6 Final discussion

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

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

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

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

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

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

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

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

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

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

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

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

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

6.1 Publications arising from this thesis

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

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

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

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

* equal first authors

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

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