02	4.74E-01	t2541	fadE	STY0354	acyl-CoA dehydrogenase
0.207	2.79E-04	6.71E-02	t2508	NA	STY0387	putative metabolite transport prc
0.473	8.91E-03	3.66E-01	t2490	NA	STY0406	putative DNA-binding transcriptic
0.496	4.07E-02	5.23E-01	t2458	tgt	STY0443	queueine tRNA-ribosyltransferase
2.106	3.03E-02	5.08E-01	t2452	yajD	STY0449	hypothetical protein
3.733	3.11E-02	5.08E-01	t2450	tsx	STY0451	nucleoside-specific channel-formi precursor
0.432	1.89E-02	4.74E-01	t2444	NA	STY0458	thiamine monophosphate kinase
2.426	1.12E-02	4.12E-01	t2437	phnV	STY0465	probable membrane component c aminoethylphosphonate transpor
2.934	4.15E-03	3.25E-01	t2436	phnU	STY0466	probable membrane component c aminoethylphosphonate transpor
3.493	4.75E-03	3.25E-01	t2425	NA	STY0477	hypothetical protein
2.714	6.80E-03	3.66E-01	t2401	NA	STY0501	putative lyase
0.258	1.94E-02	4.78E-01	t2391	rpmE2	STY0512	50S ribosomal protein L31 type E
2.534	3.32E-02	5.08E-01	t2380	ybaM	STY0524	hypothetical protein
4.647	6.85E-04	1.18E-01	t2350	sfbA	STY0558	lipoprotein
3.339	3.78E-02	5.16E-01	t2343	gcl	STY0565	glyoxylate carboligase
2.470	1.00E-02	3.90E-01	t2336	glxK	STY0573	glycerate kinase II
0.445	2.20E-02	4.79E-01	t2303	gtrA	STY0607	bactoprenol-linked glucose transl
5.088	7.42E-04	1.23E-01	t2268	cstA	STY0644	carbon starvation protein A
3.107	3.31E-03	3.18E-01	t2267	NA	STY0645	hypothetical protein
2.636	2.21E-03	2.41E-01	t2252	ybdR	STY0663	hypothetical zinc-dependant alco

0.305	5.52E-03	3.39E-01	t2251	rnk	STY0664	nucleoside diphosphate kinase re
0.127	5.69E-04	1.14E-01	t2235	tatE	STY0682	twin arginine translocase protein
0.486	1.25E-02	4.31E-01	t2230	dacA	STY0688	D-alanyl-D-alanine carboxypeptid
2.317	4.76E-02	5.48E-01	t2208	ybeJ	STY0710	glutamate and aspartate transpo
2.154	4.45E-02	5.45E-01	t2192	nagA	STY0721	N-acetylglucosamine-6-phosphat
4.895	1.11E-03	1.64E-01	t2188	ybfM	STY0725	putative outer membrane proteir
0.324	4.18E-02	5.33E-01	t2131	tolQ	STY0791	colicin uptake protein TolQ
0.344	1.41E-02	4.45E-01	t2107	ybhT	STY0813	hypothetical protein
3.596	2.17E-03	2.41E-01	t2106	modA	STY0814	molybdate transporter periplasm
2.529	2.48E-02	4.86E-01	t2105	modB	STY0815	molybdate ABC transporter perm
0.451	1.02E-02	3.93E-01	t2101	ybhC	STY0819	possible pectinesterase precursor
0.255	1.23E-02	4.31E-01	t2095	ybhB	STY0825	predicted kinase inhibitor
2.400	3.08E-02	5.08E-01	t2077	ybhN	STY0846	hypothetical protein
2.102	2.94E-02	5.08E-01	t2063	ybiI	STY0862	hypothetical protein
2.321	1.48E-02	4.49E-01	t2058	glnH	STY0868	glutamine ABC transporter peripl
0.443	3.57E-03	3.25E-01	t2055	ompX	STY0872	outer membrane protein X
4.887	9.09E-03	3.67E-01	t2036	bssR	STY0893	biofilm formation regulatory prot
2.862	1.91E-02	4.74E-01	t2016	potI	STY0913	putrescine transporter subunit:
0.397	2.29E-02	4.79E-01	t1959	NA	STY0975	membrane component of ABC su
2.082	3.85E-02	5.18E-01	t1957	serC	STY0977	phosphoserine aminotransferase
0.463	2.38E-02	4.79E-01	t1926	NA	STY1014	putative DNA methylase
0.488	2.38E-02	4.79E-01	t1924	exo	STY1016	exonuclease
0.454	3.63E-02	5.16E-01	t1923	betA	STY1017	bacteriophage recombination pro
0.386	8.32E-03	3.66E-01	t1916	NA	STY1024	putative DNA-binding protein
3.712	2.35E-02	4.79E-01	t1912	NA	STY1028	putative bacteriophage protein
0.053	2.16E-04	5.83E-02	t1800	NA	STY1156	hypothetical protein
10.025	4.72E-02	5.48E-01	t1792	NA	STY1165	putative membrane transporter
10.372	3.12E-02	5.08E-01	t1790	NA	STY1167	hypothetical protein
21.253	2.23E-02	4.79E-01	t1789	NA	STY1168	putative secreted protein
12.265	1.99E-02	4.79E-01	t1788	NA	STY1169	putative sialic acid transporter
2.410	3.38E-02	5.08E-01	t1773	NA	STY1184	hypothetical protein
2.106	1.39E-02	4.45E-01	t1767	msyB	STY1190	hypothetical protein
2.006	4.75E-02	5.48E-01	t1766	yceE	STY1191	drug efflux system protein MdtG
2.104	1.15E-02	4.20E-01	t1753	rimJ	STY1205	ribosomal-protein-S5-alanine N-
2.904	3.56E-02	5.16E-01	t1749	flgN	STY1210	flagella synthesis protein FlgN
2.560	4.78E-03	3.25E-01	t1748	flgM	STY1211	anti-sigma28 factor FlgM

2.081	2.75E-02	4.93E-01	t1740	flgH	STY1219	flagellar basal body L-ring protein
2.499	4.36E-02	5.40E-01	t1739	flgI	STY1220	flagellar basal body P-ring protein
2.261	3.08E-02	5.08E-01	t1738	flgJ	STY1221	peptidoglycan hydrolase
0.450	8.83E-03	3.66E-01	t1729	rpmF	STY1230	50S ribosomal protein L32
2.927	4.42E-03	3.25E-01	t1708	NA	STY1252	hypothetical protein
0.270	1.37E-02	4.45E-01	t1697	potC	STY1263	spermidine/putrescine ABC transport protein
0.231	4.50E-03	3.25E-01	t1643	NA	STY1320	hypothetical protein
0.053	1.14E-03	1.64E-01	t1642	NA	STY1321	hypothetical protein
0.095	1.98E-05	8.57E-03	t1641	NA	STY1322	hypothetical protein
0.249	1.68E-02	4.65E-01	t1640	NA	STY1323	hypothetical protein
0.378	7.40E-03	3.66E-01	t1620	pyrF	STY1344	orotidine 5-prime-phosphate decarboxylase
2.900	7.62E-03	3.66E-01	t1581	NA	STY1386	putative transcriptional regulator
2.225	1.10E-02	4.12E-01	t1568	NA	STY1400	conserved hypothetical DNA-binding protein
0.467	2.16E-02	4.79E-01	t1566	NA	STY1402	hypothetical protein
2.034	2.83E-02	4.99E-01	t1562	NA	STY1406	hypothetical protein
2.506	1.67E-03	2.12E-01	t1561	NA	STY1408	putative chemo-receptor protein
0.360	1.17E-02	4.25E-01	t1553	NA	STY1419	probable pyruvate-flavodoxin oxidoreductase
0.463	3.91E-02	5.22E-01	t1535	NA	STY1439	hypothetical protein
0.382	4.89E-03	3.25E-01	t1489	narY	STY1487	respiratory nitrate reductase 2 beta subunit
0.419	7.34E-03	3.66E-01	t1488	narZ	STY1488	respiratory nitrate reductase 2 alpha subunit
0.259	5.57E-03	3.39E-01	t1482	adhP	STY1493	alcohol dehydrogenase
2.902	3.08E-02	5.08E-01	t1474	NA	STY1502	putative secreted protein
2.817	3.15E-02	5.08E-01	t1470	NA	STY1507	putative aminotransferase
0.035	1.33E-05	6.39E-03	t1459	NA	STY1522	putative secreted hydrolase
0.031	5.20E-06	4.24E-03	t1458	hyaA	STY1523	uptake hydrogenase small subunit
0.199	5.60E-04	1.14E-01	t1451	NA	STY1531	putative ATP/GTP-binding protein
2.313	2.95E-02	5.08E-01	t1430	NA	STY1552	hypothetical protein
2.574	4.88E-03	3.25E-01	t1414	NA	STY1571	putative ABC transporter periplasmic domain
2.623	3.77E-02	5.16E-01	t1413	NA	STY1572	putative ABC transporter membrane domain
3.611	1.65E-02	4.65E-01	t1402	NA	STY1585	multidrug efflux system protein M
0.302	8.00E-03	3.66E-01	t1397	NA	STY1591	putative bacteriophage transcript
2.819	1.11E-02	4.12E-01	t1366	NA	STY1622	conserved bacteriophage hypothetical protein
0.239	3.66E-04	8.32E-02	t1364	NA	STY1624	hypothetical protein
0.336	9.10E-03	3.67E-01	t1360	NA	STY1628	hypothetical protein
0.127	1.37E-03	1.91E-01	t1341	NA	STY1649	outer membrane protein

0.246	1.50E-02	4.49E-01	t1329	NA	STY1661	oriC-binding nucleoid-associated
0.251	1.62E-02	4.62E-01	t1298	purR	STY1692	DNA-binding transcriptional repressor
2.072	8.50E-03	3.66E-01	t1281	ssaM	STY1707	putative pathogenicity island protein
0.270	4.86E-02	5.50E-01	t1276	ssaI	STY1712	putative pathogenicity island protein
2.387	2.35E-02	4.79E-01	t1267	sscA	STY1721	putative type III secretion system
0.100	1.60E-03	2.09E-01	t1260	ssrA	STY1728	putative two-component sensor histidine kinase
0.319	3.94E-02	5.23E-01	t1259	ssrB	STY1729	putative two-component response regulator
3.404	3.04E-02	5.08E-01	t1256	ydhZ	STY1732	hypothetical protein
2.515	2.11E-02	4.79E-01	t1255	ttrR	STY1733	putative two-component response regulator
2.954	3.85E-02	5.18E-01	t1254	ttrS	STY1735	hypothetical protein
0.222	1.10E-02	4.12E-01	t1245	lppA	STY1745	major outer membrane lipoprotein
2.588	4.49E-02	5.45E-01	t1237	sufA	STY1754	iron-sulfur cluster assembly scaffold
0.256	3.34E-02	5.08E-01	t1236	NA	STY1755	putative transporter
2.267	4.66E-02	5.48E-01	t1195	celF	STY1797	phospho-beta-glucosidase B
2.954	2.48E-02	4.86E-01	t1190	osmE	STY1802	DNA-binding transcriptional activator
3.228	1.90E-02	4.74E-01	t1149	NA	STY1850	hypothetical protein
4.817	2.68E-02	4.86E-01	t1141	NA	STY1858	hypothetical protein
0.072	3.84E-02	5.18E-01	t1132	NA	STY1867	putative lipoprotein
0.312	4.42E-02	5.43E-01	t1116	pagD	STY1880A	putative outer membrane virulence factor
0.299	4.39E-02	5.40E-01	t1107	NA	STY1891	putative pertussis-like toxin subunit
3.098	8.41E-03	3.66E-01	t1094	NA	STY1908	predicted inner membrane protein
2.605	2.61E-02	4.86E-01	t1080	NA	STY1925	hypothetical protein
2.114	4.52E-03	3.25E-01	t1076	NA	STY1929	potassium/proton antiporter
2.026	4.34E-02	5.40E-01	t1073	NA	STY1933	hypothetical protein
2.267	1.11E-02	4.12E-01	t1050	NA	STY1957	hypothetical protein
0.362	2.33E-02	4.79E-01	t1045	NA	STY1962	hypothetical protein
0.430	2.92E-02	5.08E-01	t1043	rrmA	STY1964	23S rRNA methyltransferase A
0.486	2.42E-02	4.81E-01	t1042	ftsI	STY1965	penicillin-binding protein
3.573	2.65E-02	4.86E-01	t1037	NA	STY1971	hypothetical protein
2.325	2.36E-02	4.79E-01	t1034	htpX	STY1975	heat shock protein HtpX
0.079	6.78E-03	3.66E-01	t1022	NA	STY1988	hypothetical protein
2.052	1.41E-02	4.45E-01	t1012	yjcS	STY2004	putative hydrolase
2.178	4.38E-02	5.40E-01	t1011	NA	STY2005	hypothetical protein
0.249	1.77E-02	4.74E-01	t1867	NA	STY2013	putative bacteriophage tail protein
0.308	5.46E-03	3.39E-01	t1870	NA	STY2016	putative bacteriophage protein
0.385	1.52E-02	4.49E-01	t1884	NA	STY2030	putative bacteriophage protein
0.343	2.26E-02	4.79E-01	t1892	NA	STY2038	hypothetical protein

0.492	3.41E-02	5.08E-01	t0986	znuA	STY2099	high-affinity zinc transporter peri
0.144	9.56E-03	3.82E-01	t0975	yecD	STY2110	hypothetical protein
3.977	3.11E-03	3.12E-01	t0965	flhE	STY2121	flagellar protein FlhE precursor
2.203	3.67E-02	5.16E-01	t0961	cheY	STY2125	chemotaxis regulator transmitting motor component
2.089	2.05E-02	4.79E-01	t0960	cheB	STY2126	chemotaxis-specific methyl-accepting
2.271	1.34E-02	4.42E-01	t0955	motB	STY2131	flagellar motor protein MotB
3.279	2.77E-03	2.85E-01	t0922	fliY	STY2162	cystine transporter subunit
2.433	1.30E-02	4.40E-01	t0913	yedD	STY2172	hypothetical protein
3.472	3.41E-02	5.08E-01	t0909	fliE	STY2176	flagellar hook-basal body protein
2.034	2.28E-02	4.79E-01	t0896	fliR	STY2189	flagellar biosynthesis protein FliR
5.394	1.53E-02	4.49E-01	t0895	rcaA	STY2190	colanic acid capsular biosynthesis
0.490	3.95E-02	5.23E-01	t0890	NA	STY2195	hypothetical protein
0.303	7.88E-03	3.66E-01	t0883	ompS	STY2203	outer membrane protein S1
2.043	2.14E-02	4.79E-01	t0837	pduF	STY2242	propanediol diffusion facilitator
2.732	1.74E-02	4.70E-01	t0828	pduK	STY2251	putative propanediol utilization p
2.070	3.60E-02	5.16E-01	t0811	phsC	STY2269	thiosulfate reductase cytochrome
2.223	2.82E-02	4.99E-01	t0810	phsB	STY2270	thiosulfate reductase electron tra
0.306	2.29E-02	4.79E-01	t0803	hisL	STY2279A	his operon leader peptide
2.282	1.40E-02	4.45E-01	t0763	gmd	STY2321	GDP-mannose 4,6-dehydratase
2.105	2.69E-02	4.86E-01	t0721	NA	STY2364	hypothetical protein
0.319	1.30E-02	4.40E-01	t0707	stcD	STY2378	hypothetical protein
0.246	4.35E-02	5.40E-01	t0704	stcA	STY2381	putative fimbrial subunit protein
3.165	4.57E-03	3.25E-01	t0692	yehY	STY2394	putative permease transmembrane
0.357	2.31E-02	4.79E-01	t0682	NA	STY2403	hypothetical protein
3.944	3.09E-02	5.08E-01	t0672	cdd	STY2413	cytidine deaminase
2.856	2.19E-02	4.79E-01	t0667	mgIC	STY2421	beta-methylgalactoside transport
0.298	1.91E-02	4.74E-01	t0659	sdaC	STY2430	putative L-serine dehydratase
2.355	3.59E-02	5.16E-01	t0656	cirA	STY2434	ferric iron-catecholate outer men
0.077	6.34E-03	3.65E-01	t0646	NA	STY2445	hypothetical protein
0.278	7.42E-03	3.66E-01	t0620	NA	STY2470	homolog of virulence protein msc
0.371	3.63E-02	5.16E-01	t0618	narP	STY2472	DNA-binding response regulator

0.036	9.97E-04	1.59E-01	t0597	ompC	STY2493	outer membrane porin protein C
0.429	2.90E-02	5.07E-01	t0575	NA	STY2518	hypothetical protein
0.185	2.80E-04	6.71E-02	t0569	NA	STY2524	hypothetical protein
3.172	9.89E-03	3.88E-01	t0523	NA	STY2571	putative transketolase N-termina
2.562	2.25E-02	4.79E-01	t0450	NA	STY2648	hypothetical protein
0.437	4.56E-02	5.45E-01	t0420	NA	STY2675	glutamine amidotransferase
0.469	3.68E-02	5.16E-01	t0408	amiA	STY2687	N-acetylmuramoyl-L-alanine amic
0.472	3.03E-02	5.08E-01	t0407	hemF	STY2688	coproporphyrinogen III oxidase
2.844	2.03E-03	2.41E-01	t0391	eutQ	STY2704	putative ethanolamine utilization
2.378	1.25E-02	4.31E-01	t0375	NA	STY2722	hypothetical protein
3.041	6.74E-03	3.66E-01	t0364	NA	STY2734	hypothetical protein
0.284	2.09E-02	4.79E-01	t0346	guaB	STY2752	inositol-5-monophosphate dehyd
0.311	3.33E-02	5.08E-01	t0310	suhB	STY2792	inositol monophosphatase
0.470	4.93E-02	5.50E-01	t0274	lepA	STY2829	GTP-binding protein LepA
3.023	4.36E-02	5.40E-01	t2734	srIE	STY2954	glucitol/sorbitol-specific IIBC con system
2.215	4.86E-02	5.50E-01	t2748	hych	STY2968	formate hydrogenlyase maturatic
2.650	4.73E-02	5.48E-01	t2749	hycG	STY2969	formate hydrogenlyase subunit 7
0.169	1.72E-02	4.66E-01	t2809	NA	STY3032	hypothetical protein
0.447	4.18E-02	5.33E-01	t2833	ygbE	STY3057	hypothetical protein
2.437	2.73E-02	4.91E-01	t2836	cysD	STY3060	sulfate adenylyltransferase subur
2.507	2.64E-02	4.86E-01	t2851	NA	STY3078	hypothetical protein
2.829	6.30E-03	3.65E-01	t2929	yqeF	STY3164	acetyl-CoA acetyltransferase
0.220	3.46E-02	5.14E-01	t2940	stdA	STY3177	probable fimbrial protein
2.605	4.91E-02	5.50E-01	t2994	NA	STY3233	possible ABC-transport protein, A component
0.492	6.35E-03	3.65E-01	t3014	NA	STY3255	hypothetical protein
2.425	4.81E-03	3.25E-01	t3026	nupG	STY3268	nucleoside permease
0.271	2.32E-02	4.79E-01	t3033	NA	STY3278	hypothetical protein
3.223	2.54E-02	4.86E-01	t3074	NA	STY3326	hypothetical protein
2.110	4.11E-02	5.26E-01	t3142	ygjQ	STY3402	hypothetical protein
2.232	7.90E-03	3.66E-01	t3147	NA	STY3407	hypothetical protein
2.030	8.83E-03	3.66E-01	t3151	NA	STY3411	hypothetical protein
0.177	5.05E-03	3.25E-01	t3194	NA	STY3457	putative protease
0.316	2.50E-02	4.86E-01	t3201	rpsO	STY3464	30S ribosomal protein S15
0.433	3.20E-02	5.08E-01	t3205	nusA	STY3468	transcription elongation factor Nu
0.320	3.90E-03	3.25E-01	t3206	NA	STY3469	hypothetical protein
0.490	4.57E-02	5.45E-01	t3212	folP	STY3473	dihydropteroate synthase
0.443	3.31E-02	5.08E-01	t3219	NA	STY3481	hypothetical protein
0.360	2.15E-03	2.41E-01	t3221	rpIU	STY3483	50S ribosomal protein L21
0.344	6.92E-03	3.66E-01	t3261	rpIM	STY3525	50S ribosomal protein L13
0.189	1.62E-02	4.62E-01	t3276	NA	STY3542	hypothetical protein

2.087	3.73E-02	5.16E-01	t3323	tatB	STY3585	sec-independent translocase
3.794	7.47E-03	3.66E-01	t3329	udp	STY3591	uridine phosphorylase
2.059	1.87E-02	4.74E-01	t3342	NA	STY3604	chloramphenicol-sensitive protein
0.152	3.22E-02	5.08E-01	t3356	NA	STY3618	hypothetical protein
0.436	4.99E-02	5.50E-01	t3404	cII	STY3662	regulatory protein cII
0.271	3.38E-02	5.08E-01	t3408	NA	STY3666	hypothetical protein
0.396	2.03E-02	4.79E-01	t3455	purH	STY3709	bifunctional phosphoribosylaminoimidazole
0.433	6.53E-03	3.66E-01	t3478	rplK	STY3736	formyltransferase/IMP cyclohydrolase
0.348	1.19E-02	4.26E-01	t3481	tuf	STY3739	50S ribosomal protein L11
2.093	3.62E-02	5.16E-01	t3498	yijC	STY3747	elongation factor Tu
3.458	8.76E-03	3.66E-01	t3538	NA	STY3790	DNA-binding transcriptional repressor
3.464	4.79E-02	5.50E-01	t3543	NA	STY3795	ribulose-phosphate 3-epimerase
7.367	1.86E-02	4.74E-01	t3544	NA	STY3796	putative ABC transporter permease
2.708	1.61E-02	4.62E-01	t3559	cpxP	STY3811	putative ABC transporter ATP-binding protein
2.573	7.31E-03	3.66E-01	t3585	fdoI	STY3842	periplasmic repressor CpxP
3.025	3.72E-02	5.16E-01	t3619	NA	STY3879	formate dehydrogenase-O subunit
0.200	4.49E-03	3.25E-01	t3638	rbsD	STY3897	hypothetical protein
0.367	2.69E-02	4.86E-01	t3658	glmS	STY3917	high affinity ribose transport protein
0.393	2.05E-02	4.79E-01	t3660	stgB	STY3919	D-fructose-6-phosphate amidotransferase
0.306	2.24E-03	2.41E-01	t3664	aroE	STY3924	fimbrial chaperone protein
0.361	2.16E-02	4.79E-01	t3678	yidC	STY3938	shikimate 5-dehydrogenase
0.361	2.77E-02	4.93E-01	t3680	rpmH	STY3939A	putative inner membrane protein component YidC
2.273	8.27E-03	3.66E-01	t3691	NA	STY3950	50S ribosomal protein L34
2.100	2.96E-02	5.08E-01	t3715	dsdA	STY3977	hypothetical protein
2.195	3.21E-02	5.08E-01	t3728	uhpA	STY3992	D-serine dehydratase
2.054	3.83E-02	5.18E-01	t3750	NA	STY4017	DNA-binding response regulator
0.081	1.25E-04	3.60E-02	t3757	slsA	STY4025	regulatory system with UhpB
0.485	1.97E-02	4.79E-01	t3784	rph	STY4060	putative transferase
0.462	4.07E-02	5.23E-01	t3806	waaL	STY4082	hypothetical protein
0.283	3.88E-02	5.19E-01	t3818	secB	STY4094	ribonuclease PH
2.715	1.63E-02	4.62E-01	t3841	NA	STY4118	O-antigen ligase
2.589	2.17E-02	4.79E-01	t3842	sgbE	STY4119	preprotein translocase subunit SecY
2.252	7.75E-03	3.66E-01	t3885	dppA	STY4168	putative transcriptional regulator
0.276	8.50E-03	3.66E-01	t3912	NA	STY4198	L-ribulose-5-phosphate 4-epimerase
						periplasmic dipeptide transport protein
						phage-like lysozyme

2.583	2.75E-03	2.85E-01	t3968	ugpQ	STY4258	cytoplasmic glycerophosphodiester
0.286	4.33E-02	5.40E-01	t3973	NA	STY4263	hypothetical protein
0.114	6.34E-04	1.14E-01	t4004	ompR	STY4294	osmolarity response regulator
0.056	2.47E-05	8.87E-03	t4005	envZ	STY4295	osmolarity sensor protein
5.758	5.66E-03	3.39E-01	t4017	yrfA	STY4307	hypothetical protein
0.346	1.34E-02	4.42E-01	t4033	ppiA	STY4324	peptidyl-prolyl cis-trans isomerase
0.151	4.56E-03	3.25E-01	t4057	rpsL	STY4350	30S ribosomal protein S12
0.437	1.23E-02	4.31E-01	t4059	fusA	STY4352	elongation factor G
0.280	3.78E-03	3.25E-01	t4068	rplB	STY4361	50S ribosomal protein L2
0.333	2.40E-02	4.79E-01	t4075	rplN	STY4368	50S ribosomal protein L14
0.392	4.83E-03	3.25E-01	t4076	rplX	STY4369	50S ribosomal protein L24
0.441	1.88E-02	4.74E-01	t4077	rplE	STY4370	50S ribosomal protein L5
0.339	4.58E-02	5.45E-01	t4080	rplF	STY4373	50S ribosomal protein L6
0.423	1.48E-02	4.49E-01	t4081	rplR	STY4374	50S ribosomal protein L18
0.391	1.52E-02	4.49E-01	t4085	secY	STY4378	preprotein translocase subunit SecY
0.480	3.37E-02	5.08E-01	t4088	rpsK	STY4381	30S ribosomal protein S11
2.476	4.24E-03	3.25E-01	t4100	smg	STY4393	hypothetical protein
3.748	6.23E-04	1.14E-01	t4111	aceB	STY4401	malate synthase
2.627	3.29E-02	5.08E-01	t4112	aceA	STY4402	isocitrate lyase
3.290	1.03E-03	1.59E-01	t4113	aceK	STY4403	bifunctional isocitrate dehydrogenase/kinase/phosphatase protein
2.495	4.95E-02	5.50E-01	t4114	iclR	STY4404	acetate operon transcriptional regulator
2.110	4.75E-02	5.48E-01	t4126	lysC	STY4416	aspartate kinase III
0.473	1.19E-02	4.26E-01	t4128	NA	STY4418	hypothetical protein
2.953	6.42E-03	3.65E-01	t4180	NA	STY4472	hypothetical protein
0.446	1.33E-02	4.42E-01	t4200	NA	STY4492	predicted metal dependent hydrolyase
2.413	3.14E-02	5.08E-01	t4201	NA	STY4493	arginine:agmatin antiporter
0.435	3.33E-02	5.08E-01	t4267	NA	STY4571	putative lipoprotein
2.199	1.29E-02	4.40E-01	t4271	NA	STY4574	hypothetical protein
2.136	3.79E-02	5.16E-01	t4275	NA	STY4578	hypothetical protein
2.286	4.99E-03	3.25E-01	t4289	NA	STY4595	hypothetical protein
0.230	4.67E-03	3.25E-01	t4298	NA	STY4604	hypothetical protein
2.386	2.29E-02	4.79E-01	t4308	NA	STY4614	phage baseplate assembly protein
4.234	4.93E-02	5.50E-01	t4326	NA	STY4632	hypothetical protein
0.020	4.05E-06	4.24E-03	t4344	vexE	STY4651	Vi polysaccharide export protein
0.022	2.23E-05	8.77E-03	t4345	vexD	STY4652	Vi polysaccharide export inner-membrane protein
0.008	6.11E-05	2.03E-02	t4346	vexC	STY4653	Vi polysaccharide export ATP-binding protein
0.008	8.07E-06	4.66E-03	t4347	vexB	STY4654	Vi polysaccharide export inner-membrane protein
0.019	5.38E-06	4.24E-03	t4348	vexA	STY4655	Vi polysaccharide export protein

0.019	5.89E-06	4.24E-03	t4349	tviE	STY4656	Vi polysaccharide biosynthesis pr
0.007	2.12E-06	4.24E-03	t4350	tviD	STY4659	Vi polysaccharide biosynthesis pr
0.003	5.10E-06	4.24E-03	t4351	tviC	STY4660	Vi polysaccharide biosynthesis pr
0.004	8.63E-06	4.66E-03	t4352	tviB	STY4661	Vi polysaccharide biosynthesis pr glucose/GDP-mannose dehydrog
0.004	1.02E-04	3.15E-02	t4353	tviA	STY4662	Vi polysaccharide biosynthesis pr
0.203	2.32E-02	4.79E-01	t4354	NA	STY4663	hypothetical protein
0.439	8.40E-03	3.66E-01	t4357	NA	STY4666	probable phage integrase
0.214	2.11E-03	2.41E-01	t4397	yjeM	STY4705	putative amino acid permease
0.260	4.97E-02	5.50E-01	t4398	yjeN	STY4706	hypothetical protein
0.277	5.60E-03	3.39E-01	t4445	rplI	STY4750	50S ribosomal protein L9
2.047	3.38E-02	5.08E-01	t4466	NA	STY4771	putative sugar transporter
3.688	3.23E-03	3.17E-01	t4475	NA	STY4780	hypothetical protein
0.492	1.69E-02	4.65E-01	t4479	NA	STY4784	dihydroorotase
0.482	2.62E-02	4.86E-01	t4526	cI	STY4829	phage immunity repressor protei
2.200	4.75E-02	5.48E-01	t4593	NA	STY4903	ferric iron reductase involved in f transport
0.353	1.67E-02	4.65E-01	t4604	osmY	STY4911	periplasmic protein
2.635	4.06E-02	5.23E-01	t4611	deoA	STY4919	thymidine phosphorylase
0.479	2.65E-02	4.86E-01	t4624	gpmB	STY4932	phosphoglycerate mutase
0.260	4.65E-03	3.25E-01	t2653	NA		hypothetical protein

## 9.8 Microarray Differences

Gene ID	Product	OD(600)=0.6		Experiment OD(600)=1.1		SPI-2 Inducing	
		Ratio	p-value	Ratio	p-value	Ratio	p-value
STY0012	dnaK	2.1	1.3E-11	1.0	8.9E-01	2.0	7.4E-05
STY0039	-	0.3	1.3E-04	1.0	6.9E-01		
STY0078	caiF	0.4	2.7E-04	0.7	1.0E-02		
STY0181	acnB	2.7	4.1E-09	1.3	8.2E-04	1.4	9.8E-05
STY0232	yaeG	0.3	2.6E-04	0.9	6.4E-01		
STY0307	-	2.4	4.8E-03	1.4	2.4E-01		
STY0399	prpR	3.1	8.3E-03	1.1	7.9E-01		
STY0400	prpB	8.6	6.6E-04	1.5	2.9E-02		
STY0401	prpC	7.6	1.1E-07	6.9	2.8E-01		
STY0402	prpD	10.0	1.0E-05	1.4	1.8E-01		
STY0403	prpE	4.7	2.3E-07	1.4	1.5E-04		
STY0481	cyoE	10.3	4.2E-08	2.8	1.3E-04	2.2	1.4E-04
STY0482	cyoD	8.0	1.3E-10	2.9	3.6E-12	2.0	1.9E-04
STY0483	cyoC	5.6	5.8E-08	1.6	2.2E-05	2.0	1.9E-05
STY0484	cyoB	7.9	1.1E-12	2.4	6.0E-08	2.2	2.0E-09
STY0485	cyoA	12.8	4.1E-16	2.6	5.5E-08	2.0	1.6E-05
STY0562	-	2.6	1.1E-04	1.0	7.1E-01		
STY0563	allA	4.0	5.9E-07	1.5	1.6E-06	1.0	8.8E-01
STY0564	allR	2.0	2.6E-07	1.1	2.3E-01	1.1	3.2E-01
STY0566	hyi	72.3	1.5E-03	1.6	1.4E-05		
STY0567	garR	74.3	6.1E-03	1.0	9.4E-01		
STY0568	-	10.5	3.4E-05	1.9	no replicates		
STY0569	ybbW	21.3	3.3E-02				

STY0571	allB	26.5	6.3E-03	1.0	8.9E-01		
STY0644	cstA	2.7	1.2E-08	1.6	1.5E-05	1.3	4.9E-02
STY0670	citF	2.2	9.3E-09	1.3	6.3E-03	1.0	9.9E-01
STY0707	gltL	4.7	8.1E-03				
STY0708	gltK	124.9	9.3E-03	1.9	no replicates		
STY0710	ybeJ	10.7	9.3E-07	1.5	9.7E-05	1.3	6.8E-02
STY0725	ybfM	3.1	2.0E-09	2.0	1.9E-08		
STY0773	gltA	3.2	7.7E-11	1.3	5.0E-03	1.9	8.3E-05
STY0775	sdhC	4.2	1.9E-14	1.5	1.0E-04	1.7	1.4E-02
STY0776	sdhD	4.1	1.4E-12	1.5	1.3E-05	1.7	4.0E-03
STY0777	sdhA	3.1	8.1E-08	1.7	6.8E-04	1.7	6.5E-03
STY0779	sucA	2.2	1.2E-10	1.5	1.4E-06	1.4	2.9E-02
STY0780	sucB	2.2	2.4E-09	1.5	3.5E-05	1.5	1.1E-02
STY0781	sucC	2.2	3.9E-08	1.6	2.3E-06	1.8	1.9E-04
STY0787	cydB	0.5	8.7E-04	1.0	8.8E-01	1.5	1.9E-03
STY0814	modA	2.1	2.1E-06	2.0	4.3E-06	1.3	6.5E-03
STY0815	modB	2.1	2.0E-08	1.5	6.8E-06		
STY0823	hutU	3.0	1.7E-05	1.1	3.5E-01		
STY0824	hutH	2.5	2.9E-10	1.2	1.0E-01		
STY0866	glnQ	3.4	1.6E-09	1.6	4.2E-05		
STY0868	glnH	4.3	2.5E-13	1.7	2.3E-05		
STY0880	-	3.7	5.3E-08	1.0	no replicates		
STY0881	-	2.0	3.9E-04	1.1	4.8E-01	1.7	4.2E-03
STY0962	dmsA	0.4	3.1E-04	0.8	4.7E-02		
STY1002	ompF	13.2	3.5E-19	6.3	8.2E-11	1.3	2.5E-01
STY1162	phoH	2.2	3.8E-08	1.1	4.5E-02	1.6	2.8E-03
STY1165	-	2.6	5.8E-03	0.9	no replicates		
STY1166	-	2.5	5.7E-03	1.2	1.8E-01	1.0	9.8E-01
STY1167	-	203.1	1.4E-03	1.6	4.6E-01		
STY1168	-	7.4	4.9E-07	1.2	1.4E-01		
STY1169	-	3.9	1.3E-03	1.1	2.4E-01		
STY1170	-	5.3	8.3E-06	1.1	2.2E-01	1.6	1.2E-02
STY1210	flgN	2.2	5.7E-07	2.3	1.4E-07	1.6	1.4E-02
STY1211	flgM	2.2	1.8E-07	2.5	9.9E-08	2.0	4.0E-05
STY1213	flgB	1.2	4.8E-02	1.7	3.0E-06	2.6	1.6E-04
STY1218	flgG	1.1	1.4E-01	2.0	1.0E-08	3.0	6.3E-03
STY1222	flgK	2.1	1.1E-07	2.6	4.2E-11	1.8	5.3E-03
STY1252	-	2.7	6.1E-09	2.9	4.7E-10	1.1	2.6E-01
STY1287	narK	4.2	5.3E-10	0.9	6.3E-01		
STY1288	narG	2.7	5.9E-07	0.9	5.8E-01	0.8	1.7E-01
STY1289	narH	7.1	4.6E-08	1.0	8.4E-01	1.2	2.6E-01
STY1290	narJ	10.2	1.2E-09	1.0	9.0E-01		
STY1291	narI	7.0	4.2E-08	0.9	4.6E-01	1.6	4.8E-02
STY1304	oppA	4.4	1.1E-09	1.7	2.1E-06	1.3	3.0E-01
STY1305	oppB	3.3	7.4E-09	1.5	3.0E-04	1.3	1.7E-01
STY1306	oppC	2.1	2.1E-06	1.5	6.1E-05		
STY1319	-	0.5	1.0E-03	1.1	3.9E-01	0.7	1.4E-01
STY1371	pspA	1.0	7.5E-01	1.2	1.5E-02	7.0	8.8E-16
STY1372	pspB	0.8	2.3E-01	1.0	7.2E-01	7.3	2.5E-14
STY1373	pspC	1.0	9.5E-01	1.2	5.8E-02	5.2	1.0E-14
STY1374	pspD	0.6	7.2E-03	0.9	1.5E-01	4.3	3.6E-12
STY1375	pspE	0.7	2.8E-02	0.9	2.0E-01	4.2	4.0E-08
STY1493	adhP	0.5	1.3E-03	0.5	2.0E-04	0.7	3.1E-03
STY1501	-	4.2	1.8E-09	1.9	2.7E-06	1.7	6.7E-03
STY1522	-	0.1	5.8E-05	0.1	3.7E-04	0.1	8.5E-10
STY1523	hyaA2	0.2	4.3E-04	0.2	4.2E-04	0.1	7.0E-12
STY1525	hyaB2	0.3	3.2E-02	0.3	2.4E-01		
STY1566	dmsA2	0.2	1.1E-04	0.8	3.6E-02		
STY1568	dmsC	0.5	1.3E-03	1.0	9.5E-01	0.9	2.9E-01
STY1569	-	0.4	2.9E-03	0.9	2.3E-01	1.1	6.0E-01
STY1588	pntB	0.4	4.7E-04	0.9	1.5E-01	0.9	5.8E-01
STY1610	-	2.1	4.1E-02	0.6	4.9E-01		
STY1615	-	3.9	7.2E-08	1.3	8.1E-03		

STY1649	ompS2	0.4	2.4E-01	0.3	1.6E-02		
STY1664	-	2.8	2.9E-10	1.9	1.0E-05	1.5	2.7E-03
STY1708	ssaL	1.1	5.1E-01	0.4	1.2E-03	0.9	3.8E-01
STY1711	ssaJ	1.0	8.3E-01	0.4	8.9E-04	0.4	3.2E-05
STY1712	ssaI	7.3	no replicates	0.2	3.0E-03	0.4	1.2E-03
STY1713	ssaH	1.4	no replicates	0.2	3.7E-03	0.3	5.2E-05
STY1722	sseB	0.9	5.1E-01	0.3	2.7E-05	0.5	1.4E-05
STY1728	ssrA	1.0	8.6E-01	0.3	1.2E-02	1.4	1.52E-01
STY1723	sseA	1.0	9.4E-01	0.3	1.6E-03		
STY1725	ssaD	0.4	2.9E-04	0.9	4.8E-01		
STY1815	gdhA	2.4	3.7E-08	1.3	1.5E-03		
STY1867	-	0.4	8.7E-04	0.9	2.2E-01	1.0	6.8E-01
STY1871	STY1871	1.8	2.1E-05			2.1	2.4E-09
STY1922	gdhA	4.3	1.9E-11	2.2	2.0E-08		
STY1925	-	2.2	2.6E-10	1.4	1.6E-04		
STY1927	mltE	2.6	9.9E-14	1.4	4.4E-05	1.2	1.6E-01
STY1930	dadX	7.3	2.3E-10	1.6	3.2E-04	1.5	2.6E-02
STY1931	dadA	17.0	7.1E-16	1.7	3.2E-07	1.7	4.0E-02
STY1975	htpX	2.2	2.0E-06	1.4	5.1E-03	1.7	7.4E-05
STY1991	-	4.0	2.8E-03	1.5	6.0E-04	2.0	3.1E-02
STY2141	yecI	0.4	7.7E-03	0.5	1.3E-03		
STY2159	-	2.4	7.3E-11	1.4	2.5E-03	1.4	8.9E-02
STY2167	fliC	1.1	2.9E-01			2.5	2.3E-04
STY2188	fliQ	1.5	2.2E-04			2.0	7.3E-04
STY2203	ompS1	0.4	4.4E-04	0.2	2.3E-05	0.3	5.2E-09
STY2235	cbiT	2.1	2.0E-09	1.2	7.0E-02		
STY2421	mglC	2.0	3.2E-08	1.5	1.0E-05		
STY2424	mglB	2.1	2.6E-07	1.2	4.0E-02		
STY2474	dsbE1	4.7	2.7E-15	1.3	8.7E-03		
STY2475	ccmF1	4.9	1.1E-14	1.4	2.8E-02		
STY2476	ccmE1	3.3	1.3E-07	1.2	1.0E-01		
STY2478	ccmC1	4.4	4.5E-14	1.1	6.8E-02	0.8	6.1E-01
STY2481	napC	14.2	4.0E-17	1.0	8.6E-01	1.0	8.8E-01
STY2482	napB	12.3	1.2E-13	0.9	5.3E-01		
STY2483	napH	5.3	2.3E-08	1.6	5.4E-03		
STY2484	napG	4.7	4.3E-08	1.2	3.6E-02		
STY2485	napA	8.6	6.8E-10	1.0	8.6E-01		
STY2486	yojF	4.3	5.6E-08	0.9	3.2E-01	1.0	9.5E-01
STY2487	napF	4.9	3.0E-06	0.8	1.1E-01		
STY2493	ompC	0.0	4.2E-06	0.0	2.1E-06	0.0	3.3E-05
STY2585	argT	3.5	1.2E-06	0.9	7.9E-01		
STY2621	-	2.9	9.3E-08	1.4	6.3E-03		
STY2623	fadL	2.1	1.3E-03	1.3	3.4E-01		
STY2730	-	0.5	1.3E-03	1.2	1.8E-01		
STY2763a		0.5	5.5E-03	1.3	4.5E-03	1.4	4.1E-01
STY2771	ndk	2.9	2.2E-08	1.7	7.8E-06	1.6	8.1E-03
STY2790	yfhP	1.1	6.5E-01			2.4	4.6E-03
STY2805	cadB	0.4	3.2E-04	1.9	1.6E-09	1.8	1.3E-02
STY2806	cadA	0.4	3.8E-04	2.2	2.8E-06	1.5	7.5E-02
STY2842	yfiG	2.2	2.0E-10	1.4	2.2E-05	1.6	5.3E-04
STY2901	-	0.4	1.3E-02	0.3	3.1E-03		
STY2903	tctE	2.3	1.5E-06	1.6	4.1E-02		
STY2904	tctD	3.7	2.6E-10	1.4	4.8E-05	1.0	8.3E-01
STY2906	-	2.0	4.0E-04	1.5	2.3E-01		
STY2909	-	3.1	3.9E-04	1.0	9.1E-01		
STY2912	gabT	2.4	1.4E-03	1.3	2.6E-02		
STY2974	hycB	2.1	4.4E-03	1.6	1.7E-01		
STY2987	sprB	0.4	1.3E-03	0.7	4.7E-03	1.4	1.8E-01
STY2988	sprA	0.5	4.5E-03	0.7	1.1E-02	1.4	8.0E-02
STY3004	sipF	1.1	4.0E-01	1.1	4.3E-01	2.2	1.0E-02
STY3006	sipD	0.7	1.4E-02	0.9	3.4E-01	2.0	1.1E-02
STY3009	spaT	0.7	2.3E-02	1.1	5.1E-01	2.0	5.5E-04
STY3014	spaO	0.7	5.0E-02	1.1	5.6E-01	2.5	8.9E-05

STY3016	spaM	0.7	6.2E-02	1.0	8.9E-01	2.2	1.3E-03
STY3017	spaI	0.8	7.1E-02	1.2	9.8E-02	2.3	1.4E-03
STY3018	spak	0.6	6.8E-03	0.9	4.1E-01	2.4	1.8E-03
STY3049	rpoS	2.7	1.4E-09	1.9	3.3E-08	2.0	7.4E-04
STY3097	-	0.3	1.4E-04	1.0	8.1E-01	1.3	2.6E-01
STY3098	ygcX	0.2	1.1E-04	0.9	6.9E-01	0.9	5.5E-01
STY3099	ygcY	0.2	5.5E-05	1.2	8.3E-02		
STY3100	ygcZ	0.1	9.9E-05	0.9	7.4E-01		
STY3164	yqeF	3.4	2.6E-13	1.5	2.3E-04		
STY3176	stdB	2.5	2.9E-11	1.2	7.9E-03	1.7	1.1E-05
STY3178	STY3178	1.3	7.8E-02	1.0	6.2E-01	0.1	6.1E-08
STY3179	STY3179	1.2	2.3E-01	0.9	6.0E-01	0.0	7.7E-09
STY3296	-	2.4	8.2E-08	1.0	8.8E-01		
STY3297	ordL	4.3	1.2E-04	1.0	9.6E-01		
STY3317	hybD	0.4	5.0E-04	0.7	4.1E-03		
STY3318	hybC	0.4	5.7E-04	0.6	9.4E-04		
STY3319	hybB	0.4	3.0E-03	0.6	1.7E-02		
STY3320	hybA	0.5	1.1E-03	0.7	6.6E-03		
STY3321	hybO	0.5	1.6E-03	0.6	3.9E-03		
STY3394	cheM	1.9	1.8E-06	1.8	2.1E-05	2.2	1.3E-02
STY3395	air	2.7	1.6E-08	2.5	8.4E-08		
STY3405	-	11.2	2.9E-05				
STY3406	STY3406	1.7	1.0E-04	1.4	1.6E-03	3.3	2.1E-09
STY3411	-	2.3	9.4E-08	1.4	1.0E-03	2.5	1.1E-02
STY3427	tdcB	0.5	2.7E-02	0.5	5.6E-03		
STY3429	-	0.2	1.2E-04	1.1	5.0E-01		
STY3430	garR	0.2	9.3E-05	0.9	2.4E-01		
STY3431	garL	0.1	3.6E-05	0.9	7.1E-01		
STY3432	garD	0.2	6.3E-05	1.1	1.6E-01		
STY3460	STY3460	1.4	1.8E-02	1.1	4.6E-01	2.1	2.7E-03
STY3539	mdh	2.1	1.9E-07	1.4	7.5E-04	1.7	1.1E-04
STY3558	STY3558	0.7	2.3E-02	0.8	1.1E-02	2.4	3.3E-10
STY3559	accB	0.8	1.6E-01	0.6	9.3E-04	2.3	4.6E-06
STY3572	yhdV	1.0	6.8E-01	1.0	7.7E-01	0.5	2.2E-05
STY3577	fadB	23.0	3.7E-19	1.5	2.6E-04	2.1	4.4E-02
STY3578	fadA	17.9	2.8E-19	1.4	1.7E-03		
STY3658	-	2.9	3.7E-02				
STY3671	STY3671					2.2	5.9E-03
STY3720	rsD	3.3	5.5E-11	1.2	1.2E-01		
STY3748	udhA	3.5	1.1E-09	1.5	2.4E-04	1.7	2.5E-05
STY3778	hslV	2.4	7.3E-08	1.1	1.5E-01	2.4	1.0E-04
STY3778	hslV	2.4	7.3E-08	1.1	1.5E-01	2.4	1.0E-04
STY3779	hslU	1.8	3.3E-06	1.1	5.1E-01	2.0	2.6E-04
STY3784	glpK	0.7	2.2E-02	0.8	8.9E-02	0.5	3.1E-02
STY3790	-	9.7	4.6E-14	0.9	6.2E-01	1.3	6.4E-02
STY3791	-	9.3	3.1E-13	0.9	5.4E-01	1.2	8.0E-02
STY3792	yneB	4.2	1.6E-08	0.8	1.2E-01	1.1	4.5E-01
STY3793	-	8.2	5.2E-16	0.9	1.1E-01		
STY3794	ydeZ	10.8	2.5E-11	0.8	1.4E-01		
STY3795	-	9.2	2.8E-09	0.9	2.6E-01		
STY3796	-	10.6	3.4E-09	0.6	1.6E-03		
STY3797	ydeW	4.8	2.3E-08	0.7	3.2E-02		
STY3798	ydeV	6.7	2.7E-08	1.0	9.8E-01		
STY3811	cpxP	3.2	3.6E-05	0.9	4.1E-01	0.9	2.0E-01
STY3831	yjiL	2.3	3.6E-08	1.2	2.0E-02	1.7	1.3E-02
STY3842	fdoI	1.9	9.7E-06	1.4	1.3E-03	2.2	1.7E-06
STY3879	STY3879	0.9	4.5E-01	1.3	1.1E-02	0.5	2.3E-02
STY3901	asnA	0.4	1.5E-03	0.7	1.7E-02	0.9	3.3E-01
STY4023	mgtB	1.4	2.0E-01	0.9	7.3E-01	2.1	2.6E-06
STY4025	slsA	0.0	1.4E-05	0.1	1.2E-05	0.2	2.5E-08
STY4146	-	0.1	5.6E-04	0.1	1.7E-03		
STY4168	dppA	2.2	2.5E-07	1.3	1.4E-03	1.4	8.0E-03
STY4294	ompR	0.3	1.3E-04	0.2	1.6E-05	0.4	6.5E-05

STY4299	hslR	2.1	1.1E-07	1.0	8.3E-01	1.3	1.8E-01
STY4321	nirD	2.5	9.2E-10	1.1	1.3E-01		
STY4322	nirB	2.1	1.5E-06	0.9	6.5E-01		
STY4340	yheR	2.5	3.5E-07	1.1	4.1E-01	1.4	1.7E-01
STY4400	metA	2.4	9.2E-03	1.2	1.1E-01	1.3	1.5E-01
STY4401	aceB	4.9	2.2E-07	1.8	1.0E-06		
STY4402	aceA	5.9	9.1E-08	1.7	1.4E-04	2.0	6.8E-03
STY4403	aceK	6.9	1.0E-07	1.7	8.4E-06		
STY4440	yjbO	0.8	3.1E-01	1.0	8.4E-01	4.5	1.1E-08
STY4462	yjcC	2.2	1.7E-05	1.3	1.6E-01		
STY4471	-	3.5	2.7E-10	1.3	8.2E-03		
STY4472	-	2.9	6.3E-07	1.2	7.4E-02		
STY4473	acs	4.8	1.0E-06				
STY4475	nrfA	4.8	1.8E-12	0.9	2.9E-01		
STY4476	nrfB	3.4	9.7E-12	1.2	2.7E-02		
STY4477	nrfC	4.2	1.6E-10	1.1	1.3E-01	1.0	8.5E-01
STY4478	nrfD	3.6	1.0E-08	1.1	3.2E-01		
STY4480	-	2.1	2.7E-07	1.0	8.8E-01		
STY4497	melA	0.3	1.2E-04	2.1	2.8E-07		
STY4498	melB	0.4	2.7E-04	1.4	1.7E-04		
STY4501	dcuR	3.6	3.8E-10	1.2	4.4E-02		
STY4508	dmsC	2.1	9.6E-04	1.2	1.2E-01		
STY4519	phoN	1.3	4.9E-02	1.3	1.9E-02	2.1	3.5E-06
STY4539	pilL	2.0	2.3E-02	1.1	6.2E-01		
STY4540	pilM	2.2	5.2E-03	1.2	1.4E-01		
STY4553	pilK	2.3	3.3E-03	0.7	2.7E-01		
STY4618	-	2.8	7.6E-10	1.9	2.5E-04	2.4	5.6E-02
STY4651	vexE	0.1	1.8E-05	0.0	5.7E-06		
STY4652	vexD	0.1	7.2E-05	0.0	5.9E-06		
STY4653	vexC	0.1	3.8E-02	0.0	4.7E-06	0.5	6.7E-03
STY4654	vexB	0.4	1.3E-03	0.2	5.2E-05	1.3	3.6E-01
STY4655	vexA	0.0	4.0E-05	0.0	6.5E-06		
STY4656	tvIE	0.1	4.0E-05	0.1	1.7E-05		
STY4659	tvID	0.4	3.3E-04	0.2	3.4E-05	1.1	5.1E-01
STY4660	tvIC	0.0	1.7E-05	0.0	1.3E-05		
STY4661	tvIB	0.0	2.7E-05	0.0	2.8E-05	0.5	1.5E-02
STY4662	tvIA	0.0	5.9E-06	0.0	3.2E-06	0.2	9.4E-09
STY4684	dcuA	0.1	4.7E-05	0.6	2.5E-03		
STY4685	aspA	0.4	4.5E-04	0.7	7.4E-03	0.8	4.9E-01
STY4754	cycA	2.4	1.6E-09	1.2	8.9E-02	1.2	1.4E-01
STY4757	ytfE	0.9	8.2E-01	1.0	9.6E-01	6.7	1.8E-05
STY4767	msrA	2.5	3.0E-08	1.1	5.8E-02	2.2	4.1E-07
STY4800	pyrB	0.5	4.8E-03	1.0	8.3E-01		
STY4801	-	0.3	9.4E-05	0.5	1.5E-04	0.7	1.1E-02
STY4802	-	0.3	2.1E-04	0.5	8.3E-05		
STY4803	-	0.2	5.1E-05	0.3	4.1E-05	0.9	4.5E-01
STY4804	-	0.2	4.8E-05	0.3	2.5E-05		
STY4805	-	0.2	4.6E-05	0.4	4.1E-05		
STY4806	-	0.5	2.0E-03	0.6	4.7E-04		
STY4869	STY4869	0.6	2.0E-02	1.0	6.8E-01	0.4	5.0E-03
STY4892	tsr	2.5	1.4E-09	2.7	1.2E-12		
STY4908	rimI	0.3	5.7E-04	0.7	1.0E-02		

## 9.9 chIP data

Motif

TGTWACAW	AATATATA	CTAGACTA	CTWAGGGR	GTGTSTAY	Score	Down stream gene	Function
					2.75	t0104	hypothetical
					2.27	araC	arabinose regulator
2					2.67	yabB	conserved hypothetical
			2		2.91	t0292	putative sensor kinase
					2.31	XseA	exodeoxyribonuclease large subunit
					2.57	yffB	conserved hypothetical
					2.76	pgtP	phosphoglycerate transporter
					2.32	fadL	long-chain fatty acid transport
2					2.93	fadL	long-chain fatty acid transport
							lysine arginine ornithine periplasmic protein
					2.66	argT	
					3.96	t0528	conserved hypothetical
					3.25	nuoA	NADH dehydrogenase I chain A
					2.58	nuoA	NADH dehydrogenase I chain A
					2.45	ompC	
1					2.45	ompC	
					4.58	mgIB	D-galactose binding periplasmic protein
2					2.61	rfbB	dTDP-glucose 4,6-dehydratase
2	1				6.99	ompS	
					3.34	fliC	
					2.57	t1022	conserved hypothetical
		1			2.56	t1022	conserved hypothetical
					2.38	t1080	putative membrane protein
					2.37	rplT	50S ribosomal subunit L20
				1	4.07	ssrA	SPI-2 regulator
1				1	5.41	t1320	putative proton/oligo symporter
					2.35	narU	nitrite extrusion
					2.43	narU	nitrite extrusion
2	2				6.28	csgD	curli regulator
					2.40	csgD	curli regulator
					3.22	t1790	conserved hypothetical
					5.44	t1903	putative bacteriophage proteins
1					4.84	ompX	
					3.30	sdhC	
3					4.09	sdhC	
					2.31	gltA	
					2.49	gltA	
				1	2.38	cspE	cold shock-like protein
					2.93	acrR	potential acrAB operon repressor
					4.56	rpmJ	putative 50S ribosomal subunit L36
					2.50	psiF	phosphate starvation-inducible protei
				1	2.59	yafK	putative exported protein
					2.89	yafK	putative exported protein

		2.50	t2565	hypothetical protein
		3.21	yfiA	putative sigma-54 modulation proteir
	1	2.74	-	Intergenic near virK
		3.32	sitA	SPI-1 iron transport
		2.66	hilC	SPI-1 regulator
		2.83	hilC	within gene
		2.38	hilC	middle of hilC
		2.71	rpoS	
		2.40	rpoS	
		2.98	rpoS	
		2.43	sopD	
		2.12	t2863	
		2.56	sraE	SraE_OmrA_OmrB small RNA
		3.00	stdA	probable fimbrial
		2.15	stdA	probable fimbrial
1		2.22	stdA	probable fimbrial
		2.31	-	Intergenic
		2.65	t2957	probable isomerase
		2.74	galP	galactose-proton symporter
2		3.17	galP	galactose-proton symporter
		2.70	t3046	LysR-family transcriptional regulator
	1	2.64	mug	G/U mismatch-specific DNA glycosyla
		2.62	air	aerotaxis receptor protein
1		2.83	t3144	possible drug efflux protein
		3.15	t3144	possible drug efflux protein
2		3.06	t3146	putative membrane protein
		3.44	t3197	probable amino acid permease
	1	2.35	argG	arginosuccinate synthase
		2.58	argG	arginosuccinate synthase
2		2.79	t3244	conserved hypothetical protein
		2.47	t3244	conserved hypothetical protein
		2.76	sspA	stringent starvation protein
	1	2.99	mreB	rod shaped determining protein
2		2.69	mreB	rod shaped determining protein
		3.12	-	Intergenic
		3.38	udp	uridine phosphorylase
		3.13	purH	phosphoribosylaminoimidazolecarbox
		2.94	rplJ	50S ribosomal subunit L10
		3.42	rplK	50S ribosomal subunit L11
		2.78	glpF	glycerol uptake facilitator
	1	2.64	glpF	glycerol uptake facilitator
		2.83	glpF	glycerol uptake facilitator
	1	2.63	glpK	middle of glpK, contiguous with glpF
		2.74	t3554	putative membrane protein
	1	2.74	sodA	manganese superoxide dismutase
		2.53	rhaT	l-rhamnose-proton symporter
1		2.63	fdhD	necessary for formate dehydrogenase
		2.73	glnA	glutamine synthase
		3.21	glnA	glutamine synthase

1					2.45	yihG	putative acyltransferase
					2.57	t3665	probable PTS system permease
					2.34	rpmH	50S ribosomal subunit L34
					2.36	ilvB	acetohydroxy acid synthase I and ovi
						ivbL leader peptide	
					6.98	t3725	putative carbohydrate kinase
					2.52	rmbA	conserved hypothetical upstream of r
					3.23	ttk	tetR-family transcriptional regulator
					3.40	cspA	cold shock protein A
2					5.28	dppA	periplasmic dipeptide trnsport protein
						and overlaps with pseudo	
					2.80	t3891	hypothetical protein
					2.43	yhjS	conserved hypothetical
					2.32	t3896	conserved hypothetical
					2.70	yhjH	conserved hypothetical
1					5.46	yhjE	conserved hypothetical
					2.70	rpoH	sigma-32
					2.66	rpoH	sigma-32
2					2.36	rpoH	sigma-32
				1	4.40	ompR	
		1			2.87	pckA	phosphoenolpyruvate carboxykinase
					2.62	pckA	phosphoenolpyruvate carboxykinase
					2.12	yrfE	putative NUDIX hydrolase
					2.57	yrfE	putative NUDIX hydrolase
1		1			2.37	t4222	hypothetical protein
					2.80	t4341	hypothetical protein
					3.79	t4341	hypothetical protein
1			1		14.39	tviA	viab locus
					2.42	t4357	probable phage integrase
				1	8.17	t4357	probable phage integrase
3					2.37	t4358	hypothetical protein contiguous with
						integrase	
					2.43	efp	elongation factor p
					2.98	artJ	arginine-binding periplasmic protein
					2.35	lasT	putative RNA methyltransferase
22	7	1	5	8			

misc_binding misc_binding undefined product 117836:117886 forward	Upstream of t0104 Hypothetical	
misc_binding misc_binding undefined product 122946:123001 forward	Upstream of araC arabinose regulator	◆ TGTWACAW
misc_binding misc_binding undefined product 140906:140956 forward	Upstream of yabB conserved hypothetical	◆ AATATATA
misc_binding misc_binding undefined product 332896:332951 forward	Upstream of t0292 putative sensor kinase	◆ GTGTSTAY
misc_binding misc_binding undefined product 413906:414006 reverse	Upstream of Xsea exodeoxyribonuclease large subunit	◆ CTAGACTA
misc_binding misc_binding undefined product 447516:447596 forward	Upstream of yffB conserved hypothetical	◆ CTWAGGGR
misc_binding misc_binding undefined product 532691:532821 reverse	Upstream of pgfP, phosphoglycerate transporter protein	
misc_binding misc_binding undefined product 547201:547256 reverse	Upstream of fadL long-chain fatty acid transport protein	
misc_binding misc_binding undefined product 547321:547391 reverse	Upstream of fadL long-chain fatty acid transport protein	
misc_binding misc_binding undefined product 583251:583316 forward	Upstream of argT, lysine-arginine-ornithine-binding periplasmic protein	
misc_binding misc_binding undefined product 601951:602121 forward	Upstream of t0528, conserved hypothetical	
misc_binding misc_binding undefined product 609681:609816 forward	Upstream of nuoA, NADH dehydrogenase I chain A	
misc_binding misc_binding undefined product 610106:610156 forward	Upstream of nuoA, NADH dehydrogenase I chain A	
misc_binding misc_binding undefined product 680866:680921 forward	Upstream of ompC	
misc_binding misc_binding undefined product 680996:681046 forward	Upstream of ompC	
misc_binding misc_binding undefined product 750226:750396 forward	Upstream of mglB, D-galactose-binding periplasmic protein	
misc_binding misc_binding undefined product 867191:867261 forward	Upstream of rfbB, dTDP-glucose 4,6-dehydratase	
misc_binding misc_binding undefined product 985081:985831 reverse	Upstream of ompS	
misc_binding misc_binding undefined product 1013766:1013841 forward	Upstream of fliC	
misc_binding misc_binding undefined product 1106256:1106306 reverse	Upstream of t1022, conserved hypothetical protein	
misc_binding misc_binding undefined product 1106336:1106396 reverse	Upstream of t1022, conserved hypothetical protein	
misc_binding misc_binding undefined product 1162231:1162286 reverse	Upstream of t1080, putative outermembrane protein	
misc_binding misc_binding undefined product 1283291:1283431 forward	Upstream of rpIT, 50S ribosomal subunit protein L20	
misc_binding misc_binding undefined product 1334176:1334376 reverse	Upstream of ssrA, two-component sensor kinase	
misc_binding misc_binding undefined product 1382041:1382556 reverse	Upstream of t1320, putative proton/oligopeptide symporter	
misc_binding misc_binding undefined product 1535386:1535486 forward	Upstream of narU, nitrite extrusion	
misc_binding misc_binding undefined product 1535516:1535591 forward	Upstream of narU, nitrite extrusion	
misc_binding misc_binding undefined product 1840011:1840611 forward	Upstream of csgD, curli regulator	
misc_binding misc_binding undefined product 1840701:1840761 forward	Upstream of csgD, curli regulator	

misc_binding misc_binding undefined product 1851236:1851326 reverse	Upstream of t1790, conserved hypothetical
misc_binding misc_binding undefined product 1956466:1956626 reverse	Upstream of t1903, putative bacteriophage protein
misc_binding misc_binding undefined product 2117686:2117976 reverse	Upstream of ompX
misc_binding misc_binding undefined product 2208471:2208551 reverse	Upstream of sdhC, succinate dehydrogenase
misc_binding misc_binding undefined product 2208611:2208806 reverse	Upstream of sdhC, succinate dehydrogenase
misc_binding misc_binding undefined product 2208841:2208891 forward	Upstream of gltA, citrate synthase
misc_binding misc_binding undefined product 2208951:2209026 forward	Upstream of gltA, citrate synthase
misc_binding misc_binding undefined product 2303376:2303471 reverse	Upstream of cspE, cold shock-like protein
misc_binding misc_binding undefined product 2454496:2454551 reverse	Upstream of acrR, potential acrAB operon repressor
misc_binding misc_binding undefined product 2461571:2461746 reverse	Upstream of rpmJ, putative 50S ribosomal protein L36
misc_binding misc_binding undefined product 2552046:2552101 reverse	Upstream of psiF, phosphate starvation-inducible protein
misc_binding misc_binding undefined product 2614081:2614161 forward	Upstream of yafK, putative exported protein
misc_binding misc_binding undefined product 2614221:2614296 forward	Upstream of yafK, putative exported protein
misc_binding misc_binding undefined product 2642021:2642076 reverse	Upstream of t2565, hypothetical protein
misc_binding misc_binding undefined product 2698626:2698686 forward	Upstream of yfiA, putative sigma-54 modulation protein
misc_binding misc_binding undefined product 2765436:2765501 forward	Upstream of intergenic, near virK
misc_binding misc_binding undefined product 2845046:2845131 forward	Upstream of sitA, SPI-1, iron transport
misc_binding misc_binding undefined product 2855161:2855221 forward	Upstream of hilC
misc_binding misc_binding undefined product 2855736:2855881 forward	Within hilC
misc_binding misc_binding undefined product 2856206:2856261 forward	Middle of hilC
misc_binding misc_binding undefined product 2902056:2902141 reverse	Upstream of rpoS
misc_binding misc_binding undefined product 2902306:2902376 reverse	Upstream of rpoS
misc_binding misc_binding undefined product 2902401:2902476 reverse	Upstream of rpoS
misc_binding misc_binding undefined product 2921856:2921941 forward	Upstream of sopD
misc_binding misc_binding undefined product 2942221:2942291 forward	Upstream of t2863, hypothetical protein
misc_binding misc_binding undefined product 3013351:3013406 reverse	Upstream of tnpA (t2920)
misc_binding misc_binding undefined product 3034766:3034936 reverse	Upstream of stdA, probable fimbrial protein
misc_binding misc_binding undefined product 3035011:3035061 reverse	Upstream of stdA, probable fimbrial protein
misc_binding misc_binding undefined product 3035181:3035316 reverse	Upstream of stdA, probable fimbrial protein

misc_binding misc_binding undefined product 3035011:3035061 reverse	Upstream of stdA, probable fimbrial protein
misc_binding misc_binding undefined product 3035181:3035316 reverse	Upstream of stdA, probable fimbrial protein
misc_binding misc_binding undefined product 3039506:3039556 forward	Intergenic
misc_binding misc_binding undefined product 3046486:3046536 forward	Upstream of t2957 probable isomerase
misc_binding misc_binding undefined product 3094216:3094271 forward	Upstream of galP, galactose-proton symport
misc_binding misc_binding undefined product 3094296:3094381 forward	Upstream of galP, galactose-proton symport
misc_binding misc_binding undefined product 3125301:3125356 forward	Upstream of t3046, LysR-family transcriptional regulator
misc_binding misc_binding undefined product 3221961:3222036 reverse	Upstream of mug, G/U mismatch-specific DNA glycosylase
misc_binding misc_binding undefined product 3227271:3227321 reverse	Upstream of air, aerotaxis receptor protein
misc_binding misc_binding undefined product 3234941:3235091 forward	Upstream of t3144, possible drug efflux protein
misc_binding misc_binding undefined product 3235111:3235196 forward	Upstream of t3144, possible drug efflux protein
misc_binding misc_binding undefined product 3237916:3238046 forward	Upstream of t3146, putative membrane protein
misc_binding misc_binding undefined product 3287291:3287486 reverse	Upstream of t3197, probable amino acid permease
misc_binding misc_binding undefined product 3299766:3299816 forward	Upstream of argG, arginosuccinate synthase
misc_binding misc_binding undefined product 3299911:3299986 forward	Upstream of argG, arginosuccinate synthase
misc_binding misc_binding undefined product 3330321:3330411 reverse	Upstream of t3244, conserved hypothetical protein
misc_binding misc_binding undefined product 3330446:3330496 reverse	Upstream of t3244, conserved hypothetical protein
misc_binding misc_binding undefined product 3351291:3351341 reverse	Upstream of sspA, stringent starvation protein
misc_binding misc_binding undefined product 3383646:3383721 reverse	Upstream of mreB, rod shaped determining protein
misc_binding misc_binding undefined product 3383761:3383826 reverse	Upstream of mreB, rod shaped determining protein
misc_binding misc_binding undefined product 3417121:3417226 reverse	Intergenic
misc_binding misc_binding undefined product 3430376:3430436 reverse	Upstream of udp, uridine phosphorylase
misc_binding misc_binding undefined product 3543791:3543881 forward	Upstream of purH, phosphoribosylaminoimidazolecarboxamide formyltransferase
misc_binding misc_binding undefined product 3572556:3572651 reverse	Upstream of rplJ, 50S ribosomal subunit protein L10
misc_binding misc_binding undefined product 3573836:3573941 reverse	Upstream of rplK, 50S ribosomal subunit protein L11
misc_binding misc_binding undefined product 3633856:3633961 forward	Upstream of glpF, glycerol uptake facilitator protein
misc_binding misc_binding undefined product 3633986:3634076 forward	Upstream of glpF, glycerol uptake facilitator protein
misc_binding misc_binding undefined product 3634131:3634366 forward	Upstream of glpF, glycerol uptake facilitator protein

misc_binding misc_binding undefined product 3635416:3635481 forward	Middle of glpK which is contiguous with glpF
misc_binding misc_binding undefined product 3655631:3655691 reverse	Upstream of t3554, putative membrane protein
misc_binding misc_binding undefined product 3664726:3664781 reverse	Upstream of sodA, manganese superoxide dismutase
misc_binding misc_binding undefined product 3668861:3668921 forward	Upstream of rhaT, L rhamnose-proton symporter
misc_binding misc_binding undefined product 3683026:3683086 reverse	Upstream of fdhD, necessary for formate dehydrogenase activity
misc_binding misc_binding undefined product 3713696:3713746 forward	Upstream of glnA, glutamine synthetase
misc_binding misc_binding undefined product 3713801:3713866 forward	Upstream of glnA, glutamine synthetase
misc_binding misc_binding undefined product 3725096:3725151 forward	Upstream of yihG, putative acyltransferase
misc_binding misc_binding undefined product 3773706:3773771 forward	Upstream of t3665, probable PTS system permease
misc_binding misc_binding undefined product 3790051:3790101 reverse	Upstream of rpmH, 50S ribosomal protein l34
misc_binding misc_binding undefined product 3836456:3836506 forward	Upstream of ilvB, acetohydroxy acid synthase I and overlaps ivbL leader peptide
misc_binding misc_binding undefined product 3839581:3839736 forward	Upstream of t3725, putative carbohydrate kinase
misc_binding misc_binding undefined product 3880361:3880426 reverse	Upstream of rmbA, conserved hypothetical upstream of misL (pseudo)
misc_binding misc_binding undefined product 3908531:3908616 reverse	Upstream of ttk, putative TetR-family transcriptional regulator
misc_binding misc_binding undefined product 3997551:3997671 reverse	Upstream of cspA, cold shock protein A
misc_binding misc_binding undefined product 4011051:4011421 forward	Upstream of dppA, periplasmic dipeptide transport protein and overlaps t3884 (pseudo)
misc_binding misc_binding undefined product 4019696:4019821 forward	Upstream of t3891, hypothetical protein (43aa in length)
misc_binding misc_binding undefined product 4023446:4023511 reverse	Upstream of yhjS, conserved hypothetical
misc_binding misc_binding undefined product 4023521:4023571 forward	Upstream of t3896, conserved hypothetical
misc_binding misc_binding undefined product 4040691:4040751 forward	Upstream of yhjH, conserved hypothetical
misc_binding misc_binding undefined product 4045011:4045176 reverse	Upstream of yhjE, hypothetical metabolite transport protein
misc_binding misc_binding undefined product 4093596:4093651 forward	Upstream of rpoH, RNA polymerase sigma-32 factor
misc_binding misc_binding undefined product 4093666:4093716 forward	Upstream of rpoH, RNA polymerase sigma-32 factor
misc_binding misc_binding undefined product 4093761:4093856 forward	Upstream of rpoH, RNA polymerase sigma-32 factor
misc_binding misc_binding undefined product 4154671:4154966 forward	Upstream of ompR
misc_binding misc_binding undefined product 4158646:4158701 reverse	Upstream of pckA, phosphoenolpyruvate carboxykinase
misc_binding misc_binding undefined product 4158761:4158821 reverse	Upstream of pckA, phosphoenolpyruvate carboxykinase
misc_binding misc_binding undefined product 4165156:4165206 forward	Upstream of yrfE, putative NUDIX hydrolase
misc_binding misc_binding undefined product 4165226:4165331 forward	Upstream of yrfE, putative NUDIX hydrolase

misc_binding misc_binding undefined product 4391296:4391381 reverse	Upstream of t4222, hypothetical protein
misc_binding misc_binding undefined product 4491771:4491831 forward	Upstream of t4341, hypothetical protein
misc_binding misc_binding undefined product 4491881:4491966 forward	Upstream of t4341, hypothetical protein
misc_binding misc_binding undefined product 4507511:4508051 reverse	Upstream of tvIA, viab locus
misc_binding misc_binding undefined product 4511076:4511141 forward	Upstream of t4357, probable phage integrase
misc_binding misc_binding undefined product 4511246:4511846 forward	Upstream of t4357, probable phage integrase
misc_binding misc_binding undefined product 4512831:4512881 forward	Upstream of t4358, hypothetical protein, contiguous with probable phage integrase
misc_binding misc_binding undefined product 4538976:4539061 forward	Upstream of efp, elongation factor p
misc_binding misc_binding undefined product 4555771:4555851 reverse	Upstream of artJ, probable arginine-binding periplasmic protein
misc_binding misc_binding undefined product 4790961:4791041 forward	Upstream of lasT, putative RNA methyltransferase