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	*p*-value	GC content	Annotation	*S.*Typhimurium 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)

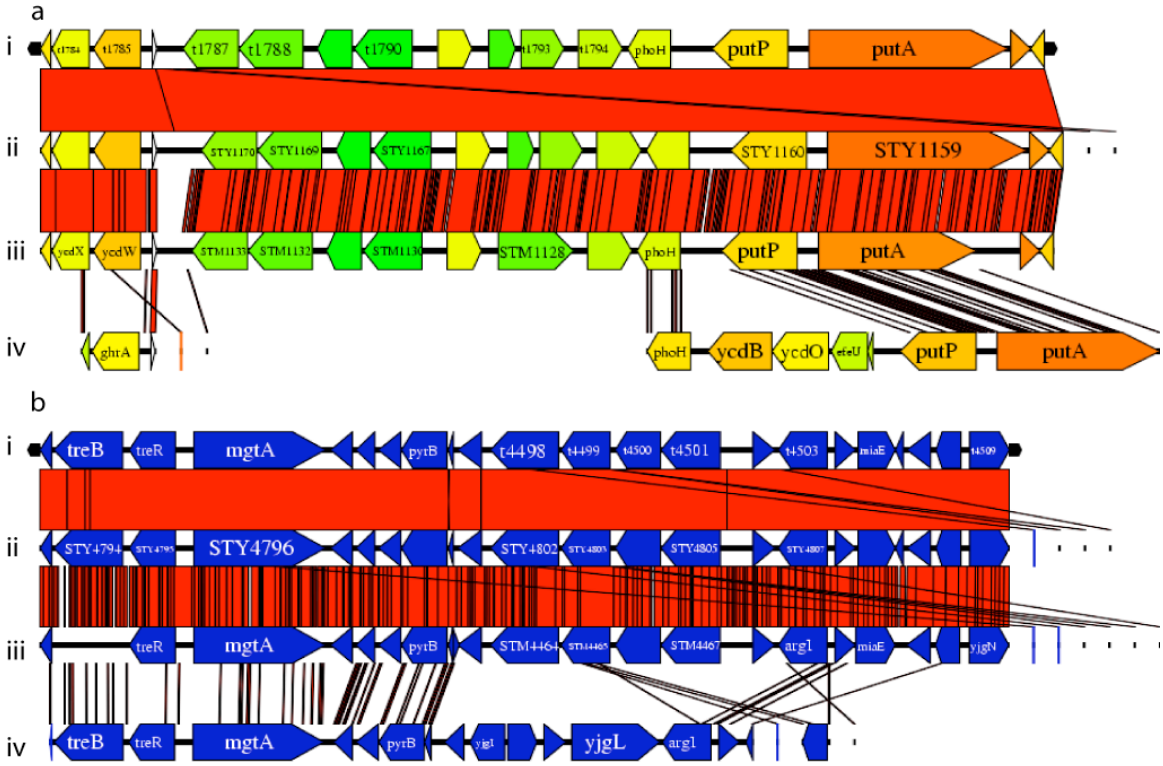

Figure 6.1. Comparative alignment of regions predicted to be OmpR regulated in *S.* Typhi.

(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).

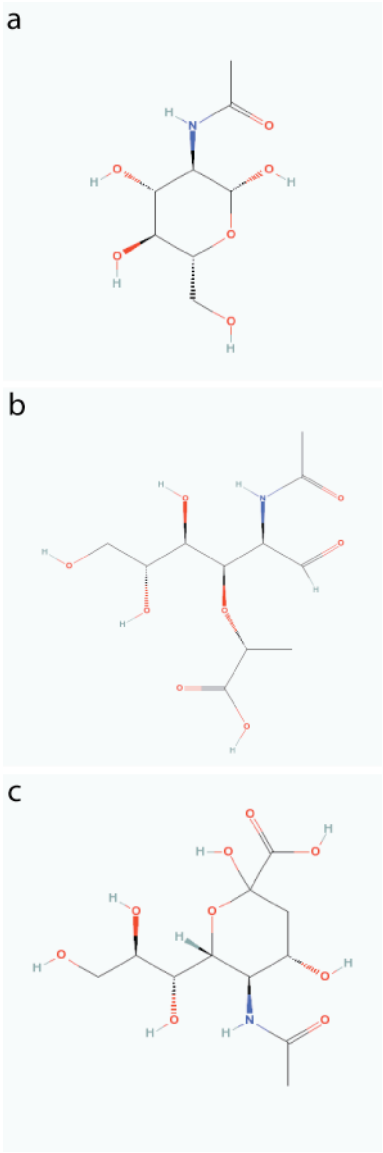

Figure 6.2 The chemical structure of related acetylated aminosugars.

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, Cholerasuis, 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 *Streptoccocus 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).

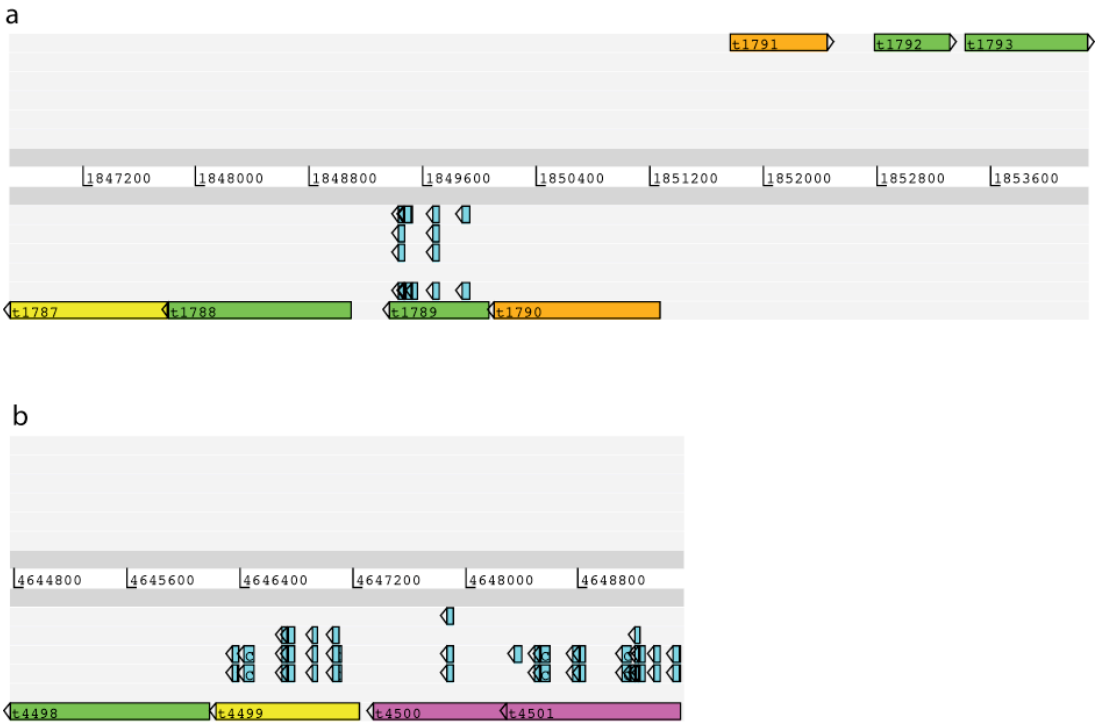

Figure 6.3 Sequenced peptides mapped to regions predicted to be OmpR-regulated.

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.

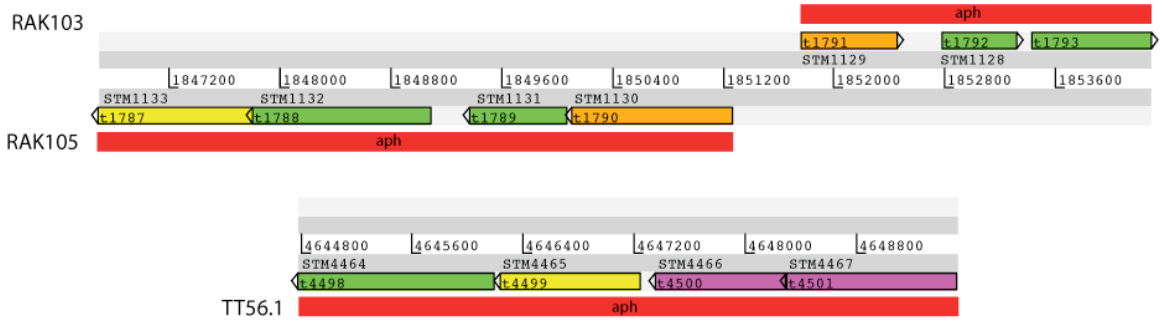

Figure 6.4 Details of mutagenesis for further phenotypic characterisation.

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.

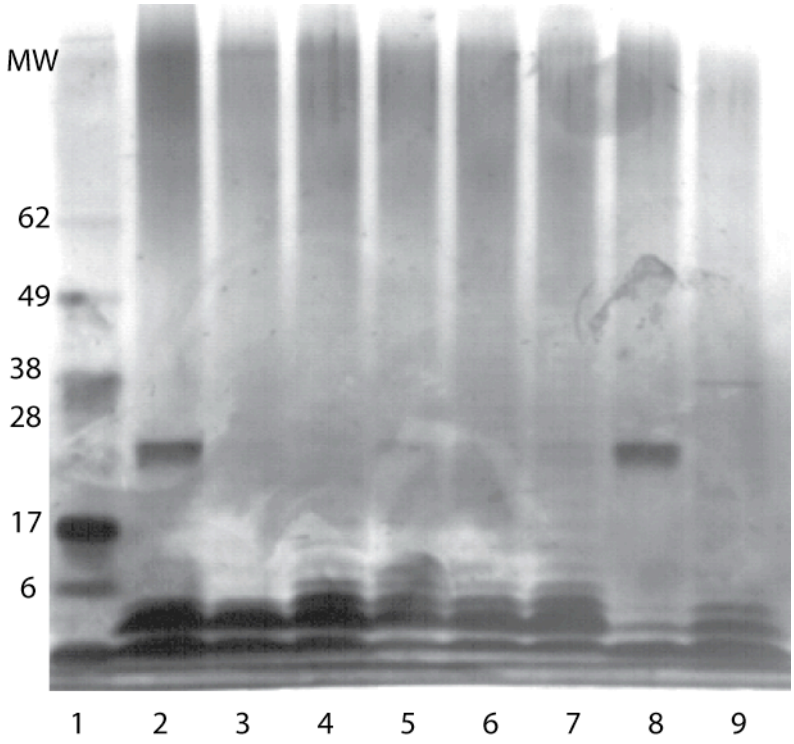

Figure 6.5 LPS silver stained gel.

MW, molecular weight, kDa; Lane 1, protein ladder; 2, BRD948; 3, BRD948 Δ*ompR*; 4, SL1344; 5, SL1344 ΔompR; 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	-'('+')	-'('+')	-'('+')
N-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 (SL1344Δ*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 SL1344Δ*phoN* and the mutants as colonies with functional PhoN are blue and without functional PhoN, are white. Ratios of the mutant: SL1344 Δ*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 Δ*phoN* as the "wild-type" strain and compared it with SL1344 (figure 6.5). We directly compared SL1344 Δ*phoN* with SL1344 Δ*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).

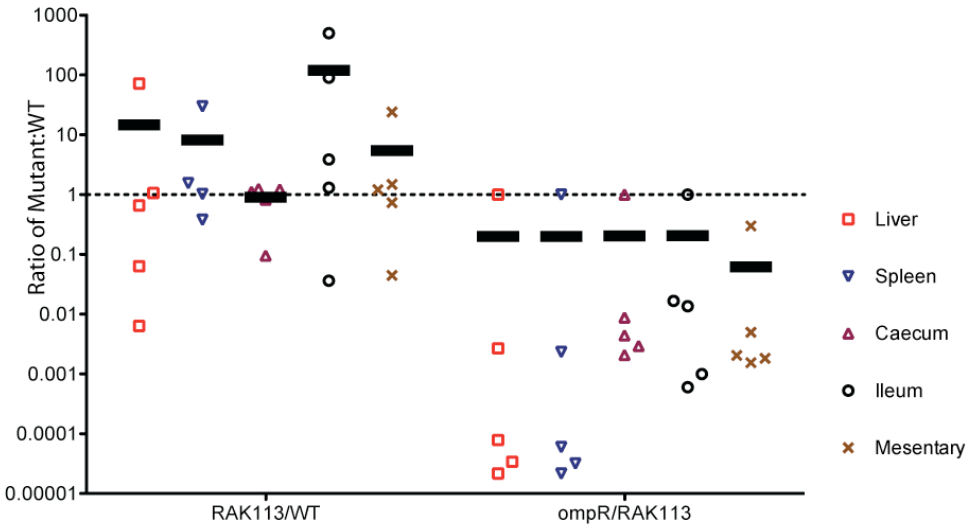

Figure 6.6 Control competitive infections.

Left group, comparison of RAK113 (SL1344 Δ*phoN*) and SL1344 wild-type; right group, comparison of SL1344Δ*ompR* and RAK113. Ratio below 1 indicates fewer bacteria survive compared with control.

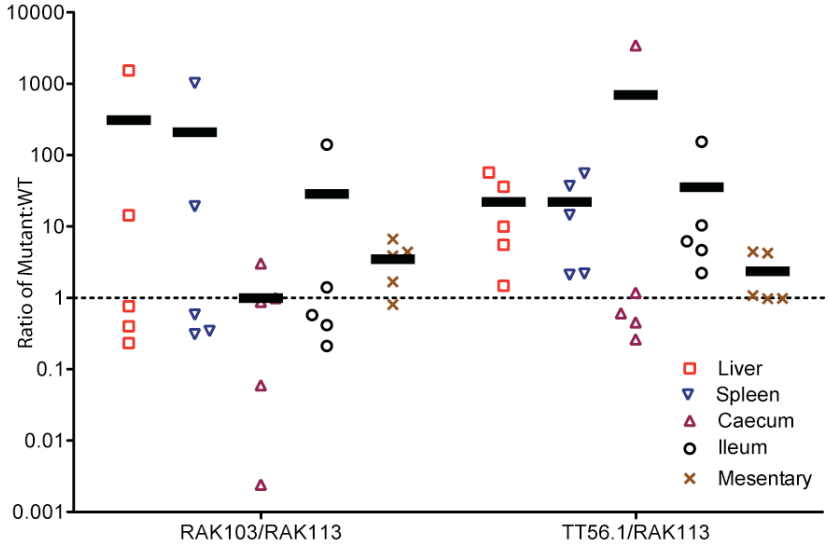

Figure 6.7 Comparison of mutants RAK103 and TT56.1 with RAK113 (SL1344 Δ*phoN*).

Left group, comparison of RAK113 (SL1344 Δ*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 Δ*phoN*, wild-type validation comparison, the system can be quite variable. However, it was also shown by the SL1344:SL1344 Δ*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*-acetlyneuraminic acid (sialic acid, NAN) transport, a *N*-acetymannosamine-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 α-(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 glycosydic 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 rational 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.