# Elucidating the factors that define host species ranges of *Salmonella enterica* serovars

Matthew Robert Hayward

August 2013

A thesis submitted for the degree of Doctor of Philosophy, from the Centre for Ecology, Evolution & Behaviour, School of Biological Sciences, Royal Holloway University of London

# Declaration of Authorship

I Matthew Robert Hayward hereby declare that this thesis and the work presented in it is entirely my own. I have clearly stated where I have used the work of others.

Signed:

Date: 11th August 2013

# Thesis abstract

*Salmonella enterica* is an important zoonotic pathogen of clinical and veterinary significance. The species is divided into seven subspecies; subspecies 1, enterica, is further divided in to over 1530 serovars based on different epitopes of two surface antigens. Clinical and veterinary isolations of *Salmonella* are frequently typed to the serovar level of classification. This epidemiological data shows that some serovars are isolated from distinct subsets of species.

In this study I have focused on two serovars, *S.* Derby and *S.* Mbandaka, which are frequently isolated from distinct subsets of livestock species in the UK. The majority of *S.* Derby isolations are from pigs and turkeys, whereas the majority of *S.* Mbandaka isolations are from cattle and chickens. To begin to identify potential mechanisms of host adaptation, I sequenced two strains of each serovar and compared their nucleotide sequences and functional annotations. This lead to the discovery of a new *Salmonella* pathogenicity island, SPI-23, in the chromosome sequence of *S.* Derby, that I go on to show is regulated in a tissue specific manner in a porcine IVOC model. Mutagenesis of the most highly up-regulated gene within SPI-23, *potR*, generated unique phenotypes that have enabled me to posit a role for SPI-23 in tropism to porcine jejunum.

To interrogate the role of metabolite utilisation in constraining colonisation of certain niches, I performed high-throughput phenotyping using Biolog phenotypic microarray technology, at ambient and body temperatures, under aerobic and anaerobic conditions. This, along with other phenotypic studies, lead me to propose a partitioned niche model

between host adaptation in the case of *S*. Derby and adaptation to persistence in the environment on soybean based feed in the case of *S*. Mbandaka.

To identify the contribution of an environment composed of a complex set of metabolites to host adaptation, I produced genome-scale metabolic reconstructions for both serovars. The models were confronted with metabolites found in porcine colon and jejunum; these I identified through metabolomics of gut sections using NMR. These models were used to observe which transport and secondary metabolic reactions contribute most to the incorporation of biomass by *S*. Derby and *S*. Mbandaka when in a porcine host.

Finally I relate the findings of these studies to a representative population of isolates, for which I have produced a phylogenetic reconstruction. I discovered two distinct lineages of *S*. Derby each with a distinct set of genotypes and phenotypes. I postulate that one lineage is adapted to turkeys and environmental persistence, and the other adapted to pathogenicity in pigs. I also show that *S*. Mbandaka is clonal in the UK, and is adapted to growth in soybean based feeds, at ambient temperatures and is adapted to environmental persistence.

## Acknowledgements

I would like to thank my supervisors Martin and Vincent for all their help planning experiments, writing and editing papers, and chapters, and for their enthusiasm and encouragement. It has been a pleasure working with you both and I thank you for giving me the opportunity to undertake this most excellent project. I look forward to our future work and collaborations.

I would also like to thank the students and staff of BAC2 for their help. You have all contributed massively to my work and my enjoyment of the past four years. I would especially like to thank Manal for offering her help and expertise in many aspects of my project.

This thesis is dedicated to my wife Sadia. We have now completed our third degree together; the past eight years of university have flown by, you have made every moment of it special and incredibly memorable. We now start on our next step, and our ultimate competition, which one gets to be professor first.

XXXXX

# *"Nothing in Biology Makes Sense except in the Light of Evolution"*

Theodosius Dobzhansky
1973

# Contents

# List of figures



The total number of incidences of *S*. Derby and *S*. Mbandaka in a subset of hosts, in the UK between 2000 and 2010. It can be clearly seen that *S*. Mbandaka is most predominantly isolated from cattle and chickens and *S*. Derby from pigs and turkeys.



The top four most common hosts for a) *S*. Derby and b) *S*. Mbandaka. For *S*. Derby the most predominant host is pig and for *S*. Mbandaka poultry. These two hosts where identified from the AHVLA and CDC statistics to be predominant hosts of isolation in the UK and USA.



Global distribution of papers used to perform a meta-analysis for a) *S*. Derby and b) *S*. Mbandaka. The scale range was calculated using natural Jenks of number of publications per country for *S*. Derby. There was no significant differences between the number of papers originating from each country ($p<0.05$). It is clear that most of the publications for both serovars originate from the USA, UK, Canada and Brazil, and in addition for *S*. Mbandaka Italy.



The proportions of total publications for the a) UK and b) USA for hosts identified from the CDC and AHVLA statistics, with the addition of animal feed. There were no publications mentioning turkeys as the host of isolation for the UK. Pigs and chickens were consistently the most common host of isolation for *S*. Derby and *S*. Mbandaka, in the literature and isolation statistics for the CDC and AHVLA.

associated proteins, green a DNA binding protein, yellow a membrane protein, brown a regulatory protein and grey a conserved hypothetical protein. Dark pink identifies an ORF that was predicted to be an effector protein by SIEVE with a p-value of 0.05 or lower.



SPI-23 four way nucleotide comparison. Four way comparison of the nucleotide sequence of SPI-23 from *S*. Derby D1*, S*. Agona SL483, *S*. Dublin CT02021853 and *S*. Gallinarum RKS5078. *S*. Derby D1 possess the largest SPI-23 (37Kb) island of the sequenced strains available on NCBI genome and the most novel in nucleotide sequence. Over 60% of the nucleotide sequence of SPI-23 in *S*. Derby D1 is unique and contains no entry on NCBI nucleotide database.



a) Association and b) invasion assays performed for 0, 5, 15, 30 and 60 minute incubation periods (+/-1SEM). There are consistently more *S*. Derby cells inside the IPEC-J2 monolayer than there are *S*. Mbandaka cells (a). Up to 10 minutes after inoculation *S*. Mbandaka strains associate with the IPEC-J2 monolayer in higher numbers then *S*. Derby strains after which *S*. Derby increases in association rate (b). By 60 minutes there are 2.5 times as many *S*. Derby cells associated to, and 4 times as many cells inside the monolayer then there are *S*. Mbandaka cells.



Fold differences in qRT-PCR expression levels of ten genes found on SPI-23, in the order they appear on the island, when exposed to porcine jejunum and colon *ex plants* relative to a no tissue control. In both tissues all the genes are up regulated relative to a no tissue control. The genes *potR, genE*, *sadZ*, *tinY* and *docB* were up-regulated to a significantly

greater degree when exposed to the porcine jejunum compared to the colon tissue relative to a no tissue control.



Comparison of morphological and structural features of the parental wild type strain *S.* Derby D1 and mutant strain *S.* Derby D1 *ΔpotR::kan.* a) Mutant cultures agglutinate forming a single cell mass after being left static at room temperature for 2 hours. b) Mutant colonies form larger and fewer colonies on LB plates then the parental wild type strain, the mutant is also surrounded by a translucent skirt which is missing from the parent strain. c) Negative electron microscopy shows that the mutant strain is heavily pilated while the parent strain has few pili and some type-1 fimbriae (both are highlighted with arrows).



Differential stain confocal microscopy of a) *S.* Derby D1 and b) *S.* Derby D1 *ΔpotR::kan* that have adhered (yellow) and invaded (red) IPEC-J2 monolayers (blue). It can clearly be seen that a fewer cells of the mutant strain a) associated to the monolayer surface then the parent strain b). The parent strain can clearly invade the monolayer, there is no evidence of this with the mutant strain.



Respiration (tetrazolium dye intensity) positively correlates with growth (absorbency $OD_{540}$), with a correlation coefficient of 0.72. The regression line explains 52% of the variance in respiration as a product of growth.

region 1 and 2, whereas D12 contains these regions. *S*. Derby lineage 2 isolates D4 and D14, both contain SPI-1 region 1 and 2 and lack all of SPI-23 with the exception of *docB*.



Variation in phenotypic characteristics and plasmid constituents in relation to a phylogenetic cladogram. Antimicrobial resistance (AMR) abbreviations correspond to: nalidixic acid (NA), tetracycline (T), neomycin (N), ampicillin (AM), sulphamethoxazole/trimethoprim (SXT), chloramphenicol (C), streptomycin (S) and sulphonamide compounds (SU). All *S*. Mbandaka isolates with the exception of M9 formed a biofilm at 25$^{o}$C within 48 hours. *S*. Derby lineage 2 isolates and D8 formed biofilms within 48 hours at 25$^{o}$C, only one lineage 1 isolate, D10, formed a biofilm, this was at 37$^{o}$C within 24 hours. *S*. Derby lineage 1 isolates contained 17 plasmids, lineage 2 isolates only contained 1. Only isolates from lineage 2 lacked plasmids. All *S*. Derby isolates with the exception of D6 and D11 were resistant to tetracycline, *S*. Mbandaka isolates were either resistant to sulphonamide compounds or were sensitive to all antimicrobial compounds.



Variation between the number of colony forming units per ml (CFU/ml) of each isolate recovered after a) association and b) invasion assays performed on IPEC-J2 monolayers. Error bars represent +/-1SEM. a) There was no lineage based trend observed in the results for the association assays. b) *S*. Derby lineage 1 isolates D1 and D2 associate in significantly greater numbers ($p< 0.05$) then all the isolates with the exception of *S*. Mbandaka M4 and M8 ($p> 0.05$). This trend may reflect an adaptation towards porcine pathogenesis, with lineage 1 isolates invading to a greater extent than lineage 2 isolates and non-clustering isolates D8 and D9.

Respiratory dynamics of isolates of *S*. Derby lineage 1 and 2, and *S*. Mbandaka on a) jejunum, b) colon and c) soybean homogenates, measured through reduction of a tetrazolium dye. a) There was no significant difference between dye intensity values of the three lineages on jejunum homogenate over the 24 hour incubation period at 37$^{o}$C. b) The dye intensity for *S*. Derby lineage 1 was significantly different ($p < 0.05$) from that of lineage 2 and *S*. Mbandaka after 13.5 and 10 hours respectively, when incubated at 37$^{o}$C. c) All strains respired on soybean homogenate, though after a period of stationary respiratory dynamics amongst *S*. Mbandaka isolates, a second period of respiration was observed. These dynamics are similar to those seen during carbon catabolite repression.

# List of tables

*S*. Derby and *S*. Mbandaka isolates were selected from background monitoring isolations, between 2000 and 2010 across the UK. Strains of *S*. Derby were isolated from pigs and turkeys, and strains of *S*. Mbandaka were isolated from chickens and cattle. Isolate number refers to the AHVLA reference number and should be used to request strains, for simplicity in this paper we refer to the strain identifiers.

# Supplementary material

**Supplementary material 1.1.....** On attached CD

Meta-analysis of global research literature regarding the host and antimicrobrial resistance profile of *S*. Derby and *S*. Mbandaka.

**Supplementary material 2.1.....** On attached CD

Differences in gene and phage complement, the genes flanking the inversion, and the CRISPR spacer sequences, between isolates D1, D2, M1 and M2.

**Supplementary material 3.1.....** On attached CD

Primer sequences used for mutant construction, qRTPCR, checking for genomic DNA contamination of RNA and confirmation of kanamycin cassette insertion at the *potR* locus.

**Supplementary material 5.1.....** On attached CD

Biomass reaction and LB virtual media compositions. Compounds are identified by the Model SEED identifier, the chemical name can be found on the Model SEED.

**Supplementary material 5.1.....** On attached CD

LB media formulation

**Supplementary material 5.2.....** On attached CD

Metabolic model for *S*. Derby D1

**Supplementary material 5.3**: **.....** On attached CD

Metabolic model for *S*. Mbandaka M1

# Abbreviations

**ACT:** Artemis Comparison Tool

**AHVLA:** Animal Health and Veterinary Laboratories Agency

**AM:** Ampicillin

**AMR:** Antimicrobial Resistance

**ANOVA:** Analysis Of Variance

**ArcGIS:** Arc Geographical Information System

**ATP:** Adenine Triphosphate

**BATMAN:** Bayesian Automated Metabolite Analyser for NMR spectra

**BGA:** Brilliant Green Agar

**BLASTn:** Basic Local Alignment Search Tool Nucleotide

**BLASTp:** Basic Local Alignment Search Tool Protein

**bp:** Base Pair

**BSAC:** British Society for Antimicrobial Chemotherapy

**C:** Chloramphenicol

**CDC:** Centre for Disease Control and prevention

**CDP:** Cytidine Diphosphate

**CDS:** Coding DNA Sequence

**CFU:** Colony Forming Unit

**CHCL3:** Chloroform

**coa:** Co-enzymeA

**COG:** Clusters of Orthologous Groups

**CRISPR:** Clustered Regularly Interspaced Short Palindromic Repeats

**CT:** Threshold Cycle

**DAPI:** 4',6-diamidino-2-phenylindole

**df:** Degrees of Freedom

**DGL:** D-Galactonolactone

**DMEM:** Dulbecco's Modified Eagle Medium

**DMT:** Drug/Metabolite Transporter

**DNA:** Deoxyribose Nucleic Acid

**dTDP:** Thymidine diphosphate

**FBA:** Flux-Balance Analysis

**FID:** Free Induction Decay

**FIGfam:** Fellowship for Interpretation of Genomes Families

**Gb:** Gigabyte

**GDP:** Guanosine Diphosphate

**GI:** Gastrointestinal

**GTP:** Guanosine Triphosphate

**1HNMR:** Proton Nuclear Magnetic Resonance

**HIB:** Brain Heart Infusion

**HPA:** Health Protection Agency

**HSD:** Honestly Significant Difference

**IPEC-J2:** Porcine Small Intestinal Epithelial Cell Line

**ITSS:** Insulin-Transferrin-Sodium Selenite

**IVOC:** *In Vitro* Organ Culture

**Kb:** Kilobyte

**KDO2:** 2-keto-3-deoxyoctonate

**KEGG:** Kyoto Encyclopaedia of Genes and Genomes

**KYA:** Thousand Years Ago

**LB:** Lysogeny Broth

**LPS**: Lipopolysaccharide

**Ltr:** Litre

**Mb:** Megabyte

**MeOH:** Methanol

**MHz:** Mega-Hertz

**MILP:** Mixed-Integer Linear Programming

**MLST:** Multi-locus Sequence Typing

**MSRV:** Rappaport Vassiliadis Medium

**N:** Neomycin

**NA:** Nalidixic Acid

**NADH:** Nicotinamide Adenine Dinucleotide

**NCBI:** National Centre for Biotechnology Information

**NI:** No Information

**NTS:** Non-typhoidal *Salmonella*

**OD:** Optical Density

**OPM:** Omnilog Phenotypic Microarray

**ORF:** Open Reading Frame

**PAI-DB:** Pathogenicity Island Database

**PBS:** Phosphate Buffered Saline

**PCR:** Polymerase Chain Reaction

**pfam:** Protein Families Database

**PhD:** Philosophical Doctorate

**PM:** Phenotypic Microarray

**qRTPCR:** Quantitative Real Time PCR

**RAM:** Random Access Memory

**RAST:** Rapid Annotation using Subsystems Technology

**RD:** Recycle Delay

**RNA:** Ribose Nucleic Acid

**ROX:** 6-Carboxyl-X-Rhodamine

**rpm:** Revolutions Per Minute

**RPMI:** Roswell Park Memorial Institute

**S:** Streptomycin

**SEM:** Standard Error of the Mean

**SNP:** Single Nucleotide Polymorphism

**SPI:** *Salmonella* Pathogenicity Island

**SYBR:** Synergy Brands

**SU:** Sulphomanides

**SXT:** Sulphamethoxazole/trimethroprim

**T:** Tetracycline

**TBE**: Tris (Hydroxymethyl) Aminomethane Borate Ethylenediaminetetraacetic Acid

**TIGR:** The Institute for Genomic Research

**tm:** Mixing Time

**tRNA:** Transfer RNA

**TTSS:** Type Three Secretion System

**UDP:** Uridine Diphosphate

**USDA:** United States Department of Agriculture

**V:** Voltage

**VMD:** Veterinary Medicines Directorate

# Chapter 1:
# **Introduction**

The arcGIS publication distribution maps were produced by Dr. Sadia E. Ahmed

## 1.1 *Salmonella enterica* serovars

*Salmonella enterica* is an important zoonotic pathogen of warm-blooded vertebrates, with both a broad host species range and geographical distribution. *S. enterica* can be divided into seven subspecies, subspecies 1, *S. enterica sub enterica*, can be further divided into over 1530 serovars based on the different epitopes of two surface antigens, the "O", lipopolysaccharide, and "H" flagellum of which there are often two phases (Grimont and Weill 2007). It has been shown that the "O" antigen is recognised by a diverse range of host specific protozoan predators. As the complement of protozoa has been found to vary between different host species, the selection pressure posed by the predators has been posited as a driving force behind the diversification of the *S. enterica* "O" antigen (Wildschutte et al. 2004; Wildschutte & Lawrence 2007). The diversity between the two "H" phase antigens has been associated with immunogenicity. The ability to switch phase allows *S. enterica* to escape detection by the adaptive immune response that has developed in response to the "H" epitope of the initial phase (Hayashi et al. 2001). Both antigens are under selection pressure from the host environment; "O" from the gut microflora and "H" from the host's immune system. Therefore, there is the potential for convergent evolution to occur towards epitopes which increase fitness in particular host species, resulting from the unique sets of selection pressures on different serovars.

## 1.2 Host and environment: association, adaptation and restriction

Salmonellosis is commonly associated with gastroenteritis, although the prognosis of the disease resulting from pathogenesis can be largely determined from identification of the host species and the *S. enterica* serovar. The relationship between different serovars and pathogenesis in different hosts has been well studied at the tissue and molecular level (Vladoianu et al. 1990; Hurme et al. 1997; Ong et al. 2010). A serovar can be classified

into one of three groups based on the proficiency of pathogenesis in a varying host range. Host adapted serovars may be transmitted between members of a single species; these serovars can be disseminated without the necessity of intermediary species (Uzzau et al. 2000). A fraction of these serovars cause systemic infections in a narrow host range and gastroenteritis in a broader host range for example *S*. Typhimurium (McCullough & Eisele 1951; Tsolis et al. 1999; Uzzau et al. 2000). Serovars such as *S*. Typhi, have only been found to cause a systemic infection in one host, in this instance humans, with few isolations from other species; these strains are considered host-restricted (Blaser & Newman 1982; Vladoianu et al. 1990; Uzzau et al. 2000). It has been shown that systemic infection is usually associated with a higher proportion of chronic carriers and long term shedders (Buchwald & Blaser 1984). The genome and plasmids of these strains frequently contain genes for manipulating the host immune system and environment to aid survival, transmission and to enrich the nutritional content of the intestine (Rubin & Weinstein 1977; Goldstein et al. 1986; Elsinghorst et al. 1989; Arricau et al. 1998). Some strains may be auxotrophic for metabolites that are abundant in the host range to which they are best adapted (Virgilio & Cordano 1981). Non-adapted serovars cause self-limiting gastroenteritis in a broad host range, the infection may be more aggressive than the adapted strains, though the mechanisms of infection are not host specific (Steinbach et al. 2000). Inoculation studies for the most economically important serovars have been carried out to observe the differences in pathogenesis amongst livestock and model species (McCullough & Eisele 1951; Lee & Falkow 1990; Jones 1994; Tsolis et al. 1999). Tissue culture has also been used to study the mechanisms of cellular infection (Lee & Falkow 1990; Jones et al. 1994; Carlson et al. 2000).

Host adaptation coupled with host restriction could result from overcrowding or intensive farming of a single variety of livestock, where food stuffs can regularly come in contact with contaminated faeces. The selection pressures are therefore specific to a single host and its surrounding environment, allowing the bacteria to specialise and out-compete other serovars, while the pressure to survive in a diverse set of environments is lost (Eswarappa et al. 2009; Mennerat et al. 2010; Pulkkinen et al. 2010). Serovars such as *S*. Gallinarum show a high degree of host restriction, in this instance the serovar has become adapted to pathogenesis in fowl. Motility, a property usually associated with increased virulence, has been lost from *S*. Gallinarum since it diverged from *S*. Enteritidis an estimated 28-42 thousand years ago (Eswaraappa et al. 2009). This could explain the host restriction of *S*. Gallinarum and the broad host range of *S*. Enteritidis. A comparative genomics study performed between *S*. Gallinarum and *S*. Enteritidis by Eswarappa et al (2009) showed that only a small subset of pathogenicity related genes have been under positive selection since the serovars diverged. It was found that 13 of the 40 genes found on *Salmonella* pathogenicity island 2, SPI-2, a genomic island which is highly conserved in *S. enterica*, were under positive selection. While only 3 of the 40 genes in SPI-1, another highly conserved genomic island, were under positive selection (Eswarappa et al. 2009). Jones et al (2001) showed that mutation of SPI-1 in *S*. Gallinarum did not affect virulence in chickens though the role of SPI-2 for survival inside macrophage was essential for virulence (Jones et al. 2001). In contrast, SPI-1 has been shown to be essential for the colonisation of the liver and spleen of chickens by *S*. Enteritidis (Rychlik et al. 2009). *S*. Enteritidis elicits an inflammatory response in chickens whereas *S*. Gallinarum does not (Eswarappa et al. 2009). Hence the divergence in nucleotide sequence of many of the SPI-1 genes in *S*. Gallinarum may pertain to the role of these genes in an alternative route of infection utilised during pathogenesis in a wider number of host species. These genes may

therefore, undergo genetic drift or negative selection as a result of exposure to a single homogenous environment.

### 1.3 Discovery of host-serovar associations

Host associations can be identified from isolation statistics compiled during background monitoring programmes, such as those performed by the UK and USA, where a particular serovar is consistently isolated from the same sub-set of host species over several reports, spanning decades and large geographical areas. Even a cursory overview of the statistics available from the CDC, HPA, AHVLA, Pulsenet and other sources identify the majority of clinical instances of Salmonellosis result from infection of just a small number of serovars; the most predominant being *S*. Enteritidis, *S*. Typhimurium, *S*. Newport, *S*. Infantis and *S*. Montevideo (CDC PHLIS 2009; AHVLA 2011; CDC 2011). The data for animals is less rich but trace back studies and source attribution often identify a zoonotic origin for human infections; two common examples are the aforementioned serovars *S*. Typhimurium, which is predominantly associated with pigs, and *S*. Enteritidis, which is predominantly associated with poultry; though both serovars are also associated with many other livestock species (AHVLA 2011). These serovars are now well studied and can be considered broad range host adapted serovars. However, the available data show an apparent bias in terms of host association between two of the most frequently isolated serovars of veterinary significance, *S*. Derby and *S*. Mbandaka (AHVLA 2011). There is also a comparative lack of research in to these two serovars, particularly at the molecular level. These observations prompted me to consider these two serovars as candidates for comparative studies which could elucidate host adaptations that potentially influence host association.

Background monitoring statistics available from the AHVLA, HPA (personal communication) and CDC PHLIS were used to identify the host species associations of these two serovars (CDC PHLIS 2009; AHVLA 2011; CDC 2011). Both the HPA and CDC statistics showed that these serovars could cause a disease in humans with 215 and 433 clinical incidences of *S*. Derby and *S*. Mbandaka respectively in the UK between 2003 and 2009. Comparably there were 1614 and 2169 clinical incidences of *S*. Derby and *S*. Mbandaka in the USA between 1999 and 2009. This may suggest that *S*. Mbandaka poses a greater public health concern. *S*. Derby and *S*. Mbandaka are consistently in the top 20 most commonly isolated serovars from livestock in the USA. From UK and USA incidences statistics it is clear that these two serovars associate with different sets of livestock species. In the UK, approximately 50% and 40% of incidences of *S*. Derby were in turkeys and pigs, respectively, and approximately 20% and 65% of incidences of *S*. Mbandaka were in cattle and chickens, respectively (Figure 1.1). Similarly in the USA approximately 80% of isolations of *S*. Derby were from pigs, while only 3% of isolations were from turkeys, 27% and 25% of *S*. Mbandaka isolations were from cattle and chickens. These host distributions have been maintained over a decade and on two continents suggesting that the mechanisms that maintain these patterns are likely to be biological properties of the host-pathogen interaction.

### 1.4 Meta-analysis of global literature

The trends drawn from clinical and veterinary isolation statistics for the UK and USA suggest that there is association between *S*. Derby and *S*. Mbandaka and distinct host species ranges. The benefit of comparing these isolation statistics is that they are geographically isolated; this reduces the likelihood of confounding factors such as point of contact contamination and changes in isolation frequency due to abrupt changes in farming

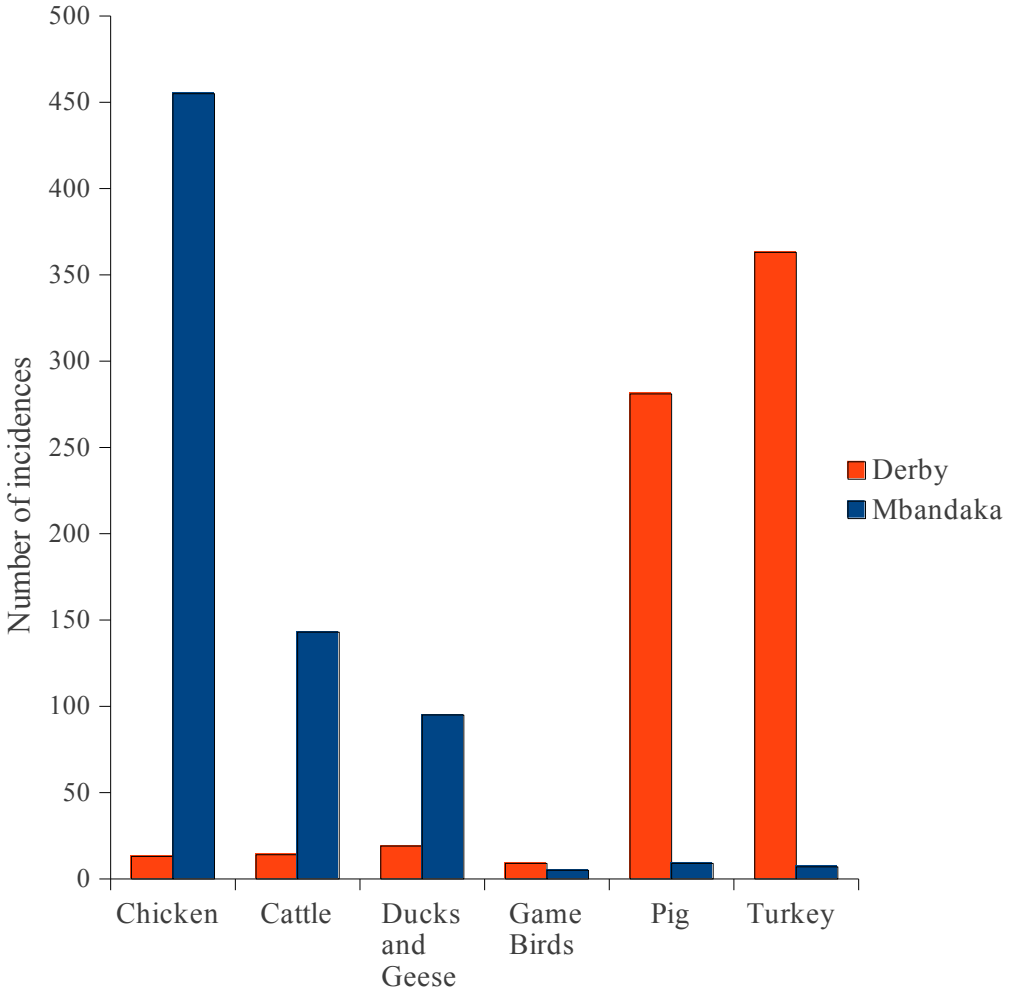



**Figure 1.1:** The total number of incidences of *S.* Derby and *S.* Mbandaka in a subset of hosts, in the UK between 2000 and 2010. It can be clearly seen that *S.* Mbandaka is most predominantly isolated from cattle and chickens and *S.* Derby from pigs and turkeys.

practice. Ideally the comparison would be done with global statistics but unfortunately few countries have the monitoring resources that the UK and USA possess. So, in an attempt to circumvent this limitation I have performed a meta-analysis of the global literature, published in English, for *S*. Derby and *S*. Mbandaka using a geographical information system ArcGIS and Google Scholar search engine (ESRI 2011; Google 2013).

Two searches were performed on 30/11/11 using Google Scholar, removing patent applications, with the search terms "*Salmonella* Derby" and "*Salmonella* Mbandaka"; this returned 2486 and 1034 unique publications respectively. I reviewed the first 1000 papers for *S*. Mbandaka and managed to access 166 with the help of AHVLA library. I then obtained the first 166 accessible papers for *S*. Derby from the first 1000 papers of the search (an excel spreadsheet containing the references, meta-data and groupings can be found in Supplementary material 1.1). From these papers, I recorded the year of the study, the antimicrobial resistance profile and the host of isolation. These data were visualised using ArcGIS and Office Libre Calc (ESRI 2011; The Document Foundation 2012). Two-tail paired t-tests using Office Libre Calc were performed, paired by country and paired by decade of publication, to check if there were publication biases towards certain years or regions for either serovar.

The meta-analysis showed that the host most associated with *S*. Derby and *S*. Mbandaka in the literature were hosts that these serovars were frequently associated with in the statistics published by CDC and AHVLA (Figure 1.2) (CDC PHLIS 2009; AHVLA 2011; CDC 2011).

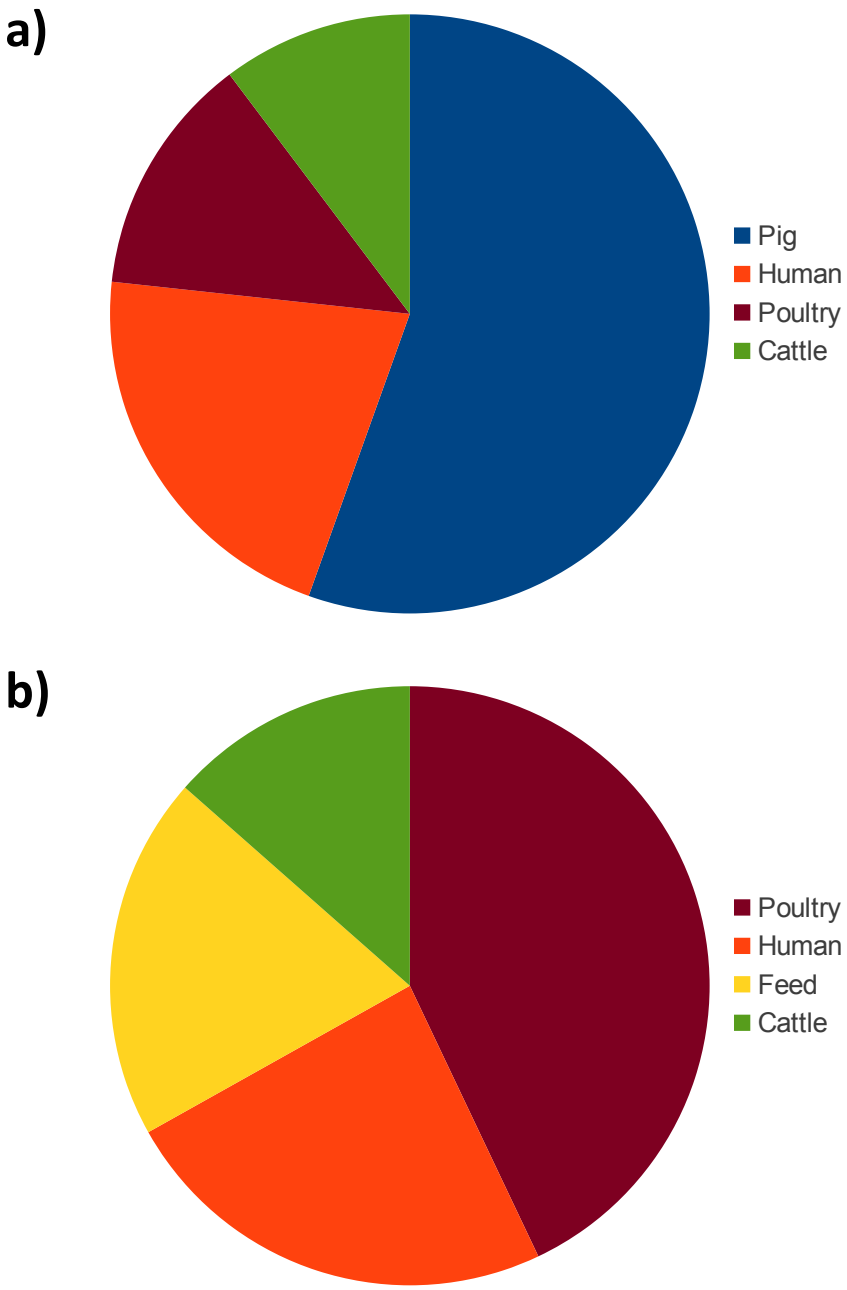



**Figure 1.2:** The top four most common hosts for a) *S.* Derby and b) *S.* Mbandaka. For *S.* Derby the most predominant host is pig and for *S.* Mbandaka poultry. These two hosts where identified from the AHVLA and CDC statistics to be predominant hosts of isolation in the UK and USA.

In total 81 papers out of 166 identify *S*. Derby as being isolated from pigs, whereas 70 papers identify *S*. Mbandaka as being isolated from poultry. For both serovars humans were the second most common host of isolation. The third host mentioned most in the global literature for *S*. Mbandaka was animal feed. Cattle are also one of the preferred hosts of *S*. Mbandaka and were the fourth most common host mentioned in the literature with 22 entries. The 3rd and 4th most commonly mentioned host in relation to *S*. Derby were poultry and cattle. These are not preferred hosts of *S*. Derby identified from the isolation statistics and this may hint at either difference in host range across the globe or a limitation of the meta-analysis approach to identifying host associations. The poultry category contains 3 papers for each serovar that identify turkey as the host of isolation. Turkey was identified from the isolation statistics as a preferred host of *S*. Derby and not *S*. Mbandaka.

There was no significant difference ($p>0.05$) between serovars for the number of papers per decade or the country of isolation (Figure 1.3). This suggests that there is no temporal or spatial publication bias. The global distribution of publication shows that the USA and Europe produce the majority of publications regarding these two serovars. In total 66% and 55% of USA and UK publications respectively (Figure 1.4), state pig as the host of isolation for *S*. Derby, this fits with the USA statistics, where the majority of isolations are from pigs. The lack of publications on turkeys for UK, may explain why this proportion is higher than 40% identified from UK isolation statistics. In both cases, *S*. Derby has been identified as being associated with pigs. The proportion of cattle and chicken isolations in USA publications, 34% and 40%, is approximately the same as the proportion of *S*. Mbandaka isolated, 27% and 25%, from these two hosts in the USA. For papers from the UK, poultry was identified in 40% of papers as the host of isolation; this is lower than the 65% identified from the isolation statistics.

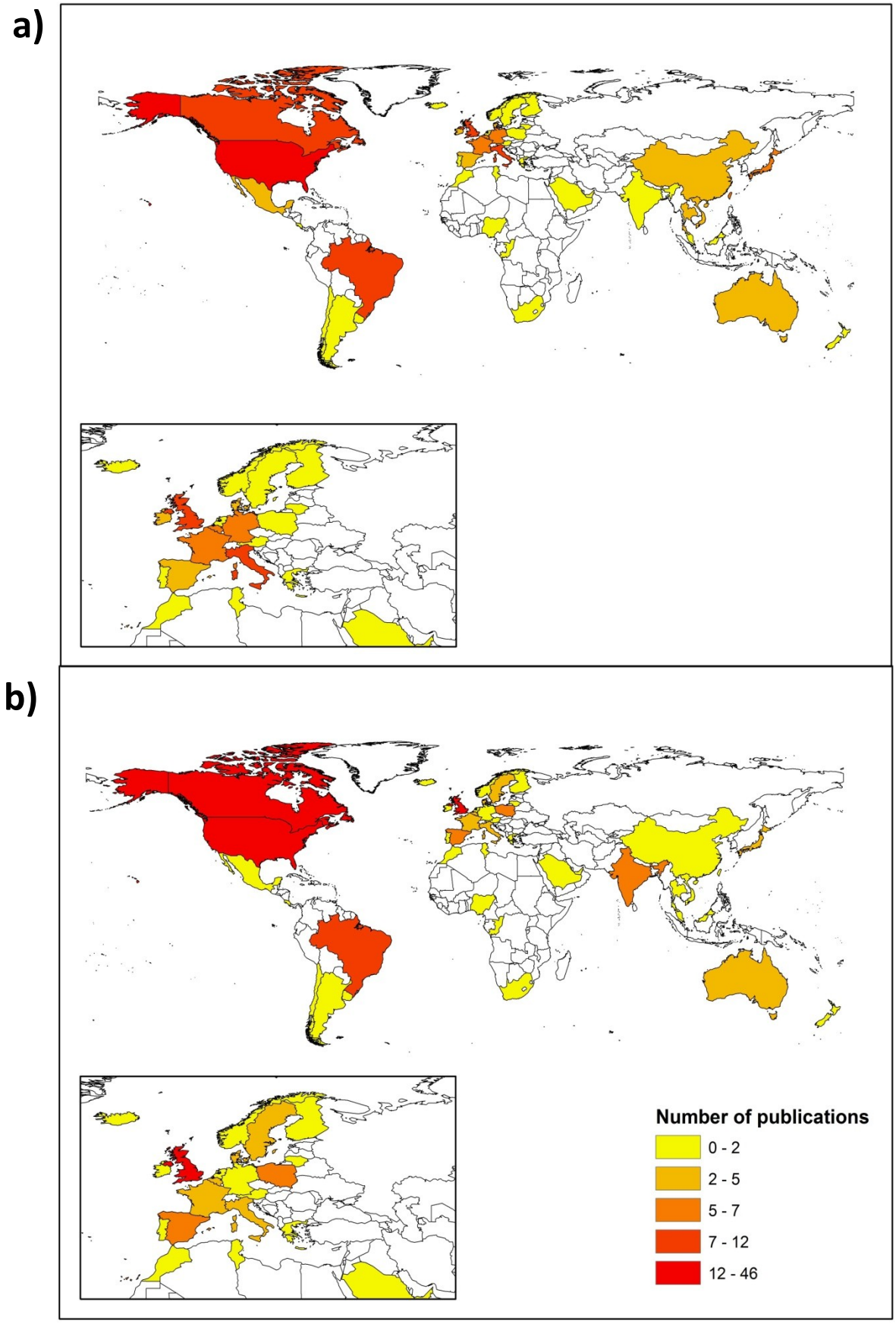

a)
b)
Number of publications
0 - 2
2 - 5
5 - 7
7 - 12
12 - 46

**Figure 1.3:** Global distribution of papers used to perform a meta-analysis for a) *S*. Derby and b) *S*. Mbandaka. The scale range was calculated using natural Jenks of number of publications per country for *S*. Derby. There was no significant differences between the number of papers originating from each country ($p<0.05$). It is clear that most of the publications for both serovars originate from the USA, UK, Canada and Brazil, and in addition for *S*. Mbandaka Italy.

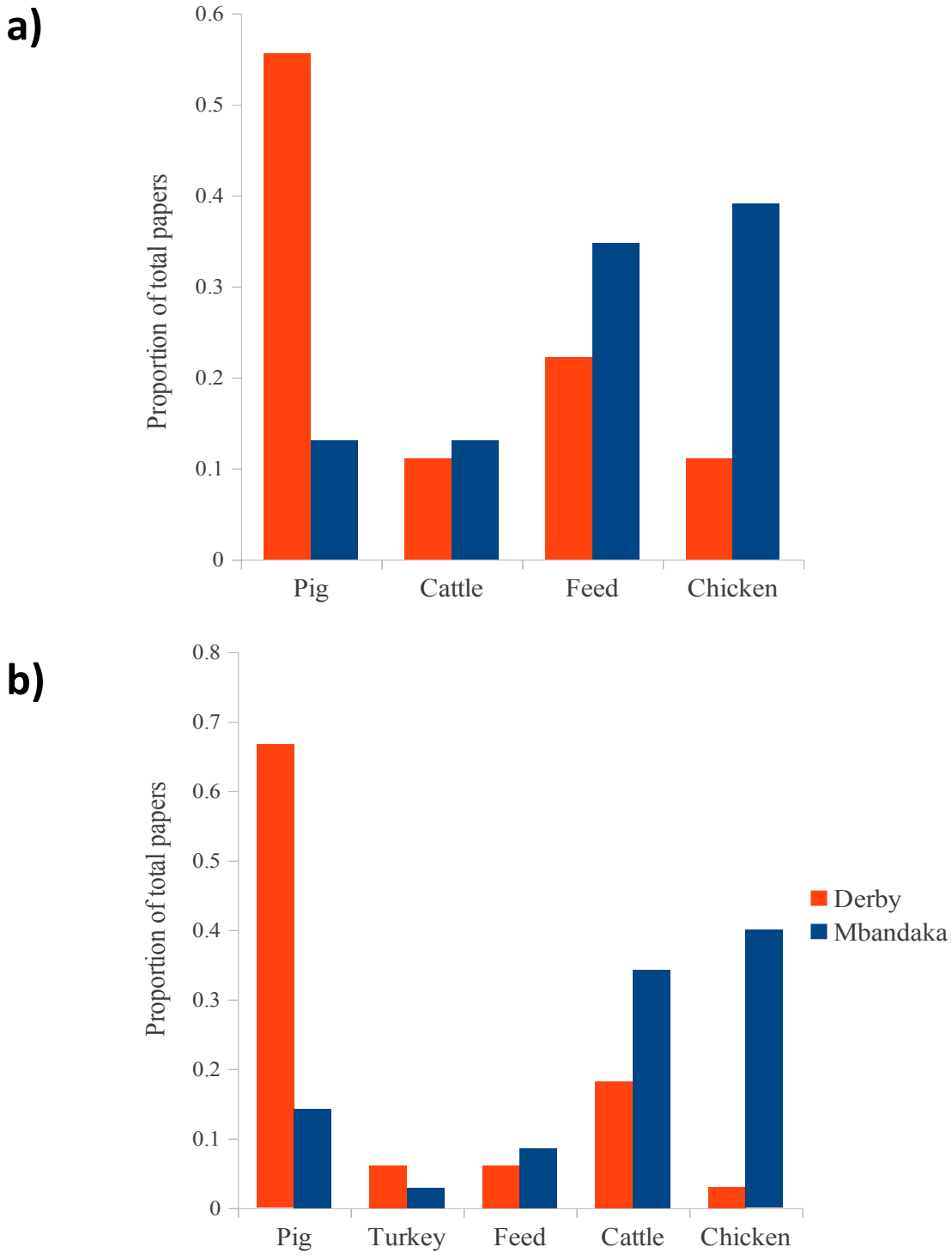



**Figure 1.4:** The proportions of total publications for the a) UK and b) USA for hosts identified from the CDC and AHVLA statistics, with the addition of animal feed. There were no publications mentioning turkeys as the host of isolation for the UK. Pigs and chickens were consistently the most common host of isolation for *S.* Derby and *S.* Mbandaka, in the literature and isolation statistics for the CDC and AHVLA.

Similarly the proportion of papers identifying cattle is much lower than may be expected if the number of publications tracked the isolation frequency. This may be explained as the USA had 45 and 46 publications for *S*. Derby and *S*. Mbandaka respectively while the UK only had 11 and 15 publications respectively. This may suggest that the higher the number of publications the more accurately the proportions of papers identified with one host more closely tracks the isolation statistics.

It is worth noting that there are a number of other metrics that could influence the number of publications originating from each country, especially in the English language, some of these being: the relative size and economic value of the livestock sector; the number of veterinary institutes and schools; the number of active researchers in the field in the country; the amount of commercial funding for health and welfare research into livestock; the source of research funding; and the potential impact factor of epidemiological papers. With these caveats identified I will now infer, cautiously, the potential differences in host association of these serovars between two geographical regions, these being, China in comparison with India, and Poland in comparison with Germany and Italy (Figure 1.5).

There were 5 papers from China, all of which were for *S*. Derby. Out of these, 3 papers identified humans, 2 pigs, 1 chicken, 1 cattle, 1 goose and 1 sheep as the host of isolation. This is quite a large distribution, though humans and pigs have a slightly higher number of publications. With further analysis we may expect to find more papers mentioning pigs. This would fit with the higher annual consumption rate of pork in China when compared to beef and chicken, which is almost ten and four times greater respectively (Vijayarashayan 2012; McMilan 2013).

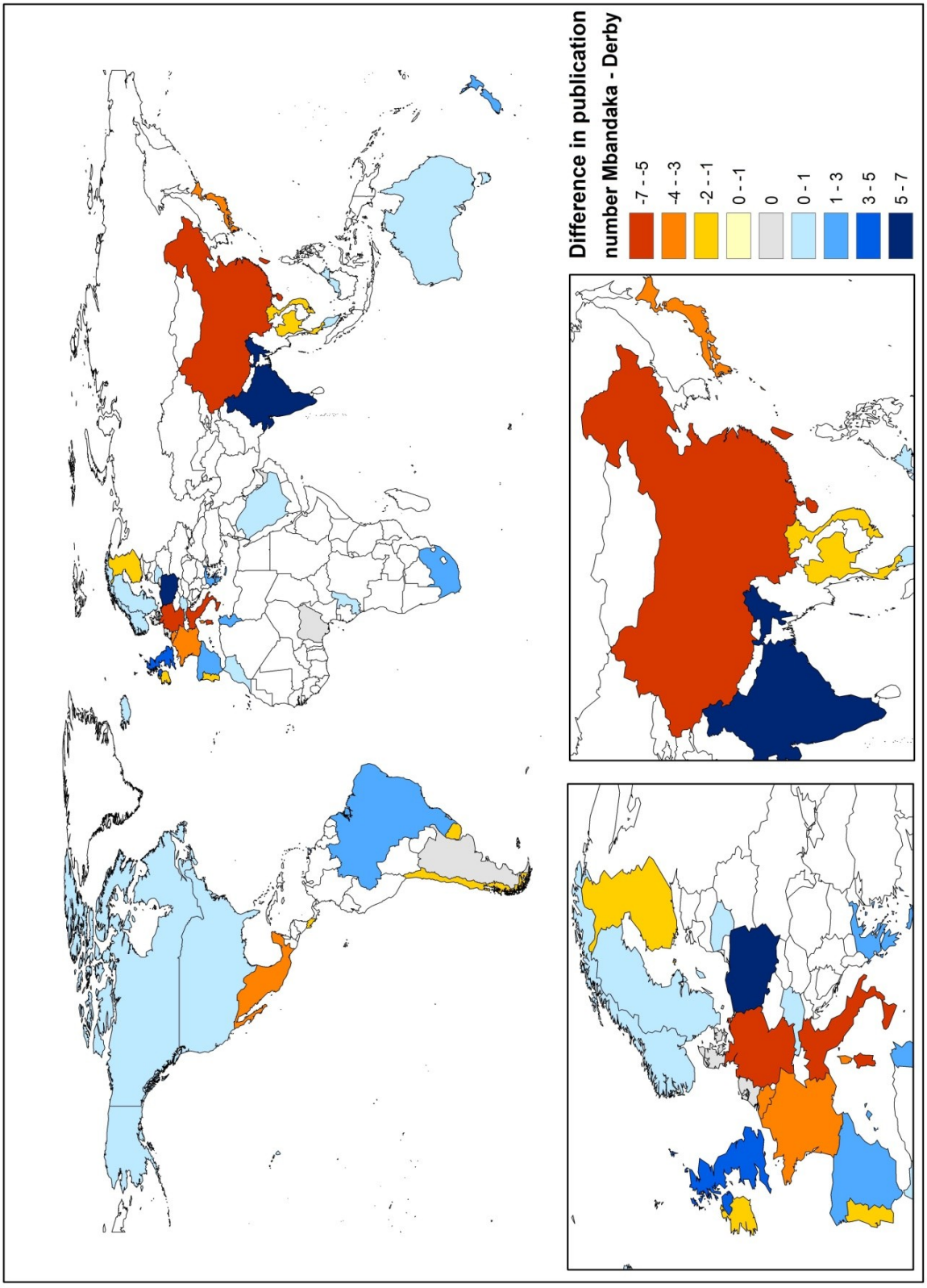



**Figure 1.5:** Publication bias are visualised as number of publications regarding *S.* Mbandaka minus *S.* Derby on a global map. Two regions with local publication biases are focused on to highlight the potential for cultural differences which could lead to biases in isolation of one serovar over another.

China is a net importer of pigs (AHVLA export test data), so it might be anticipated that endemic issues at source may be readily transferred to the importing country. It can be argued that *S*. Derby in China was already an issue and the imports may have had an additive effect. There is insufficient published data to interrogate this further and the findings for this analysis may be biased by use of English as a selection criterion for the meta-analysis.

There were 6 papers from India, all of which were for *S*. Mbandaka. Three of these papers identified chickens as the host, 2 humans, 1 cattle, 1 feed, 1 seafood, 1 egg and 2 vegetable. India has a large number of vegetarians and an estimated 65 million cows which are mainly used for agricultural means (Phaniraja & Panchasara 2009). India only has an estimated 12 million pigs (Aragon 2011). The rate of chicken consumption is on the increase, as the country becomes more economically developed, the demand for meat derived dietary protein increases (Athale et al. 2010). Cattle and chickens were the preferred hosts identified from the UK and USA statistics for *S*. Mbandaka, therefore we may expect the incidences of isolation, and hence publication, to increases as chicken consumption increases.

India and China have published on only one of the two serovars each, which may reflect their differences in dietary preferences. Expanding the meta-analysis to focus on just papers from these two regions and in their native languages could potentially reinforce this trend and show if publication bias on one of these two serovars tracks the host species population and cultural differences between these countries in their attitude towards different foods.

Another interesting pattern is that seen between these geographically close countries, Germany, Italy and Poland. Poland has six publications for *S*. Mbandaka which mention both humans and animal feed, with no papers identified for *S*. Derby. While Germany and Italy mention *S*. Derby in pigs 6 and 5 times respectively. There is one paper for *S*. Mbandaka from Germany; it identifies the host of isolation as feed. There are no papers mentioning *S*. Mbandaka from Italy. This may suggest that in Poland, *S*. Mbandaka is more of a public health concern than in Germany or Italy, and that *S*. Derby in pigs, is a bigger veterinary concern in the latter two than in Poland.

A larger meta-analysis, with serovars possessing well characterised host associations, such as *S*. Typhimurium, a broad range serovar, and *S*. Gallinarum, a host specific serovar, to name a few, could help validate the use of literature meta-analysis to identify host associations. This may also shed light on other cultural differences with regards to food preferences, availability and farming practices that potentially influence the prevalence of particular serovars in a country.

*S*. Derby and *S*. Mbandaka were identified as being resistant to 32 and 31 different antimicrobials respectively. The top four most frequently identified compounds and profiles and the number of publications for *S*. Derby were “Not Specified” 118, “Tetracycline” 39, “Chloramphenicol” 25 and “Streptomycin” 25, and for *S*. Mbandaka were “Not Specified” 39, “No Resistance” 16, “Gentamicin” 14 and “Ampicillin” 14 (Figure 1.6). This shows that there is potentially a difference in association between serovars and particular antibiotic resistance. *S*. Mbandaka was identified as being resistant to tetracycline in 13 papers, a third of the number that *S*. Derby is identified as having resistance in.

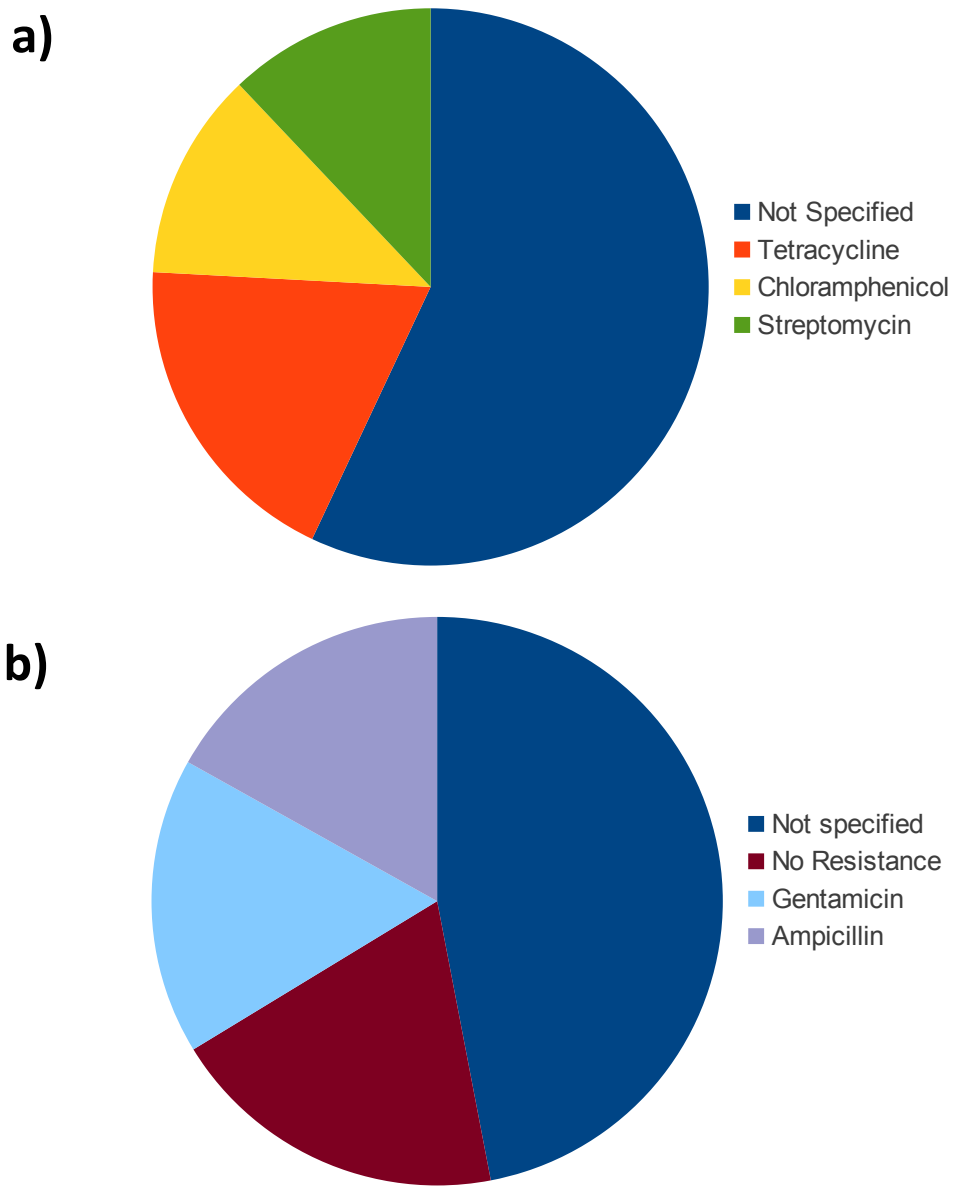



**Figure 1.6:** The majority of publications for both a) *S.* Derby and b) *S.* Mbandaka did not specify antimicrobial profiles. The most commonly reported resistance profile for *S.* Derby included tetracycline resistance, while the majority of *S.* Mbandaka publications reported antimicrobial sensitivity to all tested compounds.

The antimicrobial resistance profile was not identified in three times as many *S*. Derby papers than *S*. Mbandaka.

For *S*. Derby it would appear that from UK and USA isolation statistics and the meta-analysis of the global literature that it is most frequently isolated from pigs, whereas for *S*. Mbandaka it would appear to be poultry. There is also evidence from the meta-analysis to suggest that *S*. Mbandaka may be associated with animal feed.

### 1.5 Host association, competition and the habitable niche

The predominant isolation of one serovar from a particular host over another serovar could be the result of a higher abundance of that particular serovar due to more proficient enrichment in and around the host, or due to exclusion of the other serovar due to the lack of stress resistance to hostile conditions. These two concepts are formalised as competitive exclusion and the habitable niche (Hibbing et al. 2010). As mentioned above, a high number of publications on host association focus on the role of different mechanisms involved in pathogenesis. These adaptations do not necessarily lead to enrichment within the host and therefore to a higher relative fitness. These adaptations may not influence the dominance of one serovar over another, but rather influence the severity of the associated symptoms (Seth-Smith 2008). An example of a mechanism of host adaptation that lies outside of traditional studies is the influence of the metabolite composition of the host, and the associated environmental conditions, on the proficiency of bacterial growth (Raghunathan et al. 2009). Potentially as important as adaptations to the *in vivo* environment of the host is survival externally prior to re-infection. One such adaptation is the formation of a biofilm; this conveys a degree of resistance to desiccation, UV damage and bacterial disinfectants (Elasri & Miller 1999; Joseph et al. 2001; Lianou &

Koutsoumanis 2012). Once within a biofilm the bacteria enter a reduced metabolic state (White et al. 2010). This reduced growth, can lower the efficacy of some biocides as they target dividing cells (Gilbert et al. 1990). A subset of *S. enterica* serovars go beyond just persistence in the environment external to the gut. Some serovars have been shown recently to possess adaptations for pathogenesis in plants (Schikora et al. 2008; Schikora et al. 2012). The veterinary and clinical significance of this behaviour is that serovars that are able to colonise and become enriched in plant tissue provide a larger ingested dose of bacteria than those that enter into a period of dormancy (Schikora et al. 2008). Factors affecting the habitable niche of *S. enterica* include the ability to withstand particular antimicrobial compounds, high salinity, heavy metals and high temperatures (Barkay et al. 2003; Rychlik & Barrow 2005).

### 1.6 Systems biology

Ecology is arguably the oldest systems level study; investigating the behaviour and interactions of communities of organisms with their surroundings. Using surveys of populations and habitats, some ecologists attempt to reconstruct a complete picture of the natural world and model the effects of anthropological perturbations. The modern movement in systems biology is concerned with the cellular world; an environment populated with genes, transcripts, proteins and metabolites. Theoretical systems attempt to model the unobservable, microscopic (biochemical) and macroscopic (ecological and sociological) interactions, that allude to the causes and dynamics of biological behaviours (Westerhoff et al. 2009).

Systems biology today is synonymous with the biochemical profile, the observation of all the components of a system, this can include multiple different organisms as in the case of meta-genomics, followed by intensive data mining, bioinformatics and in some cases mathematical modelling to produce a theoretical system that accounts for emergent behaviours, such as determining growth rate as a result of identifying the metabolic genes in the organisms genome. Systems biology is currently divided into a series of systems-level studies called "-omes"; chemical species within these layers share similar chemical properties or have a similar physiological role (Röling et al 2010). For the study of habitable niche and competitive advantage, comparison of large scale profiles such as the genome, transcriptome or proteome may identify potential functional differences between serovars, which can be used to develop hypotheses and devise experiments. The remainder of this introduction will briefly cover the genome, phenome and metabolic modelling relevant to a better understanding of the molecular cell biology of *S. enterica* serovars.

### 1.7 The chromosome of *Salmonella enterica*

The chromosome of *Salmonella enterica* comprises mostly of a core sequence, with a large proportion of the genetic diversity resulting from horizontally acquired blocks of genes in the form of genomic islands and prophage (Hensel 2004; Thomson et al. 2004; Anjum et al. 2005). Genomic islands may accelerate adaptation towards particular hosts through the integration of multiple pathogenicity genes in to the chromosome, which convey new phenotypes in a modular fashion; this has occurred several times in the evolutionary history of *S. enterica*. These particular genomic islands have been called *Salmonella* pathogenicity islands (SPIs) (Marcus et al. 2000; Hensel et al. 2004). It has been postulated that the integration of the pathogenicity genes found on SPI-1 led to the divergence of *Salmonella* from *E.coli* (Groisman & Ochman 1997). SPI-1 is found in all serovars of *S. enterica* (with

the exception of *S*. Seftenberg and *S*. Lynchberg) and is highly conserved in nucleotide sequence (Matsushita et al. 1996; Porwollik et al. 2002; Hu et al. 2008). The genes found on this island code for one of two type three secretion system (TTSS) typically found on the *Salmonella* chromosome, this one is utilised during invasion of epithelial cells. SPI-2 codes for the second TTSS and is utilised during the intracellular phase of pathogenesis (Waterman & Holden 2003). There are currently 23 published *Salmonella* pathogenicity islands identified from the genomes of *S. enterica* and *S. bongori*: the 23rd, SPI-23, was discovered during this study and is described, and partially characterised in this thesis (Fookes et al. 2011; Hayward et al. 2013). The gene content of some of these islands is highly plastic, as in the case of SPI-3 found in *S*. Dublin and *S*. Typhimurium LT2 the genes between the insertion sequences are completely different. Therefore, the SPI-3 locus is defined by its insertion sequences *selC* and *mgtB* (Amavisit et al. 2003; Hayward et al. 2013).

The broad host range and variation in pathogenesis amongst serovars of *S. enterica sub. enterica* have been studied at both the tissue and molecular level (Blanc-Potard & Groisman 1997; Baumler et al. 1998; Blanc-Potard et al. 1999; Knodler et al. 2002; Morgan et al. 2004). Many of these studies focus on the interfaces between host and pathogen; much is now known about the molecular interactions that lead to bacterial uptake and modification of the host environment. These mechanisms are elicited through site directed mutation and transformation studies (O'Callaghan & Charbit 1990; Ochman & Groisman 1996; Deiwick et al. 1998; Shea et al. 1999; Figueroa-Bossi et al. 2001; Fuentes et al. 2009).

Genome-wide studies of pathogenesis determinants have been made more available with the advent of increasingly cheaper and faster genome sequencing in the past decade ( Thomson et al. 2008; Izumiya et al. 2010; Jarvik et al. 2010).

Comparative genomics of broad host range serovar *S*. Typhimurium with host restricted serovars *S*. Typhi and *S*. Paratyphi highlighted the importance of gene decay in host restriction. Both host restricted serovars cause typhoid fever and rarely gastroenteritis in a human host, whereas *S*. Typhimurium does not cause typhoid fever in humans but causes gastronetiritis in a broad range of host species including humans. It was shown that both *S*. Typhi and *S*. Paratyphi had undergone gene decay in approximately 200 genes, yet only 30 of these pseudogenes were common between serovars (McCelland et al 2004). Some of these shared pseudogenes have been shown to contribute to the gastroenterititic phase of pathogenesis and suppression of the immune response during pathogenesis of *S*. Typhimurium. One notable example is the loss of the gene encoding the effector protein GtgE from the chromosome of *S*. Typhi, which targets the Rab32 lysosyme pathway in mouse macrophages. The loss of this gene since its divergence from *S*. Typhimurium, has been linked to the inability of *S*. Typhi to colonise non-human cells (Spano et al. 2011, Spano and Galan 2011).

It may be assumed that these 200 genes have undergone genetic drift due to neutral or negative selection, whereas, in serovars such as *S*. Typhimurium, the retention of functional copies of these 30 genes may signify their importance in the colonisation and the development of gastroentertitis in a broad range of species. Conversely, comparative genomics identified SPI-7 in the chromosome sequence of *S*. Typhi which is absent from the chromosome of *S*. Typhimurium (Pickard et al 2003). This island encodes the well characterised VI capsular antigen, which has been shown to increase infectivity while

conferring resistance to complement and phagocytosis occurring through TLR4/5-mediated LPS recognition (Looney and Steigbigel 1986, Hone et al 1988, Raffatellu et al 2005). From these studies it is evident that host restriction is not solely characterised by gene decay and loss of function but also adaptation towards the host and more proficient pathogenesis through acquisition of genes.

Comparative genomic studies of *Salmonella* serovars are becoming increasingly common. This may be explained as it is more economically viable to sequence whole genomes and analyse regions of interest rather than isolate and sequence multiple smaller regions of DNA. Publication of the full genome from these studies has allowed others to utilise the data. This wealth of genomic information has highlighted more subtle and previously obscured variation in the genome of *S. enterica* serovars and phage types. For example the relatively recent discover of CRISPR operons in *S. enterica* has, in part, been studied through alignment of publicly available genome sequences (Fricke et al. 2011).

In addition to the increased availability of genome sequencing the developments in automated annotation systems has sped up the time between sequencing and hypothesis generation. Systems such as RAST (rapid annotation using subsystems technology) have pioneered the integration of genetic, functional and modelling techniques in to a single annotation pipeline (Aziz et al. 2008). Clearly, the use of such approaches to generate information on potential differences in gene functions between *S*. Derby and *S*. Mbandaka genomes will enable the development of hypotheses and experiments as a starting point from which factors influencing the different host associations and habitable niches of these two serovars can be discovered.

## 1.8 Phenome

The phenome is defined as the complete set of phenotypes an organism can produce, however, this level of detail is not yet attainable, and as many phenotypes are studied singly (e.g. adherence, invasion, etc) and are descriptive making use of a variety of approaches (e.g. biofilm formation, motility, etc) the results produced are highly subjective (Toguchi et al. 2000; Joseph et al. 2001; Hammer & Bassler 2003; Caiazza et al. 2007; Oti et al. 2008; Searle et al. 2009). The recent development of a pan-phenotyping system, the Omnilog Biolog PM, permits the use of up to 20 different 96 well plates that each contain minimal media with different sources of carbon, sulphur, nitrogen, phosphorous, dipeptides and antimicrobials per well. The technique utilises a tetrazolium dye which is irreversibly reduced by NADH and enables a consistent approach to the study of respiration and growth dynamics that has been lacking in many study systems (Bochner et al. 2001; Bochner 2009). The Biolog systems has been offered as an alternative typing method to serotyping of *S. enterica* strains; the resolution of this method has been shown to be finer than serotyping; with the ability to distinguish between individual isolates of the same serovar (Morgan et al. 2009; Pan et al. 2009. In addition to typing strains, the Biolog PM platform can also be used as a tool to corroborate putative phenotypes identified through metabolic functional annotation of genes (AbuOun et al. 2009; Mappley et al. 2012). The metabolic phenome can also be used to curate genome-scale metabolic models (AbuOun et al. 2009;Raghunathan et al 2009).

The significance of phenome profiling in the present study is that it allows the comparison of metabolite utilisation dynamics between host associated strains. This high scale profiling can then be used to identify compounds for further study. Biolog PM assays can be performed under aerobic and anaerobic conditions; this allows the study of obligate and

facilitative anaerobes (AbuOun et al. 2009; Mappley et al. 2012). The potential variation in metabolite utilisation by *S. enterica* under aerobic and anaerobic conditions is of particular interest to this study as both *S.* Derby and *S.* Mbandaka experience these environments during the stages of pathogenesis and reinfection. In addition to this the Omnilog plate reader can also stably incubate plates at temperatures between 20$^{o}$C and 45$^{o}$C for several days (Biolog 2011). This allows the comparison of host associated strains under different model environmental conditions such as aerobic ambient temperatures and anaerobic body temperatures. A difference in proficiency of metabolite utilisation in the presence of different concentrations of oxygen, and temperatures, between host associated strains, could reflect a difference in relative fitness in the environment that may contribute to their distinct host species associations. Differences in metabolite utilisation may also reflect adaptations that have evolved in response to the high level of a particular metabolite in the surrounding environment.

### 1.9 Metabolic models

Different host species possess different gut constituents and therefore different selection pressures (Goodall & Kay 1965; Fordtran & Locklear 1966; Pettersson & Åman 1986; Guilloteau et al. 1997). From our current knowledge of *E. coli*, and to a lesser extent *S. enterica*, it is possible to estimate the enzyme and transporter reactions occurring in a cell from its genome (AbuOun et al. 2009; Raghunathan et al 2009; Henry et al. 2010). These metabolic processes can be visualised as a network, converting primary external metabolites into biomass and waste. From these reconstructions constraint based models of metabolism can be produced to aid hypothesis generation for studying the relationship between the host environment and optimal growth/respiratory rate (Palsson 2006). Simulations performed with the model constructed for *E. coli* MG1655 reinforced the

previously held assumption that growth is consistently optimised towards available compounds (Edwards et al. 2001). Metabolic reconstructions of *S.* Typhimurium and *S.* Enteritidis have been produced through the adaption of the *E. coli* model (AbuOun et al. 2009; Raghunathan et al. 2009).

Genome-scale metabolic reconstructions can be compared descriptively, by identifying differences in transport and metabolic genes between serovars or by running a flux simulation with different metabolite formulations to see how metabolic flux across different branches of metabolism vary between serovars (Palsson 2006). The metabolite formulations can reflect those metabolites found in commonly used microbiological media, or the metabolite composition of an environment that may be found within the host or in the surrounding environment. Discovery of differences in metabolite utilisation may point to potential competitive advantages of one strain over another, tissue tropisms or partitioned niches defined by metabolite constituents. In this way, genome-scale metabolic models may elucidate differences in metabolite utilisation between serovars which may contribute to differences in host species associations.

## 1.10 Conclusion

*S. enterica sub. enterica* is an incredibly important human pathogen, clinically and economically; Salmonellosis is estimated to cost the US economy $2.65 billion annually; this includes losses from wages, food wastage, treatment costs and additional farm management (Frenzen et al. 1999; USDA 2012). It is likely the economic and health costs of this pathogen have influenced studies of host association; such that the majority of hypotheses generated focus on the interactions directly between pathogen and host. Though

a case can be made to focus studies on different phases of pathogenesis, including externalisation, to try and understand the factors that maintain distinct host associations. The hypotheses generated within this thesis are largely comparative between *S*. Derby and *S*. Mbandaka, therefore the conclusions drawn will frequently focus on differences between these serovars, and not the common mechanisms which aid enrichment on the farm for both serovars.

The host associations identified from the isolation statistics and meta-analysis presented in this chapter are used throughout the thesis to design experiments and to place phenotypic and genotypic discoveries into a veterinary context and allude to potential host adaptations. Viewing the differences in antimicrobial resistance profile identified from the meta-analysis of these serovars it is not unreasonable to assume that antimicrobial resistance reflects antibiotic usage in the hosts. For example, the most notable antimicrobial resistance identified from the meta-analysis is tetracycline which is used in high abundance in the UK in pigs (VMD 2012). It would be interesting to assess whether tetracycline resistance in *S*. Derby contributes to the relative fitness of the serovar compared to *S*. Mbandaka.

### 1.11 Thesis outline

Although a significant veterinary and public health issue neither *S*. Derby nor *S*. Mbandaka have been studied extensively. Due to the intriguing differences in host range there is a need for baseline genotypic and phenotypic information upon which deeper interrogation of host association and geographical differences may be assessed. In this thesis I try and establish approaches and generate preliminary data which can be used to develop hypothesis to help answer the question: "why do *S. enterica* serovars Derby and Mbandaka

have distinct host associations?" In the following sections I will give a brief list of the questions I have attempted to address in each chapter and the approaches I have used.

In chapter 2, I describe the genome sequences of two isolates of both *S*. Derby and *S*. Mbandaka, performing functional annotations and sequence alignments. This is done in an attempt to answer the question: "are there intra- and inter- serovar functional gene differences between isolates of *S*. Derby and *S*. Mbandaka?" and "how might any potential functional differences relate to host association or adaptation?" Particular attention is paid to genes relating to pathogenesis, environmental persistence and metabolism.

In chapter 3, I focus on host adaptations for pathogenesis. As was shown in chapter 1, the predominant host of isolation of *S*. Derby was pigs and turkeys, and for *S*. Mbandaka cattle and chickens. Due to the availability of a well characterised porcine jejunum derived cell line, IPEC-J2, I decided to perform association and invasion assays on cell monolayers, as a model for the pig intestine. This was performed as an initial step towards answering the question: "are host associated strains better adapted to associating to and invading host derived cells than strains that are not associated with that host?" As there are limitations in using monolayer tissues, that are highly developed cells, in an artificial environment, *in vitro* organ culture (IVOC) was exploited to study tissue tropism and gene expression of *S*. Derby when exposed to jejunum and colon ex plants. Site direct mutagenesis was used to answer the question: "do genes, for which expression correlates with tissue tropism, influence association and invasion of host cells?"

In chapter 4, I utilise Biolog PM technology to study the differences in respiratory dynamics between *S*. Derby and *S*. Mbandaka in response to various metabolites when

incubated at 25$^{o}$C aerobically and 37$^{o}$C aerobically and anaerobically. The differences are contextualised to answer the question: "how might metabolic proficiency influence host associations?" I also explore the questions: "does functional genotype reflect phenotype under varying environmental conditions?" and "how might environmentally derived cues differentially affect the metabolic potential of *S*. Derby and *S*. Mbandaka?"

In chapter 5, I attempt to answer the question: "are there differences in growth rate on porcine homogenates between *S*. Derby and *S*. Mbandaka?" and "which metabolites found in the porcine jejunum and colon might explain differences in growth rate?" as well as "what is the role of secondary metabolism in potentially determining host association?" and finally "does proficiency of growth on different tissue homogenates correlate with tissue tropism?". To start to address these questions I integrated the functional annotation from chapter 2 with the phenome profile of chapter 4 to produce a genome-scale metabolic model. This is done so that the emergent property of growth rate can be determined from the complete set of metabolic reactions identified from the functional genome annotation. I then go on to challenge these models with metabolite formulations produced through metabolomics of porcine jejunum and colon tissues.

In chapter 6, I attempt to relate the genotypic and phenotypic differences discovered in previous chapters to a larger set of isolates obtained between 2000 and 2010 from the across the UK. This is performed in an attempt to answer the question: "which genotypic and phenotypic differences discovered up to this point in the thesis are serovar-wide?" This is a very important question, as the host associations described in chapter 1 were observed at the serovar level of classification, therefore any host adaptations that contribute to these

differences would be expected to be shared by all isolates of the serovar. Another important question to consider at this point is "are isolates of *S*. Derby and *S*. Mbandaka clonal?" or "are there signs of distinct lineages within the serovar?" To address these questions, I produce a phylogenetic reconstruction and observe which phenotypes and genotypes correlate with distinct phylogenetic clades.

In chapter 7, the discussion, I attempt to combine the results of each chapter so as to allow larger conclusions to be drawn from the study, to generate new hypotheses and to suggest future work which could be performed in an initial step to furthering our understanding of "why do *S. enterica* serovars Derby and Mbandaka have distinct host associations?"

**Thesis format**

This thesis consists of five paper chapters (chapters 2-6). All five chapters reflect a single cohesive project which has been subsequently split into publications. The manuscripts have been formatted to include running section and figure numbers. I have also cross-referenced different chapter sections. The publication status is given on individual chapter title pages. I have also listed the contributions of co-authors.

# Chapter 2:
# **Comparative genomics of *Salmonella enterica* serovars Derby and Mbandaka, two prevelant serovars associated with different livestock species in the UK**

The text and figures are the same as appeared in the publication:

Hayward, M.R., Jansen, V.A. & Woodward, M.J., 2013. Comparative genomics of *Salmonella enterica* serovars Derby and Mbandaka, two prevalent serovars associated with different livestock species in the UK. *BMC Genomics*, 14(1), p.365-384.

Prof. Vincent A.A. Jansen and Prof. Martin J. Woodward supervised the work and helped write the manuscript. Genome sequencing was performed by AHVLA central sequencing unit.

## 2.1 Abstract

Despite the frequent isolation of *Salmonella enterica* sub. *enterica* serovars Derby and Mbandaka from livestock in the UK and USA little is known about the biological processes maintaining their prevalence. Statistics for *Salmonella* isolations from livestock production in the UK show that *S*. Derby is most commonly associated with pigs and turkeys and *S*. Mbandaka with cattle and chickens. Here we compare the first sequenced genomes of *S*. Derby and *S*. Mbandaka as a basis for further analysis of the potential host adaptations that contribute to their distinct host species distributions.

Comparative functional genomics using the RAST annotation system showed that predominantly mechanisms that relate to metabolite utilisation, *in vivo* and *ex vivo* persistence and pathogenesis distinguish *S*. Derby from *S*. Mbandaka. Alignment of the genome nucleotide sequences of *S*. Derby D1 and D2 and *S*. Mbandaka M1 and M2 with S*almonella* pathogenicity islands (SPI) identified unique complements of genes associated with host adaptation. We also describe a new genomic island with a putative role in pathogenesis, SPI-23. SPI-23 is present in several *S. enterica* serovars, including *S*. Agona, *S*. Dublin and *S*. Gallinarum, it is absent in its entirety from *S*. Mbandaka.

We discovered a new 37Kb genomic island, SPI-23, in the chromosome sequence of *S.* Derby, encoding 42 ORFS, ten of which are putative TTSS effector proteins. We infer from full-genome synonymous SNP analysis that these two serovars diverged, between 182kya and 625kya coinciding with the divergence of domestic pigs. The differences between the genomes of these serovars suggest they have been exposed to different stresses including, phage, transposons and prolonged externalisation. The two serovars

possess distinct complements of metabolic genes; many of which cluster into pathways for catabolism of carbon sources.

## 2.2 Introduction

*Salmonella enterica* subspecies *enterica* is an important zoonotic pathogen of warm-blooded vertebrates, with both a broad host species range and geographical distribution. The subspecies is divided into over 1530 serovars based on the different epitopes of two surface antigens, the O lipopolysaccharide, and H flagellum of which there are commonly two phases (Grimont et al. 2007).

Some serovars display association with a particular set of hosts that may be stable over many decades and large geographical distances suggesting a level of adaptation or restriction (Kingsley & Bäumler 2000). With regard to serovars *S.* Derby and *S.* Mbandaka, both serovars are isolated with similar frequency in the UK and USA. Annually compiled statistics from several sources (CDC PHLIS 2009; AHVLA 2011) (HPA personal communication) showed that, whilst both serovars can readily cause disease in people, incidences in livestock show differing host associations. In the UK, for example, approximately 50% and 40% of incidences of *S.* Derby were in turkeys and pigs, respectively, and approximately 20% and 65% of incidences of *S.* Mbandaka were from cattle and chickens, respectively (AHVLA 2011). In the USA approximately 80% of isolations of *S.* Derby were from pigs, while only 3% of isolations were from turkeys, 27% and 25% of *S.* Mbandaka isolations were from cattle and chickens. Unlike in the UK in the USA *S.* Mbandaka is isolated from pigs comprising 14% of the total (CDC PHLIS 2009). These host distributions have been maintained for over a decade and on two continents

which gives rise to at least two hypotheses. First, is it possible that the differences in host association may relate to production systems and that these serotypes possess similar functional capabilities. Second, is it possible that the differences in host association reflect functional differences between serovars or genovars therein, whereby there exist bacterially encoded mechanisms that maintain these patterns. As a starting point to tackle these opposing hypotheses, we present the first full chromosome sequence of two UK isolates of both *S.* Derby and *S.* Mbandaka. We use functional genomics to describe genome features and to identify genes that are unique with a view to gaining insights into potential genetic components that contribute to the species distributions described above.

## 2.3 Results and discussion

The chromosomes of two strains of *S.* Derby and *S.* Mbandaka were sequenced and compared with the goal of identifying potential mechanistic differences between the two serovars that could explain their skewed isolation frequencies from subsets of livestock species in the UK. Strains were obtained from background monitoring performed by the Animal Health and Veterinary Laboratories Agency (AHVLA) in the UK between 2000 and 2010. In total 28 strains were selected spanning the decade and from differing geographic points of isolation across the UK (locations not shown due to sensitivity of data). The hosts of isolation of the selected strains were chosen to reflect the two most common hosts of each serovar, for *S.* Derby these were pigs and turkeys and for *S.* Mbandaka cows and chickens (refer to Chapter 6). Two isolates of each serovar isolated from separate geographical locations, with the same host species, and identical MLST sequence types (*S.* Derby strains ST40 and *S.* Mbandaka strains ST900) were chosen for full genome sequencing. We recognised that in the absence of information regarding the pan-genome of the population, that by comparing just two isolates of each serovar, we

could potentially infer, incorrectly, that differences in gene complement between isolates of the same serovar isolated from different hosts were adaptations to these different hosts. The selection was therefore made with the aim of better understanding the genomic differences between strains which would typically be considered clonal. *S*. Derby strains D1 and D2 were both isolated in 2008 from porcine hosts. *S*. Mbandaka M1 and M2 were isolated from cattle in 2008 and 2009 respectively. No research has previously been performed on these strains.

### 2.3.1 General genome features of *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2

*S.* Derby strains D1 and D2 possessed chromosomes of 4.86 Mb nucleotides in length with a GC skew of 51.16% and 51.46% respectively. The RAST annotation system predicted that the chromosome sequence of *S*. Derby D1 encodes 4720 genes and the sequence of D2 4717 genes. The chromosome of *S.* Mbandaka strains M1 and M2 were both 4.72 MB nucleotides in length with a GC skew of 51.91% and 52.01% respectively. These were predicted to encode 4616 and 4619 genes respectively. Interestingly all four chromosomes contained different numbers of RNA coding sequences, D1 contained 69, D2 contained 73, M1 contained 74 and M2 contained 75 (Figure 2.1). RNA sequences are frequently sites for integration of horizontally acquired DNA sequences, in some cases leading to duplication of the RNA (Amavisit et al. 2003; Lu et al. 2003; Pickard et al. 2003). The difference in the number of RNAs in each genome could reflect a difference in evolutionary potential of each chromosome.

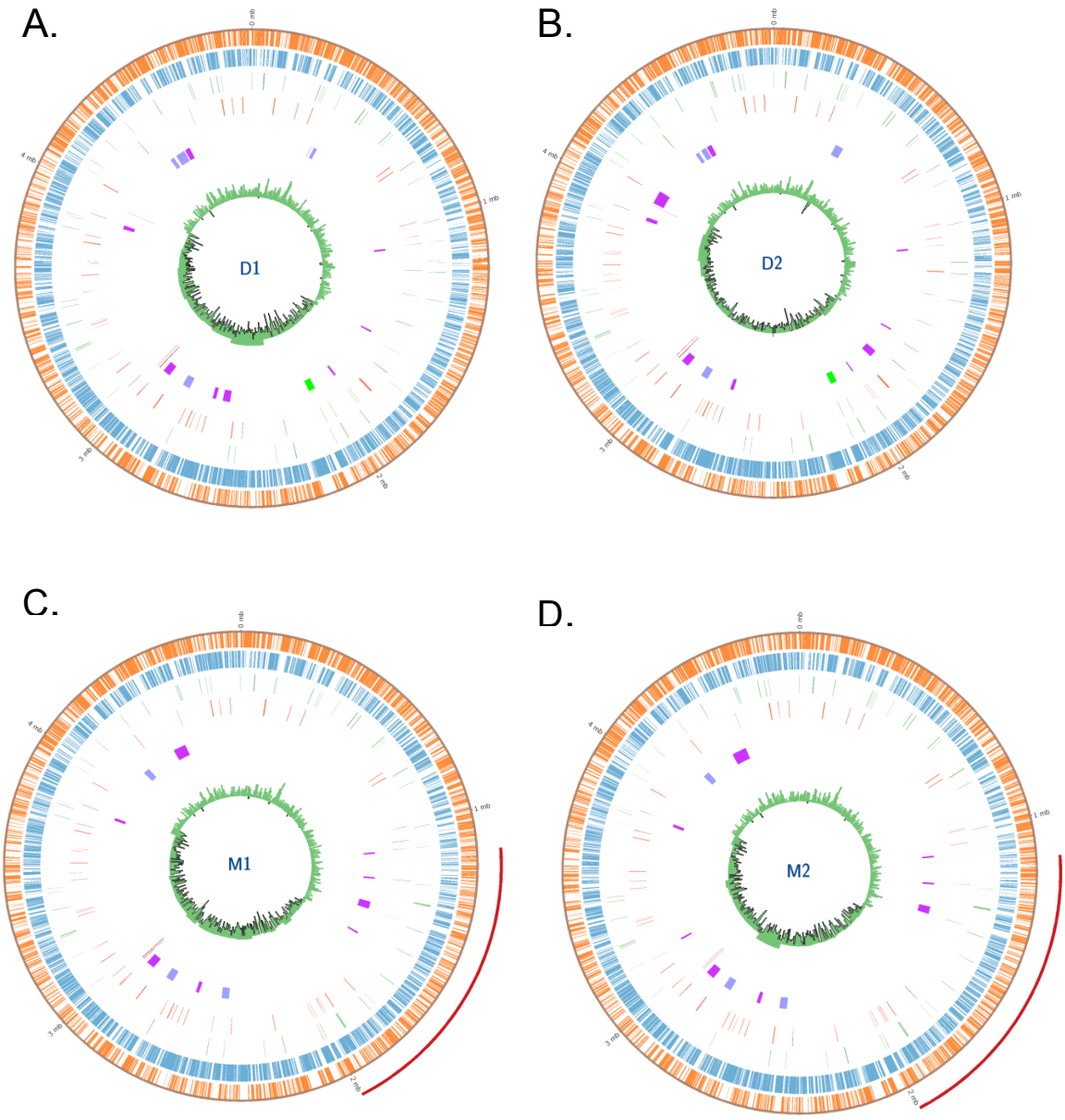



**Figure 2.1**: Chromosome annotation maps of newly sequenced isolates; **a**) *S.* Derby D1, **b**) *S.* Derby D2, **c**) *S.* Mbandaka M1 and **d**) *S.* Mbandaka M2. The two outer tracks represent ORFs identified by RAST in forward and reverse DNA strands. The third and fourth tracks display non-coding RNAs in forward and reverse confirmation. The fifth track shows, SPI-23 (green), prophage (mauve), *Salmonella* pathogenicity islands (purple) and CRISPR operons (burgundy). The sixth track shows the GC base composition under a 1000 bp moving window. The additional red track outside of the *S.* Mbandaka maps indicates the location of a unique 860kbp sequence inversion.

2.3.2 *S.* Mbandaka contains a large sequence inversion

*S.* Mbandaka contains a 860Kb sequence inversion between a mobile element protein and tRNA-ser-GGA (located between base 1086415 and 1947250 of M1, and 1132370 and 1992477 of M2) which was also found in *S.* Choleraesuis SC-B67, and was absent from *S.* Derby (Figure 2.1) and other sequenced *S. enterica* serovars including *S.* Agona SL483, *S.* Dublin CT02021853, *S.* Enteritidis P125109, *S.* Gallinarum 28791 and *S.* Typhimurium LT2 and SL1344. This region codes for 909 genes identified by the RAST gene caller.

Large sequence inversions have a significant impact on the transcript composition of the cell during replication, as those genes closer to the origin of replication are present in duplicate for a longer period of time than those genes closer to the terminus of replication (Schmid & Roth 1987; Liu & Sanderson 1996). The effects of increased gene dosage during replication are most noticeable when bacteria are growing at an optimal rate (Couturier & Rocha 2006). In *Escherichia coli* DNA replication from the origin of replication to terminus of replication takes 22 minutes during a 40 minute cell cycle when grown in LB broth at 37 °C (Hill et al. 2012). If we apply this duration to the inversion found in *S.* Mbandaka M1 and M2, where almost a quarter of the chromosome is in a different orientation to *S.* Derby D1 and D2, then there is an 8.6 minute difference between gene duplication events of the genes adjacent to the sites of inversion. These genes are therefore in duplicate and the other genes in singlet for 21% of the cell cycle. In *S.* Derby the ten genes closest to the mobile genetic element signifying the start of the inverted sequences do not pertain to a common mechanism. Though interestingly, amongst these ten genes is a permease of the drug/metabolite transporter (DMT) superfamily which in *S.* Mbandaka occupies the very furthest gene in the inversion. The ten genes at the terminus of the inversion in the chromosome of *S.* Derby D1 and D2 comprise of two operons, the

Csg-curli operon (four genes) and Ycd-swarming operon (five genes). The most interesting aspect of these two operons is that they are associated with two diametrically opposed mechanisms; the curli operon is associated with biofilm development in a sessile population and the swarming operon is associated with directed movement of the bacterial population, both using quorum sensing (Toguchi et al. 2000; Hammer & Bassler 2003). Both are population scale emergent phenotypes of gene regulation at a single cell level. *csgD* found in the curli operon regulates curli expression and cellulose secretion, the main components of biofilms in *Salmonella enterica* (Brombacher et al. 2006). The ten genes at the centre of inversion and therefore in similar dosage throughout replication in *S.* Derby and *S.* Mbandaka are genes for formate dehydrogenase alpha, beta and gamma subunits which form a single transmembrane enzyme (Jormakka et al. 2003). Also contained within this region is a permease of the drug/metabolite transporter (DMT) superfamily. The ten genes at either end and in the centre of the sequence inversion can be found listed in the supplementary materials (Supplementary material 2.1).

### 2.3.3 Summary of functional annotation

The chromosome of *S.* Derby is 140 Kb longer than that of *S.* Mbandaka, coding for 100 additional genes. RAST annotation was performed on 9/10/12 and achieved 67% coverage with FIGfam subsystems of the *S.* Derby D1 and D2 chromosomes and 68% of the *S.* Mbandaka M1 and M2 chromosomes (Aziz et al. 2008). FIGfam clusters genes based on protein sequence similarity. From this the function of a novel gene may be inferred. These genes are then clustered into hierarchical subsystems that display increasing functional breadth (Overbeek et al. 2005). As the database has developed the subsystem coverage of the four genomes presented here has markedly increased by 8-9% over the same annotation performed on the 9/10/11. The most recent annotation is available through the following

RAST IDs: D1 [RAST: 28144.16], D2 [RAST: 28144.17], M1 [RAST: 192954.16] and M2 [RAST: 192954.17] so that the chromosome can be re-annotated as the RAST databases are updated. The number of hypothetical genes between the annotations remained constant for each chromosome. In all cases almost a quarter of the genome annotation was found to be of hypothetical gene status. *S.* Derby contains 96 hypothetical/putative proteins which share less than 90% amino acid sequence homology with ORFs in *S.* Mbandaka. *S.* Mbandaka contains 155 unique hypothetical/putative proteins.

### 2.3.4 Intra-serovar differences in the complement of functionally unique genes

The majority of the diversity between strains of the same serovars was in the complement of phage associated genes. All intra-serovar differences in gene complement can be found listed in the supplementary materials (Supplementary material 2.1). Between the *S.* Derby isolates, D1 contains a single unique gene for an aconitate hydratase 2 (EC 4.2.1.3) associated with glyoxylate bypass. D2 contains 11 unique genes, of which five are associated with phage, the remaining are associated with metabolism. A single gene which is associated with the ribosome at stationary growth phase is absent from D1. There is less diversity between isolates of *S.* Mbandaka. M1 contains a single additional gene which encodes a phage tail fibre protein. M2 contains six additional genes, two for cytochrome-c biosynthesis, two phage genes and a gene encoding a 2,5-diketo-D-gluconic acid reductase B (EC 1.1.1.274).

2.3.5 Inter-serovar differences of functionally unique genes

Genes pertaining to metabolite utilisation, prophage, CRISPR spacers and *Salmonella* pathogenicity islands will be dealt with separately. The following summarises the genes that do not fit into these categories. All inter-serovar differences in gene complement can be found listed in the supplementary materials (Supplementary material 2.1).

2.3.6 *Salmonella* Derby

*S.* Derby contains 16 genes that are functionally unique. Of these 16 genes 12 are distributed between two operons. One is the *mer* operon conveying mercury resistance. This consists of five genes including *merC* and *merT* transport proteins, which actively take up toxic mercuric cations ($Hg^{+2}$) for subsequent reduction to non-toxic metallic mercury ($Hg^{0}$) (Barkay et al. 2003). Mecuric cations enter the food chain from several sources, including fish, poultry and meat. Animals feed is frequently supplemented in the UK with fish meal which is high in mercury. Fish meal contains the second highest concentration of mercury per kg in animal feed/pet food in Europe, with fish oil containing the highest concentration (Alexander et al. 2008). The second operon, is a CRISPR operon made up of seven genes, one of putative status and *cse1*-*cse4*, *cas1*, *cas2* and *cas5e*. Of the remaining four genes one is the gene for an UDP-galactopyranose mutase (EC 5.4.99.9) which is associated with the biosynthesis of alpha-D-galactofuranose, a component of the O-antigen in *Salmonella enterica* groups B, C2, D and E (Stevenson et al. 1994). This fits with the classification of *S.* Derby into group B and *S.* Mbandaka into group C1.

2.3.7 *Salmonella* Mbandaka

*S.* Mbandaka contains a cluster of four type VII secretion system Yad fimbrial chaperone proteins. Two genes with the same function from the HtrE fimbrial cluster sit approximately 100 KB away. A further 2 MB away at 4.5U sits a cluster of three beta-fimbriae genes also associated with type VII secretion. Three sialic acid metabolism genes associated with capsule production, *nanC*, *nanM* and a hypothetical gene, are clustered around 1U. Two cell death toxin-antitoxin genes, *phd* and *doc,* are unique to M1 and M2, and may be involved in plasmid addiction systems. Two genes associated with reduction in mutation rate due to exposure to bile salts are absent from *S.* Derby. These genes, *umuC* and *umuD* are part of the SOS DNA repair response and form DNA polymerase V. It has been shown in *E. coli* that in the absence of *umuC* genomic lesions are not repaired correctly by DNA polymerase III and can leave frame shift mutations which lead to pseudogene formation. DNA polymerase V has a higher rate of single nucleotide mutations than DNA polymerase III (Reuven et al. 1998; Merritt & Donaldson 2009). This could lead to a higher rate of pseudogene formation in *S.* Mbandaka strains and SNP formation in *S.* Derby strains. However, this would need to be confirmed through further analyses.

There are only seventeen genes that are unique in function to either *S.* Derby or *S.* Mbandaka that are not clustered. Of these seventeen genes *S.* Mbandaka contains seven unique genes related to biogenesis of cytochrome-c, specifically the maturation of the molecule, and are spread across the chromosome. The genes *ccmB*, *ccmC* and *ccmD* convey the heme-b group to the product of CcmE, a monotopic membrane protein (Sanders et al. 2010). The products of *ccmF*, *ccmG* and *ccmH* complex with CcmE to convey the heme-b group to the apocytochrome-c precursor of cytochrome-C (Reid et al. 2001; Ren et al. 2002). Though these genes are ubiquitous amongst Gram negative bacteria, strains of *E.*

*coli* have been discovered that lack the *ccm* operon and yet are able to synthesis cytochrome-c containing heme-b (Sinha & Ferguson 1998).

### 2.3.8 Differences in metabolic gene complement between *S.* Derby and *S.* Mbandaka

Fourteen genes were identified by RAST subsystem annotations as being involved in primary or secondary metabolism which were found to differ between *S.* Derby and *S.* Mbandaka. Six of these genes belong to *S.* Mbandaka are associated with D-galactonate catabolism, this includes uptake, regulation and processing into central carbon metabolism. *S.* Derby contains six genes for the uptake and catabolism of six different carbon sources, this comprises an asparagine synthetase (EC 6.3.5.4), a hydroxyaromatic non-oxidative decarboxylase protein D (EC 4.1.1.-), a protein fumarylacetoacetate of the hydrolase family, phosphatase NagD predicted to act in N-acetylglucosamine utilization subsystem, an aconitate hydratase 2 (EC 4.2.1.3), a galactose-specific IIA component (EC 2.7.1.69) and the large subunit of a glycerol dehydratase reactivation factor.

### 2.3.9 Metabolic pathways

The biological significance of the differences in metabolic genes was elaborated through construction of metabolic models from the genome sequences using SEEDmodel (Henry et al. 2010). These differences were then elaborated in context of the surrounding reactions. Metabolic reconstructions curated with phenotypic data are under way to better understand the effect of secondary metabolism on the optimal growth rate of *S.* Derby D1 and *S.* Mbandaka M1 (refer to sections 5.4.2 and 5.4.8 ).

2.3.9.1 Alanine, aspartate and glutamate metabolism map 00250 created 1/6/12

*S.* Derby lacks a single gene, an aspartate—ammonia ligase (EC6.3.1.1) for the conversion of L-aspartate to L-asparagine. The same reaction is achievable through two additional reactions utilising an asparaginase/glutaminase (EC3.5.1.38) and an L-asparaginase (EC3.5.1.1) which are also present in *S.* Mbandaka.

2.3.9.2 Galactose metabolism map 00052 created 31/5/12

The three genes encoding products needed to feed D-galactonate into glycolysis (EC 4.2.1.6, EC 2.7.1.58 and EC 4.1.21) by conversion to D-glyceraldehyde-3P are present on the chromosome of *S.* Mbandaka and absent from that of *S.* Derby. There are no alternative routes from D-galactonate to glycolysis.

2.3.9.3 Nitrogen metabolism map 00910 created 21/8/12

A gene coding for the enzyme L-glutamine amido-ligase that converts L-glutamine to L-glutamate using one molecule of $H_2O$ in the process (EC 6.3.5.4) is missing from the chromosome of *S.* Derby D1. All strains contain a gene that catalyses the same reaction but with the requirement of a molecule of $NADP^+$ as opposed to one of $H_2O$ (EC 1.4.1.13).

2.3.9.4 Starch and sucrose metabolism map 00500 created 9/7/12

A single reaction is missing from *S.* Mbandaka in this map for the conversion of alpha-D-Glucose-1-P to CDP-glucose (EC 2.7.7.33); there is no route to this compound other than this on the map. The CDP-glucose then leads into amino sugar and nucleotide sugar metabolism map 00520 created 19/1/10. In this map there is an additional reaction from

CDP-glucose leading to CDP-4-keto-6-deoxy-D-Glucose missing in *S*. Mbandaka. This reaction is catalysed by the enzyme RfbG, a CDP-glucose 4,6-dehydratase (EC 4.2.1.45) which is found in *Salmonella enterica* groups A, B, C2, C3, D1 and D2 and required for binding of the O antigen to the core oligosaccharide (Xiang et al. 1993; Manning et al. 1995). *S*. Mbandaka is a member of *S. enterica* group C1.

2.3.9.5 Streptomycin biosynthesis pathway map 00521 created 27/12/10

Two steps from D-glucose-1-P are present in both serovars (EC 2.7.7.24 and EC 4.2.1.46), following on from the terminal product of this reaction, two additional steps that lead to dTDP-L-rhamnose are missing in *S.* Mbandaka (EC 5.1.3.13 and EC 1.1.1.133). dDTP-L-Rhamnose feeds directly into novobiocin biosynthesis, diverted out of the streptomycin biosynthesis pathway. *S*. Mbandaka is left with a product which feeds into polyketide sugar unit biosynthesis (Pathway 00523, created 14/3/12, polyketide sugar unit biosynthesis).

2.3.10 *Salmonella* pathogenicity islands

The chromosome of *Salmonella enterica* comprises largely of a core sequence punctuated with horizontally acquired sequences (Chan et al. 2003). The complement of genomic islands within the chromosome of *Salmonella enterica* can vary amongst isolates of the same serovar (Saroj et al. 2008; Seth-Smith 2008). It has been postulated that the acquisition of horizontally acquired genes into a *Salmonella* pathogenicity island (SPI) led to the divergence of *Salmonella* from *Escherichia coli* (Ochman & Groisman 1996; Groisman & Ochman 1997). *Salmonella* pathogenicity island 1 (SPI-1) is found in all serovars of *S. enterica* (with the exception of *S.* Seftenberg and *S*. Litchfield) and is highly conserved (Ochman & Groisman 1996; Matsushita et al. 1996; Hu et al. 2008). There are

currently 22 published *Salmonella* pathogenicity islands identified from the genomes of *Salmonella enterica* and *Salmonella bongori* (Fookes et al. 2011). The gene content of some of these islands is highly plastic, as exemplified by the different gene complement of SPI-3 found in *S.* Dublin CT02021853 and *S.* Typhimurium LT2 (Blanc-Potard & Groisman 1997). The *Salmonella* pathogenicity islands are well characterised in terms of genetic composition and putative function but less so, with notable exceptions, for their role in pathogenicity (Marcus et al. 2000; Hensel 2004). Hence differences in SPI complement and gene content of D1, D2, M1 and M2 chromosomes may hint at mechanisms that maintain their respective host species range.

2.3.10.1 Complete or absent *Salmonella* pathogenicity islands

SPIs 2 and 4 found in the genome of *S.* Choloreaesuis SC-B67 and SPI-18 from *S.* Typhi CT18 are complete in the genomes of *S.* Derby D1 and D2, and *S.* Mbandaka M1 and M2. SPI-7, 8, 10, 15, 16, 17, 19, 20, 21 and 22 were absent from both *S.* Derby D1 and D2, and *S.* Mbandaka M1 and M2 genomes.

2.3.10.2 Variation in SPI-1of *S.* Derby and *S.* Mbandaka

SPI-1 in *S.* Mbandaka M1 and M2 shares 100% nucleotide sequence identity with *S.* Typhumirum LT2 with the addition of two ORFs coding for hypothetical proteins found in the SPI-1 of *S.* Choleraesuis SC-B67, SC2837 and SC2838 which are absent in *S.* Derby D1 and D2. *S.* Derby D1 and D2 lack three genes from SPI-1 of *S.* Typhimurium LT2, STM2901, STM2902 and STM2903 (Table 2.1). SIEVE an online server for the prediction of TTSS effector proteins, found that the *S.* Mbandaka M1 and M2 contained an ORF with 98% amino acid sequence homology with SC2837 from *S.* Choleraesuis SC-B67, is a

likely candidate for an effector protein with a p-value of 0.003. With reference to well-characterised effector proteins, all four isolates contain intact versions of *sopB* and *sopE*. The two putative cytoplasmic proteins found in SPI-1 of *S.* Typhimurium LT2, STM2901 and STM2902 and here in *S.* Mbandaka M1 and M2 and not D1 and D2 are unlikely candidates for effector proteins with p-values of 0.142.

2.3.10.3 Variation in SPI-3 between other serovars and *S.* Derby and *S.* Mbandaka

SPI-3 is highly variable, between *S.* Typhimurium 14028 and *S.* Choleraesuis SC-B67 the only region of homology is the insertion sequence tRNA-*selC*. SPI-3 from *S.* Derby D1 and D2 is an amalgamation of 19 SPI-3 genes from *S.* Typhimurium 14028, *S.* Dublin, *S.* Choleraeasuis SC-B67 and *S.* Typhi CT18. *S.* Mbandaka M1 and M2 also contain a unique SPI-3 gene complement, containing 12 genes found in *S.* Typhimurium 14028, *S.* Choleraesuis SC-B67 and *S.* Typhi CT18. Unlike *S.* Derby D1 and D2, *S.* Mbandaka M1 and M2 have no SPI-3 genes in common with *S.* Dublin. STY4039 previously unique to *S.* Typhi CT18 is present in *S.* Mbandaka M1 and M2 and absent from *S.* Derby D1 and D2 (Table 2.1). The main region of variation between *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2 SPI-3 is at the start of the island where the complete *S.* Dublin SPI3 is present, this was shown previously for *S.* Derby 9813031, 0010160 and 0010158 (Amavisit et al. 2003). This region contains seven genes relating to the adhesion structures, pili and fimbriae. *S.* Mbandaka M1 and M2 contains *rhuM* found in the SPI-3 of *S.* Typhimurium 14028; this sequence is absent from the SPI-3 of *S.* Derby D1 and D2. *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2 share 10 SPI-3 genes in common; this complement of genes is unique to these two serovars. Both serovars contain five virulence genes present in the SPI-3 of *S.* Typhimurium 14028 and *S.* Choleraesuis SC-B67.

**Table 2.1:** Comparison of previously published genomic islands that distinguish between D1, D2, M1 and M2 in their gene complement. SPI-6 distinguishes between *S*. Derby and *S*. Mbandaka strains. This island has also been extensively studied for a role in pathogenicity, and displays the largest amount of diversity in gene complement between *S*. Derby and *S*. Mbandaka. NI identifies an element in the table for which there is No Information.**SPI**.

| | **Gene** | **D1** | **D2** | **M1** | **M2** | **Descriptor** | Role in pathogenesis |
|---|---|---|---|---|---|---|---|
| CS54 | *ratA* | - | + | + | + | Ribosome association toxin | No role in virulence (Kingsley et al. 2003) |
| | *sivI* | - | + | + | + | Outer membrane protein | No role in virulence (Kingsley et al. 2003) |
| | *sivH* | - | + | + | + | Invasin-like | Colonising peyers patch (Mouse) (Kingsley et al. 2003) |
| SPI-1 | STM2901 | - | - | + | + | Putative cytoplasmic protein | NI (McCelland et al. 2001) |
| | STM2902 | - | - | + | + | Putative cytoplasmic protein | NI (McCelland et al. 2001) |
| | STM2903 | - | - | + | + | Putative cytoplasmic protein | NI (McCelland et al. 2001) |
| | SC2837 | - | - | + | + | Hypothetical protein | NI (Ciu et al 2005) |
| | SC2838 | - | - | + | + | Hypothetical protein | NI (Ciu et al 2005) |
| SPI-3 | Pseudo | + | + | - | - | NI | NI (Amavisit et al 2003) |
| | *yadC* | + | + | - | - | Fimbrial like protein | Stress response (Amavisit et al. 2003; Chiu et al 2005) |
| | *yadK* | + | + | - | - | Fimbrial like protein | NI (Amavisit et al 2003) |
| | *yadL* | + | + | - | - | Fimbrial like protein | NI (Amavisit et al 2003) |
| | *yadN* | + | + | - | - | Fimbrial like protein | NI (Amavisit et al 2003) |
| | *htrE* | + | + | - | - | Porin/ fimbrial assembly | High temperature resistance above 50 °C (Raina et al. 1993) |
| | *ecpD1* | + | + | - | - | Pilin chaperone | Expressed with increasing temp above 22 °C (Raina et al. 1993) |
| | *ecpD2* | + | + | - | - | Pilin chaperone | NI (Amavisit et al 2003) |
| | Pseudo | + | + | - | - | NI | NI (Amavisit et al 2003) |
| | *rhuM* | - | - | + | + | Cytoplasmic protein | Epithelial migration (Tenor et al. 2004) |
| | STY4039 | - | - | + | + | EnvR binding site | No role in virulence (Parkill et al. 2001) |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| SPI-6 | STY0296 | + | + | - | - | Hypothetical protein | No role in virulence (Parkill et al. 2001) |
| | STY0300 (*sirA*) | - | - | + | + | Transcription factor | Regulates SPI1 and flagellu (Johnston et al. 1996; Teplitski et al. 2003) |
| | STY0301 (*safC*) | - | - | + | + | Outer membrane usher protein | Up regulated during intracellular replication (Klump & Fuchs 200) |
| | STY0302 (*sciM*) | - | - | + | + | Hemolysin-coregulated protein | NI (Ong et al. 2010) |
| | STY0303 (*sciN*) | - | - | + | + | Outer membrane lipoprotein | Need for Type VI secretion, biofilm formation (Aschtgen et al. 2008) |
| | STY0307 | - | - | + | + | Hypothetical protein | NI (Ong et al. 2010) |
| | STY0311 | - | - | + | + | Mannosyl-glycoprotein | NI (Ong et al. 2010) |
| | STY0312 | - | - | + | + | Hypothetical protein | NI (Ong et al. 2010) |
| | STY0319 | - | - | + | + | Rhs-family protein | NI (Ong et al. 2010) |
| | STY0320 | - | - | + | + | Putative cytoplasmic protein | NI (Ong et al. 2010) |
| | STY0321 | - | - | + | + | Rhs1 protein | NI (Ong et al. 2010) |
| | STY0322 | - | - | + | + | Hypothetical protein | NI (Ong et al. 2010) |
| | STY0323 | - | - | + | + | Hypothetical protein | NI (Ong et al. 2010) |
| | *safA* | - | - | + | + | Fimibrial usher protein | No effect in virulence in mice (Folkesson et al. 1999; Ong et al. 2010) |
| | Pseudo | - | - | + | + | NI | NI (Parkill et al. 2001) |
| | *safB* | - | - | + | + | Periplasmic fimbrial chaperone protein | NI (Parkill et al. 2001) |
| | *safC* | - | - | + | + | Outer membrane usher protein | Up regulated during intracellular replication (Klump and Fuchs 2006) |
| | *safD* | - | - | + | + | Fimibrial usher protein | No effect in virulence in mice (Folkesson et al. 1999) |
| | STY0338 | - | - | + | + | Periplasmic binding protein | NI (Parkhill et al. 2001; Ong et al. 2010) |
| | Pseudo | - | - | + | + | NI | NI (Parkill et al. 2001) |
| | *sinR (pagN*) | - | - | + | + | HTH transcription factor | No effect in mice (Groisman et al 1993; Folkesson et al. 1999) |
| | NI | + | + | - | - | Rhs-family protein | NI |
| | NI | + | + | - | - | Rhs-family protein | NI |
| | NI | + | + | - | - | Phosphotriesterase | NI |
| | NI | + | + | - | - | Hypothetical protein | NI |

| | | | | | | |
|---|---|---|---|---|---|---|
| NI | + | + | - | - | Hypothetical protein | NI |
| orf7 (*Photorhabdus* ) | + | + | - | - | SinR-like, HTH transcription factor | NI (Waterfield et al. 2002) |
| NI | + | + | - | - | Putative cytoplasmic protein | NI |
| Pseudo | + | + | - | - | NI | NI |
| Pseudo | + | + | - | - | NI | NI |
| *tcfA* | + | + | - | - | Fimbrial protein | Increased expression with increased salinity, non virulence in INT-407 cells (Bishop et al. 2008) |
| *tcfB* | + | + | - | - | Fimbrial protein | Increased IgG-tcfB in patients with *S.* Typhi (Harris et al. 2006) |
| *tsaC* | + | + | - | - | Fimbrial usher protein | No effect on adhesion to mice monolayers (Ghosh et al. 2011) |
| *tcfD* | + | + | - | - | Fimbrial protein | NI (Parkill et al. 2001) |
| *rnhA-dnaQ* like | - | + | - | - | DNA polymerase 3 epsilon subunit ribonuclease H | NI (Barbe et al. 2004) |
| NI | - | + | - | - | Ribonuclease HI | NI |
| *gloB* like | - | + | - | - | Hydroxyacylglutathione hydrolase | methylglyoxal degradation (Parkill et al. 2001) |
| *mltD* | - | + | - | - | Membrane-bound lytic murein transglycosylase D | Enchance virulence in *Vibria anguillarum* in zebrafish (Xu et al 2011) |
| NI | - | + | - | - | Methyltransferase UbiE/COQ5 | Ubiquinone/ menaquinone biosynthesis (Poon et al. 2000) |
| *yafD* | - | + | - | - | AP like endonuclease | Egg albumen resistance (Lu et al. 2003) |
| NI | - | + | - | - | Putative drug efflux pump | NI |
| NI | - | + | - | - | Hypothetical oxidoreductase | NI |
| *dkgB* | - | + | - | - | 2,5 didehydrogluconate reductase B | Detoxing response to hypersomatic *E. Coli* (Shabala et al. 2009) |

Both serovars lack the virulence gene *mgtC* which is present in *S.* Typhimurium 14028, *S.* Choloraesuis SC-B67 and *S.* Typhi CT18. In *S.* Typhimurium LT2 and 14028 *mgtC* was shown to be essential for intra-macrophage survival (Blanc-Potard & Groisman 1997).

2.3.10.4 Variation in SPI-5 between other serovars and *S.* Derby and *S.* Mbandaka

It has previously been shown that SPI-5 from *S.* Derby 9813031, 0010160, 0010158 and *S.* Ohio 9815932, 9714920, 9714922 contain an additional unnamed ORF, this ORF was present in both *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2 (Amavisit et al. 2003).

2.3.10.5 Variation in SPI-6 between *S.* Derby and *S.* Mbandaka

SPI-6 is found in *S.* Typhi CT18, is 57Kb in length and contains 59 genes. Between *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2, 24 genes that were not found in other islands on PAI-DB were identified by Glimmer3 (Table 2.1) (Delcher et al. 2007). The annotations here were taken from NCBI BLASTn results, many of which were hypothetical or putative in description. SPI-6 also shows the largest variation between *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2 outside of prophage and SPI-23 nucleotide sequences. SPI-6 in D2 had 8 unique genes at the C-terminus of the positive strand that were not found in the other isolates. This contains an AP-like endonuclease gene related to egg albumin resistance, *yafD* (Lu et al. 2003). *S.* Derby has been isolated from inside of eggs while *S.* Mbandaka has been shown to grow slower than other *S. enterica* serovars in albumin (Messens et al. 2004; Betancor et al. 2010). The remainder of the island showed no variation amongst isolates of the same serovar. Seven *S.* Typhi CT18 genes were absent from both serovars, these were STY0300-STY303, STY0342, STY0350 and STY03351. *S.* Derby D1 and D2 SPI-6 contained 8 genes from *S.* Typhi CT18 that were absent from *S.*

Mbandaka M1 and M2. *S.* Mbandaka M1 and M2 SPI-6 contained 20 genes from *S.* Typhi CT18 that were absent from *S.* Derby D1 and D2.

The variation in the gene complement of SPI-6 in *S.* Derby and *S.* Mbandaka is of particular interest with regards to host adaptation. *S.* Derby possess the gene *sirA* which corresponds with ORF STY0300 in *S.* Typhi CT18, that codes for a transcription factor linked with regulation of the TTSS encoding SPI-1 when inside a mammalian host. Interestingly mutants for *sirA* in *S.* Typhimurium LT2 were attenuated in a bovine gastro-enteritis model, but were still proficient at causing typhoid fever in a mouse model (Johnston et al. 1996; Ahmer et al. 1999; Teplitski et al. 2003; Ong et al. 2010). The SPI-6 of *S.* Mbandaka also contains a gene *sciN* which is an outer membrane lipoporotein essential for biofilm formation in *E. coli* which is absent from *S.* Derby (Ahmer et al. 1999).

2.3.10.6 Variation in SPI-9 from *S.* Typhi CT18

The alignment between SPI-9 of *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2 showed 100% sequence homology. SPI-9 from *S.* Typhi CT18 contains four genes as do the islands in *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2, though there is a difference in ORF length. STY2875 is at the start of the island and is 10.8 Kb in length, in both *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2 an additional region of 595 bp is found between bases 3056 and 3057. The other three ORFs are truncated at the beginning of each sequence by 162 bp.

2.3.10.7 Variation in SPI-11 from *S.* Derby and *S.* Mbandaka

The same eight genes from SPI-11 of *S.* Cholereaesuis SC-B67 are absent in *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2. One of these genes is the effector protein *sopB*, which has been implicated in fluid secretion in calf ileal loops and is essential for enteropathogenicity of *S*. Dublin although, as previously mentioned, a homolog to this gene was found elsewhere on the chromosome (Galyov et al. 1997; Norris et al. 1998). SPI-11 also encodes the gene *pagC,* an envelope protein which increases survival within mouse macrophage (Gunn et al. 1995).

2.3.10.8 Variation in SPI-12 between *S.* Derby and *S.* Mbandaka

SPI-12 is an 11 Kb island first identified in *S.* Choleraesuis SC-B67. The island is inserted at a tRNA-Pro. The insertion sequence was present in both *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2, though no genes were adjacent to this site. Alignment of the whole SPI-12 island from *S.* Choleroeaesuis SC-B67 with D1, D2, M1 and M2 identified homologs for all the genes in each sequence, not in a single unit, but spread across the chromosome.

2.3.10.9 Variation in CS54 between *S.* Derby and *S.* Mbandaka

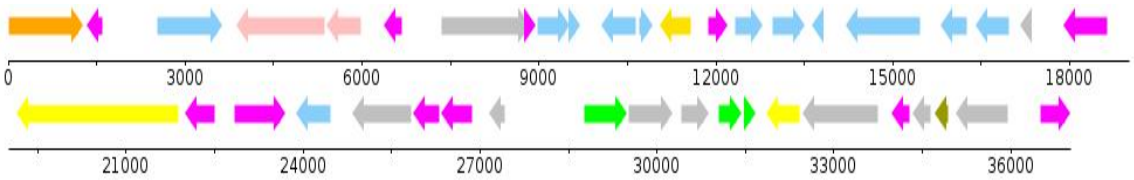
CS54 identified from *S.* Typhimurium 14028 is associated with virulence and shows variation between isolates of the same serovar. All isolates lack the virulence genes *shdA* and *ratB*, and the untested gene *ratC*. M2 lacks the whole island with only the insertion sequences present (Table 2.1). This region in D1, D2 and M1 contains three genes *ratA*, *sivI* and *sivH* previously identified in CS54 of *S.* Typhimurium 14028. CS54 was previously described in *S.* Derby strain De1, in this instance *ratB* was also found, a gene

essential for the colonisation of the cecum in BALB/c mice by *S.* Typhimurium IR715 (Kingsley et al. 2003).

2.3.10.10 A novel *Salmonella* pathogenicity island designated SPI-23

A new genomic island with a putative role in pathogenesis, SPI-23, was discovered in this study on the chromosome of D1 and D2 between bases 2027348–2065972 and 2052685–2089962 respectively flanked by tRNA-asn (GTT) and a hypothetical protein, *docB* (Figure 2.2). SPI-23 is composed of 42 ORFS with an overall GC composition of 38% differing largely from the 51% of the *S.* Derby genome. SPI-23 is completely missing from *S.* Mbandaka. Of the 42 ORFs 28 were of hypothetical status, of which, 17 contained no homology with an entry in the NCBI nucleotide database (accessed on 1/10/12).

The island contains two genes, *potR* and *talN*, both implicated in type IV secretion and the production of pili. There is a single gene, *zomB*, predicted here to encode a lipoprotein. We also find five DNA binding proteins, *furB*, *lamE*, *halF*, *mstR* and *numT* and two putative membrane protein *bigM* and *putM*. SPI-23 contains a single NUDIX hydrolase, a very ubiquitous protein family involved in a multitude of regulator processes (McLennan 2006).




| Gene name | Function |
|---|---|
| gooN | integrase |
| sanA | Putative type 3 effector protein |
| redB | hypothetical protein |
| potR | PilV-like protein |
| talN | Putative type IV pilin protein precursor |
| kayT | Putative type 3 effector protein |
| pikL | FIG00639151: hypothetical protein |
| janE | Putative type 3 effector protein |
| bunJ | hypothetical protein |
| minE | hypothetical protein |
| penE | hypothetical protein |
| comA | hypothetical protein |
| zomB | Putative lipoprotein (94%) |
| chlE | Putative type 3 effector protein |
| newT | hypothetical protein |
| rosF | hypothetical protein |
| chrD | hypothetical protein |
| troN | FIG00638315: hypothetical protein |
| dinO | hypothetical protein |
| hawK | hypothetical protein |
| furB | hypothetical protein |
| yuaM | Putative type 3 effector protein |
| bigM | Putative membrane protein |
| genE | Putative type 3 effector protein |
| shaU | Putative type 3 effector protein |
| arnE | hypothetical protein |
| tigR | FIG01047911: hypothetical protein |
| dumE | Putative type 3 effector protein |
| sadZ | Putative type 3 effector protein |
| rexT | hypothetical protein |
| lamE | Putative DNA-binding protein |
| eviL | FIG01045788: hypothetical protein |
| deaD | hypothetical protein |
| numT | putative DNA-binding protein (histone-like protein hlp-II) |
| halF | Partial putative DNA-binding protein (histone-like protein hlp-II) |
| putM | FIG00787128: hypothetical protein |
| pltO | hypothetical protein |
| tinY | Putative type 3 effector protein |
| lilM | FIG01047641: hypothetical protein |
| mstR | Putative phage regulatory protein |
| marT | FIG01047637: hypothetical protein |
| docB | Putative type 3 effector protein |

**Figure 2.2**: SPI-23 from *S.* Derby D1 and D2. SPI-23 is a putative pathogenicity island, in *S.* Derby it is 37 Kb long and contains 42 ORFS. Gene colours reflect putative function; orange, identifies a phage protein, blue a novel hypothetical protein, light pink identifies pili associated proteins, green a DNA binding protein, yellow a membrane protein, brown a regulatory protein and grey a conserved hypothetical protein. Dark pink identifies an ORF that was predicted to be an effector protein by SIEVE with a p-value of 0.05 or lower.

SIEVE effector protein predictor identified ten ORFs (*sanA, janE, chlE, yuaM, genE, shaU, dumE, sadZ, tinY* and *docB*) in SPI-23 of *S*. Derby D1 and D2 with a p-value of 0.05 or lower corresponding to a Z-Score of 1.5 or higher (Table 2.2) (Samudrala et al. 2009). *docB,* encoding a putative effector protein was identified by RAST as a putative endoprotease and was found here to be conserved in *S.* Derby D1 and D2, *S*. Mbandaka M1 and M2, *S*. Agona SL483, *S*. Dublin ct02021853, *S*. Gallinarum SGG1, *S*. Enteritidis P125109, *S.* Newport SL254, and *S*. Typhimurium LT2. The functional prediction of *docB* fits with the function of other type III secretion effector proteins which have a cysteine protease activity (Shao et al. 2002; Dean 2011). The high number of potential type III secretion system effector proteins makes SPI-23 a strong candidate for classification as a pathogenicity island. The acquisition of this sequence could be responsible for the modulation of the host's cell, cytoskeleton, immune response and intracellular signalling (Waterman & Holden 2003). Though it is not possible to determine here if SPI-23 plays a role in defining the host range of *S.* Derby, the high number of potential effector proteins has identified it as a very interesting region for future experimental study of host adaptation.

**Table 2.2**: Comparison of the annotation results for SPI-23 from different *S. enterica* serovars. This table shows the comparative structure and gene content of SPI-23 in the chromosome of different serovars. SIEVE Z-scores above 1.5 indicate a potential type III effector protein. Functions are taken from RAST, or where no function was given, the highest hit on NCBI BLASTn. Provisional gene names are given for ORFs in SPI-23 of *S.* Derby; this does not conflict with existing gene names, which have been used where possible.

| Gene name | *S.* Derby function | Sieve z Score | *S.* Agona function | Sieve z Score | *S.* Dublin function | Sieve z Score | *S.* Gallinarum function | Sieve z Score |
|---|---|---|---|---|---|---|---|---|
| *gooN* | Phage intergrase | 0.27 | Phage intergrase | −0.05 | Phage intergrase | 0.27 | Phage intergrase | 0.27 |
| *sanA* | Exported protein | 1.56 | Exported protein | 1.56 | Exported protein | 1.40 | Exported protein | 1.56 |
| *redB* | No Matches | 1.43 | - | - | - | - | - | - |
| - | - | - | hypothetical protein | 0.84 | - | - | - | - |
| - | - | - | - | - | threonine operon leader | 1.09 | threonine operon leader | 1.09 |
| - | - | - | - | - | hypothetical protein | 1.21 | hypothetical protein | 0.73 |
| - | - | - | hypothetical protein | 1.31 | - | - | - | - |
| - | - | - | - | - | hypothetical protein | 1.68 | hypothetical protein | 2.06 |
| - | - | - | - | - | hypothetical protein | 2.44 | hypothetical protein | 2.44 |
| - | - | - | RelA/SpoT | 0.25 | - | - | - | - |
| - | - | - | hypothetical protein | 0.71 | - | - | - | - |
| *potR* | prepilin-type N- cleavage/methylation domain protein | 0.90 | prepilin-type N- cleavage/methylation domain protein | 0.72 | - | - | - | - |
| - | - | - | - | - | Pil-v like | 1.82 | Pil-v like | 1.82 |
| - | - | - | - | - | Pil-v like | 0.75 | Pil-v like | 0.75 |
| *talN* | Putative type 4 pilin protein | 1.36 | Putative type 4 pilin protein | 1.03 | Putative type 4 pilin protein | 1.36 | Putative type 4 pilin protein | 1.36 |
| *kayT* | Conserved Hypothetical | 1.44 | Conserved Hypothetical | 1.71 | Conserved Hypothetical | 2.40 | Conserved Hypothetical | 1.66 |
| *pikL* | Hypothetical 91% homology | 0.38 | Hypothetical 91% homology | 0.38 | Hypothetical 91% homology | 1.44 | Hypothetical 91% homology | 0.33 |
| *janE* | No Matches | 1.51 | - | - | - | - | - | - |
| - | - | - | hypothetical protein | 1.06 | - | - | - | - |
| - | - | - | hypothetical protein | 0.98 | - | - | - | - |
| - | - | - | - | - | hypothetical protein | 0.49 | hypothetical protein | 0.49 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *bunJ* | No Matches | 1.35 | - | - | - | - | - | - |
| *minE* | No Matches | 0.13 | - | - | - | - | - | - |
| *penE* | No Matches | 1.25 | - | - | - | - | - | - |
| *comA* | No Matches | 1.40 | - | - | - | - | - | - |
| *zomB* | Putative lipoprotein (94%) | −0.24 | Putative lipoprotein (94%) | −0.05 | Putative lipoprotein (94%) | 1.22 | Putative lipoprotein (94%) | 1.22 |
| - | - | - | hypothetical protein | 0.69 | hypothetical protein | 0.89 | hypothetical protein | 2.61 |
| - | - | - | hypothetical protein | 1.22 | hypothetical protein | 0.30 | hypothetical protein | −0.34 |
| - | - | - | hypothetical protein | 1.17 | - | - | - | - |
| - | - | - | hypothetical protein | 1.46 | - | - | - | - |
| - | - | - | hypothetical protein | 0.52 | - | - | - | - |
| - | - | - | hypothetical protein | 0.98 | - | - | - | - |
| - | - | - | hypothetical protein | 0.98 | - | - | - | - |
| - | - | - | hypothetical protein | 1.13 | - | - | - | - |
| - | - | - | - | - | hypothetical protein | 0.36 | hypothetical protein | 1.30 |
| - | - | - | - | - | hypothetical protein | 0.70 | hypothetical protein | 0.70 |
| - | - | - | - | - | hypothetical protein | 0.16 | hypothetical protein | 0.16 |
| *chlE* | No Matches | 2.02 | - | - | - | - | - | - |
| *newT* | No Matches | 0.82 | - | - | - | - | - | - |
| *rosF* | No Matches | 0.75 | - | - | - | - | - | - |
| *chrD* | No Matches | 0.15 | - | - | - | - | - | - |
| *troN* | No Matches | 0.50 | - | - | - | - | - | - |
| *dinO* | No Matches | 0.72 | - | - | - | - | - | - |
| *hawK* | No Matches | 1.26 | - | - | - | - | - | - |
| *furB* | Hypothetical | 0.39 | - | - | - | - | - | - |
| *yuaM* | No Matches | 1.67 | - | - | - | - | - | - |
| *bigM* | Putative membrane protein (89%) | 1.44 | - | - | - | - | - | - |
| *genE* | No Function | 2.06 | - | - | - | - | - | - |
| *shaU* | No Matches | 1.66 | - | - | - | - | - | - |
| *arnE* | No Matches | 0.82 | - | - | - | - | - | - |
| *tigR* | Conserved Hypothetical | 1.29 | - | - | - | - | - | - |
| *dumE* | No Matches | 1.92 | - | - | - | - | - | - |
| *sadZ* | Hypothetical 88% | 1.58 | - | - | - | - | - | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *rexT* | Pentatricopetide 90% | 0.85 | - | - | - | - | - | - |
| *lamE* | Putative DNA-binding protein | 0.27 | Putative DNA-binding protein | 1.40 | Putative DNA-binding protein | 1.48 | Putative DNA-binding protein | 1.48 |
| *eviL* | Conserved Hypothetical | 0.68 | Conserved Hypothetical | 0.27 | Conserved Hypothetical | 1.49 | Conserved Hypothetical | 0.27 |
| *deaD* | Conserved Hypothetical | 1.33 | Conserved Hypothetical | 0.68 | Conserved Hypothetical | 0.85 | Conserved Hypothetical | 0.85 |
| *numT* | putative DNA-binding protein (histone-like protein hlp-II) | 1.31 | putative DNA-binding protein (histone-like protein hlp-II) | 1.43 | putative DNA-binding protein (histone-like protein hlp-II) | 0.25 | putative DNA-binding protein (histone-like protein hlp-II) | 0.25 |
| - | - | - | - | - | DNA-binding protein H-NS | 0.57 | DNA-binding protein H-NS | 0.68 |
| - | - | - | - | - | hypothetical protein | 1.33 | hypothetical protein | 1.33 |
| - | - | - | - | - | hypothetical protein | 1.31 | hypothetical protein | 1.31 |
| - | - | - | - | - | hypothetical protein | 0.98 | hypothetical protein | 0.98 |
| - | - | - | - | - | hypothetical protein | 1.18 | hypothetical protein | 1.07 |
| - | - | - | - | - | hypothetical protein | 0.93 | - | - |
| - | - | - | - | - | hypothetical protein | 0.73 | hypothetical protein | 0.73 |
| *halF* | Partial putative DNA-binding protein (histone-like protein hlp-II) | 0.98 | - | - | - | - | - | - |
| *putM* | Putative membrane protein | 0.93 | - | - | - | - | - | - |
| *pltO* | Conserved Hypothetical | 0.73 | - | - | - | - | - | - |
| *tinY* | Conserved Hypothetical | 2.23 | - | - | - | - | - | - |
| *lilM* | Hypothetical 90% | 1.41 | - | - | - | - | - | - |
| *mstR* | Putative phage regulatory protein | 0.74 | Putative phage regulatory protein | 1.90 | Putative phage regulatory protein | 1.71 | Putative phage regulatory protein | 1.71 |
| *marT* | Hypothetical 99% | −0.14 | - | - | - | - | - | - |
| - | - | - | hypothetical protein | 0.71 | - | - | - | - |
| - | - | - | hypothetical protein | 0.74 | - | - | - | - |
| *docB* | Putative endoprotease 99% | 1.81 | Putative endoprotease 99% | 1.81 | Putative endoprotease 99% | 1.81 | Putative endoprotease 99% | 1.81 |
| - | - | - | hypothetical protein | −0.14 | - | - | - | - |
| - | - | - | hypothetical protein | 1.11 | - | - | - | - |
| - | - | - | TPR domain protein, putative component of TonB system | 0.56 | - | - | - | - |
| - | - | - | hypothetical protein | 0.49 | - | - | - | - |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| - | - | - | hypothetical protein | 0.75 | - | - | - | - |
| - | - | - | putative P4-type integrase | 1.18 | - | - | - | - |
| | - | - | hypothetical protein | 1.30 | - | - | - | - |
| - | - | - | hypothetical protein | 0.93 | - | - | - | - |
| - | - | - | hypothetical protein | −0.05 | - | - | - | - |
| - | - | - | hypothetical protein | 0.53 | - | - | - | - |

2.3.10.11 Comparison of SPI-23 from *S.* Agona SL483, *S.* Dublin ct02021853 and *S.* Gallinarum SGG1 with SPI-23 found in *S.* Derby D1 and D2

There were no genes between *gooN* and *docB* in *S.* Enteritidis P125109 even though NCBI BLASTn showed 100% sequence homology with 17 genes from SPI-23 of *S.* Derby D1 and D2. A four way comparison between SPI-23 excised from the genomes of *S.* Agona SL483, *S.* Dublin CT02021853, *S.* Gallinarum SGG1 and *S.* Derby D1 was performed (Figure 2.3). The differences in SPI-23 between *S.* Agona SL483 and *S.* Derby D1 and D2 are dispersed across the island in four sections (Table 2.2). *S.* Agona SL483 contains seventeen unique genes and lacks twenty two genes when compared to the SPI-23 of *S.* Derby D1 and D2. All of the genes unique to *S.* Agona SL483 with the exception of three are of hypothetical status, *relA* a GDP/GTP pyrophosphokinase, a putative component of the TonB system and a P4-type intergrase. SPI-23 in *S.* Agona SL483 contains only four genes that are likely candidates for type III secretion system effector proteins (*sanA, kayT, mstR* and *docB*). All four genes are identical in nucleotide sequence to that of *S.* Derby D1. Serovars *S.* Dublin CT02021853 and *S.* Gallinarum SGG1 have identical sequences for SPI-23. There are fifteen genes in the SPI-23 of these two serovars that are not found in either *S.* Derby D1 and D2 or *S.* Agona SL483 and fifteen which are found in all five serovars. Only two of the hypothetical genes found in *S.* Agona SL483 and not *S.* Derby D1 and D2 are found in the SPI-23 of *S.* Dublin CT02021853 and *S.* Gallinarum SGG1. The SPI-23 of *S.* Dublin CT02021853 and *S.* Gallinarum SGG1 contains four unique genes that are not of hypothetical status. This comprises two *pilV*-like proteins, a DNA binding protein HNS and a threonine operon leader protein. Both *S.* Dublin CT02021853 and *S.* Gallinarum SGG1 contain eight putative type III secretion system effector proteins, three of these are unique genes to these two sequences and are absent from *S.* Derby D1, D2 and *S.* Agona SL483.

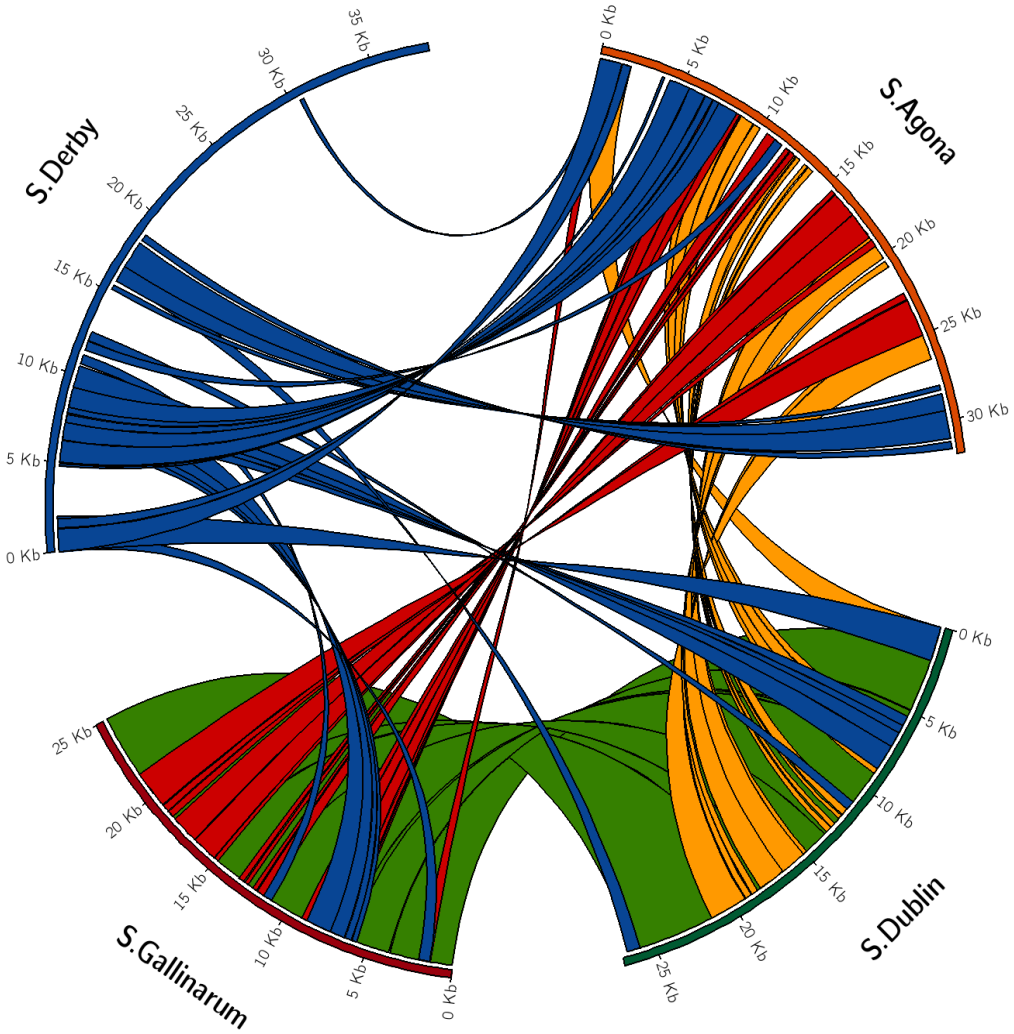



**Figure 2.3**: SPI-23 four way nucleotide comparison. Four way comparison of the nucleotide sequence of SPI-23 from *S*. Derby D1, *S*. Agona SL483, *S*. Dublin CT02021853 and *S*. Gallinarum RKS5078. *S*. Derby D1 possess the largest SPI-23 (37Kb) island of the sequenced strains available on NCBI genome and the most novel in nucleotide sequence. Over 60% of the nucleotide sequence of SPI-23 in *S*. Derby D1 is unique and contains no entry on NCBI nucleotide database.

Two putative effector proteins are different between the SPI-23 sequences of *S.* Dublin CT02021853 and *S.* Gallinarum SGG1, *sanA* is present in *S.* Dublin CT02021853 but not identified as a putative effector protein and similarly a hypothetical gene in *S.* Dublin CT02021853 and *S.* Gallinarum SGG1. Interestingly SIEVE predicted the gene *kayT* as a type III secretion system effector protein from the amino acid sequences of *S.* Agona SL483, *S.* Dublin CT02021853 and *S.* Gallinarum SGG1 but not that of *S.* Derby D1 or D2. Similarly *sanA* is identified as a candidate effector protein in all sequences with the exception of *S.* Dublin CT02021853.

### 2.3.11 Prophage

Bacteriophages are viruses that infect bacteria, integrating into the bacterial genome in order to replicate; in this form they are known as prophage. As a result of phage insertion the genome gains a substantial amount of foreign sequence, much of which encodes phage structural proteins. However, some phage carry cargo genes which convey a pathological advantage to the recipient (Boyd et al. 2012). The process of lysogenic conversion prevents the prophage from destroying the host through maturation of progeny. The cargo genes and prophage remnants are therefore retained within the bacterial lineage, undergoing genetic mutation, drift and selection (Canchaya et al. 2003).

PHAST identified distinct complements of intact prophage and remnant prophage regions between *S.* Derby and *S.* Mbandaka (Zhou et al. 2011). All isolates contain four phage regions, sharing only the remnants of a BcepMu phage in common. This remnant is identical in all strains, suggesting that the integration and degradation of this phage predates the split between *S.* Derby and *S.* Mbandaka. *S.* Mbandaka isolates contain the same

prophage regions in the same locations along the chromosome. These comprise one intact prophage, resembling phage P2, two questionable prophage, similar to L413c and Epsilon34 and one incomplete prophage BcepMu. *S.* Derby isolates differed on the location of the prophage within the chromosome and the number of genes in all four phage regions. No ambiguous bases were identified in these regions. The partial prophage resembling SFV contains one additional ORF in D1 than in D2 and occupies the same region that the complete prophage of SFV occupies in D2. Whereas the complete copy of SFV in D1 occupies the position of the complete prophage in D2 and contains one fewer ORF. The BcepMu partial in D1 contains two additional ORFs than that found in D2. In D1 the intact prophage resembling ST64B comprises three additional ORFs than that found in D2, they occupy the same chromosomal region. ST64B is of particular interest as its homolog in *S.* Typhimurium SL1344 contains a gene with homology to a type III secreted effector protein SseK2, mutants of which have shown to have reduced pathogenicity in a bovine model (Brown et al. 2011). *S.* Derby contains an intact version of IN0, a transposon identified from *Pseudomonas aeruginosa*.

### 2.3.12 *S.* Derby and *S.* Mbandaka contain unique CRISPR spacer sequences

CRISPR operons convey an adaptive immunity against plasmids and bacteriophage to a broad range of archaeal and bacterial species. This is achieved through integration of unique regions of foreign DNA into the prokaryotic chromosome. Subsequent expression of these fragments interfere with foreign nucleic acid, through complementation (Sorek et al. 2008; Fricke et al. 2011). The spacer sequences within a CRISPR operon reflect the historical interaction between the lineage of a strain and foreign DNA elements. The efficacy of invasion and ecological distribution of bacteriophage, transposons and plasmids have been found to associate with particular hosts and environments (Corpet 1986; Rotger

& Casadesús 1999; Qu et al. 2008; Clokie et al. 2011). Hence the different genomic complement of prophage and CRISPR operon elements in *S.* Derby and *S.* Mbandaka could reflect their particular niche or even define their niche within a specific group of livestock species.

*S.* Derby D1 and D2 contain four CRISPR operons each, with 34 and 35 spacers respectively. *S.* Mbandaka M1 contained two CRISPR operons with 25 spacers. M2 contains three CRISPR operons with 27 spacers. With the exception of two spacers, the sequences are completely unique to each serovar. *S.* Derby isolates contain four CRISPR spacer operons, the smallest contains only one sequence with the largest containing 25 spacers. D2 contains two additional spacer sequences and half of a much larger spacer than D1. *S.* Mbandaka isolates differ on the number of spacers they each contain; M1 contains two operons while M2 contains three. The majority of spacers are homologous between the isolates, with M2 containing four additional spacers. M2 CRISPR operon 2 and 3 contain all of the spacer sequences in M1 CRISPR operon 1. All spacer sequences can be found in the supplementary materials (Supplementary material 2.1).

We have already shown that *S.* Derby strains contain seven functionally unique CRISPR operon proteins. The lack of functional homolog in *S.* Mbandaka, leaves it without a functioning CRISPR operon. CRISPRdb shows here that *S.* Mbandaka strains contain arrays of CRISPR spacer sequences; these may be remnant from when *S.* Mbandaka had a fully functioning CRISPR operon. *S.* Mbandaka may now be susceptible to those phage and plasmid for which it once had resistance; this could reflect the loss of positive selection pressure on the operon from the surrounding environment.

2.3.13 Estimating the time since the divergence of *S.* Derby and *S.* Mbandaka

Whole genome alignment and SNP calling across CDS nucleotide sequences was used to estimate the years since divergence of D1, D2, M1 and M2. Interestingly the time since the divergence of *S.* Derby and *S.* Mbandaka is estimated at between 182,291 and 625,000 years, based on an average of the Ks values for all four pair-wise comparisons of *S.* Derby and *S.* Mbandaka isolates ranging between 0.015 and 0.019. The divergence of *S.* Derby and *S.* Mbandaka coincides with the estimated time of the divergence of all domesticated pig species, approximately 500,000 years ago (Giuffra et al. 2000). The time since the split between D1 and D2 was estimated at between 350 and 1200 years ago based on 31 synonymous SNPs spread across 923506 synonymous positions. The isolates M1 and M2 are estimated to have diverged between 1271 to 4357 years ago based on 118 synonymous SNPs spread across 965114 synonymous positions.

## 2.4 Conclusions

We estimate here that *S.* Derby D1 and D2 diverged from *S.* Mbandaka M1 and M2 between 182kya and 625kya, during this period these serovars appear to have adapted towards two distinct ranges of host species. Comparative functional genomics has alluded to several mechanisms that could contribute towards distinct host adaptations of *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2. Most noteworthy of these differences are the diversity in SPI-6 gene complement and the discovery in the chromosome sequence of *S.* Derby, of a new 37Kb genomic island, SPI-23, encoding 42 ORFS, ten of which are putative TTSS effector proteins. The absence of functional homologs to several CRISPR operon genes in the chromosome sequences of *S.* Mbandaka may reduce the fitness of the serovar in environments laden with actively integrative foreign genetic elements. The increased gene dosage of the Csg-biofilm operon and the Ycd-swarming operon in *S.*

Mbandaka could make the implementation of these two behaviours more readily achievable. Both of these behaviours are considered stress responses. *S.* Mbandaka also possesses an operon pertaining to the uptake and metabolism of D-galactonate into glycolysis which is absent from the chromosome of *S.* Derby.

The genetic background in which the function of the genes discussed here have been characterised is non-isogenic to the chromosome of *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2. Due to the different context these genes are found in, firm conclusions on the function of these genes in these specific serovars can only be formed through further biological experimentation.

## 2.5 Methods

### 2.5.1 Bacterial strains and culturing

The original isolates were stored at RT on Dorset egg slopes from which bead stocks were made with HIB + 30% glycerol, samples were frozen to −80 °C in 2010 and remained frozen throughout the study. Unless stated otherwise, strains were grown for 16 hours aerobically either on LB agar plates or in liquid broth vigorously agitated at 220 rpm.

### 2.5.2 DNA extraction, genome sequencing and assembly

DNA was extracted from 3 ml overnight cultures as per manufacturer's instructions (Invitrogen EasyDNA kit). Sequencing was performed by the AHVLA Central Sequencing Unit, Weybridge. A Roche GSFLX titanium 454 pyrosequencer was used to produce rapid and paired-end libraries for whole genome DNA preparations of *S.* Derby D1 and D2 and *S.*

Mbandaka M1 and M2. Roche protocols were used in all stages of sequencing. Paired-end library inserts were between 4 Kb to 9Kb, containing 20,000 to 89,000 reads each. The rapid libraries contained between 69,000 and 173,000 reads each. Sequences were assembled *de novo* using Newbler v2.5. Scaffolds were reordered in ACT v9.0 in reference to a DoubleACT v2 comparison file of each genome with D1; D1 was chosen as the assembly consisted of a single scaffold (HPA 2004; Carver et al. 2005). The final sequences were then formatted so as to begin at the gene *thrL*, in line with other published *Salmonella enterica* genomes.

### 2.5.3 Automated annotation, metabolic model construction and comparative genomics

Genomes were annotated using the RAST annotation system performed on 9/10/12, backfilling of gaps and automatic error fixing were enabled. Functional comparisons were implemented using the SEED genome viewer v2 (Aziz et al. 2008). An automated metabolic reconstruction was also produced from the complete genome sequence using the ModelSEED server v1.0 (Henry et al. 2010). Differences in *S.* Derby and *S.* Mbandaka models were identified through gene overlays on top of KEGG maps (Ogata et al. 1999). Reciprocal BLASTing was implemented in SEED genome viewer for each ORF that differed between isolates to identify functional homologs. The genomes were also compared through sequence homology. The population of "hypothetical" and "putative" genes were aligned with a cut off of 90% bi-directional amino acid sequence homology.

2.5.4 Mobile genetic elements

SPIs were identified from the genomes of *S.* Derby and *S*. Mbandaka through alignment of the insertion sequences with the newly acquired genomes. SPIs for the serovars *S*. Choleraesuis B67 (SPI-1, 2, 3, 4, 11, 12) and 1240 (SPI-11), *S.* Derby (SPI-5) isolate not specified, *S.* Gallinarum SGG1 (SPI-13) and SG8 (SPI-14) (Shah et al. 2005), *S*. Dublin isolate not specified (SPI-3), *S*. Typhi CT18 (SPI-3, 6, 7, 8, 9, 10) *S*. Typhimurium LT2 (SPI-1, 5, 18) and 14028 (CS54, SPI-3 ) were acquired from PAI-DB website (Yoon et al. 2007), these were aligned using DoubleACT with the newly isolated islands. From previously published annotated genomes, SPI-22 from *S*. Bongori (Fookes et al. 2011) and SPI-15 , 16 and 17 from *S*. Typhi CT18 were excised from the genome (Vernikos & Parkhill 2006). SPI-19, 20 and 21 were excised from the genome of *S. enterica* subspecies *arizonae* (IIIa) serotype 62:z4,z23:- (Blondel et al. 2009). In most cases the SPI from *S*. Derby and *S*. Mbandaka could be completely annotated through alignment using DoubleACT with the existing SPI. Where gaps were present the BLAST facility was first used on PAI-DB, when no results were obtained, the RAST annotation and NCBI BLASTn were used to annotate the genes, and extensive literature research was used to assign a putative role in pathogenesis. Prophage were identified and categorised as intact, questionable and partial using PHAST (Altschul et al. 1997; Zhou et al. 2011). CRISPR spacers were identified using CRISPRfinder (Grissa et al. 2007a). CRISPRdb BLAST facility was used to see if the spacers found in the newly sequenced genome were found within other bacterial species (Grissa et al. 2007b). Spacer sets of the newly sequenced strains were also cross compared to elicit the historical differences in exposure to phage that has occurred since their divergence. Hypothetical proteins found in SPI-1 and SPI-23 were tested for potential roles as TTSS effector through implementation of SIEVE-SVM based TTSS effector protein predictor (Samudrala et al. 2009). A Z-score above 1.5 was taken to reflect a good indicator of a type III effector protein. SPI-23 was identified and

extracted from the publicly available genomes for *S*. Agona SL483, *S*. Dublin CT02021853 and *S.* Gallinarum RKS5078. The sequences were annotated using RAST and SEIVE. Sequences were compared using DoubleACT and ACT.

2.5.5 Estimation of years since *S*. Derby and *S*. Mbandaka diverged

For each genome (D1, D2, M1 and M2), nucleotide sequences of the CDS identified in the RAST annotation were converted into a single concatenated FASTA file using Artemis (Rutherford et al. 2000; Aziz et al. 2008). Sequences were aligned in Mauve genome aligner (Darling et al. 2004). Aligned sequence blocks were reassembled from the Mauve alignment. A multiFASTA file was made for each combination of the four genomes. DNAsp was used to identify the synonymous and non-synonymous positions and SNPs (Librado & Rozas 2009). The years since the isolates diverged was estimated as described by Foster et al. 2009 using the following formula (Foster et al. 2009):

$$X = \frac{Ks}{(2ZY)}$$

Where X is the years since divergence, Ks is the proportion of synonymous SNPs to synonymous sites, Z is the mutation rate per generation and Y the number of generations per year. The mutation rate for *S*. Typhi has been estimated at 1.6 x $10^{-10}$ mutations per nucleotide per generation, calculated over 20,000 generations (Barrick et al. 2009). This is very close to the calculation for the rate of mutation in *E. coli* of 1.4 x $10^{-10}$ per nucleotide per generation estimated over 20,000 generations (Lenski et al. 2003). The rate of mutation has been shown to be highly variable and dependent on the mutation rate phenotype of the lineage (Barrick et al. 2009). Here we use both rates calculated between *S.* Typhi generations as the lower limit and *E. coli* generations as the upper limit, with the assumption that the true rate sits somewhere between these two. This is based on the

assumption that the variation in the mutation rate correlates with the phylogenetic relationship between the strains (Fukushima et al. 2002; van Cuyck et al. 2011). The number of generations per year (Y) is taken from the estimate produced for a wild population of *E. coli* of between 100 and 300. The denominator is multiplied by two as it applies to the number of SNPs between two genomes (Ochman et al. 1999). To estimate the time of divergence between *S.* Derby and *S.* Mbandaka an average was taken of the four possible Ks values for each of the four pair wise comparisons.

### 2.5.6 Visualisation of sequence architecture

Genomic maps were constructed using CIRCOS circular visualization of data tool v 0.56 (Krzywinski et al. 2009). A program for calculating GC skew in R v2.11.0 using the library SeqinR v3.0-6 was modified from R graphical manual example "fragment of *E. coli* chromosome" (Charif & Lobry 2007; R Core Team 2012; RGM 2012). The GC skew was calculated under a 1 kb window at a 200 bp interval. The RAST annotation files were de-constructed into four tracks, forward and reverse coding DNA and RNA. The SPI-23 comparison maps were constructed from modified DoubleACT outputs for each combination of *S.* Derby D1 SPI-23 and the genomes of *S.* Agona SL483. *S.* Dublin CT02021853 and *S.* Gallinarum RKS5078, with a 100 bp cut-off for width between non-homologous sequences.

# Chapter 3:
# **Characterisation of *Salmonella* pathogenicity island 23 found in *S.* Derby**

The text and figures are the same as those that were submitted on 24/7/13 to ASM Infection and Immunity. Manuscript number IAI00905-13.

Dr. Monika A. Tchórzewska helped plan and perform IVOC experiments. Prof. Roberto La Ragione performed the post mortems. Dr. Manal AbuOun helped plan and troubleshoot the mutant construction. Mr William A. Cooley and Mr David J. Everest performed the microscopy. Prof. Vincent A. A. Jansen, Prof. Martin J. Woodward and Dr. Liljana Petrovska supervised the work. The manuscript was written and edited with the help of Dr. Manal AbuOun, Prof. Vincent A. A. Jansen and Prof. Martin J. Woodward.

## 3.1 Abstract

*Salmonella enterica* serovars Derby and Mbandaka are isolated from different groups of livestock species in the UK. *S*. Derby is predominantly isolated from pigs and turkeys and *S*. Mbandaka is predominantly isolated from cattle and chickens. Alignment of the genome sequences of two isolates of each serovar led to the discovery of a new putative *Salmonella* pathogenicity island, SPI-23, in the chromosome sequence of *S*. Derby isolates. SPI-23 is 37kb in length and contains 42 ORFs, ten of which are putative type III effector proteins. In this study we use porcine jejunum derived cell line IPEC-J2 and *in vitro* organ culture (IVOC) of porcine jejunum and colon, to characterise the association and invasion rates of *S.* Derby and *S*. Mbandaka, and tissue tropism of *S*. Derby respectively. We show that *S*. Derby invades and associates to an IPEC-J2 monolayer faster than *S*. Mbandaka, and that *S*. Derby preferentially attaches to porcine jejunum over colon. We also show that nine genes across SPI-23 are up-regulated to a greater degree in the jejunum compared to the colon. Furthermore, we constructed a mutant of the highly up-regulated, *pilV*-like gene, *potR*, and found that it produces an excess of surface pili compared to the parent strain which form a strong agglutinating phenotype that prevents association and invasion of IPEC-J2 monolayers. We conclude that SPI-23 is a pathogenicity island and suggest that it may play a role in tissue tropism of *S*. Derby.

## 3.2 Introduction

*Salmonella enterica sub enterica* is an important zoonotic pathogen of warm blooded vertebrates, including humans and livestock. The symptoms of Salmonellosis include chronic gastroenteritis, affecting a wide range of host species and caused primarily by broad host range serovars, and an often fatal typhoid fever affecting a narrow range of host species, caused primarily by host limited or restricted serovars (Baumler et al. 1998; Suar et

al. 2006). The discovery of host ranges was made possible by the large diversity in the complement of surface antigens, O and H, and the routine serotyping of clinical and veterinary isolates. International serotyping has provided statistics that have elucidated links between certain serovars and a defined host species range (Kingsley & Bäumler 2000; CDC PHLIS 2009; AHVLA 2011; CDC 2011).

In previous work we compared the genomic sequences of two isolates of two serovars of *S. enterica* subspecies *enterica*, *S*. Derby and *S*. Mbandaka, which display different host species biases in the UK: *S*. Derby is prominently isolated from pigs (40%) and turkeys (50%) and *S*. Mbandaka from chickens (65%) and cattle (20%) (AHVLA 2011; Hayward et al. 2013) (refer to chapter 1). Alignment of the nucleotide sequences of the chromosomes of *S*. Derby with the chromosomes of *S*. Mbandaka led to the discovery of a new putative *Salmonella* pathogenicity island (SPI) in isolates of *S*. Derby, designated SPI-23 (Hayward et al. 2013) (refer to section 2.3.10.10). SPI-23 is 37kb in length and encodes 42 genes of which ten were identified by the online tool SEIVE as potential type III effector proteins (Samudrala et al. 2009). SPIs are an exemplar of modular evolution whereby large regions of DNA encoding genes for increased pathogenicity evolve in other bacterial lineages and are conveyed to the recipient cell through a single integration event. An example of this is the acquisition by *Salmonella* genus of the highly conserved SPI-1, which encodes a type three secretion system. This integration event has been posited as the pivotal moment that led to the divergence of the *Salmonella* genus from *Escherichia coli* (Baumler et al. 1998).

BLASTn alignment with a non-redundant nucleotide database found that 17 of the 42 genes were unique in nucleotide sequence to the SPI-23 found on the chromosome of *S*. Derby.

SPI-23 was also discovered in the chromosome sequences of serovars *S*. Agona, *S*. Dublin and *S*. Gallinarum, though in these sequences the islands contained fewer putative pathogenicity genes (Hayward et al. 2013) (refer to section 2.3.10.11).

Given the UK isolation and serotyping data, and the large number of unique putative pathogenicity genes, we hypothesized that SPI-23 is associated with the host bias demonstrated by *S*. Derby. We show that SPI-23 is regulated in a tissue specific fashion and that a knock-out mutant of the most up-regulated gene, *potR*, results in agglutination in static culture. We discuss the possible role that *potR* and other SPI-23 genes may play in tissue tropism.

## 3.3 Methods and materials

### 3.3.1 Strains and culturing

The strains used in this study, *S*. Derby D1 and D2 and *S*. Mbandaka M1 and M2, were selected as their chromosome sequences have been previously annotated and compared (Hayward et al. 2013) (refer to chapter 2). The strains were isolated from different geographical locations during background monitoring performed on livestock in the UK between 2000 and 2010. *S*. Derby isolates D1 and D2 were obtained from pigs in 2008 and *S.* Mbandaka isolates M1 and M2 were obtained in 2008 and 2009 from cattle (Hayward et al. 2013) (refer to chapter 2).

Unless stated otherwise, the strains were grown at 37$^{o}$C for 16 hours aerobically either on LB agar plates or in LB broth agitated vigorously at 220rpm. For porcine IVOC and IPEC-

J2 assays mid-log cultures were used; this was achieved through inoculation of 20ml of pre-warmed LB with 200µl of a 16 hour overnight culture, followed by incubation for 3 hours at 37$^{o}$C with agitation at 220rpm. A bead stock was made of the *potR* mutant, *S.* Derby D1 *ΔpotR*::kan directly from a planktonic culture. Due to the strong agglutinating phenotype and sensitivity to prolonged periods at cold temperatures, beads were made directly from fresh planktonic cultures grown for 16 hours in LB broth at 37$^{o}$C with agitation at 220rpm. From the culture, 100µl of planktonic culture was mixed in a 2ml CryoVial (Fisher Scientific, USA) containing 1.5ml of HIB + 30% glycerol and quickly placed into a -80$^{o}$C freezer, to freeze the culture in its planktonic state.

### 3.3.2 Culturing and infection of IPEC-J2 monolayers

Association and invasion assays using the porcine jejunum derived IPEC-J2 monolayers were performed in triplicate on three separate occasions, based on the method described by Searle et al. (2009) (Searle et al. 2009). IPEC-J2 (passage 70-72) cells were seeded at 1.6x10$^{5}$cells/ml into 24 well plates and cultured using IPECs media consisting of: Dulbecco's Modified Eagle's Medium (DMEM; Sigma, UK) supplemented with 5% foetal bovine serum (Sigma, UK), 1% 2mM l-glutamine (Sigma, UK), 1% sodium pyruvate (Sigma, UK) and 1% ITSS (Sigma, UK). Following 48 hours incubation at 37$^{o}$C (100% confluence) the monolayers were washed three times with Hank's balanced salt solution (Sigma, UK). Each washed monolayer was inoculated separately with 1ml of *S.* Derby D1 and D2, *S.* Mbandaka M1 and M2, *Escherichia coli* K12 DH5*α* (non-invading control) and a blank PBS control. This was performed in triplicate for each strain. For these studies, inocula were cultured to mid-log phase, standardised to an $OD_{540nm}$ of 1.2 in PBS, and further diluted 1:20 in IPECs medium. Inoculated monolayers were incubated statically, at 37$^{o}$C in 5% $CO_2$. The assay was performed for five different incubation periods, 0, 5, 15,

30 and 60 minutes to identify variation in association and invasion rate between *S*. Derby D1 and D2 and *S*. Mbandaka M1 and M2. After incubation, monolayers were washed a further three times with Hank's balance salt solution. To the plates designated for studying invasion, 1ml of IPECs medium containing 1% gentamycin (Sigma, UK) was added to each well, to kill external, but not internal, bacteria. Invasion plates were incubated for a further two hours at 37$^{o}$C in 5% $CO_2$. After incubation the monolayers were washed a further three times with Hank's balance salt solution (Sigma, UK).

Monolayers of both preparations were disrupted with magnetic stirrers for 10 minutes in 1ml of 1% TritonX-100 (Sigma, UK) diluted in PBS. Serial dilutions between $10^{0}$–$10^{-8}$ for each preparation including the initial standardized cultures were made in PBS, plated on to LB agar and incubated at 37$^{o}$C for 16 hours. Colony forming units per ml were determined. T-tests were performed between the number of colony forming units per ml (CFU/ml) enumerated for all combinations of strains at each time point. To the average values of all the time steps a linear model was fitted using OfficeLibre Calc. The data were represented as bargraphs using OfficeLibre Calc (The Document Foundation 2012).

### 3.3.3 Porcine *in vitro* organ culture and tissue association assay

*In vitro* organ culture (IVOC) was performed as previously described by Collins et al. (Collins et al. 2010). On two separate occasions, three, six-week old, cross bred commercial pigs were stunned and euthanized through exsanguination. Jejunum and colon tissues were immediately removed from each pig and stored in separate Duran (Duran group, Germany) bottles containing 300ml of pre-chilled IVOC medium which comprised RPMI 1640 medium (Sigma, UK) containing 10% foetal bovine serum (Sigma, UK), 0.25%

lactalbumin hydrosylate (Sigma, UK), 75mM mercaptoethanol (Sigma, UK), 0.2µg $ml^{-1}$ hydrocortisone (Sigma, UK) (1:1 chloroform/ethanol), 0.1µg $ml^{-1}$ ITSS (Sigma, UK), and 2mM L-glutamine and L-aspartate (Sigma, UK). Jejunum and colon tissues were opened through a single transverse incision, followed by washing through submersion in PBS. The clean tissues were maintained in IVOC medium. Tissues were cut with a scalpel into $2cm^2$ squares; sections were stored in IVOC medium until all tissues had been processed. Sections were mounted on CellCrowns (Scaffdex, Finland) with the mucosal side facing inwards, so as to form the base of the compartment. Crowns were placed into 24 well plates containing 1ml of IVOC medium.

Association assays were performed on the mounted jejunum and colon tissues on two separate occasions. Mid-log cultures of *S.* Derby D1 and D2 were standardised in PBS to an $OD_{540nm}$ of 1.2, from which a 1:20 dilution was made in IVOC medium pre-warmed to 37°C. Mounted tissues were inoculated in quadruplicate for RNA expression and association assays with 1ml of each inoculum. A no tissue control (just bacteria) was also included for the RNA samples. Plates were subsequently incubated for 30 minutes at 37°C with 5% $CO_2$. After incubation tissues were washed in 2ml of sterile PBS; samples for RNA were added to 1ml Tri-reagent (Sigma, UK), samples for enumeration were added to 10ml of PBS.

Samples for enumeration were homogenised using a D-7801 hand held homogenizer equipped with emulsifier blades (Ystral, Germany). Homogenates were diluted in PBS in a series between $10^{-1}$ to $10^{-4}$ and were plated onto BGA (Oxoid, UK) supplemented with 1µg/ml novobiocin (Sigma, UK) to enrich for *S. enterica* (Hoben et al. 1973). Plates were

incubated for 16 hours at 37°C, prior to enumeration. T-tests were performed between the number of colony forming units per ml (CFU/ml) enumerated for all combinations of strains and tissues.

### 3.3.4 RNA extraction and quantitative RT-PCR (qRT-PCR)

Porcine colon and jejunum tissue sections infected with *S*. Derby D1 and blank controls, as well as no tissue controls were stored in Tri-reagent at -80$^{o}$C until RNA purification was performed. Samples were thawed on ice and homogenised using a Retsch MM301 bead homogeniser with 2.5mm steel beads for 2 minutes at 30S$^{-1}$ (Retsch, Germany). RNA was extracted using the Tri-reagent (Invitrogen, UK) chloroform method as per manufacturer's instructions. The final pellets were re-suspended in nuclease free water. Genomic DNA was removed using two treatments with DNA free (Ambion, UK) as per manufacturer's instructions and confirmed by the absence of a band after gel electrophoresis of a PCR reaction for the housekeeping gene *aroC* (Supplementary material 3.1). qRT-PCR reactions were performed using one-step Brilliant II SYBR Green with low ROX master mix (Stratagene, UK) as per manufacturer's instructions. In brief, 25µl reactions were performed consisting of, 12.5µl of Brilliant II SYBR green with low ROX master mix, 8.5µl of nuclease free water, 1µl of reverse transcriptase/RNase block enzyme, 1µl of RNA (200ng), and 1µl (200nM) of forward and reverse primers specific to the genes *gmk, sanA, potR, talN, chlR, genE, shaU, dumE, sadZ, tinY* and *docB* (Supplementary material 3.1). For all primer sets and treatments, control reactions were performed, comprising either whole cell DNA extract of *S*. Derby D1, RNA with no reverse transcriptase or reverse transcriptase with no nucleotide template. Reactions were performed on a Stratagene Mx3000P, the parameters of the reaction were in accordance with the 2-step reaction recommended by Stratagene for the one-step Brilliant II SYBR Green with low ROX

master mix with an annealing temperature of 52°C. The threshold cycle (CT) was calculated for each reaction by the Stratagene Mx3000P software V2.0 (Stratagene, UK) and standardised to the reading for the constitutively expressed guanylate kinase gene, *gmk*, as previously described (Bohez et al. 2006; López-Garrido & Casadesús 2010). The efficiency of each PCR primer set was tested using dilutions of DNA template in nuclease free water to 1:10, 1:100, 1:1000; this was then used to adjust the relative concentration of transcript to the *gmk* control. Results are given as fold changes relative to the readings for the no tissue control. T-tests were performed on the fold change values between jejunum and colon, p-values below 0.05 were deemed to be significant.

### 3.3.5 Mutant construction and validation

The gene *potR* was deleted from the chromosome of *S.* Derby D1 using the Quick & Easy *E. coli* Gene Deletion Kit (Genebridges, Germany), as per manufacturer's instructions. The following experiment was also performed for the putative type III effector protein gene *genE*, in this case no cells were recovered after several attempts at the initial recombination step.

In brief, a 10ml overnight culture of *S.* Derby D1 was made electrocompetent through five successive rounds of centrifugation at 4000rpm for 10 minutes at 4°C followed by re-suspension in ice cold distilled water and incubation on ice for 20 minutes. A final round of centrifugation was performed, this time the pellet was re-suspended in 200µl of distilled water containing 10% glycerol. Cells were stored on ice until used.

To delete the gene *potR* from SPI-23, a kanamycin resistance cassette under a pGB2 promoter (Genebridges, Germany) was produced so that it was flanked by 50bp of sequence homologous to the regions either side of the gene, as per manufacturer's instructions. In brief, the site specific homology arms were introduced into the cassette by designing 20bp primers (Sigma, UK) for the 5' and 3' ends of a FRT-pPGK-pGB2-neo-FRT cassette (Genebridges, Germany) with an additional region of 50bp that was homologous to either the 5' or 3' sequence adjacent to *potR* (Supplementary material 3.1). PCR products were purified through gel electrophoresis, staining with ethidium bromide and were excised from the gel and purified using a QIAquick gel extraction kit (Qiagen, USA) as per manufacturer's instructions. The concentration of the purified cassette was determined using a Nanodrop N1000, spectrophotometer (Nanodrop, USA).

Electroporation was performed twice, once to introduce a plasmid carrying Red/ET recombinase under a temperature dependent promoter (Genebridges, Germany), and secondly to introduce the cassette with homology arms for site specific recombination with *potR*. In both cases, 2µl of nuclease free water containing approximately 200ng of DNA (either cassette or plasmid) was added to 40µl of electrocompetent cells in a chilled 1mm gap electroporation cuvette. Electroporation was performed with the following parameters: charging voltage 1.8kV, resistance 600 ohms, pulse time of 4.2m/s and capacitance timing 25µF. Selection of, and recombination in, transformed strains was performed as per manufacturer's instructions (Genebridges, Germany). The correct insertion of the kanamycin cassette and deletion of *potR* was confirmed through colony PCR of the cassette chromosome margins and the central 1kb of the gene (Supplementary material 3.1). The mutant strain is from here on referred to as *S*. Derby D1 *ΔpotR*::kan.

### 3.3.6 Negative staining of *S.* Derby D1 parent and *ΔpotR*::kan mutant

Negative staining was performed as follows, a 50µl of planktonic bacterial suspension was air dried onto a glow discharged formvar/carbon coated support grid. Grids were then negatively stained with 2% phosphotungstic acid (pH 6.6) and immediately examined under a FEI Tecnai 12BT transmission electron microscope (FEI, USA) at 80kV.

### 3.3.7 Differential staining and confocal microscopy of invaded and associated bacteria to IPEC-J2 monolayers

IPEC-J2 monolayers were cultured on cover slips. The monolayers were exposed to *S.* Derby D1 and *S.* Derby D1 *ΔpotR*::kan as described above for the association assay. The monolayers were incubated with the inoculum for 4 hours to allow for a large number of bacteria to adhere and invade. After the incubation step, the inoculum was removed and the infected cells were fixed with 4% paraformaldehyde for 16 hours at 4$^{o}$C, instead of being homogenised with 1% TritonX-100 as in the above described association assays. Cover slips were removed from fixative and washed three times in PBS before being blotted with tissue paper and placed monolayer side up in a fresh 12 well plate. Goat anti-*Salmonella* O4 antibodies were added to the centre of each cover slip and incubated at room temperature under tin foil for 45 minutes, at which point the cover slips were removed from the well, washed three times in PBS and added to a fresh 12 well plate. This was repeated three times with the antibodies, donkey anti-goat labelled with Alexa 488 (Invitrogen, UK), goat anti-*Salmonella* O4, and donkey anti-goat labelled with Alexa 555 (Invitrogen, UK). Between the applications of antibodies Alexa 488 and the second application of goat anti-*Salmonella* O4, the IPEC-J2 cells were made permeable to antibodies through 10 minutes incubation with 0.3% TritonX-100 diluted in PBS. After the final antibody was washed off the coverslips were dried with tissue paper and placed monolayer side down on to 5µl drops of

DAPI (Vector Laboratories, UK) on glass slides. The slides were incubated at room temperature under tin foil for 10 minutes, after which a thin layer of nail varnish was painted around the edge of the cover slips. Slides were viewed on a Leica TCS SP2 AOBS confocal system (Leica, Germany) attached to a Leica DM IRE2 microscope (Leica, Germany) equipped with ArKr laser excitation (488nm), HeNe laser excitation (543) and a diode laser (405nm). Oil-immersion objective lenses (40x and 63x) were used, and imaging parameters were selected to optimise resolution.

It is worth noting that scanning electron microscopy was performed with IPEC-J2 monolayers exposed to *S*. Derby strains D1 and the mutant D1 *ΔpotR*::kan for 30 minutes. Unfortunately no cells were seen on the surface of the IPEC-J2 monolayers for either mutant or parent strains.

### 3.3.8 Characterising colony morphology and culture agglutination

*S*. Derby D1 parent and D1 *ΔpotR*::kan mutant were cultured from a frozen stock in 2ml of pre-warmed LB at 37$^{o}$C with agitation. From these cultures 100µl was plated onto dried LB plates and incubated for 16 hours at 37$^{o}$C. The remainder of the culture was left at room temperature for 2 hours before the cultures were gently agitated by hand and photographed.

### 3.3.9 Bioinformatics

The conserved domains and homologous sequences of the amino acid sequence of *potR* were identified using BLASTp against a non-redundant protein database (using default settings, Accessed 5/5/13).

## 3.4 Results

### 3.4.1 *S.* Derby associates and invades IPEC-J2 monolayers faster than *S.* Mbandaka

*S.* Derby is predominantly associated with pigs, while *S.* Mbandaka is rarely isolated from pigs in the UK (refer to chapter 1). To test the hypothesis that *S.* Derby and *S.* Mbandaka possess adaptations to distinct host species ranges we used a well characterised porcine jejunal cell line IPEC-J2 as a model for the porcine intestine, a site associated with the invasion of *Salmonella enterica* (Green & Brown 2006; Schierack et al. 2006; Collins et al. 2010). The rate of invasion and association from the very start of contact was studied through varying lengths of incubation of *S.* Derby and *S.* Mbandaka isolates with IPEC-J2 monolayers.

Between the points of initial exposure at time 0 to 5 minutes post inoculation, *S.* Mbandaka associated in significantly greater numbers ($p<0.05$) with the monolayer than *S.* Derby (Figure 3.1). From 15 minutes onwards there were significantly more *S.* Derby associated to the monolayer than there were *S.* Mbandaka ($p<0.05$), with the exception of *S.* Derby D1 and *S.* Mbandaka M2 at 15 minutes ($p>0.05$).

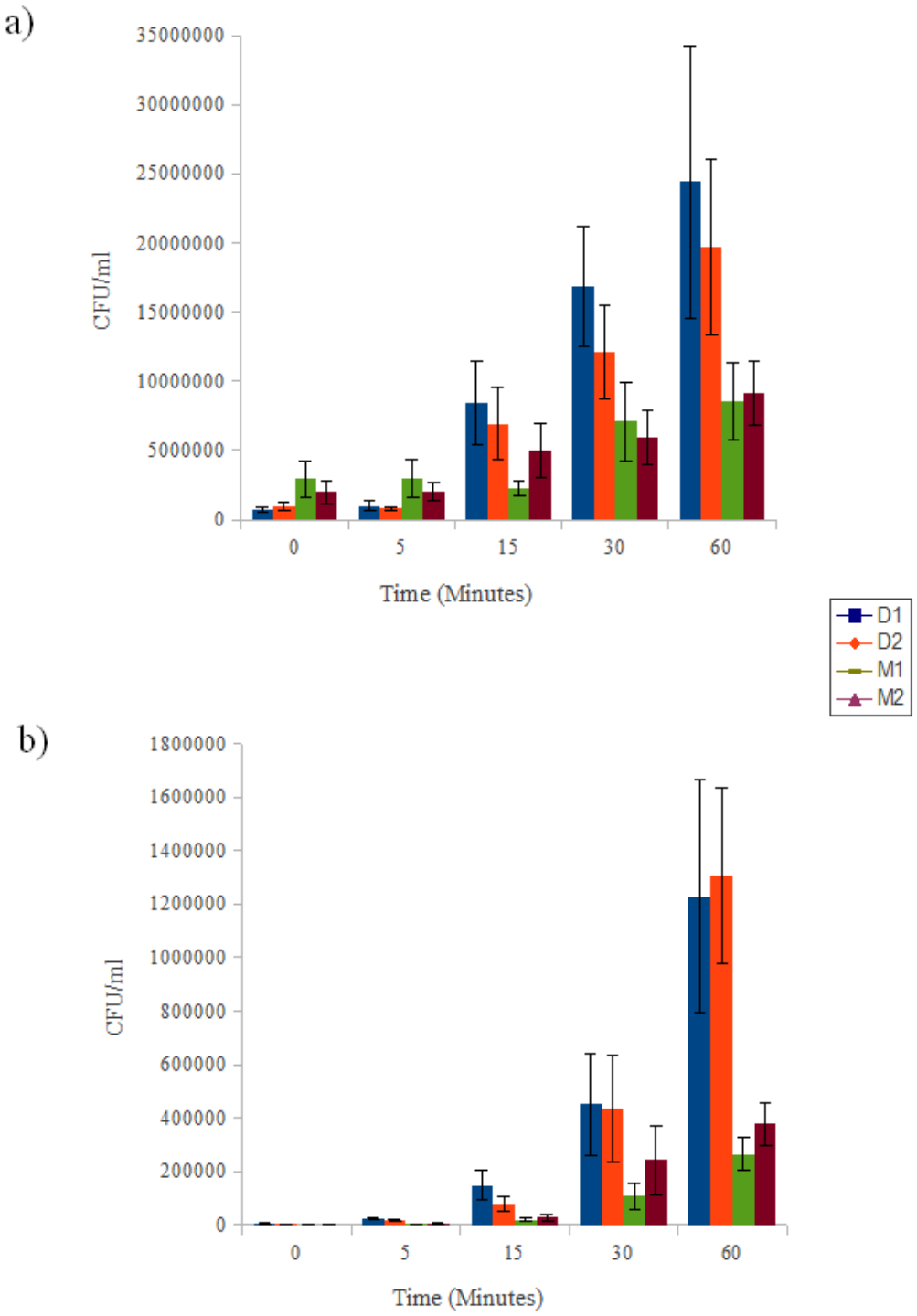



**Figure 3.1:** a) Association and b) invasion assays performed for 0, 5, 15, 30 and 60 minute incubation periods (+/-1SEM). There are consistently more *S.* Derby cells inside the IPEC-J2 monolayer than there are *S.* Mbandaka cells (a). Up to 10 minutes after inoculation *S.* Mbandaka strains associate with the IPEC-J2 monolayer in higher numbers then *S.* Derby strains after which *S.* Derby increases in association rate (b). By 60 minutes there are 2.5 times as many *S.* Derby cells associated to, and 4 times as many cells inside the monolayer then there are *S.* Mbandaka cells.

### 3.4.2 *S*. Derby associates in greater numbers to pig jejunum than colon

The jejunum and colon are commonly associated with *Salmonella* invasion (Wood et al. 1989; Santos et al. 2009). To assess if there was preferential attachment by *S*. Derby to either of these tissues, porcine IVOC association assays were performed. Both *S*. Derby strains associated in significantly greater numbers with jejunum when compared to colon ($p< 0.05$). *S*. Derby D2 associated with both the jejunum and colon in significantly greater numbers than *S*. Derby D1 ($p<0.05$). There were 2.5 times as many *S*. Derby D1 cells associated with the jejunum than the colon. There were 1.5 times as many S. Derby D2 cells associated to jejunum than colon tissues. In previous work we showed that there is very little genetic diversity between the genomes of these two isolates. All genes that differed between these two strains had metabolic functions or were phage structural proteins. We found no genetic differences between these strains that would suggest that D2 would associate to both tissues in higher numbers than D1 (Hayward et al. 2013) (refer to chapter 2).

### 3.4.3 *Salmonella* pathogenicity island 23 found on the chromosome of *S*. Derby D1 is up-regulated in both the porcine colon and jejunum IVOC preparations

Due to the larger difference between the number of *S*. Derby D1 cells associated with the jejunum and colon tissues compared to that of *S*. Derby D2, this strain was selected to be studied further for potential differences in the expression of putative pathogenicity island, SPI-23, when exposed to jejunum and colon tissues.

The gene *potR* and the putative type three effector protein genes *genE*, *sadZ*, *tinY* and *docB* were up-regulated to a significantly greater degree ($p < 0.05$) in jejunum when compared to colon, with fold changes from the no tissue control of between 21.6 and 74.4 (Figure 3.2). The putative type III effector protein, *sanA*, was the only gene expressed to a significantly greater degree in colon than in jejunum ($p<0.05$). The fold differences in expression levels from no tissue controls of the putative pilin protein gene *talN* and the putative type III effector protein genes *chlR*, *shaU* and *dumE* were not significantly different between the jejunum and colon ($p > 0.05$). The largest significant fold change was observed for the gene *docB*, a putative type III effector protein, which was up-regulated 74.4 times more in the jejunum than in the colon. *docB* was shown previously to be highly conserved in SPI-23 of *S.* Agona, *S.* Dublin and *S.* Gallinarum, and was also found in *S.* Mbandaka which lacks SPI-23 (Hayward et al. 2013) (refer to section 2.3.10.11).

There was a significantly greater number ($p < 0.05$) of S. Derby inside IPEC-J2 cells than S. Mbandaka at all time points, with the exception of S. Derby D2 and S. Mbandaka M2 at 30 minutes ($p>0.05$). After 60 minutes there were approximately 2.5 times as many S. Derby cells associated with the monolayer and approximately 4 times as many cells internalised than S. Mbandaka cells. The average association rate of S. Derby isolates to the monolayers was 3.34 times that of S. Mbandaka ($R^2$ for linear regressions D1 = 0.94, D2 = 0.93, M1 = 0.93 and M2 = 0.72). The average invasion rate of S. Derby isolates in to the monolayers was 3.65 times that of S. Mbandaka isolates ($R^2$ for linear regressions D1 = 0.74, D2 = 0.79, M1 = 0.84 and M2 = 0.78).

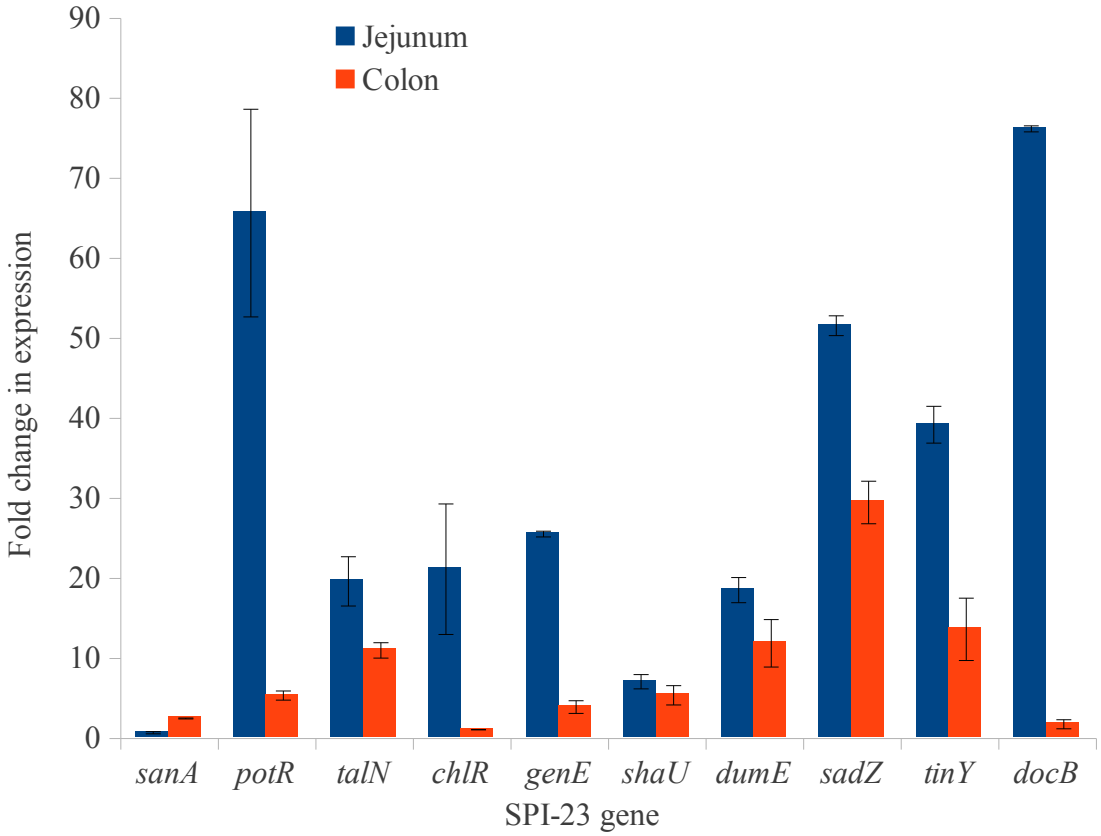



**Figure 3.2:** Fold differences in qRT-PCR expression levels of ten genes found on SPI-23, in the order they appear on the island, when exposed to porcine jejunum and colon *ex plants* relative to a no tissue control. In both tissues all the genes are up-regulated relative to a no tissue control. The genes *potR, genE*, *sadZ*, *tinY* and *docB* were up-regulated to a significantly greater degree when exposed to the porcine jejunum compared to the colon tissue relative to a no tissue control.

The *pilV*-like gene *potR,* previously shown to be unique in amino acid sequence to the SPI-23 of *S.* Derby D1 and D2*,* has the second largest difference between tissues, with a 13 fold greater expression level in the jejunum compared to colon (Hayward et al. 2013) (refer to section 2.3.10.10). The smallest significant difference between fold change from the no tissue control between the jejunum and colon treated cells was in the gene *sanA,* with a 3.5 fold greater level of transcription in colon samples than in jejunum.

### 3.4.4 Sequence features of *potR* and the phenotype of *S.* Derby D1 *ΔpotR::kan*

Due to the novelty of the gene *potR* in the SPI-23 of *S.* Derby, the high degree of up-regulation of the gene during exposure to jejunal tissue and the preferential association of the isolate to jejunal tissue, we postulated that this gene may be involved in tissue tropism.

To identify the role *potR* may play in tissue tropism, we initially identified the genes conserved protein domains. *potR* encodes 495 amino acid residues, sharing 97% sequence homology with the hypothetical protein B386_RS10215 found in the genome sequence of *Salmonella* Bareilly strain 06-0784 and 89% with the putative *pilV*-like partial protein ZP_12137808 from the genome sequence of *Salmonella* Hvittingfoss strain A4-620. BLASTp showed the first 405 amino acids of *potR* consist of a multi-domain region containing a shufflon domain (5-405aa) with a homology E-value of $6.22e^{-17}$ with the pfam entry pfam04917, a *pulG* pseudopilin motif domain (1-134aa) with a homology E-value of $9.65e^{-04}$ with COG entry COG2165, a *gspG* type II export sequence motif domain (5-56aa) with a homology E-value of $8.88e^{-03}$ with TIGR entry TIGR0171 and a type IV pilin methylation domain (1-26aa) with homology E-value of $3.83e^{-03}$ to pfam entry pfam13544. The region between amino acids 405 and 495 had no identifiable conserved domain.

We subsequently produced a knock-out mutant of the entire *potR* gene, replacing the gene with a kanamycin cassette. The first phenotypic difference observed between *S*. Derby D1 and *S*. Derby D1 *ΔpotR*::kan was during initial culturing in LB broth. Both strains formed planktonic cultures when grown with agitation at 220 rpm. Whereas if both strains were left to sit at room temperature for two hours, all the planktonic bacteria of the mutant strain agglutinated and formed a layer at the base of the culture vessel, while the parent strain remained in suspension (Figure 3.3a).

Re-suspension of the agglutinated mutant culture reduced recovery of colony forming units to the extent where a culture left for more than 2 hours was not always sub-culturable, we suspect that this is related to the strong agglutinating phenotype. If the culture was left at 4$^{o}$C on a plate or in liquid culture for several hours it was no longer recoverable by sub-culture. We would ideally have excised the kanamycin cassette and complemented the mutant strain with a plasmid born copy of *potR,* yet this was not possible due to the strains sensitivity to cold temperature interfering with the preparation of electrocompetent cells and the strong agglutinating phenotype interfering with prolonged incubation periods. Association and invasion assays were also performed with the mutant and parent strain, yet the inability to re-suspend cultures interfered with enumeration.

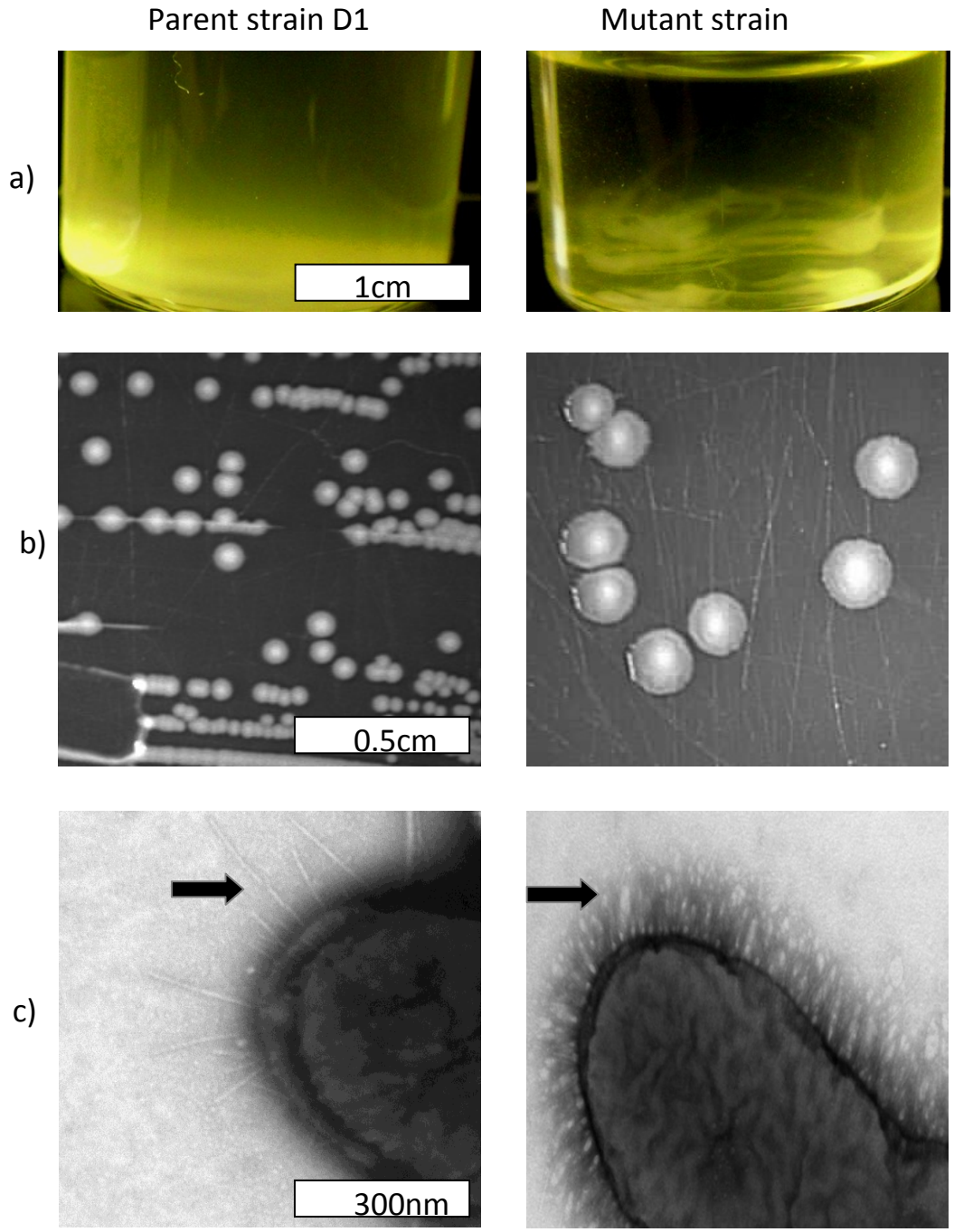



**Figure 3.3:** Comparison of morphological and structural features of the parental wild type strain *S.* Derby D1 and mutant strain *S.* Derby D1 *ΔpotR::kan.* a) Mutant cultures agglutinate forming a single cell mass after being left static at room temperature for 2 hours. b) Mutant colonies form larger and fewer colonies on LB plates then the parental wild type strain, the mutant is also surrounded by a translucent skirt which is missing from the parent strain. c) Negative electron microscopy shows that the mutant strain is heavily pilated while the parent strain has few pili and some type-1 fimbriae (both are highlighted with arrows).

Colony morphology also distinguished the mutant from the parent strain. Plating planktonic cultures of parent and mutant strains onto LB agar plates resulted in the formation of fewer and larger colonies by the mutant strain (Figure 3.3b). Both parent and mutant colonies were of smooth morphology. Yet half of the diameter skirting the outer side of the mutant colonies was translucent, this was absent from the parent colonies which were opaque to the margins.

Negative stain microscopy of overnight planktonic cultures of *S*. Derby D1 and *S*. Derby D1 Δ*potR*::kan showed clearly that the mutant strain displayed a much higher number of pili on its cell surface when compared to the parent strain (Figure 3.3c). The parent strain appeared to express classical type 1 fimbriae whereas these structures were lacking from the surface of the mutant strain.

To further evaluate the role *potR* plays in the interaction between *S*. Derby and the porcine jejunum, confocal microscopy was performed on IPEC-J2 monolayers after 4 hours of exposure to mid-log cultures of *S*. Derby D1 and *S*. Derby D1 *ΔpotR*::kan. Inspection of confocal preparations showed that the parent strain adhered in much greater number than the mutant (Figure 3.4). It was also observed that the parent strain was able to invade the monolayer, whereas no evidence of invasion was seen in the mutant preparations.

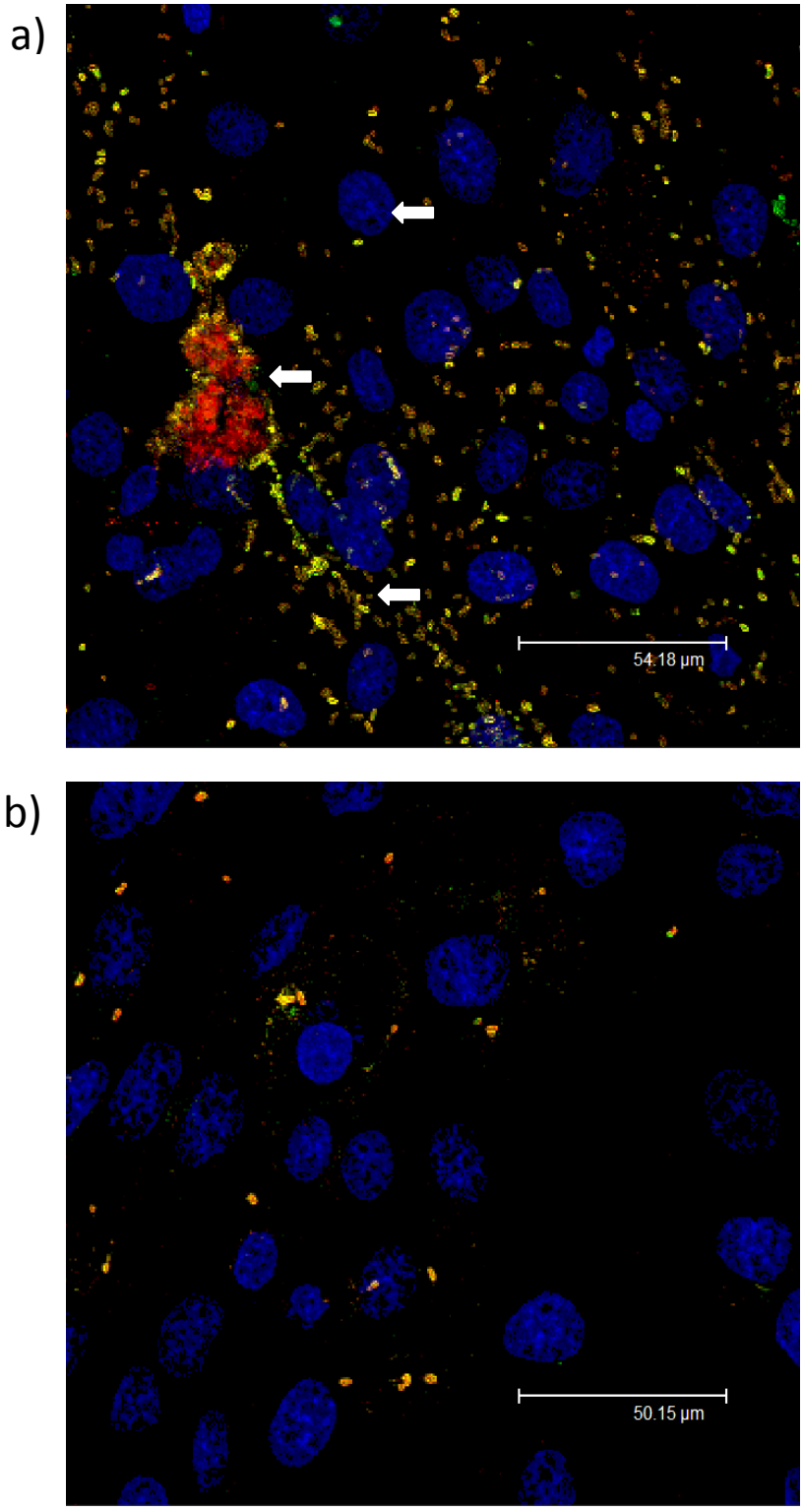



**Figure 3.4**: Differential stain confocal microscopy of a) *S*. Derby D1 and b) *S*. Derby D1 *ΔpotR::kan* that have adhered (yellow) and invaded (red) IPEC-J2 monolayers (blue). It can clearly be seen that a fewer cells of the a) mutant strain associated to the monolayer surface then the b) parent strain. The parent strain can clearly invade the monolayer, there is little evidence of this with the mutant strain.

### 3.5 Discussion

SPI-23 was discovered in the chromosome of *S.* Derby D1 and D2 through nucleotide sequence alignment with the chromosome sequence of *S.* Mbandaka M1 and M2 and other *S. enterica* serovars (Hayward et al. 2013) (refer to section 2.3.10.10). We found that SPI-23 consisted of 42 genes, ten of which were putative type III effector proteins, of which eight were unique to *S.* Derby (Hayward et al. 2013) (refer to section 2.3.10.10). Type III effector proteins are important pathogenicity factors secreted through the type III secretion system in to a host cell where they modulate cell signalling, in some instances pacifying the hosts immune system or aiding in invasion of, or translocation across, the intestinal epithelial barrier (Waterman & Holden 2003; Hapfelmeier et al. 2004). Here we determined the expression level of 10 of the 42 genes from the SPI-23 of *S.* Derby D1, eight putative type III effector proteins *sanA, chlR, genE, shaU, dumE, sadZ, tinY* and *docB*, a *pilV*-like gene *potR* and a putative pilin gene *talN* when exposed to colon and jejunum tissues relative to a no tissue control. All of the genes studied were up-regulated. We showed, using jejunum and colon derived IVOC that the porcine jejunum, was the preferred site of association for both *S.* Derby D1 and D2 and that it elicited a significantly greater increase in expression level of the genes *potR, genE*, *sadZ*, *tinY* and *docB* when compared to the colon. Interestingly, these genes are unique to *S.* Derby with the exception of *docB* which is found on the SPI-23 of *S.* Agona, *S.* Dublin and *S.* Gallinarum (Hayward et al. 2013) (refer to section 2.3.10.11).

These data collectively indicate that *potR, genE*, *sadZ* and *tinY* may contribute to host and tissue tropism. The novelty of these genes and the significantly greater level of transcripts when exposed to jejunum, led to the selection of *potR* for further study. When a BLASTp alignment was performed with a non-redundant protein library, the gene *potR*, though

similar to *pilV*, was found to be unique in amino acid sequence for the majority of the sequence outside of the shufflon domain. To characterise the phenotype of this gene, we created a *potR* knock-out mutant. The colony morphology of the *potR* mutant strain, *S.* Derby D1 *ΔpotR::kan,* is similar to that seen when the *pilV*A' C terminus region of the *pilV* shufflon of *E. coli* strain C is knocked-out (Yoshida et al. 1998). *S.* Derby D1 *ΔpotR::kan* also formed a strong agglutinating phenotype when left stationary, which caused cells to sediment on the base of the culture vessel; this phenotype interfered with the removal of the kanamycin resistance cassette, and gene complementation experiments.

In *S.* Typhi the PilV protein is located to the tip of the pilus, attaching to a larger PilS protein, causing the pili to detach from the cell surface in a particular conformation of the shufflon (Tam et al. 2006). It was suggested by Morris et al. (2003) that the aggregating phenotype seen with the down regulation of *pilV,* found on SPI-7 of *S.* Typhi, may increase the number of cells invading the small intestine in humans (Morris et al. 2003; Seth-Smith 2008). They proposed that large clumps of aggregated cells would invade increasing the intracellular load of the pathogen, though they found no evidence of mutation in *pilV* affecting the number of cells invading INT407 human cells (Morris et al. 2003). We show here that the *pilV*-like gene, *potR* knock-out produces a highly self-aggregating phenotype that led to lower numbers of cells adhering and invading a porcine jejunal monolayer.

The results presented here would suggest the converse to the hypothesis put forward by Morris et al. (2003), namely that the bacteria are likely to be planktonic in porcine jejunum which potentially allows a greater degree of adhesion and invasion, as cells can cover a larger surface area, through the formation of detachable, self-aggregating, pili (Morris et al.

2003). A similar hypothesis has been proposed for an aggregative strain of *S.* Typhimurium which is less pathogenic than a non-aggregating strain in a mouse model (White et al. 2008).

When the bacteria migrate through to the colon, *potR* is presumably down-regulated, hence we may postulate that the bacteria would exhibit an agglutinating phenotype, prior to exiting the body. If this is the case, this phenotype may aid in the manifestation of density dependent phenotypes such as biofilm formation, an adaptation for environmental persistence (White et al. 2006).

We have shown that the tight coupling of increased association to porcine jejunum derived monolayers of *S.* Derby compared to *S.* Mbandaka, the preferential association to porcine jejunum and the greater up-regulation of SPI-23 genes in the jejunum compared to the colon may indicate that SPI-23 contributes to tissue tropism. This hypothesis is supported here by the characterisation of the gene *potR.* We have shown that *potR* is integral for adhesion to, and invasion of, porcine jejunum derived monolayers by *S.* Derby.

# Chapter 4:
# Metabolite composition of host and environment may lead to niche partition between two *Salmonella enterica* sub enterica serovars

The text and figures are the same as those that are due to be submitted to npg ISMEJ; Multidisciplinary journal of microbial ecology.

Dr. Manal AbuOun helped plan and troubleshoot Biolog PM assays. Prof. Vincent A.A. Jansen and Prof. Martin J. Woodward supervised the work and helped write the manuscript.

## 4.1 Abstract

*Salmonella enterica* is a zoonotic pathogen of clinical and veterinary significance, and for this reason much of the research on factors influencing host range have focused at the host-pathogen interface. Less attention has been paid to the surrounding environment, *ex vivo*, as a potential niche for the pathogen. Here we look at two serovars, *S.* Derby and *S.* Mbandaka, both isolated from distinct sets of host species in the UK. We posit here that *S.* Mbandaka displays adaptations towards the external environment, while *S.* Derby is better adapted for pathogenicity, particularly in a porcine host. We use Biolog phenotypic microarray technology to test the proficiency of utilisation of a large number of simple metabolites at ambient and body temperature, with and without oxygen. We show that *S.* Mbandaka is more metabolically competent, and uses more proficiently several metabolites at ambient temperatures which are found in soybean based animal feeds. From the results presented here we formulate a hypothesis of a partitioned niche model for the farm environment in which *S.* Derby and *S.* Mbandaka could co-exist, one in an animal host and the other in animal feed. These results may also suggest that differences in host association between serovars is in part determined by differences in the regulation of metabolic genes, which results in variation in metabolic competency, under the same, and under different environmental conditions.

## 4.2 Introduction

The non-typhoidal Salmonellas (NTS) cause a severe gastrointestinal disease in man and are often associated with ingestion of contaminated foods. NTS can colonise a wide range of vertebrates, including various livestock species, often leading to a persistent infection that, by faecal shedding, results in environmental contamination of the farm and beyond. Veterinary, clinical and environmental isolations of *S. enterica* are routinely serotyped in

the UK and USA, for the somatic "O" antigen and the flagella "H" antigen; combinations of different epitopes of these antigens allow strains to be classified in to over 1530 different serovars (Grimont et al. 2007). Some of these serovars are host specific, whereas others can colonise a wide range of host species (Baumler et al. 1998). Some of these pathogens can also persist in soils and on animal feeds. Evidence is also growing to suggest that certain *S. enterica* serovars/biotypes can be phytopathogenic, possibly indicating adaptation to environments other than the infected animal host (Purvis et al. 2005; Schikora et al. 2008; Papadopoulou et al. 2009; Schikora et al. 2012). From the mounting evidence it is clear that many individual members of *S. enterica* have evolved to have a very broad host range.

It is unlikely that the differences in host range can be explained exclusively by random distribution or geographical barriers, and that wider, ecological aspects of *Salmonella* biology play a role in their respective epidemiology. This concept is supported by host species distribution trends amongst serovars that are similar between the UK and USA (CDC PHLIS 2009; AHVLA 2011; CDC 2011). Host barriers are important factors influencing infection, colonisation and persistence; which in part have driven the diversification and genetic radiation of *S. enterica* serovars (Rabsch et al. 2002; Hensel 2004; Thomson et al. 2008). Consequently, much of the study of host adaptation has concentrated on the role of novel or unique complements of virulence factors that contribute to the colonisation of different animal host species (Baumler et al. 1998; Norris et al. 1998; Kingsley et al. 2003; Lu et al. 2003).

Here we study two serovars that are not specific to one host, but rather differ in their respective host species range. *Salmonella enterica* serovars Derby and Mbandaka are

isolated consistently from two distinct groups of livestock in the UK and the USA; in the UK, *S*. Mbandaka is predominantly isolated from cattle and chickens while *S*. Derby is most frequently isolated from pigs and turkeys (CDC PHLIS 2009; AHVLA 2011; CDC 2011). However, plant based feed is frequently identified as the route by which *S*. Mbandaka enters farms (AHVLA 2011) This therefore raises the possibility of *S*. Mbandaka possessing adaptations, not only to cattle and chickens, but also to plants.

In previous work, we compared the genome sequences of two strains of *S*. Derby and *S*. Mbandaka, identifying several potential differences in metabolite utilisation (Hayward et al. 2013) (refer to sections 2.3.8 and 2.3.9). Here, to elaborate on the functional genomic studies, we decided to profile a wide range of metabolic phenotypes, using Biolog phenotypic microarray technology, not only to verify functional genomic predictions but also to investigate the hypothesis that model environmental conditions influence metabolite utilisation. With the aim of addressing this hypothesis, we wished to test whether metabolite utilisation is dependent on the environment in which the pathogen is found, and if so, is it possible to develop this hypothesis, in which these differences may act as adaptations to animal hosts or the surrounding environment and the results we present here suggest this may indeed be the case.

## 4.3 Methods and Materials

### 4.3.1 Determining the metabolic phenome of *S*. Derby and *S*. Mbandaka

The phenome is the summation of the phenotypes of an organism. Here, for practical purposes, we define the metabolic phenome as the summation of the metabolic phenotypes that are measured by a set of Biolog phenotypic microarrays under different conditions.

Analysis of the metabolic phenome of *S*. Derby and *S*. Mbandaka using phenotypic microarray technology was performed as previously described for aerobic and anaerobic conditions (AbuOun et al. 2009; Mappley et al. 2012). Phenotypic microarray plates, PM1-PM4 and PM6 were purchased from Biolog Inc, Hayward CA. Experiments were carried out in duplicate and on two separate occasions. In brief, isolates *S*. Derby D1 (originally isolated from pig) and *S*. Mbandaka M1 (originally isolated from cattle), previously studied by Hayward et al (2013), were grown on LB agar plates incubated at 25$^{o}$C, 37$^{o}$C aerobically and 37$^{o}$C anaerobically for 16 hours (Hayward et al. 2013) (refer to chapter 2). Cultures were standardised to a turbidity of 85% against a Biolog standard in M9 media (formula described by Sambrook et al. (2002)) without glucose. PM3 and PM4 were supplemented with 8mg/ml D-gluconic acid sodium salt as a carbon source, sources of nitrogen, sulphur and phosphate were removed from M9 media where appropriate (Sambrook et al. 2002). The inoculation solution was also supplemented with 3µg/ml thiamine. Phenotypic microarrays were read at 15 minute intervals for 72 hours during incubation in an Omnilog incubator.

### 4.3.2 Analysis of Biolog phenotypic microarrays

Results were exported using Omnilog-PM Kinetics software v1.3 (Biolog). All subsequent analysis was performed using R statistical language v2.15.2 utilising the packages OPM v0.8 and Grofit v1.1 (Buddruhs et al. 2012; Kahm 2012; R Core Team 2012). A separate spline model was fitted to the respiratory response, using the do_aggr() function from the OPM package, for each well under each condition. The models were fitted to the first 48 hours of reads for the 25$^{o}$C aerobic and 37$^{o}$C anaerobic plates, and for the first 24 hours for the 37$^{o}$C aerobic plates. By these time points, a respiratory response was observed if one was to be observed over the 72 hour incubation period, this was done for practical reasons,

as it greatly reduced the time taken to fit the models. The model parameters μ (the slope of the curve), λ (the duration of lag-phase) and A (the maximum reading over the duration of the assay), were extracted for further analysis. Significant differences between the serovars were identified through performance of a two-tailed Student's t-test with a p-value of 0.01 or lower. Under conditions showing no significant differences between strains a positive respiratory response was identified when a logistic curve could be fitted. A flat line signified no respiratory response.

## 4.4 Results

The strains selected for this study were shown elsewhere to be from the most prevalent MLST sequence type of each serovar in livestock in the UK between 2000 and 2010 (refer to Chapter 6). The functional annotations of the genome sequences of *S.* Derby D1 and *S.* Mbandaka M1 were compared previously for metabolic gene complement (Hayward et al. 2013) (refer to sections 2.3.8 and 2.3.9). To see if these genetic differences, as well as potentially unidentified regulatory differences, resulted in phenotypic differences, the respective respiratory responses of the two strains were tested against a panel of 474 metabolites under three experimental conditions, reflecting the approximate temperature and oxygen composition *ex vivo* and *in vivo,* in a pig (Kraatz 2011).

### 4.4.1 Overview of the metabolic phenomes

In total, 26 test conditions showed a significant difference for at least one of three curve parameters between *S.* Derby D1 and *S.* Mbandaka M1. Only one compound, D-galactonolactone, could be linked directly to a difference in gene complement between *S.* Derby D1 and *S.* Mbandaka M1 previously described by Hayward et al (2013) (Hayward et

al. 2013) (refer to sections 2.3.8). Of the 26 significantly different test conditions, 18 were either for faster initial utilisation, a steeper gradient at log phase or a greater degree of dye reduction by *S.* Mbandaka (Table 4.1). Only three metabolites, under all test conditions, showed a significant difference for more than one model parameter. There was no significant difference between *S.* Derby and *S.* Mbandaka in the utilisation of sulphur sources under any of the test conditions. Significant differences in nitrogen metabolism were only observed at 37$^{o}$C under aerobic conditions. No metabolite was significantly different between serovars across all three test conditions.

### 4.4.2 Differences in metabolite utilisation between *S.* Derby D1 and *S.* Mbandaka M1 in an aerobic environment at an ambient temperature

For the purposes of this study, incubation at 25$^{o}$C under aerobic conditions equates with the *ex vivo* environment. In total 14 of the 26 significantly different respiratory kinetic parameters were observed under these conditions. Of these 14 differences, 11 showed a faster or greater respiratory kinetic for *S.* Mbandaka, the majority of these were for carbon sources.

The largest difference between strains in log-phase gradient, and the largest differences between maximum readings found in this study were observed for the carbon source, D-galactonolactone (DGL). *S.* Mbandaka achieved 16 times the maximum dye reduction, and had a log phase gradient 40 times steeper than that of *S.* Derby when grown on DGL. Interestingly neither strain respired on DGL at 37$^{o}$C aerobically or anaerobically. Hence utilisation of this compound was temperature dependent.

**Table 4.1:** Summary of significant differences in μ (duration of lag-phase), λ (slope gradient) and A (maximum dye reading) parameters fit to logistic respiratory curves between D1 and M1at 25°C and 37°C aerobic and anaerobic. The symbols (D) and (M) signify respiratory parameters which under competitive conditions would favour D1 and M1 respectively. The symbol (+) signifies a logistic respiratory curve for both D1 and M1 with parameters that were not significantly different and the symbol (-) signifies no respiratory response from either strain.

| | | 25°C Aerobic | | | 37°C Aerobic | | | 37°C Anaerobic | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Metabolite | μ | λ | A | μ | Λ | A | μ | λ | A |
| *Salmonella* Mbandaka | D-Galactonolactone (c) | M | + | M | - | - | - | - | - | - |
| | D-Glucosaminic Acid (c) | + | + | M | - | - | - | - | - | - |
| | D-Saccharic Acid (c) | M | + | + | + | + | + | + | + | + |
| | D-Trehalose (c) | M | + | + | + | + | + | + | + | + |
| | Fumaric Acid (c) | + | + | M | + | + | + | - | - | - |
| | Guanosine-2'-Monophosphate (p) | + | + | M | + | + | + | + | + | + |
| | Maltotriose (c) | M | + | + | + | + | + | + | + | + |
| | N-Acetyl-D-Glucosamine (c) | M | + | + | + | + | + | + | + | + |
| | N-Acetyl-β-D-Mannosamine (c) | + | + | M | + | + | + | - | - | - |
| | Succinic Acid (c) | + | + | M | + | + | + | - | - | - |
| | L-Serine (n) | + | + | + | + | M | + | + | + | + |
| | D-Fucose (c) | - | - | - | - | - | - | M | + | + |
| | Dihydroxy-Acetone (c) | - | - | - | - | - | - | + | M | + |
| | Thymidine-3'-Monophosphate (p) | - | - | - | + | + | + | + | + | M |
| | B-D-Allose (c) | - | - | - | - | - | - | M | + | + |
| *Salmonella* Derby | Glutamine-Glutamine (dP) | + | D | + | + | + | + | + | + | + |
| | Mucic Acid (c) | + | + | D | + | + | + | + | + | + |
| | Alanine-Lysine (dp) | - | - | - | D | + | + | - | - | - |
| | Asparagine-Glutamine (dp) | + | + | + | + | + | D | + | + | + |
| | Glycine (n) | - | - | - | D | + | + | + | + | + |
| | Nitrite (n) | - | - | - | + | + | D | + | + | + |
| | Glycylalanine (dp) | + | + | + | + | + | + | + | D | + |
| Both | D-Tagatose (c) | - | - | - | - | - | - | M | + | D |
| | D-Melibiose (c) | M | D | + | + | + | + | + | + | + |
| | Mono-Methyl Succinate (c) | + | D | M | + | + | + | - | - | - |

*S.* Mbandaka achieved 14.5 times the maximum dye reduction of *S.* Derby on D-glucosaminic acid as a carbon source. The duration of lag-phase and the log-phase slope were not significantly different. As with DGL there was no respiratory response to D-glucosamic acid at either 37$^{o}$C aerobically or anaerobically. Therefore utilisation of this compound was also temperature dependent.

The slope of the respiratory response to the carbon sources N-acetyl-D-glucosamine, D-saccharic acid, D-trehalose and maltotriose were significantly greater for *S.* Mbandaka with slopes between 1.30 and 2.23 times steeper than that of *S.* Derby. *S.* Mbandaka achieved a greater maximum dye reduction value on the carbon sources succinic acid, fumaric acid and N-acetyl-beta-D-mannosamine with 1.25 to 4.55 fold that of *S.* Derby. Utilisation of carbon sources succinic acid, fumaric acid and N-acetyl-beta-D-mannosamine was oxygen dependent since both serovars respired at 37$^{o}$C, under aerobic but not anaerobic conditions.

*S.* Mbandaka achieved a maximum dye reduction 1.5 times greater than *S.* Derby on the phosphate source, guanosine-2'-3'-cyclic monophosphate. Both serovars showed a positive respiratory kinetic at 37$^{o}$C aerobically and anaerobically.

Initial utilisation of D-melibiose after *S.* Mbandaka inoculation took 1.76 times longer than *S.* Derby, but once started it achieved a log-phase gradient 1.76 times steeper. Both serovars respired on D-melibiose at 37$^{o}$C aerobically and anaerobically. *S.* Mbandaka was 1.33 times slower than *S.* Derby at initiating the use of mono-methyl-succinate, but achieved a

maximum respiratory response 1.54 times greater. It can therefore be assumed that utilisation of mono-methyl-succinate was oxygen dependent.

The only carbon source at 25$^{o}$C to elicit a greater respiratory response from *S.* Derby was mucic acid which achieved 1.64 times the maximum dye reduction that *S.* Mbandaka achieved. Both serovars respired on mucic acid at 37$^{o}$C under aerobic and anaerobic conditions.

### 4.4.3 Differences in metabolite utilisation between *S.* Derby D1 and *S.* Mbandaka M1 at 37$^{o}$C in an aerobic environment

For the purpose of this study, an aerobic environment at 37$^{o}$C was chosen as a model for the upper gastrointestinal (GI) tract of a porcine host. The Biolog system uses 96 well plate format, and is incubated without shaking, and is regarded to be moderately hypoxic (Bochner personal communication). These conditions may represent the oxygen tension observed in the upper GI tract of a pig (Kraatz 2011). *S.* Derby and *S.* Mbandaka were significantly different for six metabolites under these conditions. All differences were in nitrogen and dipeptide metabolism.

*S.* Derby achieved 3.44 times the maximum dye reduction of *S.* Mbandaka when utilising the nitrogen source nitrite. Nitrite metabolism was found to be temperature dependent since both serovars respire on the metabolite at 37$^{o}$C anaerobically and not at 25$^{o}$C aerobically. *S.* Derby has a lag-phase 2.82 times shorter on glycine as a nitrogen source than *S.* Mbandaka. Utilisation of glycine was also found to be temperature dependent. The lag-phase of *S.* Derby on L-serine was 1.93 times shorter than *S.* Mbandaka, both respired on the

metabolite at 25°C aerobically and 37°C anaerobically. *S.* Derby had a log-phase gradient 2.78 times steeper on alanine-lysine dipeptides than *S.* Mbandaka. Neither respired on alanine-lysine dipeptides at 25°C aerobically or 37°C anaerobically. *S.* Derby was able to achieve a maximum dye reduction 1.14 times greater than *S.* Mbandaka on asparagine-glutamine dipeptides. Both serovars respired on asparagine-glutamine dipeptides at 25°C and 37°C under aerobic conditions.

### 4.4.4 Differences in metabolite utilisation between *S.* Derby D1 and *S.* Mbandaka M1 at 37°C in an anaerobic environment

An anaerobic environment at 37°C reflects conditions which the pathogen may encounter during passage through the majority of the distal GI tract. Similar conditions have been found in the mucus of the porcine intestine, a barrier the pathogen navigates through to gain access to the underlying epithelial cells (Kraatz 2011). Six differences were observed between strains under these conditions.

*S.* Derby and *S.* Mbandaka differ significantly on the utilisation of four carbon sources when grown anaerobically at 37°C. *S.* Derby lags behind *S.* Mbandaka in the utilisation of dihydroxy-acetone with a 1.61 times longer lag-phase. For D-tagatose, *S.* Mbandaka began to utilise this compound with a log-phase gradient 2.26 times steeper than that of *S.* Derby. However, *S.* Derby achieved 1.05 times the maximum dye reduction of *S.* Mbandaka. Neither serovar respired on dihydroxy-acetone, or D-tagatose at 25°C or 37°C aerobically. The lag-phase of *S.* Derby, before utilisation of the dipeptide glycylalanine, was nine times

shorter than *S*. Mbandaka. Both serovars respired on glycylalanine at 25$^{o}$C and 37$^{o}$C aerobically.

## 4.5 Discussion

In this study we profiled the respiratory dynamics of two *S. enterica* serovars commonly associated with different hosts in the UK when exposed to a wide range of metabolites. We did this to investigate if resource utilisation is a potential adaptation to the host which distinguishes the serovars. We also investigated the effect on respiratory dynamics of two temperatures and the presence and absence of oxygen. This was done in an attempt to see how model *in vivo* and *ex vivo* environments influence metabolite utilisation, and if this may act to define the niches occupied by the pathogen. We have shown that differences in metabolite utilisation between the serovars are largely associated with the environmental conditions the pathogen is exposed to.

The parameters, duration of lag-phase, rate of utilisation and maximum respiratory response represent three biologically distinct phases of bacterial response to the utilisation and depletion of available metabolites. In a closed system, these parameters are influenced by gene regulation, metabolic rate, resource abundance and toxic bi-product build up (Chang et al. 2002; Rolfe et al. 2012). Respiration may occur independently of growth, though the logistic respiratory dynamics of *S. enterica* in the well of a phenotypic microarray almost certainly reflects logistic growth dynamics. Conversely a linear respiratory dynamic would most likely represent a stable population density and a constant rate of metabolism, as respiration is measured cumulatively through the reduction of a tetrazolium dye by NADH. This dynamic could be seen in wells where metabolic flux does not lead to biomass

incorporation or where the media, oxic or temperature conditions have a bacteriostatic effect on the culture. In the absence of respiration, no cellular division can occur (Aguilera & Benitez 1988). Under these assumptions the conclusions drawn from the logistic respiratory kinetics observed here may be extrapolated to incorporate the ability of the cells to grow on these metabolites.

### 4.5.1 *S.* Mbandaka M1 is better adapted to the utilisation of more metabolites under ambient conditions than *S.* Derby D1

As mentioned previously, *S.* Mbandaka frequently enters the farm environment as a contaminant of plant based animal feed (Jones et al. 1982; Davis 2003; Papadopoulou et al. 2009; Wierup & Häggblom 2010). With the data presented in this study it is now possible to generate hypotheses regarding differences in metabolic competences of *S.* Mbandaka, that are lacking in *S.* Derby, which may act as adaptations to plant based animal feeds.

Eleven of the eighteen differences which favour *S.* Mbandaka are in sugar metabolism under ambient conditions; these sugars are readily utilised by the porcine host and would not be expected to survive digestion in high abundance, with the exceptions of DGL, D-glucosaminic acid and mono-mehtyl succinate (KEGG maps for *Sus scrofa*: 00020 created 31/5/12, 00052 created 31/5/12, 00053 created 30/7/12, 00500 created 9/7/12, 00520 created 25/1/13) (Ogata et al. 1999). The metabolites that *Sus scrofa* can metabolise are D-saccharic acid, succinic acid, D-trehalose, D-mellibiose, mono-methyl succinate and fumaric acid which are all components of the soybean metabolome (SoyMetDB accessed 9/6/13) (Joshi et al. 2010). Soybeans are one of the main ingredients in cattle, pig and poultry feeds in the UK and USA (Csáky & Fekete 2004; Food Standards Agency 2013). From this we can posit that the utilisation of these metabolites would be a competitive

advantage possessed by *S*. Mbandaka over *S*. Derby. This advantage could be observed in unconsumed animal feed, at an ambient temperature, under aerobic conditions, where an abundance of these sugars can be found. This hypothesis is strengthened as the literature shows an association between *S*. Mbandaka and soybean based feed (Hald & Wingstrand 2006; AHVLA 2011). This makes growth curves and competitive studies on soybean homogenates a very interesting avenue to explore in relation to adaptation of *S*. Mbandaka to livestock feeds. As a contaminant of animal feed it may suggest that *S*. Mbandaka is exposed to a larger livestock host range then *S*. Derby, and that there are potential host or environmental barriers which prevent colonisation, and subsequent isolation from, that broader range of host species.

### 4.5.2 *S*. Derby D1 is better adapted to the utilisation of glycine and dipeptide glycylalanine when exposed to model host conditions than *S*. Mbandaka M1

In this study, lag-phase was typically much longer than log-phase (data not shown). In the case of a metabolite such as glycylalanine dipeptide, incubated at 37$^{o}$C anaerobically, *S*. Derby reached plateau phase of respiratory dynamics before *S*. Mbandaka had started to respire. This is most likely due to the time taken for a change in gene expression. Interestingly both serovars were able to use glyclalanine at 25$^{o}$C and 37$^{o}$C aerobically without any significant differences in respiratory parameters, this could suggest that *S*. Derby is better adapted, i.e. quicker, at utilising this metabolite *in vivo* when in an anoxic environment. Whereas the monopeptide form of glycine elicited a much steeper log-phase gradient in *S*. Derby than *S*. Mbandaka when incubated at 37$^{o}$C aerobically. This may suggest more proficient utilisation of glycine in the upper GI tract. Interestingly glycine is used as a sweetener and pH buffer in animal feed, and therefore may be in high abundance in the GI tract (Azelis 2013). A competition study on glycine and glycylalanine between the

two serovars at the three test conditions used in this study is needed to comment on the competitive advantage a much shorter lag-phase or steeper log-phase gradient gives *S.* Derby when in a mixed culture in different environments.

The differences in metabolite utilisation between *S.* Derby and *S.* Mbandaka at ambient and body temperatures may reflect adaptations, as metabolite utilisation has become coupled with potential environmental cues distinguishing *in vivo* and *ex vivo* environments. In this way the farm could be viewed as a potentially partitioned niche, whereby two serovars are better adapted to different phases of the cycle of pathogenesis and reinfection.

### 4.5.3 Differences between metabolic phenome and functional genome comparisons of *S.* Derby D1 and *S.* Mbandaka M1

D-galactonolactone (DGL) is synthesised by bacteria from D-galactonate through a reversible reaction that produces a lactone ring. Three reactions convert DGL into D-glyceraldehyde-3-P which feeds directly into glycolysis (KEGG map 00052, created 31/5/12) (Ogata et al. 1999). It was previously shown that the chromosome sequence of *S.* Mbandaka M1 contains an operon of five functionally unique genes for the uptake and metabolism of D-galactonate that are absent from the chromosome sequence of *S.* Derby D1 (Hayward et al. 2013) (refer to sections 2.3.8). We can infer in light of the results presented here, that this pathway is utilised in a temperature dependent manner in *S.* Mbandaka M1, though expression studies are a requisite to confirm this.

Comparative functional genomics also identified several other incomplete metabolic pathways in either *S.* Derby D1 or *S.* Mbandaka M1 that were complete in the other. None

of these differences in metabolic pathways manifest here as differences in proficiency of metabolism, with the exception of DGL. There were approximately 251 hypothetical genes that possessed less than 90% bidirectional amino acid sequence homology distinguishing the chromosome sequences of *S.* Derby D1 and *S.* Mbandaka M1; these genes have no putative function and therefore could be potential metabolic genes with distinct functions and hence could explain the unforeseen metabolic differences described here (Hayward et al. 2013) (refer to section 2.3.3). Gene expression studies may also identify which genes relating to transport, primary and secondary metabolite reactions, correlate with the differences observed here, in metabolite utilisation between different environmental conditions. This will be further investigated in future work through the construction of genome-scale metabolic models for *S.* Derby D1 and *S.* Mbandaka M1, which will be curated using the data presented here (refer to section 5.5.4).

In this study we show that utilisation of some metabolites by two serovars displaying distinct host associations are effected by differences in temperature or oxygen concentration. From these results we hypothesise that the farm habitat may be partitioned by the barriers, additional to those relating to pathogenesis, formed by differences between host and feed in metabolite composition, oxygen concentration and temperature.

# Chapter 5:
# Comparison of genome-scale metabolic reconstructions of *S. enterica* serovars isolated from distinct host ranges in the UK

The text and figures are the same as will be submitted for publication to npg Molecular Systems Biology; with the exception of the section discussing future curation, these steps will be performed prior to submission.

Dr. Jonathan Swann supervised, helped perform metabolite extractions and performed the 1H NMR. Dr Sadia E. Ahmed helped produce the R code for the extraction of curve parameters. Prof. Andrzej M. Kierzek is at present integrating the dry growth weight parameters, he has also helped troubleshoot the original curation of the metabolic models. Prof. Vincent A. A. Jansen and Prof. Martin J. Woodward supervised the work and helped write the manuscript.

## 5.1 Abstract

*Salmonella enterica* is a pathogen that can colonise and cause disease in a wide range of organisms, including vertebrates, invertebrates and in some cases plants. This large host range represents a diverse set of environments which the pathogen may encounter, requiring mechanisms to survive, colonise and proliferate. One highly variable element of these environments is the metabolite composition. Here we focus on two serovars of *S. enterica* that are associated with distinct subsets of livestock species; *S.* Derby is predominantly isolated from pigs and turkeys; *S.* Mbandaka is predominantly isolated from chickens and cattle. We produce genome-scale metabolic models from previously published functional genome annotations, in an attempt to better understand how the emergent property of growth, that is the product of thousands of interacting metabolic reactions, may contribute to host associations. To challenge these models we also profile the metabolic composition of porcine jejunum and colon, mucus and tissue. We show that *S.* Derby out-competes *S.* Mbandaka on porcine jejunum and colon homogenates. In addition to this we also find that the four separate metabolomes comprise different complements of core metabolites. This leads us to develop testable hypothesis on the possible contribution of these differences in metabolite constituents in determining host association, and in the case of *S.* Derby, tissue tropism.

## 5.2 Introduction

*Salmonella enterica sub enterica* is a zoonotic pathogen that consists of over 1530 different serological types (Grimont et al. 2007). These serogroups display varying degrees of host association and adaptation (Kingsley & Bäumler 2000; AHVLA 2011). The mechanisms of host adaptation between *S. enterica* serovars are typically viewed as occurring at the interface between the host and pathogen, and therefore, primarily focus on

the mechanisms for colonisation of the host tissue, or systemic infection (Baumler et al. 1998). Considering that the successful colonisation of the gut requires replication, and can be viewed in terms of the fitness of the pathogen within a given host environment (Kingsley & Bäumler 2000), it is equally plausible that the proficiency of utilisation of host derived metabolites by the pathogen, either from host exudates, pre-digested or digested food, could prove pivotal to successful colonisation of the host and persistence within the host population. Therefore the ability to utilise or out-compete commensals or other pathogens for host derived metabolites, allows us to view the distinct niches formed by the metabolite composition of different hosts, as a potential influences on host distribution (Rohmer et al. 2011).

Serovars *S*. Derby and *S*. Mbandaka were previously identified as two serovars that colonise different groups of host species in the UK (Hayward et al. 2013) (refer to chapter 1). *S*. Derby was predominantly isolated from pigs and turkeys and *S*. Mbandaka from cattle and chickens. The functional annotations of the genome sequences of two isolates of each serovar were compared (Hayward et al. 2013) (refer to chapter 2). Comparison of genes with metabolic functions identified potential differences between the serovars in their complement of primary metabolite transport reactions and secondary metabolic reactions. The effect of differences in primary transport reactions were studied directly using the Biolog phenotypic microarray (PM) platform (refer to chapter 4). Yet it was not possible to derive the significance of differences in secondary metabolism from respiratory dynamics. In addition to this complication, the reactions under study do not occur in isolation, and the host environment is not comparable to the defined complement of nutrients found in the well of a Biolog PM, but rather is a very complex environment

where many metabolites are bioavailable, and multiple metabolic pathways are simultaneously active.

The majority of metabolic genes that differed between *S.* Derby and *S.* Mbandaka isolated were implicated in secondary metabolic processes. To better understand the cumulative effects of these multiple differences on cellular growth and to see if they conveyed a benefit in the porcine intestinal tract, genome-wide metabolic models were constructed for *S.* Derby D1 and *S.* Mbandaka M1, and tested with metabolomes generated for porcine jejunum and colon tissue and mucus, obtained through NMR spectroscopy. We show that the metabolite composition of jejunum and colon tissue and mucus are unique and, therefore, pose four potentially distinct niches. We show that *S.* Derby is more proficient at growing on porcine colon than *S.* Mbandaka, and that in mix culture competition, on porcine jejunum and colon homogenates, *S.* Derby out competes *S.* Mbandaka.

## 5.3 Methods and Materials

### 5.3.1 Strains and culturing

*S.* Derby D1 and *S.* Mbandaka M1 have been previously compared at the genomic level and through phenotypic profiling using Biolog PM (Biolog, USA) technology at ambient and host body temperature, under aerobic and anaerobic conditions (refer to chapter 2 and section 4.3.1) (Hayward et al. 2013). Unless stated otherwise strains were cultured in LB medium from bead stocks stored at -80$^{o}$C, for 16 hours at 37$^{o}$C with agitation at 220rpm.

### 5.3.2 Relating respiration to growth

To allow the metabolic models for biomass incorporation (growth) to be curated by Biolog phenotypic microarray data, we needed to establish the relationship between growth and respiration. To do this, two Biolog (Biolog, USA) PM1 plates for each strain, were set up to measure respiration and growth at 37$^{o}$C aerobically using methods modified from those used in chapter 4. In brief, two Biolog PM1 plates per strain, one to measure growth and one to measure respiration, were inoculated with a standardised culture of *S*. Derby D1 and *S*. Mbandaka M1 made in M9 media (formulated as described Sambrook et al (2002)) (Sambrook et al. 2002); the plate designated to measure respiration were supplemented with 1% Dye mix A (Biolog, USA), a tetrazolium dye which is reduced by NADH. The plates designated to measure respiration were incubated in a Omnilog reader (Biolog, USA) at 37$^{o}$C and read at a 15 minute interval for 24 hours. Plates designated to measure growth were incubated in a FLUOstar spectrophotometer (BMG labtech, Germany) at 37$^{o}$C, absorbency was measured at $OD_{600}$, at a 15 minute interval for 24 hours. Mid-log values for growth and respiration were plotted against each other and a linear regression was fitted using Office Libre Calc (The Document Foundation 2012).

### 5.3.3 Model creation

The initial models of *S*. Derby D1 and *S*. Mbandaka M1 were automatically generated by the modelSEED (Henry et al. 2010) from the functional annotation produced by the RAST (Aziz et al. 2008) previously described by Hayward et al (2013) (Hayward et al. 2013) (refer to chapter 2). This initial model was visualised on KEGG map overlays (Ogata et al. 1999; Hayward et al. 2013). Gap filling was performed using a complete media formulation, whereby flux was allowed to move through all transport reactions.

### 5.3.4 Model curation

The biomass metabolite is a list of metabolites which contribute to the dry weight of the cell. The biomass reaction is a list of reactions which feed in to the biomass metabolite, this incorporation leads to cellular growth. Both of these elements are automatically generated by the modelSEED server during the metabolic reconstruction (Henry et al. 2010). The metabolites which make up the biomass metabolite were modified to reflect the list used for the metabolic model constructed by Raghunathan (2009) for *S*. Typhimurium (Raghunathan et al. 2009). The contribution of these metabolites to the biomass metabolite were weighted in accordance with Raghunathan (2009) for *S*. Typhimurium (Raghunathan et al. 2009). Backfilling of the reactions essential to produce these metabolites was performed using KEGG metabolic pathways and surreyFBA (Gevorgyan et al. 2011). The model was run in surreyFBA against a formulation for LB media (formula can be found in Supplementary material 5.1), a medium which both strains are known to grow on (Gevorgyan et al. 2011). The flux to each metabolite which contributed to the biomass metabolite was maximized under simulation with the metabolites found in LB medium, to guarantee that all biomass metabolites could be reached from the transport reactions, and to confirm that the network was fully connected. The network was tested for connectivity and essential genes using surreyFBA. Flux balance analysis was performed with LB to visualize the connectivity of the network and subsequent flux through particular metabolites. Internal flux loops were also removed, by performing a MILP to optimise the smallest subset of reactions that contribute to ATP production in the absence of extracellular metabolites. These reactions were removed until no reactions full filling this criteria remained.

Biolog phenotypic microarray data previously described in chapter 4, performed at 37$^{o}$C aerobically, was used to determine flux across the transport reactions. This was achieved by fitting a logistic curve to the data using R statistical language and the package OPM. The mid-log gradient was used as a surrogate for flux across transport reactions, as the transport reactions are the only section of the model at which flux can be constrained without influencing the flux across other metabolic reactions. Flux was allowed to flow across transport reactions which correspond to metabolites, which elicited a positive respiratory response under Biolog PM conditions, when incubated statically at 37$^{o}$C aerobically, and were present in the porcine gut metabolome, described below.

5.3.5 Metabolite extraction from porcine colon and jejunum

Tissue samples were collected at the same time as the samples used in section 3.3.3. In brief, on two separate occasions, three, six-week old, cross bred commercial pigs were stunned and euthanized through exsanguination. Jejunum and colon tissues were immediately removed from each pig and stored in distilled water. Colon and jejunum were removed and flash frozen using dry ice and methanol. Samples were stored at -80 $^{o}$C. Three tissue cross-sections weighing 30mg were taken from each tissue type and added to separate microcentrifuge tubes. From each tissue type, 30mg of mucus was removed from the tissue using the back of a scalpel and added to separate microcentrifuge tubes in triplicate for each tissue. To each sample 300µl of CHCL3/methanol (2:1) was added. The samples were then homogenized using a TissueLyser LT (Qiagen, Germany) homogeniser with glass pellets. To the homogenate, 300µl of distilled water was added and mixed through vortexing. Samples were spun at 13,000rpm for 10 minutes. The aqueous and organic phases were separated into two separate micocentrifuge tubes. To both phases 300µl of CHCL3/MeOH and 300µl of distilled water were added, mixed through vortexing

and spun down at 13,000rpm for 10minutes. Organic phase samples were dried in a fume hood and the aqueous phase samples were dried in a speedvac (Thermo Scientific, USA). Once dry all samples were stored at -40$^{o}$C.

### 5.3.6 1H NMR spectroscopic analysis

For each sample, a standard one-dimensional NMR spectrum was acquired using a Bruker Avance III 700 spectrometer operating at 700.13 MHz equipped with a TCI cryoprobe (Bruker Germany), with water peak suppression using a standard pulse sequence (recycle delay (RD)-90o-t1-90o-tm-90o-acquire free induction decay (FID)). The 90o pulse length was set at 7.7s, t1 set to 3s, a 2s RD and a mixing time (tm) of 10ms. For each spectrum, 8 dummy scans were followed by 128 scans with an acquisition time per scan of 2.91s, readings were collected in to 64K data points with a spectral width of 20ppm.

### 5.3.7 Fitting of metabolites to NMR spectra

Seventy two core and gut pig metabolites identified by Merrifield et al 2011 (Table 5.1) were fit to NMR spectra using R statistical language and the package BATMAN (Merrifield et al. 2011; Hao et al. 2012; R Core Team 2012). The region between 1.0 – 8.0 ppm was analysed for each spectrum, from this, the region between 4.6 and 5.2 ppm was removed as it contains the resonance peak for water. The parameters used to run BATMAN for each spectrum were as follows: the number of iterations = 5000, the lower limit for spectrum data point =0, the normalisation factor = 50000, down sampling = 3, the speed for random number generation = 25, the burn-in period = 2000 iterations, the results were saved every 50 generations, the start temperature = 1000, all other settings were as default.

**Table 5.1**: Relative abundance of 72 core metabolites in porcine jejunum, jejunum mucus, colon and colon mucus identified from NMR spectra using, R programming language, package BATMAN. Those values shaded in were not significantly different (p>0.05) from 0.

| | Colon | | Colon Mucus | | Jejunum | | Jejunum Mucus | |
|---|---|---|---|---|---|---|---|---|
| **Metabolite** | **Average** | **95% Confidence** | **Average** | **95% Confidence** | **Average** | **95% Confidence** | **Average** | **95% Confidence** |
| 3-Aminoisobutanoic acid | 0.146 | 0.047 | 0.145 | 0.092 | 0.063 | 0.028 | 0.565 | 0.373 |
| 2-Hydroxybutyric acid | 0.708 | 0.661 | 1.711 | 0.126 | 0.689 | 0.100 | 0.696 | 0.504 |
| 3-Phosphoglyceric acid | 12.437 | 1.288 | 25.893 | 18.109 | 42.353 | 26.618 | 28.957 | 17.728 |
| Acetone | 0.499 | 0.243 | 0.497 | 0.315 | 0.747 | 0.214 | 2.088 | 1.082 |
| Alpha-D-Glucose | 4.058 | 2.555 | 5.969 | 4.493 | 1.470 | 0.641 | 2.921 | 1.274 |
| Alpha-Hydroxyisobutyric acid | 0.322 | 0.229 | 0.261 | 0.121 | 2.471 | 1.406 | 0.456 | 0.231 |
| Ascorbic acid | 0.053 | 0.016 | 0.027 | 0.002 | 0.014 | 0.005 | 0.031 | 0.006 |
| Beta-Alanine | 1.979 | 0.945 | 1.896 | 0.656 | 1.295 | 0.500 | 0.504 | 0.301 |
| Beta-D-Glucose | 0.852 | 0.079 | 1.663 | 0.851 | 0.862 | 0.151 | 0.427 | 0.115 |
| Betaine | 0.035 | 0.022 | 0.038 | 0.030 | 0.019 | 0.017 | 0.021 | 0.002 |
| Cholesterol | 0.217 | 0.114 | 0.355 | 0.154 | 0.286 | 0.113 | 0.566 | 0.438 |
| Choline | 0.399 | 0.121 | 0.875 | 0.120 | 0.493 | 0.259 | 0.222 | 0.168 |
| Citric acid | 0.243 | 0.027 | 3.107 | 1.436 | 1.553 | 0.312 | 0.639 | 0.294 |
| Citrulline | 3.729 | 1.750 | 3.890 | 0.817 | 3.907 | 2.788 | 0.705 | 0.232 |
| Creatine | 1.394 | 0.558 | 0.449 | 0.211 | 0.868 | 0.316 | 0.592 | 0.407 |
| D-Alanine | 1.583 | 0.334 | 0.376 | 0.283 | 0.668 | 0.216 | 1.199 | 0.731 |
| D-Arginine | 0.872 | 0.287 | 0.740 | 0.422 | 0.223 | 0.098 | 0.378 | 0.264 |
| D-Cysteine | 11.917 | 3.420 | 1.999 | 1.460 | 1.875 | 2.290 | 6.583 | 2.328 |
| Dimethylamine | 0.522 | 0.205 | 1.103 | 0.051 | 0.470 | 0.264 | 0.416 | 0.021 |
| Dimethylglycine | 0.482 | 0.175 | 0.774 | 0.107 | 0.810 | 0.537 | 0.399 | 0.127 |
| Ethanolamine | 3.487 | 2.306 | 0.644 | 0.468 | 0.339 | 0.240 | 1.681 | 1.122 |
| Glutathione | 2.875 | 2.757 | 1.834 | 0.769 | 2.684 | 0.811 | 3.319 | 2.673 |
| Glycerol | 0.028 | 0.011 | 0.033 | 0.007 | 0.037 | 0.016 | 0.028 | 0.014 |
| Glycerophosphocholine | 0.208 | 0.051 | 0.213 | 0.052 | 0.442 | 0.250 | 0.309 | 0.273 |
| Glycine | 1.413 | 0.028 | 3.506 | 2.219 | 6.990 | 4.346 | 2.191 | 0.960 |
| Glycolic acid | 3.007 | 1.787 | 4.991 | 2.196 | 3.023 | 1.144 | 3.079 | 1.347 |
| Guanidinoacetic acid | 0.350 | 0.193 | 0.894 | 0.767 | 0.361 | 0.100 | 0.380 | 0.280 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Hippuric acid | 2.931 | 1.744 | 2.541 | 1.295 | 0.279 | 0.138 | 1.194 | 0.211 |
| Hypotaurine | 0.291 | 0.190 | 1.930 | 0.420 | 0.404 | 0.050 | 0.615 | 0.402 |
| Indoleacetic acid | 1.792 | 0.255 | 2.822 | 1.747 | 1.154 | 0.116 | 1.994 | 0.453 |
| Isovaleric acid | 2.631 | 1.099 | 4.946 | 2.233 | 4.203 | 1.986 | 2.844 | 1.093 |
| Isovalerylglycine | 3.403 | 0.369 | 2.479 | 1.450 | 2.233 | 1.261 | 0.387 | 0.148 |
| Ketoleucine | 1.825 | 1.080 | 6.853 | 3.061 | 2.292 | 0.546 | 4.812 | 3.975 |
| L-Arabitol | 2.641 | 0.908 | 2.310 | 1.225 | 0.523 | 0.085 | 0.359 | 0.075 |
| L-Asparagine | 0.501 | 0.368 | 0.434 | 0.360 | 0.537 | 0.235 | 0.145 | 0.027 |
| L-Glutamine | 1.075 | 0.548 | 4.267 | 6.879 | 4.653 | 2.397 | 0.639 | 0.376 |
| L-Histidine | 0.265 | 0.145 | 5.351 | 3.637 | 0.488 | 0.252 | 1.777 | 1.423 |
| L-Isoleucine | 0.299 | 0.167 | 0.329 | 0.168 | 0.598 | 0.487 | 0.699 | 0.435 |
| L-Leucine | 31.748 | 12.677 | 25.093 | 5.603 | 26.001 | 7.633 | 24.259 | 3.423 |
| L-Lysine | 0.133 | 0.036 | 0.114 | 0.076 | 0.217 | 0.193 | 0.157 | 0.113 |
| L-Malic acid | 0.518 | 0.259 | 3.862 | 1.757 | 1.678 | 0.631 | 0.555 | 0.291 |
| L-Ornithine | 0.763 | 0.606 | 1.446 | 1.278 | 0.495 | 0.197 | 0.235 | 0.016 |
| L-Proline | 0.709 | 0.343 | 3.068 | 2.590 | 2.195 | 0.713 | 1.882 | 0.416 |
| L-Serine | 0.050 | 0.046 | 0.057 | 0.034 | 0.025 | 0.014 | 0.036 | 0.002 |
| L-Threonine | 1.331 | 0.458 | 1.950 | 0.405 | 1.596 | 0.564 | 0.327 | 0.035 |
| L-Tryptophan | 3.458 | 1.191 | 4.322 | 2.829 | 2.676 | 0.462 | 3.340 | 2.227 |
| L-Valine | 1.083 | 0.827 | 1.105 | 0.485 | 0.616 | 0.567 | 1.712 | 0.600 |
| Methionine sulfoxide | 1.590 | 0.663 | 3.197 | 0.673 | 1.824 | 0.456 | 2.203 | 1.216 |
| Methylguanidine | 1.161 | 0.637 | 1.671 | 0.544 | 1.289 | 0.435 | 2.252 | 0.925 |
| Methylmalonic acid | 0.187 | 0.087 | 0.220 | 0.110 | 0.569 | 0.148 | 0.194 | 0.129 |
| Myoinositol | 2.734 | 0.795 | 1.837 | 0.505 | 1.491 | 0.146 | 3.069 | 2.468 |
| N-Acetyl-L-aspartic acid | 0.158 | 0.065 | 0.454 | 0.282 | 0.349 | 0.100 | 0.327 | 0.103 |
| N-Acetyl-L-phenylalanine | 0.474 | 0.221 | 1.292 | 0.392 | 0.958 | 0.523 | 0.478 | 0.206 |
| Nicotinuric acid | 5.278 | 0.754 | 6.204 | 3.109 | 2.972 | 1.410 | 3.358 | 1.683 |
| Oxoglutaric acid | 0.366 | 0.229 | 1.368 | 0.548 | 0.274 | 0.107 | 0.888 | 0.158 |
| P-Cresol glucuronide | 2.426 | 0.700 | 2.049 | 0.786 | 3.553 | 1.556 | 5.600 | 3.337 |
| P-Hydroxybenzoate | 3.282 | 1.786 | 2.939 | 0.626 | 2.882 | 1.319 | 2.653 | 0.667 |
| p-Hydroxyphenylacetic acid | 1.824 | 0.207 | 1.869 | 0.534 | 0.741 | 0.199 | 0.661 | 0.355 |
| P-Hydroxyphenylacetylglycine | 3.251 | 0.878 | 4.146 | 1.476 | 4.434 | 1.298 | 3.779 | 3.601 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Phenyl acetate | 1.567 | 0.430 | 1.568 | 0.853 | 0.123 | 0.146 | 1.396 | 0.568 |
| Phenylacetylglycine | 1.209 | 0.730 | 1.609 | 0.384 | 1.567 | 0.313 | 1.541 | 0.556 |
| Pyruvic acid | 1.392 | 0.368 | 0.973 | 0.452 | 3.362 | 2.202 | 0.746 | 0.150 |
| Sarcosine | 0.768 | 0.437 | 2.056 | 0.619 | 1.334 | 0.459 | 1.339 | 0.885 |
| Succinic acid | 0.747 | 0.307 | 0.618 | 0.219 | 1.382 | 0.262 | 1.805 | 0.647 |
| Tartaric acid | 2.302 | 0.805 | 2.174 | 1.301 | 3.198 | 0.393 | 17.874 | 6.620 |
| Taurine | 1.415 | 0.225 | 0.800 | 0.794 | 0.270 | 0.163 | 0.316 | 0.197 |
| Trehalose | 0.527 | 0.484 | 0.832 | 0.386 | 0.332 | 0.262 | 0.265 | 0.118 |
| Trigonelline | 0.517 | 0.143 | 1.423 | 0.749 | 0.857 | 0.170 | 1.560 | 0.420 |
| Trimethylamine | 0.328 | 0.228 | 0.560 | 0.301 | 0.677 | 0.154 | 0.268 | 0.034 |
| Trimethylamine oxide | 0.269 | 0.113 | 0.334 | 0.044 | 0.282 | 0.050 | 0.252 | 0.029 |
| Uracil | 1.902 | 1.277 | 2.402 | 0.903 | 2.411 | 0.502 | 4.753 | 0.959 |
| Uridine | 1.954 | 0.386 | 0.545 | 0.356 | 1.417 | 0.600 | 2.236 | 0.838 |

The analysis was performed on a server with CentOS 6.21 (The CentOS project) operating system and 16 intel Xeon processors (Intel USA), with 128Gb of RAM. The output from BATMAN included the relative concentration of each peak to the area of the spectrum. Averages and standard deviations of relative concentration were calculated for each metabolite from three biological repeats. A one-tailed T-test was performed for each metabolite from each tissue against 0, to determine if the relative abundance was significantly different from 0 ($p<0.05$).

### 5.3.8 Virtual media formulation and model simulation

Virtual media formulations were constructed from the porcine jejunum and colon metabolomes, described above. Metabolites with a relative abundance significantly different from 0 ($p<0.05$) were included. The virtual media was fed in to the model through the opening of transport reactions, this meant that only metabolites with a corresponding transport reaction were allowed to enter the model. As the flux entering the model has not curated, the optimal flux parameter across the biomass reaction was either 0, corresponding to no growth, or 1000 corresponding to growth. For the four virtual media, the log-phase gradient of the respiratory response for each metabolite, described in section 4.3.2, was compared between strains through paired t-tests, a p-value below 0.05 was deemed significant.

### 5.3.9 Production of porcine jejunum and colon homogenates

Porcine jejunum and colon tissues were collected as described above. Tissues were cleaned in distilled water to remove faecal matter, after which 5g of tissue was weighed and placed into 10ml of distilled water. Tissues were homogenised using a D-7801 hand held

homogenizer equip with emulsifier blades (Ystral, Germany). The homogenates were diluted 1:10 in distilled water and sterilised by filtration through a 22nm, 1Ltr, Stericup filter (Merck Millipore, USA).

### 5.3.10 Measuring respiratory dynamics on porcine jejunum and colon

Respiratory dynamics were measured in triplicate on three separate occasions, as follows. Overnight cultures were standardised to an $OD_{540}$ of 1.2. A 1:5,000 dilution of each isolate was made in distilled water. Wells of a blank Biolog PM plate (Biolog, USA), were filled with 100µl of one of the two homogenates. To each well 1µl of Dye mix A (Biolog, USA) and 1µl of diluted culture were added. Plates were incubated at 37$^{o}$C and read every 15 minutes for 24 hours using an Omnilog reader (Biolog, USA). T-tests were performed between strains for each time point, for each assay, a p-value less than 0.05 was deemed significant.

### 5.3.11 Spent media assays with porcine colon and jejunum homogenates

Spent media assays were performed to see if any cellular exudates of one strain could inhibit growth of the other strain, and to see if there were metabolites left in solution that the other strain could utilise. Spent media was produced from porcine jejunum and colon homogenates. This was achieved by standardising overnight cultures to an $OD_{540}$ of 1.2 in PBS, from this 10µl was added to 1ml of each homogenate in triplicate and incubated for 24 hours at 37$^{o}$C, so that the growth dynamic in the culture had reached stationary phase. Cultures were centrifuged at 2000rpm and filter sterilised through a 20nm Minisart syringe filter (Satorius Stedim, UK). For each porcine tissue spent media, 200µl of spent media were added to a 96 well FALCON plate (Fischer Scientific, USA), assay were performed

with a 10% spent media made in fresh homogenate and a 100% spent media. Plates were subsequently inoculated with 10µl of a 1:5000 dilution made in distilled water of an overnight culture grown in tissue homogenate, this was done for both spent media of the same strain and that of the other strain. Plates were incubated at 37$^{o}$C for 24 hours and read every 15 minutes in a FLUOstar spectrophotometer. T-tests were performed for each time point, for each assay, where a p-value less than 0.05 was deemed significant.

5.3.12 Competition studies on porcine colon and jejunum homogenates

Competition studies were performed on porcine jejunum and colon homogenates. Overnight cultures were standardised to an $OD_{540}$ of 1.2, and further diluted, 1:5000 in distilled water; 30µl of standardised culture was added to 3ml of each homogenate in triplicate and incubated at 37$^{o}$C for 24 hours with agitation at 220rpm. After incubation, 1ml of culture was standardised to an $OD_{540}$ of 1.2 in distilled water and subsequently diluted 1:5000 in distilled water, from this 30µl was added to 3ml of fresh homogenate. This was repeated three times. For each time point serial dilutions between $10^{0} – 10^{-8}$ of the remaining 2ml of each 24 hour culture were made in PBS. From these dilutions 100µl was plated on to LB agar plates and incubated for 24 hours at 37$^{o}$C. The dilution that formed distinct colony forming units was replicate plated on to an LB agar plate supplement with 3ug/ml tetracycline (Sigma, UK). *S*. Derby D1 is resistant to tetracycline and *S*. Mbandaka M1 sensitive (refer to Chapter 6). Both plates were enumerated for colony forming units (CFUs) and the proportion of colonies that were *S*. Derby was determined.

## 5.4 Results

### 5.4.1 Mid-log growth positively correlates with mid-log respiration

The mid-log growth and respiratory response for each metabolite, was viewed as an individual data point linking respiration to growth, for this reason, the values for both *S.* Derby and *S.* Mbandaka were combined (Figure 5.1). There was a positive correlation between mid-log respiration and growth values, with a correlation coefficient of 0.72. We regressed respiration and growth (slope = 212, $r^2 = 0.51$, df = 147, $p < 0.05$), finding that 51% of the variance in respiration was explained by growth. This was expected as not all respiration occurs as a result of growth, but all growth is dependent on respiration.

### 5.4.2 Comparison of genome-scale metabolic reconstructions

Initial metabolic model reconstructions were automatically generated by the modelSEED from the genome annotations performed by RAST on 9/10/12, described by Hayward et al (Aziz et al. 2008; Henry et al. 2010; Hayward et al. 2013) (refer to chapter 2). The automated model incorporated 1177 and 1172 genes, representing 1579 and 1584 reactions for *S.* Derby D1 and *S.* Mbandaka M1 respectively (the models can be found in Supplementary materials 5.2 and 5.3). These reactions utilised 1297 metabolites for *S.* Derby and 1294 metabolites for *S.* Mbandaka. *S.* Derby had the ability to process the metabolites, L-cysteine, CDP-glucose, CDP-ethanolamine, dTDP-rhamnose, dTDP-4-oxo-L-rhamnose, UDP-D-galacto-1,4-furanose, Monogalactosyl-monoacylglycerol, phospatidylglycerol in the cytoplasm of the cell which *S.* Mbandaka did not.

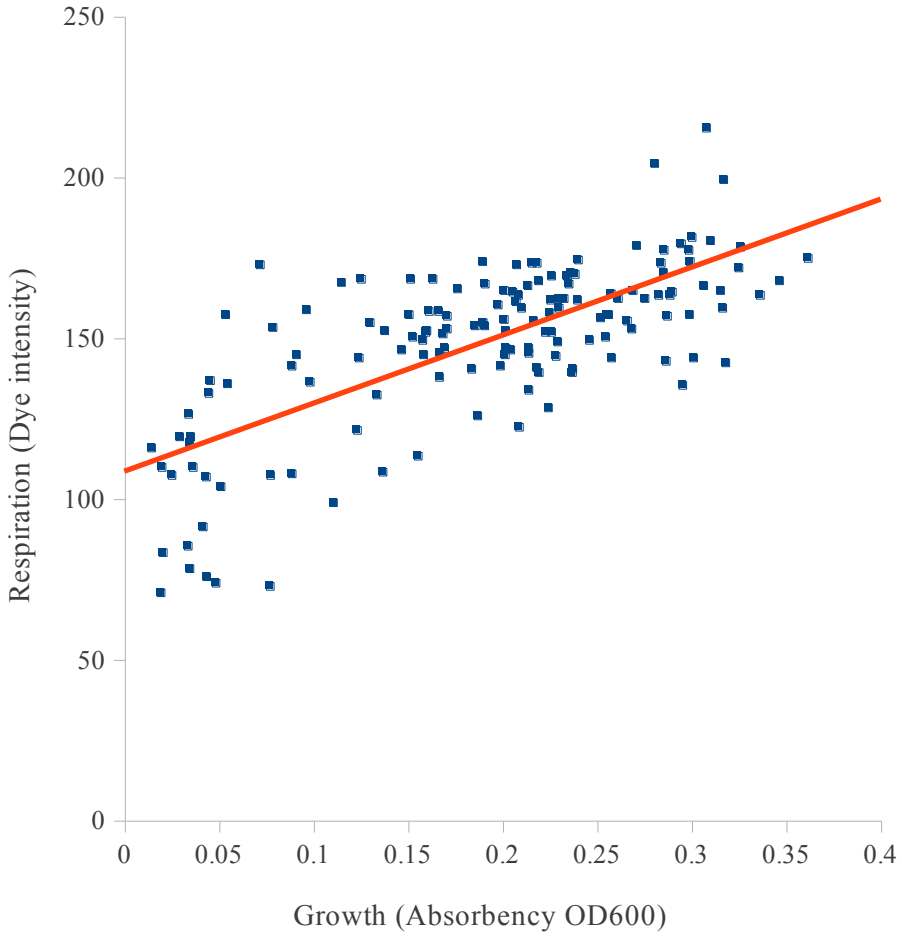



**Figure 5.1**: Respiration (tetrazolium dye intensity) positively correlates with growth (absorbency $OD_{540}$), with a correlation coefficient of 0.72. The regression line explains 52% of the variance in respiration as a product of growth.

*S.* Mbandaka contained the reactions to process the metabolites, CDP-choline, D-galactonate, 2-dehydro-3-deoxy-D-galactonate 6-phosphate, Fe(III)dicitrate in the cytoplasm of the cell that *S.* Derby did not.

The *S.* Derby model contained 149 transport reactions, all of which were found in *S.* Mbandaka. The *S.* Mbandaka model contained four additional transport reactions for the metabolites: Fe(III)di-citrate, Glycine-Phenylalanine, Phosphate and D-galactonate. The operon for D-galactonate utilisation was previously shown to be in the chromosome sequence of *S.* Mbandaka M1 and not *S.* Derby D1 (Hayward et al. 2013) (refer to section 2.3.8). Both models required 4 gap filling reactions; these were incorporated to connect all biomass metabolites to the network; these were identified using a complete media formulation. The essential transport reactions identified using SurreyFBA were the same for both serovars, and were for the uptake of the metabolites/ micronutrients: magnesium, zinc, copper, calcium, chloride, cobalt, potassium, manganese, spermidine, thiamin and folate.

Manual curation was required to edit the biomass reaction automatically generated by the modelSEED. The biomass reaction was edited to include the metabolites described for *S.* Typhimurium by Raghunathan et al. (2009) (Raghunathan et al. 2009). The additional metabolites which required back filling reactions were; acetyl-coA, 1,2-diacyl-sn-glycerol, cardiolipin, lipopolysaccharide, phosphatidylethanolamine, phospatidylglycerol, phosphatidylserine, peptidoglycan polymer (n subunits), peptidoglycan polymer (n-1 subunits) and LPS core + KDO2-lipidA. All biomass metabolites were checked for connectivity to the network. This was achieved by optimising flux to a particular biomass

metabolite through the back filling reaction when the model was challenged with an external LB formulation (Supplementary material 5.1).

### 5.4.3 Respiration on porcine jejunum and colon

There was no significant difference in respiratory dynamics ($p>0.05$) between *S*. Derby D1 and *S*. Mbandaka M1 at any time point when cultured in homogenised porcine jejunum (Figure 5.2). Up until 21.75 hours there was no significant difference ($p>0.05$) in respiratory dynamics between *S*. Derby D1 and *S*. Mbandaka M1 cultured on porcine colon homogenate (Figure 5.2), after this point the respiratory response of *S*. Derby D1 was significantly greater ($p<0.05$) than that of *S*. Mbandaka M1. By the end of the 24 hour period *S*. Derby D1 had reduced 1.6 times more dye then *S*. Mbandaka M1. Comparing between test conditions for a single serovar, showed that *S*. Derby respired to a significantly greater ($p<0.05$) degree for the majority of the incubation period on porcine jejunum compared to porcine colon. Whereas, respiratory values for *S*. Mbandaka were only significantly ($p<0.05$) greater after 20 hours of incubation on porcine jejunum compared to porcine colon.

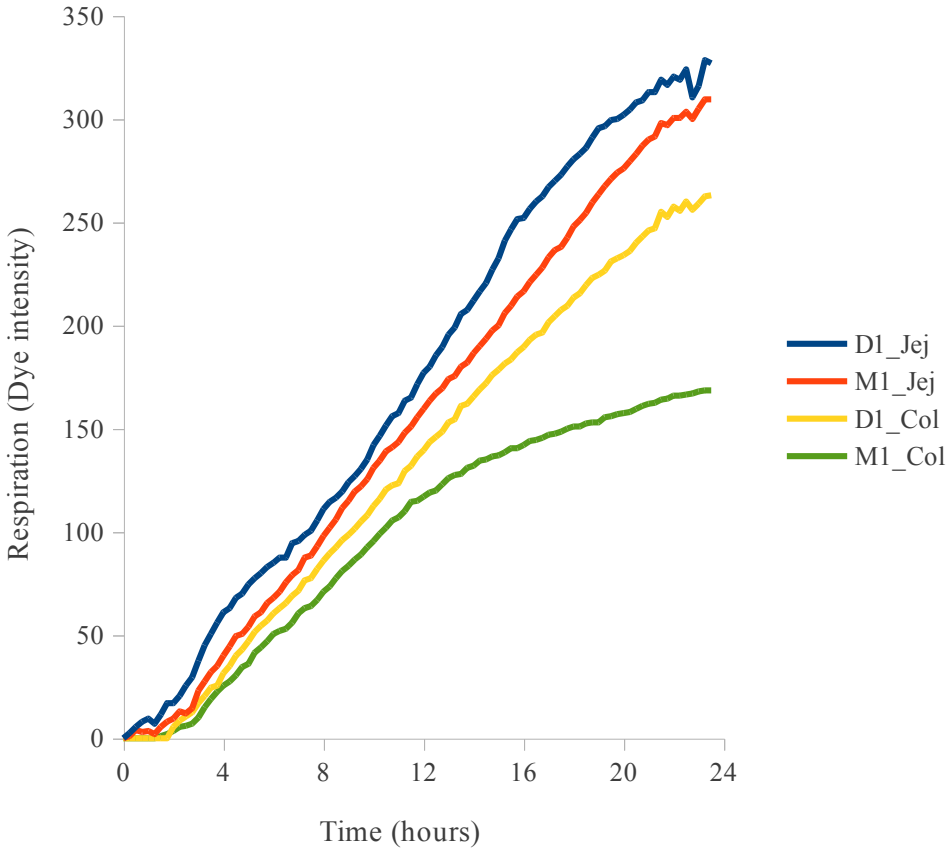



**Figure 5.2**: Respiratory dynamics on porcine colon and jejunum. There is no significant ($p>0.05$) differences between respiratory response (tetrazolium dye intensity) of *S.* Derby and *S.* Mbandaka on porcine jejunum homogenate. There is a significant difference ($p<0.05$) in respiratory dynamics on porcine colon after 21.75 hours. After 24 hours *S.* Derby had reduced 1.6 times the dye that *S.* Mbandaka had on porcine colon homogenate. Between conditions for the same serovar, for the majority of the incubation *S.* Derby respires to a significantly greater ($p<0.05$) extent on the porcine jejunum compared to the colon. For the final 4 hours of incubation *S.* Mbandaka respires to a significantly greater ($p<0.05$) extent on the porcine jejunum compared to the colon, achieving a maximum dye reduction 1.8 times greater.

5.4.4 Spent media has no effect on the growth of *S.* Derby or *S.* Mbandaka

There was no significant effect ($p>0.05$) on the growth of either strain when exposed to porcine jejunum or colon homogenate comprising 10% spent media of the other strain, when compared to their own. This suggests that there are no inhibitory compounds excreted in to the spent media by either strain. There was also no significant effect ($p>0.05$) on growth of 100% spent media of either homogenate on strains complemented with each other's spent media compared to their own. This suggests that there are no useful metabolites left in the spent media.

5.4.5 *S.* Derby out competes *S.* Mbandaka in both porcine colon and jejunum homogenates

Over each passage the proportion of *S.* Derby cells to *S.* Mbandaka cells in the cultures increased (data not shown). There was no significant difference between the proportions of *S.* Derby to *S.* Mbandaka CFUs, at any passage, in porcine jejunum compared to porcine colon ($p>0.05$). By passage four *S.* Mbandaka made up less than 6% of the recovered CFUs obtained from mixed cultures of *S.* Derby and *S.* Mbandaka.

5.4.6 Porcine jejunum and colon metabolome

We attempted to fit 72 core pig metabolites in triplicate to NMR spectra of porcine jejunum, jejunum mucus, colon and colon mucus, using the automated program BATMAN. We performed one tailed t-tests to see if the relative abundance of these metabolites was significantly different from 0. We found that porcine colon contained 62 metabolites, colon mucus 60, the jejunum 63 and jejunum mucus 59 (Table 5.1). There are few metabolites that are unique to only one component of the porcine intestine studied here, yet there are many that are absent in one or two components. The relative abundance of metabolites L-asparagine and L-ornithine was significantly different ($p<0.05$) inthe porcine jejunum tissue

and jejunum mucus and not ($p>0.05$) in the porcine colon and colon mucus. The relative abundance of the metabolite L-lysine was significantly different ($p<0.05$) from 0 in the porcine colon and colon mucus and not ($p>0.05$) in the porcine jejunum and jejunum mucus. The relative abundance of the metabolites guanidinoacetic acid and L-histidine were significantly different ($p<0.05$) from 0 in porcine tissues and not ($p>0.05$) between either of the the porcine mucus types. The relative abundance of the metabolites L-valine and trehalose were significantly different ($p<0.05$) from 0 in the porcine mucus and not ($p>0.05$) in the tissues. The remaining metabolites for which the relative abundance was significantly different from 0 did not follow a trend in intestinal section or component.

### 5.4.7 Virtual media

At this point in model curation, no transport reactions will be added to the models; this will be done at a later point when flux rates can be adjusted. Therefore transport reactions represent the compounds the model can utilise (Table 5.2). Metabolites from the porcine jejunum and colon metabolomes that correspond to a transport reaction in the models can be utilised. Therefore the virtual media, corresponds to metabolome components with a transport reaction. At present, *S.* Derby and *S.* Mbandaka do not differ on the utilisation of the compounds in the virtual media (Table 5.3). This will have a more significant effect on flux rate when the transport reactions are weighted for mid-log phase gradients. More metabolites will also be incorporated when the model is curated for the inclusion of additional metabolites identified using Biolog phenotypic microarray data (refer to chapter 4).

**Table 5.2:** The complete set of transport reactions found in the *S*. Derby D1 and *S*. Mbandaka M1 automated genome-scale metabolic model reconstruction produced by the modelSEED, and the corresponding Biolog PM well the metabolite is tested in. Shaded cells represent a Biolog PM well which elicited a logistic respiratory dynamic. PS (positive) represents a tissue which has a relative abundance significantly different ($p<0.05$) from 0, and NG (negative) a tissue which does not ($p>0.05$). There is a separate section at the bottom of the table for *S*. Mbandaka alone, contains the four additional transport reactions mentioned in section 5.4.2.

| | | Well and plate | Derby | Mbandaka | Colon | C olon mucus | Jejunu | Jejunum mucus |
|---|---|---|---|---|---|---|---|---|
| **Trimethylamine** | EX_cpd00441_e | x | X | X | | | | |
| **Trimethylamine N-oxide** | EX_cpd00811_e | x | X | X | | | | |
| **2-keto-3-deoxygluconate** | EX_cpd00176_e | x | X | X | | | | |
| **2-Oxoglutarate** | EX_cpd00024_e | x | X | X | | | | |
| **Adenosine** | EX_cpd00182_e | PM1 E12 | GRO | GRO | | | | |
| **Ala-Gln** | EX_cpd11587_e | PM3 H02 | GRO | GRO | | | | |
| **Ala-His** | EX_cpd11584_e | PM3 H05 | GRO | GRO | | | | |
| **ala-L-asp-L** | EX_cpd11593_e | PM3 H01 | GRO | GRO | | | | |
| **ala-L-glu-L** | EX_cpd11586_e | PM3 H03 | GRO | GRO | | | | |
| **ala-L-Thr-L** | EX_cpd11582_e | PM3 H07 | GRO | GRO | | | | |
| **Ala-Leu** | EX_cpd11583_e | PM3 H06 | GRO | GRO | | | | |
| **Allantoin** | EX_cpd01092_e | PM3 G05 | NO | NO | | | | |
| **Aminoethanol** | EX_cpd00162_e | x | X | X | | | | |
| **Amylotriose** | EX_cpd01262_e | x | X | X | | | | |
| **Glycine betaine** | EX_cpd00540_e | x | X | X | | | | |
| **Ca2+** | EX_cpd00063_e | x | X | X | | | | |
| **Cadaverine** | EX_cpd01155_e | x | X | X | | | | |
| **Carnitine** | EX_cpd00266_e | PM2 H05 | NO | NO | | | | |
| **Vitamin B12s** | EX_cpd00635_e | x | X | X | | | | |
| **Cadmium** | EX_cpd01012_e | x | X | X | | | | |
| **Citrate** | EX_cpd00137_e | PM1 F02 | GRO | GRO | PS | PS | PS | PS |
| **Hydrochloride** | EX_cpd00099_e | x | X | X | | | | |
| **Carbon dioxide** | EX_cpd00011_e | x | X | X | | | | |
| **Cobalt** | EX_cpd00149_e | x | X | X | | | | |
| **Copper** | EX_cpd00058_e | x | X | X | | | | |
| **Cys-Gly** | EX_cpd01017_e | PM4 F09 | GRO | GRO | | | | |
| **Cytidine** | EX_cpd00367_e | PM3 F04 | GRO | GRO | | | | |
| **Cytosine** | EX_cpd00307_e | PM3 F05 | NO | NO | | | | |
| **D-Alanine** | EX_cpd00117_e | PM1 A09 | GRO | GRO | PS | NG | PS | PS |
| **D-Fructose** | EX_cpd00082_e | PM1 C07 | GRO | GRO | | | | |
| **D-Galacturonate** | EX_cpd00280_e | PM1 H10 | NO | NO | | | | |
| **D-Glucarate** | EX_cpd00609_e | PM1 B05 | GRO | GRO | | | | |
| **D-Glucose** | EX_cpd00027_e | PM1 C09 | GRO | GRO | PS | PS | PS | PS |
| **D-glucose-6-phosphate** | EX_cpd00079_e | PM1 C01 | GRO | GRO | | | | |
| **D-Mannitol** | EX_cpd00314_e | PM1 B11 | GRO | GRO | | | | |
| **D-Mannose** | EX_cpd00138_e | PM1 A11 | GRO | GRO | | | | |
| **D-mannose-6-phosphate** | EX_cpd00235_e | PM4 D02 | GRO | GRO | | | | |
| **D-Methionine** | EX_cpd00637_e | PM4 G08 | NO | NO | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **D-Mucic acid** | EX_cpd00652_e | PM1 F08 | GRO | GRO | | | | |
| **D-Ribose** | EX_cpd00105_e | PM1 C04 | GRO | GRO | | | | |
| **D-Serine** | EX_cpd00550_e | PM3 C08 | GRO | GRO | | | | |
| **Deoxyadenosine** | EX_cpd00438_e | PM1 E11 | GRO | GRO | | | | |
| **Deoxycytidine** | EX_cpd00654_e | x | X | x | | | | |
| **Deoxyguanosine** | EX_cpd00277_e | x | X | x | | | | |
| **Deoxyinosine** | EX_cpd03279_e | x | X | x | | | | |
| **Deoxyuridine** | EX_cpd00412_e | x | X | x | | | | |
| **Dulcose** | EX_cpd01171_e | x | X | x | | | | |
| **Iron (2+)** | EX_cpd10515_e | x | X | x | | | | |
| **Iron (3+)** | EX_cpd10516_e | x | X | x | | | | |
| **Ferrichrome** | EX_cpd03724_e | x | X | x | | | | |
| **Folate** | EX_cpd00393_e | x | X | x | | | | |
| **Formate** | EX_cpd00047_e | PM1 B10 | NO | NO | | | | |
| **Fumarate** | EX_cpd00106_e | PM1 F05 | GRO | GRO | | | | |
| **gamma-Aminobutyric acid** | EX_cpd00281_e | PM3 G08 | NO | NO | | | | |
| **Galactose** | EX_cpd00108_e | PM1 A06 | GRO | GRO | | | | |
| **gamma-butyrobetaine** | EX_cpd00870_e | x | X | x | | | | |
| **N-Acetyl-D-glucosamine** | EX_cpd00222_e | PM3 E11 | GRO | GRO | | | | |
| **Glucuronate** | EX_cpd00164_e | PM1 B05 | GRO | GRO | | | | |
| **gly-asn-L** | EX_cpd11581_e | PM3 H08 | GRO | GRO | | | | |
| **gly-asp-L** | EX_cpd11589_e | PM1 F01 | NO | NO | | | | |
| **Gly-Cys** | EX_cpd15603_e | PM6 E04 | GRO | NO | | | | |
| **Gly-Gln** | EX_cpd11580_e | PM3 H09 | GRO | GRO | | | | |
| **gly-glu-L** | EX_cpd11592_e | PM3 H10 | GRO | GRO | | | | |
| **Gly-Leu** | EX_cpd15604_e | PM6 E07 | GRO | GRO | | | | |
| **Gly-Met** | EX_cpd11591_e | PM6 E09 | GRO | GRO | | | | |
| **gly-pro-L** | EX_cpd11588_e | PM6 E11 | GRO | GRO | | | | |
| **Gly-Tyr** | EX_cpd15606_e | PM6 F03 | GRO | GRO | | | | |
| **Glycerol** | EX_cpd00100_e | PM1 B03 | GRO | GRO | PS | PS | PS | PS |
| **Glycerol-3-phosphate** | EX_cpd00080_e | PM1 B07 | GRO | GRO | | | | |
| **Glycine** | EX_cpd00033_e | PM2 G05 | NO | NO | | | | |
| **Guanosine** | EX_cpd00311_e | PM3 F07 | NO | NO | | | | |
| **Hydrogen** | EX_cpd00067_e | x | X | x | | | | |
| **Water** | EX_cpd00001_e | x | X | x | | | | |
| **Thiosulfate** | EX_cpd00268_e | x | X | x | | | | |
| **Mercury(2+)** | EX_cpd00531_e | x | X | x | | | | |
| **Hypoxanthine** | EX_cpd00226_e | x | X | x | | | | |
| **Inosine** | EX_cpd00246_e | PM1 F12 | GRO | GRO | | | | |
| **Potasium** | EX_cpd00205_e | x | X | x | | | | |
| **L-Alanine** | EX_cpd00035_e | PM1 G05 | GRO | GRO | | | | |
| **L-alanylglycine** | EX_cpd11585_e | PM1 G06 | GRO | GRO | | | | |
| **L-Arabinose** | EX_cpd00224_e | PM1 A02 | GRO | GRO | | | | |
| **L-Arginine** | EX_cpd00051_e | PM3 A08 | GRO | GRO | | | | |
| **L-Aspartate** | EX_cpd00041_e | PM1 A07 | GRO | GRO | | | | |
| **L-Fucose 1-phosphate** | EX_cpd01912_e | x | X | x | | | | |
| **L-Glutamate** | EX_cpd00023_e | PM1 B12 | GRO | GRO | | | | |
| **L-Glutamine** | EX_cpd00053_e | PM1 E01 | GRO | GRO | PS | NG | PS | PS |
| **L-Histidine** | EX_cpd00119_e | PM2 G06 | NO | NO | | | | |
| **L-Idonate** | EX_cpd00573_e | x | X | x | | | | |
| **L-Inositol** | EX_cpd00121_e | x | X | x | | | | |
| **L-Isoleucine** | EX_cpd00322_e | PM2 G09 | NO | NO | | | | |
| **L-Lactate** | EX_cpd00159_e | PM1 B09 | GRO | GRO | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **L-Leucine** | EX_cpd00107_e | PM3 B05 | NO | NO | | | | |
| **L-Lysine** | EX_cpd00039_e | PM3 B06 | NO | NO | | | | |
| **L-Malate** | EX_cpd00130_e | PM1 G12 | GRO | GRO | PS | PS | PS | PS |
| **L-Methionine** | EX_cpd00060_e | PM2 G12 | GRO | GRO | | | | |
| **L-methionine R-oxide** | EX_cpd11576_e | x | X | x | | | | |
| **L-Methionine S-oxide** | EX_cpd01914_e | PM4 G11 | GRO | GRO | PS | PS | PS | PS |
| **L-Phenylalanine** | EX_cpd00066_e | PM2 H02 | NO | NO | | | | |
| **L-Proline** | EX_cpd00129_e | PM1 A08 | GRO | GRO | PS | NG | PS | PS |
| **L-Rhamnose** | EX_cpd00396_e | PM1 C06 | GRO | GRO | | | | |
| **L-Serine** | EX_cpd00054_e | PM1 G03 | GRO | GRO | NG | PS | PS | PS |
| **L-Threonine** | EX_cpd00161_e | PM1 G04 | NO | NO | | | | |
| **L-Tryptophan** | EX_cpd00065_e | PM3 B12 | NO | NO | | | | |
| **L-Tyrosine** | EX_cpd00069_e | PM3 C01 | NO | NO | | | | |
| **L-Valine** | EX_cpd00156_e | PM3 C02 | NO | NO | | | | |
| **Maltohexaose** | EX_cpd01329_e | x | X | x | | | | |
| **Maltose** | EX_cpd00179_e | PM1 C10 | GRO | GRO | | | | |
| **Melibiose** | EX_cpd03198_e | PM1 C11 | GRO | GRO | | | | |
| **met-L-ala-L** | EX_cpd11590_e | PM3 H12 | GRO | GRO | | | | |
| **Magnesium** | EX_cpd00254_e | x | X | x | | | | |
| **Manganese** | EX_cpd00030_e | x | X | x | | | | |
| **Myristic acid** | EX_cpd03847_e | x | X | x | | | | |
| **N-Acetyl-D-glucosamine** | EX_cpd00122_e | PM1 A03 | GRO | GRO | | | | |
| **Sodium** | EX_cpd00971_e | x | X | x | | | | |
| **Ammonia** | EX_cpd00013_e | x | X | x | | | | |
| **Nickel** | EX_cpd00244_e | x | X | x | | | | |
| **Nicotinamide ribonucleotide** | EX_cpd00355_e | x | X | X | | | | |
| **Nitrate** | EX_cpd00209_e | PM3 A04 | NO | NO | | | | |
| **Nitrite** | EX_cpd00075_e | PM3 A03 | GRO | NO | | | | |
| **Oxygen** | EX_cpd00007_e | x | X | x | | | | |
| **Octadecanoate** | EX_cpd01080_e | x | X | x | | | | |
| **Ornithine** | EX_cpd00064_e | PM3 C12 | NO | NO | | | | |
| **Palmitate** | EX_cpd00214_e | x | X | x | | | | |
| **Pyroracemic acid** | EX_cpd00644_e | x | X | x | | | | |
| **Lead** | EX_cpd04097_e | x | X | x | | | | |
| **Phenylpropanoate** | EX_cpd03343_e | x | X | x | | | | |
| **Pyrophosphoric acid** | EX_cpd00012_e | x | X | x | | | | |
| **Putrescine** | EX_cpd00118_e | PM3 D11 | NO | NO | | | | |
| **Sorbitol** | EX_cpd00588_e | PM1 B02 | GRO | GRO | | | | |
| **Spermidine** | EX_cpd00264_e | x | X | x | | | | |
| **Succinate** | EX_cpd00036_e | PM1 A05 | GRO | GRO | PS | PS | PS | PS |
| **Sulfate** | EX_cpd00048_e | x | X | x | | | | |
| **Thiamin** | EX_cpd00305_e | x | X | x | | | | |
| **Thymidine** | EX_cpd00184_e | PM1 C12 | GRO | GRO | | | | |
| **Thyminose** | EX_cpd01242_e | x | X | x | | | | |
| **Trehalose** | EX_cpd00794_e | PM1 A10 | GRO | GRO | NG | PS | NG | PS |
| **Uracil** | EX_cpd00092_e | PM3 F10 | NO | NO | | | | |
| **Urea** | EX_cpd00073_e | PM3 A05 | NO | NO | | | | |
| **Uridine** | EX_cpd00249_e | PM1 D12 | GRO | GRO | PS | PS | PS | PS |
| **Vitamin B12** | EX_cpd03424_e | x | X | x | | | | |
| **Vitamin B12r** | EX_cpd00423_e | x | X | x | | | | |
| **Xanthine** | EX_cpd00309_e | PM3 G01 | NO | NO | | | | |
| **Xanthosine** | EX_cpd01217_e | PM3 G02 | NO | NO | | | | |
| **Xylose** | EX_cpd00154_e | PM1 B08 | GRO | GRO | | | | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Zinc** | EX_cpd00034_e | x | X | X | | | | |
| | | | | | | | | |
| **Mbandaka only** | | | | | | | | |
| **D-Galactonate** | | PM1 C02 | NO | GRO | | | | |
| **Fe(III)dicitrate** | | x | X | x | | | | |
| **Gly-Phe** | | PM6 E10 | GRO | GRO | | | | |
| **Phosphate** | | PM4 A02 | GRO | GRO | | | | |

**Table 5.3**: Virtual media formulations for porcine colon, colon mucus, jejunum and jejunum mucus. Metabolites in the virtual media, correspond to metabolites that were present in the porcine metabolome, contained a transport reaction in one of the models and elicited a logistic curve respiratory response in a Biolog PM well. The log-phase gradient is also shown as a surrogate for flux rate across the corresponding transport reaction.

| | | | **Log phase gradient** | | **Tissue metabolite composition** | | | |
|---|---|---|---|---|---|---|---|---|
| **Metabolite** | **Reaction** | **Well and plate** | ***S.* Derby** | ***S.* Mbandaka** | **Colon** | **Colon mucus** | **Jejunum** | **Jejunum mucus** |
| Citrate | EX_cpd00137_e | PM1 F02 | 42 | 44 | + | + | + | + |
| D-Alanine | EX_cpd00117_e | PM1 A09 | 25 | 22 | + | - | + | + |
| D-Glucose | EX_cpd00027_e | PM1 C09 | 75 | 83 | + | + | + | + |
| Glycerol | EX_cpd00100_e | PM1 B03 | 56 | 49 | + | + | + | + |
| L-Glutamine | EX_cpd00053_e | PM1 E01 | 38 | 25 | + | - | + | + |
| L-Malate | EX_cpd00130_e | PM1 G12 | 31 | 34 | + | + | + | + |
| L-Methionine S-oxide | EX_cpd01914_e | PM4 G11 | 9 | 12 | + | + | + | + |
| L-Proline | EX_cpd00129_e | PM1 A08 | 37 | 39 | + | - | + | + |
| L-Serine | EX_cpd00054_e | PM1 G03 | 76 | 59 | - | + | + | + |
| Succinate | EX_cpd00036_e | PM1 A05 | 45 | 41 | + | + | + | + |
| Trehalose | EX_cpd00794_e | PM1 A10 | 64 | 54 | - | + | - | + |
| Uridine | EX_cpd00249_e | PM1 D12 | 72 | 69 | + | + | + | + |
| Trimethylamine | EX_cpd00441_e | X | x | X | - | + | + | + |
| Trimethylamine N-oxide | EX_cpd00811_e | X | x | X | + | + | + | + |

The virtual media at present contains 14 metabolites these are; citrate, D-alanine, D-glucose, glycerol, L-glutamine, L-malate, L-methionine S-oxide, L-proline, L-serine, succinate, trehalose, uridine, trimethylamine and trimethylamine N-oxide. Of these metabolites, 12 are found on the Biolog plates PM1 and PM4, trimethylamine and trimethylamine N-oxide are not found on PM1-4 or PM6. All 14 metabolites are present in the porcine jejunum mucus metabolome. The relative abundance of trehalose was not significantly different ($p<0.05$) from 0 in the porcine jejunum tissue metabolome. The colon tissue contained 11 metabolites found in the virtual media. The relative abundance of metabolites L-serine, L-trehalose and trimethylamine were not significantly different from 0 ($p>0.05$) in the porcine colon tissue metabolome. The colon mucus also contained 11 metabolites. The relative abundance of metabolites D-alanine, L-glutamine and L-proline were not significantly different from 0 ($p>0.05$) in the porcine colon mucus metabolome. The essential transport reactions, identified in section 5.4.2, were added to all media formulations.

### 5.4.8 Model predictions

"NB: The results presented in the following section may be subject to change during future curation steps. The curation steps are laid out in detail in the Discussion section of this chapter. These steps will be completed before this paper is submitted for publication."

The optimised flux parameters, for the biomass metabolite, for both models on LB and porcine colon tissue, colon mucus, jejunum tissue, jejunum mucus virtual media were 1000, the maximum value achievable. Performing a paired t-test between the log-phase gradients of the respiratory response to the metabolites in the virtual media, showed that there was no

significant difference between the two models for these parameters in the colon tissue, colon mucus or jejunum tissue virtual media. There was however a significant difference between strains on porcine jejunum mucus. Trehalose is the only metabolite that distinguishes porcine jejunum tissue from porcine jejunum mucus.

## 5.5 Discussion

Here we present preliminary metabolic models for *S*. Derby D1 and *S*. Mbandaka M1 and the data needed to curate and test these models in future work. This is done in an effort to better understand how the metabolic constitution of the pathogen's environment may influence host association, and potentially niche partitioning. This was achieved using a "systems biology" approach whereby functional gene annotations were integrated to produce genome-scale metabolic models of *S*. Derby D1 and *S*. Mbandaka M1. These models were curated using metabolic phenomes, and were challenged with four different porcine derived metabolomes.

### 5.5.1 *S*. Derby grows faster and out competes *S*. Mbandaka on porcine colon homogenate

*S*. Derby is frequently isolated from pigs in the UK, whereas *S*. Mbandaka is rarely isolated from pigs. We showed in previous work that *S*. Derby invades and associates to a porcine jejunum cell line IPEC-J2 at a much faster rate than *S*. Mbandaka (refer to section 3.4.2). We described a new *Salmonella* pathogenicity island, SPI-23, found in *S*. Derby and not *S*. Mbandaka, that shows signs of tissue tropism in its regulation, and contains a gene *potR*, which is essential for association and invasion of a porcine jejunum derived cell monolayers (refer to sections 2.3.10.10 and 3.4.4). From these results we posited that *S*. Derby was adapted to pathogenicity in a porcine host, when compared to *S*. Mbandaka. We

also showed that *S*. Derby displayed tropism towards porcine jejunum over porcine colon using *in vitro* organ culture (IVOC) (refer to section 3.4.2). Here we have also shown evidence in support of this tissue tropism as *S*. Derby respired to a significantly greater ($p<0.05$) extent on porcine jejunum compared to porcine colon homogenates. We also showed that *S*. Derby is more proficient at utilising pig derived metabolites then *S*. Mbandaka, as in porcine tissue mixed culture, where there was no signs of inhibitory compounds in spent media, *S*. Derby out competed *S*. Mbandaka. These results suggest that proficiency of metabolite utilisation may contribute to host distribution and tissue tropism. This may add a metabolic component to the previously posited host adaptations (refer to chapter 4) possessed by *S*. Derby to a porcine host which are missing from *S*. Mbandaka.

### 5.5.2 *S*. Mbandaka has the potential to utilise more compounds than *S*. Derby

Based on the transport reactions incorporated into the genome-scale metabolic models it would suggest that *S*. Mbandaka is capable of utilising four additional compounds then *S*. Derby. With the exception of D-galactonate, the other three compounds were not identified on Biolog PM plates to distinguish the two serovars. Though it was shown in chapter 4 that the variation in metabolite utilisation is largely dependent on temperature and oxygen concentration, therefore it is possible that under different conditions these differences in transport reactions would distinguish the serovars.

### 5.5.3 *S*. Derby and *S*. Mbandaka only use a small subset of host derived metabolites

Of the 72 core metabolites fit to the porcine NMR spectrum, there were only 12 corresponding transport reactions present in the automated genome-scale reconstruction produced by modelSEED. Paired t-tests showed that there was a significant difference

between strains in log-phase gradients for metabolites found in the porcine jejunum mucus virtual media. The only difference between porcine jejunum and jejunum mucus was the metabolite trehalose. This favoured *S*. Mbandaka, and conflicts with the respiratory curves shown here to display no significant differences between serovars when grown on homogenised porcine jejunum. This may be due to the absence of transport reactions which were not automatically added to the model, and may change under further curation. It may also be due to differences in internal secondary metabolic pathways; the role of these reactions will be addressed in future work.

### 5.5.4 Future model curation steps

Future curation is required to produce model predictions which estimate the dry weight incorporation rate of the biomass components. These predictions can be compared against experimental growth curves. When the experimental data corroborates the models it will be possible to investigate the flux across transport and secondary reactions which contribute most to the differences we have observed between *S*. Derby and *S*. Mbandaka when competing and growing on porcine colon homogenate. It will also allow us to identify which metabolites that differ between the tissues that influence the flux towards the biomass metabolite of the models.

In addition to standard curation steps we also propose a new novel curation step that utilises the Biolog PM data acquired in chapter 4. We propose using the log-phase gradient of *S*. Derby D1 and *S*. Mbandaka M1 respiring at 37$^{o}$C, in an aerobic environment, when cultured in Biolog PM1-4 and PM6, to constrain flux across the transport reactions. In brief, in previous work (refer to section 4.3.2), Biolog PM data was analysed using R

statistical language with the libraries OPM and Grofit, through the fitting of logistic models to all respiratory responses. The parameter for the log-phase gradient will be used to weight the flux across transport reactions allowing metabolites to enter the model. This will be achieved by comparing the log-phase gradient value to that of an internal standard. For example, in the case of carbon sources, these are typically restrained to values estimated for D-glucose. We propose weighting the transport reactions for other carbon sources by the difference in log-phase gradient between the metabolite and D-glucose, found in a Biolog PM1 plate well A09. We do this with the logic that biomass incorporation is occurring most efficiently during log-phase of growth, and as we have shown here this correlates with respiration. We are unable to weight internal branches of the model as they are used by different metabolites entering the model, therefore the only place we can weight the model is at the point where the metabolite enters the pathogen from the well of a Biolog phenotypic microarray. Using the log-phase gradient, as opposed to the area under the curve, also allows us to accommodate for catabolite repression over the 48 hour period of the assay, as a delay in utilization does not affect the log-phase gradient.

We will also produce new transport reactions and forward fill these reactions in to the models for metabolites that the strains are found to respire on from the Biolog PM data and that lack a transport reaction in the automated genome-scale reconstruction produced by modelSEED. The reactions that will be added will need corresponding genes in the genomes of the strains. These reactions will be identified from KEGG pathways; potential enzymes identified from these pathways, will be BLASTn aligned with the genome, sufficient homology will lead to the incorporation of the reaction in to the model.

In addition to curating the models presented here, we also propose curating the models for the Biolog PM data acquired at 25°C under aerobic conditions (refer to section 4.5.1). This would be done in the same way as described here for the 37°C models. These additional models will then be tested with a virtual media produced from the online soybean metabolome available at http://www.soymetdb.org/. This will be done as it was shown in section 4.5.1 and in the next chapter section 6.4.8, that *S.* Mbandaka can utilise metabolites abundant in soybean, at 25°C, that *S.* Derby does not. This is of particular interest as *S.* Mbandaka is a common contaminant of soybean based animal feed in the UK. Hopefully the production of these models, and curation for 25°C will shed light on reactions which have aided in adaptation to feed by *S.* Mbandaka and not *S.* Derby.

# Chapter 6:
# The population structure, genotypes and phenotypes of host adapted serovars *S.* Derby and *S.* Mbandaka in the UK

The text and figures are the same as will be submitted to ASM Journal of Bacteriology. Dr. Liljana Petrovska supervised the work. Prof. Vincent A. A. Jansen and Prof Martin J. Woodward supervised the work and helped write the manuscript.

## 6.1 Abstract

*Salmonella enterica* is an important zoonotic pathogen of both people and livestock. The pathogen is routinely typed into serovars based on the complement of different epitopes of two surface antigens. Some of these serovars are consistently isolated from a specific sub-set of livestock species. The serovars *S*. Derby and *S*. Mbandaka are frequently isolated from livestock in the UK, yet rarely are they isolated from the same host species. *S*. Derby is predominantly isolated from pigs and turkeys, and *S*. Mbandaka from cattle and chickens. Serotyping has been shown to be an unreliable indicator of genotype, as some serovars have been shown to be polyphyletic. In this study we produce a phylogenetic reconstruction of a representative population of isolates of *S*. Derby and *S*. Mbandaka, isolated from across the UK between 2000 and 2010, and show how different genotypes and phenotypes correlate with the population structure. We show that *S*. Derby is formed of two lineages. One that is associated with both pigs and turkeys (lineage 1) and the other is associated with turkeys (lineage 2). We show that more *S*. Derby lineage 1 cells invade a porcine jejunum derived monolayer after 30 minutes incubation than *S*. Derby lineage 2 cells. We also show that the presence of the full *Salmonella* pathogenicity island, SPI-23, is associated with *S*. Derby lineage 1 whereas the ability to form a biofilm at 25$^{o}$C is associated with *S*. Derby lineage 2 and *S*. Mbandaka. The results of this study also strengthen the previous explored hypothesis that the different host ranges of *S*. Derby and *S*. Mbandaka are due to adaptations to pathogenesis as well as utilisation of metabolites abundant in their respective host niches.

## 6.2 Introduction

*Salmonella enterica* subspecies *enterica* is an important zoonotic pathogen of warm-blooded vertebrates, with both a broad host species range and geographical distribution.

The subspecies can be divided into over 1530 serovars based on the different epitopes of two surface antigens (Grimont and Weill 2007). Epidemiological reports on isolations of *S. enterica* serovars from different livestock species in the UK have revealed interesting trends in association between particular serovars and defined subsets of livestock species (AHVLA 2011; Hayward et al. 2013) (refer to chapter 1). Associations between host species and serovars may reflect the acquisition, by the pathogen, of host adaptations during its evolutionary history (Kingsley & Bäumler 2000). These adaptations could be for pathogenesis and subsequent proliferation within the host, utilisation of metabolites abundant in the host or adaptations for persistence within the environment surrounding the host.

In previous work we characterised two strains of *S.* Derby D1 and D2 and *S.* Mbandaka M1 and M2, serovars that show evidence of differential host association. We identified several potential mechanisms pertaining to host adaptation from comparative functional genomics (Hayward et al. 2013) (refer to chapter 2). We demonstrated that *Salmonella* pathogenicity island 23, SPI-23, an island discovered in the genomes of *S.* Derby, and absent from the genomes of *S.* Mbandaka, potentially plays a role in tissue tropism to porcine jejunum over porcine colon (refer to section 3.5). We also studied the metabolite utilisation of *S.* Derby D1 and *S.* Mbandaka M1 using Biolog phenotypic microarray technology at ambient and porcine body temperatures, under aerobic and anaerobic conditions. This data lead us to propose that *S.* Mbandaka was better adapted to utilising metabolites found in soybean based feeds, and posit a partitioned niche model of the farm environment (refer to section 4.5.1). These observations have strengthened the hypothesis that *S.* Derby isolates D1 and D2 and *S.* Mbandaka isolates M1 and M2 are adapted to distinct host species and niches. We are, however unable to speculate on the contribution

of these observations to the stable host distributions observed during background monitoring for *Salmonella* serotypes in livestock production in the UK, as classification is performed at a serogroup level.

Here we identify the population structure of a representative set of isolates obtained between 2000 and 2010 across the UK. We perform several phenotypic and genotypic experiments to see how differences between the sequenced isolates identified in previous work relate to the phylogenetic structure of the population of isolates from the UK (Hayward et al. 2013). We also aimed to identify potential adaptations to the host and the external environment.

## 6.3 Methods and Materials

### 6.3.1 Bacterial Strains and Culturing

Strains were selected from background monitoring isolations that were acquired by the AHVLA between 2000 and 2010 (Table 6.1). Fourteen strains of each serovar were selected, spanning the decade and different geographical regions across the UK. The host of isolation was also specified, and was chosen to reflect the two most common hosts of isolation, for *S.* Derby this was pigs and turkeys, for *S.* Mbandaka cows and chickens (AHVLA 2011). Strains were originally stored at room temperature on dorset egg slide from the time of isolation until 2010, at which point bead stocks were made with HIB + 30% glycerol, samples were frozen to -80 $^{o}$C and remained frozen throughout the study. Unless stated otherwise, strains were grown for 16 hours aerobically on either LB agar plates or in liquid broth vigorously agitated at 220rpm.

**Table 6.1**: *S*. Derby and *S*. Mbandaka isolates were selected from background monitoring isolations, between 2000 and 2010 across the UK. Strains of *S*. Derby were isolated from pigs and turkeys, and strains of *S*. Mbandaka were isolated from chickens and cattle. Isolate number refers to the AHVLA reference number and should be used to request strains, for simplicity in this paper we refer to the strain identifiers.

| Strain identifier | Serovar | Isolate number | Year | Host |
|---|---|---|---|---|
| D1 | Derby | 4052-08 | 2008 | Pig |
| D2 | Derby | 4166-08 | 2008 | Pig |
| D3 | Derby | 1300-00 | 2000 | Pig |
| D4 | Derby | 7725-00 | 2000 | Turkey |
| D5 | Derby | 523-02 | 2002 | Pig |
| D6 | Derby | 8278-02 | 2002 | Turkey |
| D7 | Derby | 1728-04 | 2004 | Pig |
| D8 | Derby | 3380-04 | 2004 | Turkey |
| D9 | Derby | 6479-06 | 2006 | Pig |
| D10 | Derby | 8292-06 | 2006 | Turkey |
| D11 | Derby | 1617-08 | 2008 | Pig |
| D12 | Derby | 5440-08 | 2008 | Turkey |
| D13 | Derby | L0296-10 | 2010 | Pig |
| D14 | Derby | 3315-10 | 2010 | Turkey |
| M1 | Mbandaka | 3611-08 | 2008 | Cattle |
| M2 | Mbandaka | 5431-09 | 2009 | Cattle |
| M3 | Mbandaka | 1045-00 | 2000 | Cattle |
| M4 | Mbandaka | 8768-00 | 2000 | Chicken |
| M5 | Mbandaka | 3161-02 | 2002 | Chicken |
| M6 | Mbandaka | 8766-02 | 2002 | Chicken |
| M7 | Mbandaka | 727-04 | 2004 | Cattle |
| M8 | Mbandaka | 1223-04 | 2004 | Chicken |
| M9 | Mbandaka | 243-06 | 2006 | Chicken |
| M10 | Mbandaka | 1789-06 | 2006 | Cattle |
| M11 | Mbandaka | 1331-08 | 2008 | Cattle |
| M12 | Mbandaka | 4826-08 | 2008 | Chicken |
| M13 | Mbandaka | 1015-10 | 2010 | Cattle |
| M14 | Mbandaka | 2826-10 | 2010 | Chicken |

### 6.3.2 DNA extraction, MLST and PCR

DNA was extracted from 3ml overnight cultures of each strain using a EasyDNA kit (Invitrogen, UK) as per manufacturer's instructions. The MLST (Multi-Locus Sequence Typing) protocols used, and the identification of sequence types, were performed as described on http://mlst.ucc.ie/mlst/dbs/Senterica (accessed on 15.8.12). PCR to confirm the presence or absence of SPI-1 regions between STM2901-STM2903 and SC2837-SC2838 and for the SPI-23 genes *gooN, potR, talN, chlE, bigM, shaU, sadZ* and *docB* were performed using HotStarTaq master mix (Qiagen, USA). PCR cycle conditions were as follows: 15 minutes at 95$^{o}$C, followed by 35 cycles of, 30 seconds at 95$^{o}$C, 30 seconds at 52$^{o}$C and 1 minute at 72$^{o}$C, followed by 10 minutes at 72$^{o}$C. PCR products were confirmed through size comparison to a 1Kb DNA ladder (Promega, USA), after gel electrophoresis was performed, for 45 minutes, at 100V on a 1% agarose gel, followed by staining in ethidium bromide, and imaging under a UV light in a Gene Genius imaging system (Syngene, UK) (primer sequences can be found in Table 6.2).

### 6.3.3 Phylogenetic reconstruction

The sequences for the seven housekeeping loci used in the MLST scheme were concatenated and entered in to MrBayes v3.1.2 with partitions and unlinked substitution rates for each gene loci (Ronquist & Huelsenbeck 2003). Two independent Metropolis-coupled Markov chain Monte Carlo (MCMCMC) analyses were carried out on a generalised time-reversible model with invgamma rates, consisting each of 3 heated and 1 cold chain, to allow good mixing of parameters.

**Table 6.2:** Primer sequences used to test the presence or absence of SPI-23 and SPI-1 regions 1 and 2.

| Primer Name | Sequence | Amplicon Length | Sequence of origin | Start position |
|---|---|---|---|---|
| SPI1_R1_F | ctgccagcaggtgaactatt | 534 | M1 | 2935546 |
| SPI1_R1_R | aaagcttcatgcaggtcatc | | | 2936080 |
| SPI1_R2_F | tcttcggccatatgattgtt | 637 | M1 | 2939573 |
| SPI1_R2_R | acatgcctgtggagttcaat | | | 2940210 |
| gooN_F | atgtcttcccggatataggc | 517 | D1 | 2027994 |
| gooN_R | ctggcggatctctttcagta | | | 2028511 |
| potR_F | gacattacggcatcaggaac | 684 | D1 | 2032269 |
| potR_R | gcattagctccacagcattt | | | 2031585 |
| talN_F | acgtcagcatccagctttac | 224 | D1 | 2033380 |
| talN_R | cgcaatcttcacacactctg | | | 2033156 |
| chlE_F | ccacaaacaaccgaacagat | 218 | D1 | 2039475 |
| chlE_R | gaacacatatttcggcatca | | | 2039693 |
| bigM_F | gctgctcacaatcttcctgt | 375 | D1 | 2048635 |
| bigM_R | ccgctaataatgggttgatg | | | 2048260 |
| genE_F | cgcagtaaaacaggctcaat | 374 | D1 | 2050357 |
| genE_R | tcagatcctgacgtggagtt | | | 2049983 |
| tinY_F | atgcggagctttttaactca | 226 | D1 | 2062654 |
| tinY_R | ctctgccagaacggtgtagt | | | 2062428 |
| docB_F | cactccaccgaaagaagaaa | 282 | D1 | 2065588 |
| docB_R | attagacgccagcttgtcac | | | 2065870 |

Seven million iterations, sampling a tree at a 1000 tree interval, were required to achieve a desirable standard deviation of branch split frequency below 0.01. Tracer v1.4 (found at http://evolve.zoo.ox.ac.uk/software.html) was used to visualise the parameters pertaining to frequency of base substitution and used to identify an adequate burn-in period. The first 3000 trees were discarded, the consensus tree was summarised from the remaining 4000 trees, leaving an effective sample size of 569.8, a stable solution of -5176 and a standard error of the mean of 0.3674. Bootstrap values are the mean of the prior probability densities, the branch lengths represent the average number SNPs per partition. SNPs between MLST sequence concatemers were identified using DNAsp (Librado & Rozas 2009).

### 6.3.4 Association and Invasion of IPEC-J2 monolayers

Association and invasion assays were carried out as previously described in section 3.3.2 using a porcine jejunum derived cell line, IPEC-J2. IPEC-J2 (passage 70-72) cells were seeded at $1.6x10^5$cells/ml and cultured using IPECs media consisting of: Dulbecco's Modified Eagle's Medium (DMEM; Sigma, UK) supplemented with, 5% foetal bovine serum (Sigma, UK), 1% 2mM L-Glutamine (Sigma, UK), 1% Sodium Pyruvate (Sigma, UK) and 1% ITSS (Sigma, UK). After two days cells had reached 100% confluence, at which point the monolayers were washed three times with Hank's Balanced Salt Solution (Sigma, UK). Mid-log cultures of strains D4, D8, D9, D12, M4 and M8 plus previously characterised strains, *S*. Derby D1 and D2 and *S*. Mbandaka M1 and M2 and *Escherichia coli* K12 DH5*α* (refer to section 3.3.2) were produced by diluting overnight cultures 1:100 in LB broth followed by incubation at 37$^o$C for 3 hours with agitation at 220rpm. Cultures were standardised to an $OD_{540}$ of 1.2 in PBS and diluted 1:20 in IPECs medium. Washed monolayers were inoculated with 1ml of standardised culture and incubated statically, at

37°C in 5% $CO_2$ for 30 minutes. After which they were washed a further three times with Hank's Balance Salt Solution. The preparations for testing invasion were treated with a 1% gentamycin (Sigma, UK) solution made up in IPECs medium, and incubated for a further two hours at 37°C in 5% $CO_2$. After the incubation period, monolayers, of both preparations, were washed a further three times with Hank's Balance Salt Solution. Monolayers were disrupted with magnetic stirrers in 1ml of 1% TritonX (Sigma, UK) made up in PBS. Serial dilutions of each preparation including the initial inoculum were made in PBS from $10^{0}$–$10^{-8}$ and were plated on to LB agar and incubated at 37°C for 16 hours. Colony forming units (CFUs) were enumerated. ANOVA and Tukey's HSD tests were performed in R statistical language, where a p-value less than 0.05 was deemed significant (R Core Team 2012).

### 6.3.5 Respiratory dynamics of isolates grown in homogenates of porcine jejunum and colon, and soybean

Porcine jejunum and colon tissues were collected at the same time as tissues were taken for the experiments described in section 3.3.3. In brief, on two separate occasions, three, six-week old, cross bred commercial pigs were stunned and euthanized through exsanguination. Jejunum and colon tissues were immediately removed from each pig and stored in distilled water. Tissues were cleaned in distilled water to remove faecal matter, after which 5g of tissue was weighed and placed into 10ml of distilled water. Parallel to this 5g of soybeans (Natco, UK) were soaked in distilled water for 24 hours, rinsed, and placed in to 10ml of distilled water. Tissues and soybeans were homogenised using a D-7801 hand held homogenizer equip with emulsifier blades (Ystral, Germany). The homogenates were diluted 1:10 in distilled water and sterilised by filtration through a 22nm, 1 litre, Stericup filter (Merck Millipore, USA).

Respiratory dynamics were measured in triplicate on three separate occasions, as follows. Overnight cultures were standardised to an $OD_{540}$ of 1.2. A 1:5,000 dilution of each isolates was made in distilled water. Plates were set up in duplicate for each strain. The wells of a blank Biolog PM plate (Biolog, USA), were filled with 100µl of one of the three homogenates. To each well of the Biolog PM plate 1µl of dye mix A (Biolog, USA) and 1µl of diluted culture were added. Plates were incubated at 37$^{o}$C and read every 15 minutes for 24 hours for porcine homogenates and incubated at 25$^{o}$C and read every 15 minutes for 48 hours for soybean homogenates using an Omnilog reader (Biolog, USA). T-tests were performed for each time point, where a p-value less than 0.05 was deemed significant.

### 6.3.6 Antimicrobial resistance

Disc-diffusion assays were performed by the AHVLA reference laboratory Weybridge, in accordance with BSAC guidelines (www.bsac.org.uk), to identify antimicrobial sensitivity to 16 compounds; nalidixic acid, tetracycline, neomycin, ampicillin, furazolidone, ceftazidime, sulphamethoxazole/trimethoprim, chloramphenicol, amikacin, amoxicillin/clavulanic acid, gentamycin, streptomycin, sulphonamide compounds, ceftaxime, apramycin and ciprofloxacin (AHVLA 2011).

### 6.3.7 Plasmid profile

Plasmid profiling was performed as described previously by Kado and Liu (Kado & Liu 1981). In brief, 1ml of overnight liquid culture for each strain was pelleted at 13000rpm for 5 minutes. Alkaline lysis and chloroform-phenol phase separation was used to isolate genomic and plasmid DNA. Separation through gel electrophoresis was carried out at 150V for 5 hours on a 0.8% TBE agarose gel; the temperature of the buffer was maintained

at 4°C. Plasmids were sized through direct comparison to *Escherichia coli* 39R and a supercoiled ladder (Sigma Aldrich, UK) using GelAnalyzer 2010a (downloaded from http://www.gelanalyzer.com/).

### 6.3.8 Temperature and time dependent biofilm formation

Crystal violet based biofilm assays were performed in triplicate as described previously (Woodward et al. 2000). In brief, overnight cultures of all strains with the inclusion of positive biofilm forming control *S*. Enteritidis 27655R and negative non-forming control *S*. Enteritidis 27655S, were grown in LB medium without salt and were standardised to an $OD_{540}$ of 1.2. From this 1µl was added to a 96 well plate (Iwaki, Japan) containing 200µl of LB medium without salt. Plates were incubated statically at 25°C and 37°C for 24 and 48 hours. After which the inoculum was removed and the plates were washed three times with distilled water. Each well was filled with 230µl of 1% crystal violet (Pro-lab diagnostics, UK) solution made in water and incubated at room temperature for 30 minutes. Plates were washed a further 3 times to remove unbound crystal violet. Plates were dried at 65°C for 1 hour, before 200µl of acetone was added to each well to lyse cells, suspending the bound crystal violet in solution. Plates were read at 570nm in a MRX revelation (Dynex Magnellan Biosciences, USA). An $OD_{570}$ above 0.3 was considered to reflect biofilm formation.

### 6.3.9 Temperature dependent curli expression

Congo red identification of curli expression was carried out in duplicate as described by Collinson et al (1993) (Collinson et al. 1993). Overnight cultures were standardised to an $OD_{540}$ of 1.2, and subsequently diluted 1:10000, from this 5µl was spotted, in triplicate,

onto LB desalt agar plates containing 0.05% congo red (Sigma Aldrich, UK). Plates were incubated at 25°C, 37°C and 42°C for 72 hours. Colony morphology and colouration were recorded.

6.3.10 Temperature dependent motility

Modified Semi-solid Rappaport-Vassiliadis (MRSV), a semi-solid sloppy media containing 0.3% agar was stab inoculated with a single colony of each strain. Plates were incubated at 25°C, 37°C and 42°C for 16 hours. Motile strains were identified through the formation of a white halo, whereas non-motile strains were identified when a solid spherical colony was observed around the point of inoculation.

## 6.4 Results

6.4.1 Population structure of UK isolates of *S.* Derby and *S.* Mbandaka

The population structure in the UK of *S.* Derby and *S.* Mbandaka was investigated through phylogenetic reconstruction using MLST sequence concatemers of isolates collected between 2000 and 2010 *S.* Derby and *S.* Mbandaka isolates all clustered into groups based on serogroup. Between *S.* Derby and *S.* Mbandaka there were 58 SNPs that distinguished the serovars over the 3336bps of the MLST sequence concatemers. *S.* Mbandaka isolates formed a single, clonal, lineage with sequence type ST900.

*S.* Derby isolates formed two distinct lineages, lineage 1 was isolated from pigs and turkeys (D1, D2, D3, D5, D6, D10, D11 and D13) possessing the sequence type ST40 with the exception of one isolate, D7, with a ST90 sequence type (Figure 6.1). Isolates D9 and D8 do not cluster with either grouping and possessed ST types, 678 and 1554 respectively.

The second lineage is formed of isolates taken from turkeys (D4, D12 and D14), with sequence type ST71. The two lineages of *S*. Derby were distinguished by 39 SNPs over the 3336bps of the MLST concatemers. This accounts for 67% of the SNPs that distinguish *S*. Derby from *S*. Mbandaka isolates.

6.4.2 Distribution of SPI-23 amongst UK isolates of *S*. Derby and *S*. Mbandaka

SPI-23 was present in all *S.* Derby lineage 1 isolates and D9 (Figure 6.2) whereas the island was missing in part, or in whole, from isolates of lineage 2 and D8. PCR showed that isolates D8 and D12 contained the genes *genE, tinY* and *docB*, if the island is intact between these genes this would account for 18 of the 42 genes on SPI-23. The isolates D4 and D14 lack the entire island with the exception of the gene *docB*. The whole of SPI-23 was missing from *S.* Mbandaka isolates with the exception of the gene *docB,* a putative type III effector protein, which was previously shown to be highly conserved, and was found in *S.* Derby D1 and D2, and *S.* Mbandaka M1 and M2 as well as serovars *S*. Dublin, *S.* Gallinarum and *S*. Enteritidis (Hayward et al. 2013) (refer to section 2.3.10.11).

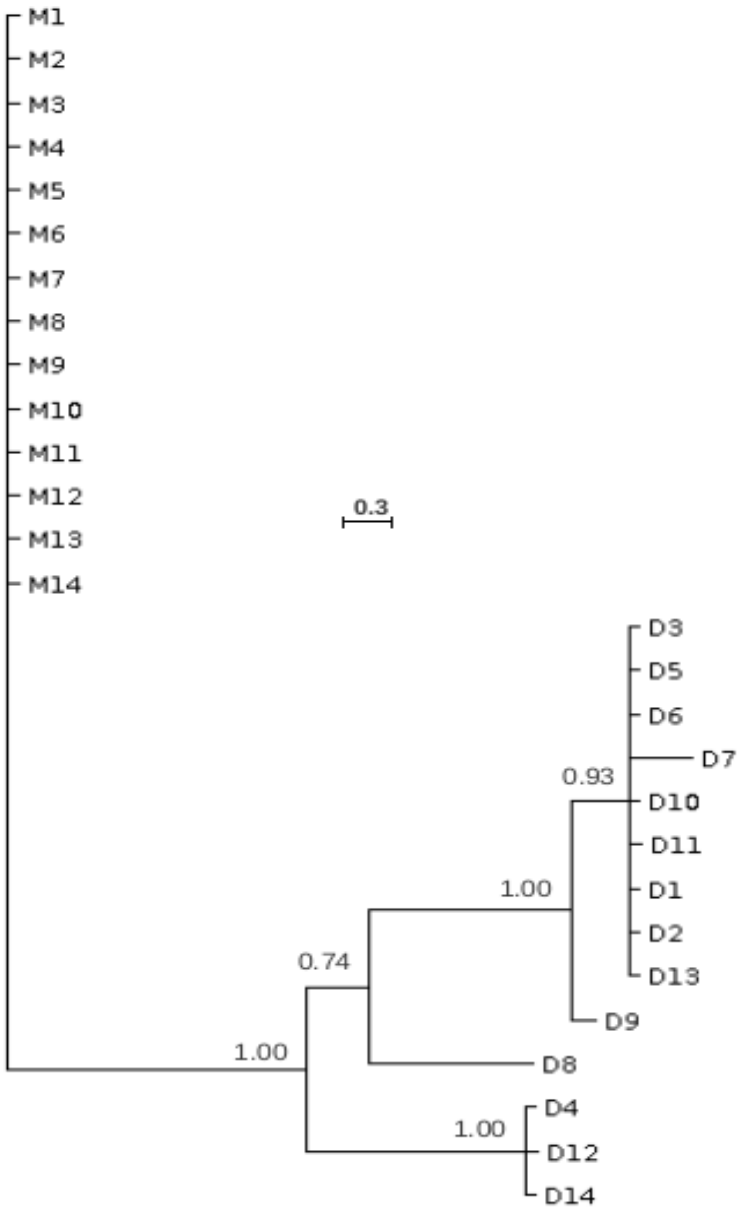
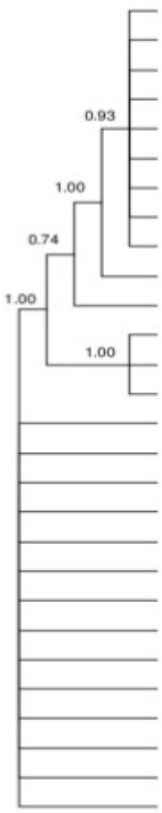



**Figure 6.1:** Phylogenetic distance tree of UK isolates of *S*. Derby and *S*. Mbandaka. The branch lengths represent the average number of SNPs per partition. Bootstrap values are the mean of the prior probability densities for each node. *S*. Mbandaka is formed of one single clonal population. Whereas *S*. Derby is formed of two distinct lineages, lineage 1 (D3, D5, D6, D7, D10. D11. D1. D2. D13) and lineage 2 (D4, D12 and D14). There are also two non-clustering isolates of *S* .Derby D8 and D9.




| Isolate | ST | SPI-1 G1 | SPI-1 G2 | SPI-23 *gooN* | SPI-23 *potR* | SPI-23 *talN* | SPI-23 *chlE* | SPI-23 *bigM* | SPI-23 *genE* | SPI-23 *tinY* | SPI-23 *docB* |
|---|---|---|---|---|---|---|---|---|---|---|---|
| D3 | 40 | - | - | x | x | x | x | x | x | x | x |
| D5 | 40 | - | - | x | x | x | x | x | x | x | x |
| D6 | 40 | - | - | x | x | x | x | x | x | x | x |
| D7 | 90 | - | - | x | x | x | x | x | x | x | x |
| D10 | 40 | - | - | x | x | x | x | x | x | x | x |
| D11 | 40 | - | - | x | x | x | x | x | x | x | x |
| D1 | 40 | - | - | x | x | x | x | x | x | x | x |
| D2 | 40 | - | - | x | x | x | x | x | x | x | x |
| D13 | 40 | - | - | x | x | x | x | x | x | x | x |
| D9 | 678 | - | - | x | x | x | x | x | x | x | x |
| D8 | 1554 | - | - | - | - | - | - | - | x | x | x |
| D4 | 71 | x | x | - | - | - | - | - | - | - | x |
| D12 | 71 | x | x | - | - | - | - | - | x | x | x |
| D14 | 71 | x | x | - | - | - | - | - | - | - | x |
| M3 | 900 | x | x | - | - | - | - | - | - | - | x |
| M4 | 900 | x | x | - | - | - | - | - | - | - | x |
| M5 | 900 | x | x | - | - | - | - | - | - | - | x |
| M6 | 900 | x | x | - | - | - | - | - | - | - | x |
| M7 | 900 | x | x | - | - | - | - | - | - | - | x |
| M8 | 900 | x | x | - | - | - | - | - | - | - | x |
| M9 | 900 | x | x | - | - | - | - | - | - | - | x |
| M10 | 900 | x | x | - | - | - | - | - | - | - | x |
| M11 | 900 | x | x | - | - | - | - | - | - | - | x |
| M12 | 900 | x | x | - | - | - | - | - | - | - | x |
| M1 | 900 | x | x | - | - | - | - | - | - | - | x |
| M2 | 900 | x | x | - | - | - | - | - | - | - | x |
| M13 | 900 | x | x | - | - | - | - | - | - | - | x |
| M14 | 900 | x | x | - | - | - | - | - | - | - | x |

**Figure 6.2:** Presence and absence of SPI-23 genes and SPI-1 region 1 and 2 in relation to a phylogenetic cladogram. All *S*. Derby lineage 1 isolates and non-clustering isolate D9 lack SPI-1 region 1 and 2, and contain a full version of SPI-23. None clustering isolate D8 and lineage 2 isolate D12 contain the SPI-23 regions between *genE* and *docB*. D8 lacks SPI-1 region 1 and 2, whereas D12 contains these regions. *S*. Derby lineage 2 isolates D4 and D14, both contain SPI-1 region 1 and 2 and lack all of SPI-23 with the exception of *docB*.

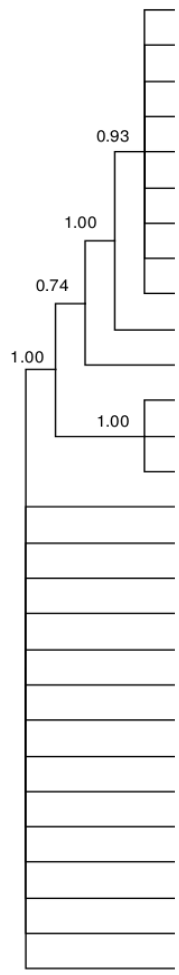
### 6.4.3 Diversity of SPI-1 amongst UK isolates of *S.* Derby and *S.* Mbandaka

SPI-1 region 1 containing the genes STM2901, STM2902 and STM2903 from *S.* Typhimurium LT2 and region 2 containing the genes SC2837, a putative type III effector protein, and SC2838 from *S.* Choleraesuis B67 were absent from lineage 1 isolates of *S.* Derby but present in lineage 2(Figure 6.2) (Hayward et al. 2013) (refer to section 2.3.10.2). All isolates of *S.* Mbandaka contained both regions.

### 6.4.4 Diversity in biofilm formation, motility and curli expression amongst UK isolates of *S.* Derby and *S.* Mbandaka

After 48 hours incubated at 25$^{o}$C all isolates of *S.* Mbandaka, with the exception of M9, formed a biofilm (Figure 6.3). Three isolates of *S.* Mbandaka formed biofilms at 37$^{o}$C, M6 and M13 after 48 hours, and M7 after 24 hours. All lineage 2 *S.* Derby isolates and D8 formed biofilms at 25$^{o}$C after 48 hours, none of the lineage 1 isolates formed biofilms at 25$^{o}$C. Lineage 1 isolate D10 was the only isolate of *S.* Derby to form a biofilm at 37$^{o}$C.

All test isolates bound to congo red indicating that the strains expressed curli when cultured at 25$^{o}$C, 37$^{o}$C and 42$^{o}$C for 72 hours on LB desalt agar (Collinson et al. 1993). All isolates of *S.* Derby and *S.* Mbandaka were motile at 25$^{o}$C, 37$^{o}$C and 42$^{o}$C with the exception D11 and M9 which were not motile at 25$^{o}$C and 42$^{o}$C, and D12 which was not motile at 42$^{o}$C.




| Isolate | Biofilm 24hr | Biofilm 48hr | Plasmids (Kb) | AMR |
|---|---|---|---|---|
| D3 | - | - | 3.4, 5.7, 7.6 | T, TM, SU |
| D5 | - | - | 3.6, 3.7, 7.6 | T, N |
| D6 | - | - | 4.4 | - |
| D7 | - | - | 4.4 | T, S, SU |
| D10 | 37°C | 37°C | 118 | T, S, SU |
| D11 | - | - | 3.7, 3.9, 5.9 | NA, SXT, S, SU |
| D1 | - | - | 3.2, 4, 4.2 | T, S, SXT |
| D2 | - | - | 5.2 | T, S, SXT |
| D13 | - | - | 3.6, 3.8, 5.9, 90 | T, SU, SXT |
| D9 | - | - | 2, 3.7, 3.9, 4, 5.9 | T, SXT, C, SU |
| D8 | - | 25°C | 92, 166 | T, S, SU |
| D4 | 25°C | 25°C | 53 | T, AM |
| D12 | - | 25°C | - | T, S, SU |
| D14 | - | 25°C | - | T, S, SU |
| M3 | - | 25°C | 116 | - |
| M4 | - | 25°C | 2. 5.1, 8, 69 | TM, SU |
| M5 | - | 25°C | 116 | T, TM, S, SU |
| M6 | - | 25°C, 37°C | 4, 101 | SXT, S, SU |
| M7 | 37°C | 25°C, 37°C | - | - |
| M8 | - | 25°C | 44 | T, SXT, S, SU |
| M9 | - | - | 2.2, 3.2, 4, 7.2, 195 | SXT, SU |
| M10 | 25°C | 25°C | - | - |
| M11 | 25°C | 25°C | - | - |
| M12 | 25°C | 25°C | - | T, C, S, SU |
| M1 | - | 25°C | - | - |
| M2 | - | 25°C | - | - |
| M13 | 25°C | 25°C, 37°C | - | T, AM, S, SU |
| M14 | - | 25°C | 2.4 | T, AM, SXT, S, SU |

**Figure 6.3:** Variation in phenotypic characteristics and plasmid constituents in relation to a phylogenetic cladogram. Antimicrobial resistance (AMR) abbreviations correspond to: nalidixic acid (NA), tetracycline (T), neomycin (N), ampicillin (AM), sulphamethoxazole/trimethoprim (SXT), chloramphenicol (C), streptomycin (S) and sulphonamide compounds (SU). All *S.* Mbandaka isolates with the exception of M9 formed a biofilm at 25°C within 48 hours. *S.* Derby lineage 2 isolates and D8 formed biofilms within 48 hours at 25°C, only one lineage 1 isolate, D10, formed a biofilm, this was at 37°C within 24 hours. *S.* Derby lineage 1 isolates contained 17 plasmids, lineage 2 isolates only contained 1. Only isolates from lineage 2 lacked plasmids. All *S.* Derby isolates with the exception of D6 and D11 were resistant to tetracycline, *S.* Mbandaka isolates were either resistant to sulphonamide compounds or were sensitive to all antimicrobial compounds.

6.4.5 Association and invasion of IPEC-J2 monolayers

Porcine jejunum derived cell line, IPEC-J2, was used here as a porcine model for studying association to, and invasion of, porcine jejunum by representative isolates of *S.* Derby from lineage 1 and 2, and *S.* Mbandaka. Monolayers and isolates were incubated together for 30 minutes at 37$^{o}$C, a temperature chosen to reflect that of a pig's intestine (Mount 1959).

An ANOVA of the number of cells associated to the monolayer showed a significant difference between isolates ($F = 7.743$, $df = 10$ & $282$, $p < 0.001$). Tukey's HSD showed that; *S.* Mbandaka isolate M4 associated to the monolayer in significantly greater numbers ($p < 0.05$) than all isolates with the exception of *S.* Derby D1 (Figure 6.4). *S.* Derby D1 associates in significantly greater numbers ($p < 0.05$) than lineage 2 isolates, non-clustering isolate D8, and *S.* Mbandaka isolates M2 and M8. *S.* Derby D4 associated in greater numbers ($p < 0.05$) to the monolayer than the other lineage 2 isolates D8 and D12.

An ANOVA of the number of cells that had invaded the monolayer showed a significant difference between strains ($F = 10.43$, $df = 10$ & $281$, $p < 0.001$). Tukey's HSD showed that unlike the association assays, the proficiency of invasion differed between *S.* Derby lineages 1 and 2 (Figure 6.4). *S.* Derby lineage 1 isolates D1 and D2, invaded the monolayers in significantly greater numbers ($p < 0.05$) than all lineage 2 isolates tested (D4, D8, D12 and D14) with approximately 18 times as many cells internalised, as well as the non-clustering isolate D9 and *S.* Mbandaka isolates M1 and M2 with approximately 13 and 6 times as many cells internalised respectively.

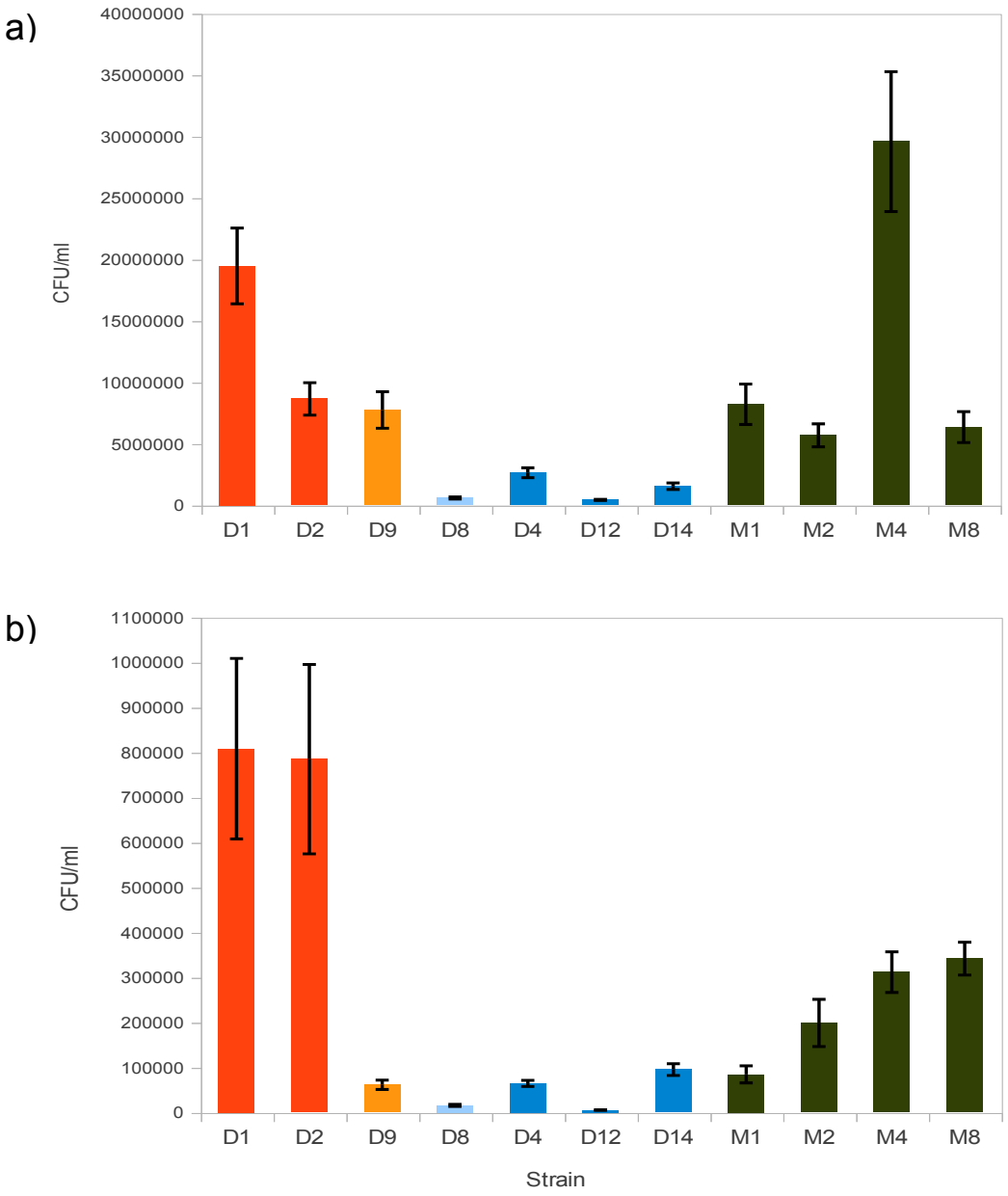



**Figure 6.4:** Variation between the number of colony forming units per ml (CFU/ml) of each isolate recovered after a) association and b) invasion assays performed on IPEC-J2 monolayers. Error bars represent +/-1SEM. a) There was no significant lineage based trend observed in the results for the association assays. b) *S.* Derby lineage 1 isolates D1 and D2 associate in significantly greater numbers ($p < 0.05$) than all the isolates with the exception of *S.* Mbandaka M4 and M8 ($p > 0.05$). This trend may reflect an adaptation towards porcine pathogenesis, with lineage 1 isolates invading to a greater extent than lineage 2 isolates and non-clustering isolates D8 and D9.

*S*. Mbandaka isolate M4 also invaded in significantly greater numbers ($p < 0.05$) than the non-clustering isolate D9, M1 and M2 as well as the *S*. Derby lineage 2 isolates, with the exception of D14. *S*. Mbandaka isolate M8 invades in significantly greater numbers ($p < 0.05$) than isolates D4, D12 and D8 from *S*. Derby lineage 2, as well as non-clustering D9 and *S*. Mbandaka isolates M1 and M2.

### 6.4.6 Plasmid constitution of UK isolates of *S*. Derby and *S*. Mbandaka

Though the plasmid nucleotide sequences will not be analysed in this paper, the number and size of plasmids possessed by the characterised samples gives a guage to the size and plasticity of the extra-chromosomal gene pool. Neither *S*. Derby lineage 1 or 2 nor *S*. Mbandaka contained a constitutive plasmid (Figure 6.3) such as the serotype associated virulence plasmids found in *S*. Typhimurium, *S*. Enteritidis, *S*. Dublin, *S*. Choleraesuis and others (Rotger & Casadesús 1999). Seven isolates of *S*. Mbandaka, M3 to M9 and M14, possessed one to five plasmids with sizes ranging from 2.2kb to 116kb. *S*. Derby has a much larger plasmid repertoire; only two isolates lacked plasmids, D12 and D14, both from lineage 2. Twenty three out of 28 plasmids found in *S*. Derby isolates where less than 8kb, the other five were between 53kb and 166kb. Isolate D8 contained two large plasmids at 96Kbp and 166Kbp, this is equivalent in size to 5% of the chromosome sequence. Isolate M9 contained five plasmids, this was the largest number of plasmids found in a single isolate, four were smaller than 8kb, and the fifth was the largest plasmid found in this study at approximately 195kb.

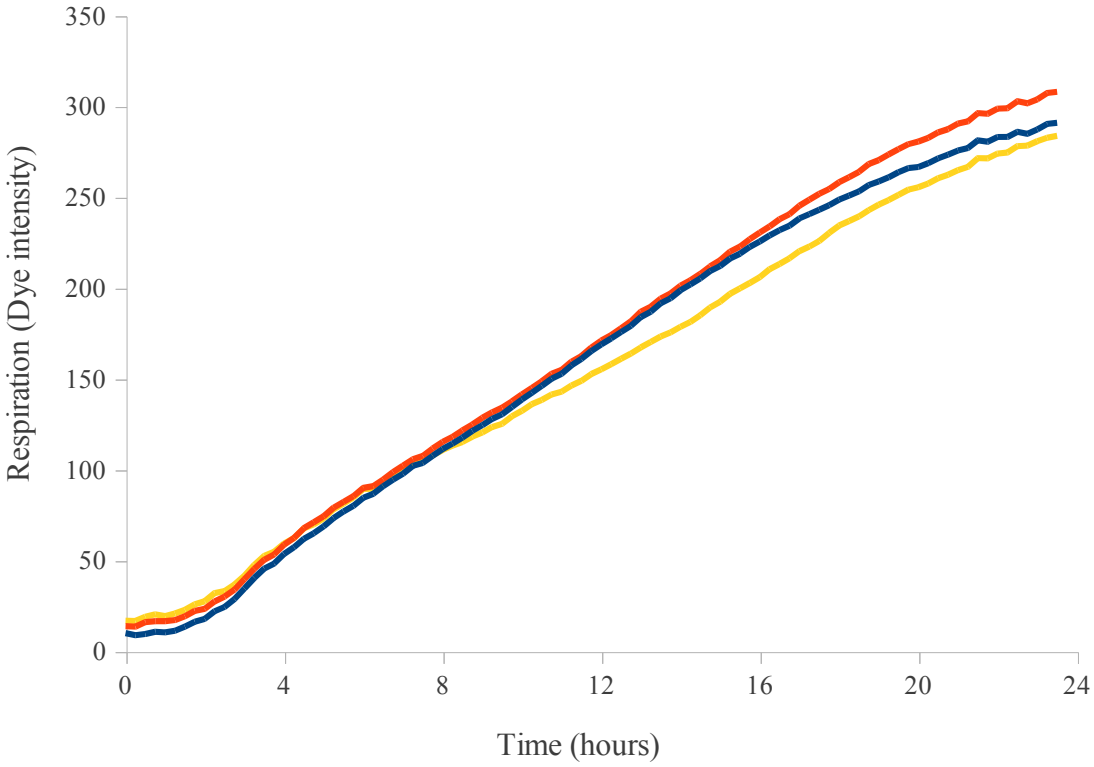
### 6.4.7 Diversity in antimicrobial resistance profiles amongst UK isolates of *S.* Derby and *S.* Mbandaka

Six isolates of *S.* Mbandaka were sensitive to all 16 antimicrobial compounds (Figure 6.3). The other 8 isolates of *S.* Mbandaka were resistant to sulphonamide compounds. Five isolates of *S.* Mbandaka were resistant to tetracycline. All *S.* Derby isolates were resistant to tetracycline with the exception of D6 and D11. Eight *S.* Derby isolates from both lineages were resistance to streptomycin. Four isolates from lineage 1 and D9 were resistant to sulphamethoxazole. Only one isolates of *S.* Derby, D6, was sensitive to all 16 test compounds. One isolate of both *S.* Derby and *S.* Mbandaka, D9 and M12 respectively, were resistant to cephalosporin. Only one isolate, D11, was resistant to naladixic acid. Isolate D5 was the only isolate resistant to neomycin. All isolates displaying antimicrobial resistance, were resistant to two or more compounds.

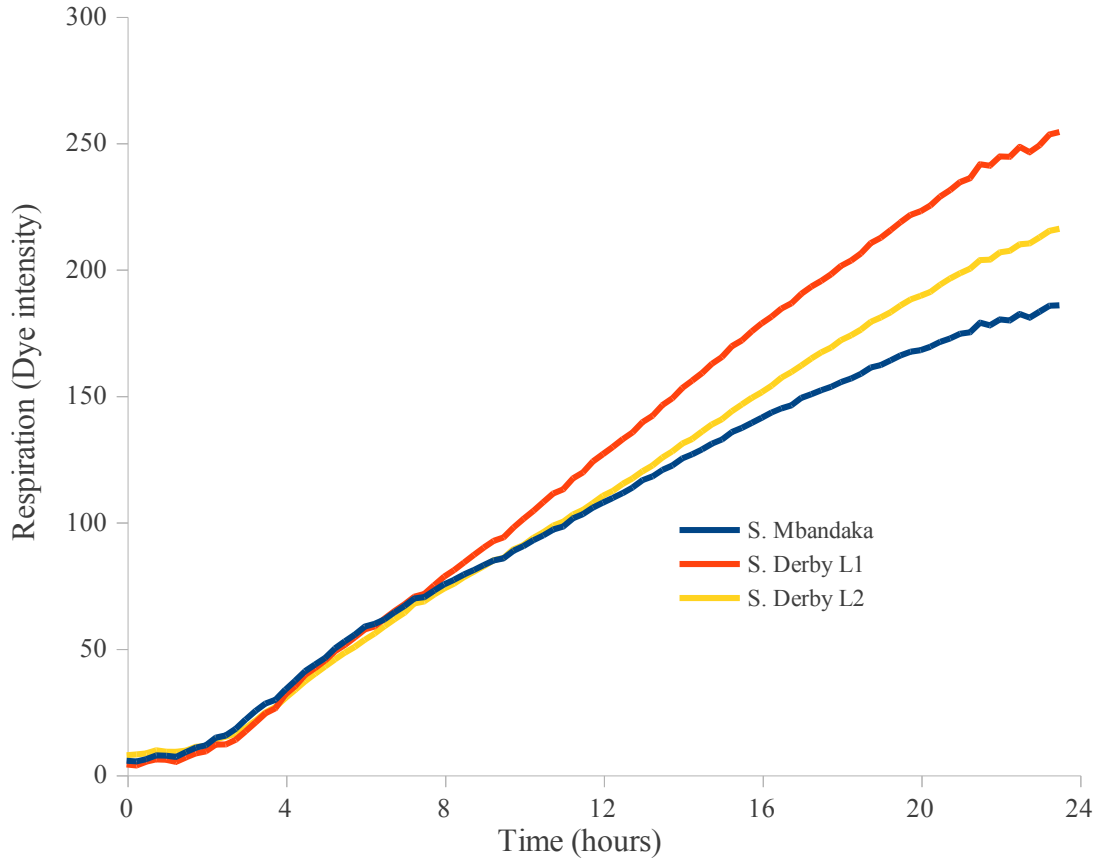
### 6.4.8 Variability in utilisation of soybean and porcine jejunum and colon homogenates by *S.* Derby and *S.* Mbandaka isolates

For each homogenate, averages for the three distinct lineages were produced, lineage 1 (D1, D2, D3, D5, D6, D7, D10, D11 and D13), lineage 2 (D4, D12 and D14) and *S.* Mbandaka. All isolates respired on porcine jejunum (Figure 6.5). Absorbance values on jejunum between lineages were not significantly different ($p > 0.05$) at any time point. Whereas on porcine colon, the dye intensity for lineage 1 of *S.* Derby was significantly different ($p < 0.05$) from that of lineage 2 and *S.* Mbandaka isolates after 13.5 and 10 hours respectively, achieving, by the end of the 24 hour period, 1.4 and 1.18 times the dye reduction of *S.* Mbandaka and lineage 2 isolates respectively.

a)


Respiration (Dye intensity)
350
300
250
200
150
100
50
0
0
4
8
12
16
20
24
Time (hours)

b)


Respiration (Dye intensity)
300
250
200
150
100
50
0
0
4
8
12
16
20
24
Time (hours)
S. Mbandaka
S. Derby L1
S. Derby L2

c)

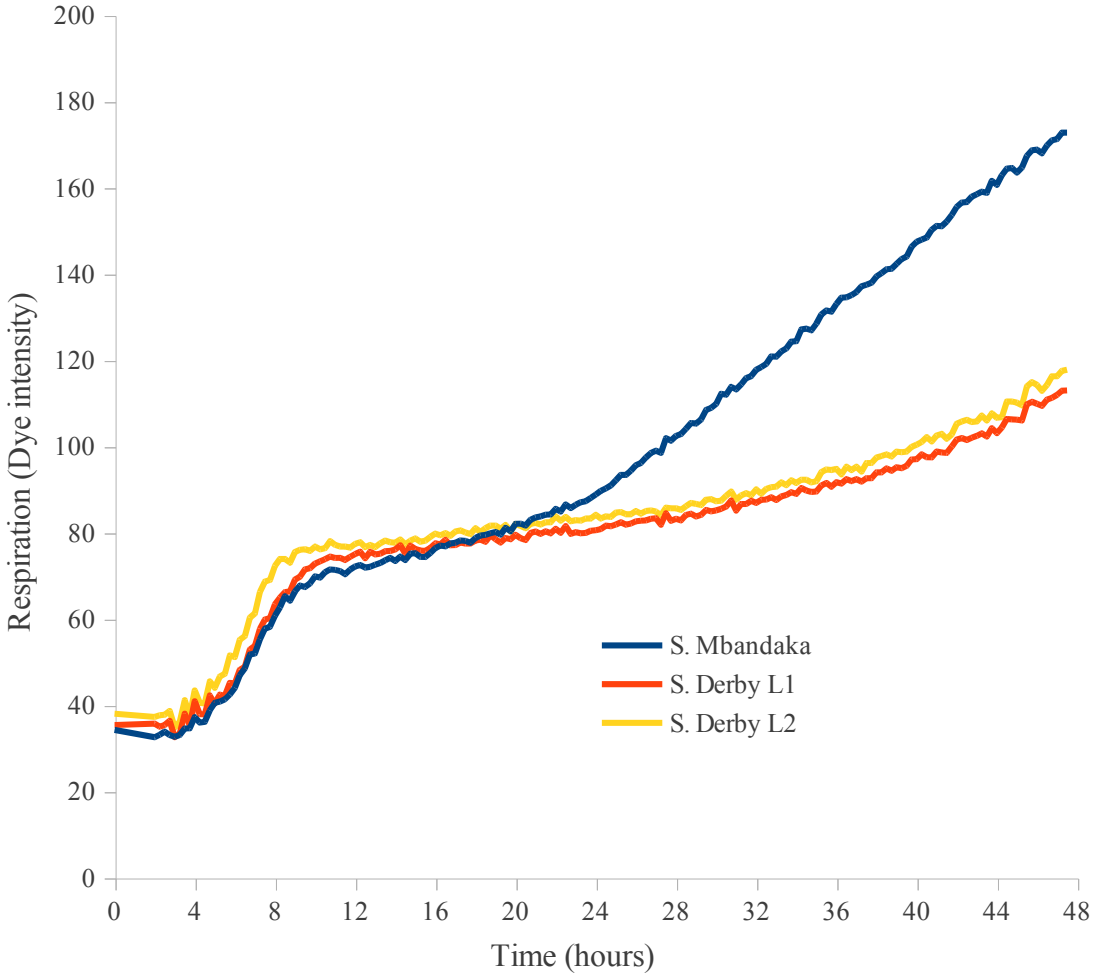



**Figure 6.5**: Respiratory dynamics of isolates of *S*. Derby lineage 1 and 2, and *S*. Mbandaka on a) jejunum, b) colon and c) soybean homogenates, measured through reduction of a tetrazolium dye. a) There was no significant difference between dye intensity values of the three lineages on jejunum homogenate over the 24 hour incubation period at 37°C. b) The dye intensity for *S*. Derby lineage 1 was significantly different ($p < 0.05$) from that of lineage 2 and *S*. Mbandaka after 13.5 and 10 hours respectively, when incubated at 37°C. c) All strains respired on soybean homogenate, though after a period of stationary respiratory dynamics amongst *S*. Mbandaka isolates, a second period of respiration was observed. These dynamics are similar to those seen during carbon catabolite repression.

When grown on soybean homogenate *S*. Mbandaka isolates displayed similar dynamics of growth, plateau, growth; dynamics typically associated with carbon catabolite repression. During the exponential phase the curves of *S*. Derby lineage 1 and 2 and *S*. Mbandaka were not significantly different ($p > 0.05$) whereas after 23.25 and 27.25 hours lineage 1 and 2 were significantly different ($p < 0.05$) from *S*. Mbandaka respectively, as *S*. Mbandaka enters the second more linear phase respiration. The final dye reduction achieved by the *S*. Mbandaka lineage was 1.4 fold greater than that achieved by both *S*. Derby lineages 1 and 2.

## 6.5 Discussion

In this study we performed several genotypic and phenotypic assays on 14 isolates of both *S*. Derby and *S*. Mbandaka isolated between 2000 and 2010 from across the UK; and observed how the results of these assays corresponded to the population structure determined by a phylogenetic reconstruction. This was done in an attempt to see which phenotypic properties, that potentially represent host adaptations are associated with particular genovars of each serovar. These adaptations may help explain why *S*. Derby and *S*. Mbandaka are consistently isolated from distinct ranges of livestock species in the UK.

### 6.5.1 Population structure and host range

There is a large diversity in the epitopes of the surface antigens "O" and "H" (Grimont & Weill 2007). Studies have suggested that this diversity is the result of different environmentally derived selection pressures. The "O" antigen is recognised by a diverse range of host specific protozoan predators. As the complement of protozon varies amongst different hosts the selection pressure posed by the predators has driven the diversification of the "O" antigen (Wildschutte et al. 2004; Wildschutte & Lawrence 2007). The diversity

between the two "H" phase antigens has been associated with immunogenicity (Bellanti 1963; Sbrogio-Almeida & Mosca 2004). It has been proposed that the ability to switch phase allows *Salmonella* to escape detection by the adaptive immune response of the host that has developed in response to the initial phase epitope (Iino & Enomoto 1966; McQuiston et al. 2008). Both antigens experience selection pressure from the host environment, "O" the gut microflora and "H" the host's immune system. We may therefore view convergent evolution towards certain epitopes, and therefore certain serogroups, as potential host adaptations. If this is the case, then the evolution of host adaptation amongst *S. enterica* isolates, cannot be studied at the serogroup level of classification alone; we must also have an understanding at the genovar level, to interpret the results of phenotypic studies in relation to host adaptation.

Identifying broad range host adaptations in *S*. Derby and *S*. Mbandaka has been the main objective of this study, and of several recent papers published by our group (Hayward et al. 2013). The results of phylogenetic reconstructions suggest that *S*. Derby comprises at least two distinct lineages in the UK, with lineage 1 associated with pigs and turkeys and lineage 2 associated with turkeys. It would be interesting to test this hypothesis further by identifying the proportion of porcine and turkey isolates that belong to these two lineages across the UK. The scale of phenotypic difference between the two lineages would suggest a high degree of genetic difference. We showed that a high number of SNPs had formed between the two *S.* Derby lineages since their divergence. We previously estimated the divergence of *S*. Derby D1 and D2 from *S*. Mbandaka M1 and M2 at between 182kya and 625kya (Hayward et al. 2013) (refer to section 2.3.13). SNP differences between lineage 1 and lineage 2 account for 67% of the SNPs between *S*. Derby and *S*. Mbandaka. If we assume a constant substitution rate then we can estimate the split of the two lineages as

occurring between 121kya and 419kya. Whole genome alignment, SNP analysis and estimation of the time since divergence between lineage 2 isolate D4 and Lineage 1 isolates D1 and D2 is the aim of future work.

*S*. Mbandaka isolates used in this study, and potentially all isolations made in the UK, come from a single lineage with no nucleotide polymorphisms across the seven house-keeping genes. Phenotypic and genotypic diversity amongst strains was low; from this we can assume that isolates of this serovar are adapted to both cattle and chickens with no clear specialist lineage. The dominance of one genovar could reflect a founder effect, resulting from recent colonisation of a small number of livestock hosts, or eradication of other lineages through extensive antibiotic use, or similar selection pressure.

### 6.5.2 The presence of SPI-23 and not SPI-1 region 1 and 2 correlates with increased invasion of IPEC-J2 monolayers

Interestingly those strains most proficient at associating to the IPEC-J2 monolayer within 30 minutes, were *S*. Derby D1 and *S*. Mbandaka M4, of these two isolates D1 contains SPI-23 and lacks SPI-1 regions 1 and 2, and M4 lacks SPI-23 and possess SPI-1 region 1 and 2. With the exception of *S*. Derby D4 associating in significantly greater number than the other *S*. Derby lineage 2 isolates D8 and D12, there was little significant difference between *S.* Derby and *S.* Mbandaka isolates in the numbers associating to IPEC-J2 monolayers. This trend was not mirrored in the number of cells invading the IPEC-J2 monolayer.

*S*. Derby lineage 1 isolates D1 and D2 invaded IPEC-J2 monolayers in significantly greater numbers than isolates from *S*. Derby lineage 2 and *S*. Mbandaka M1 and M2. In previous

work we showed the same trend in invasion between *S*. Derby D1 and D2, and *S*. Mbandaka M1 and M2; these strains have been included here as controls, but also to allow direct comparison with a wider population of isolates. *S.* Derby D1 and D2 isolates both contain SPI-23, while *S*. Mbandaka isolates M1 and M2 both lack SPI-23 (Hayward et al. 2013) (refer to section 2.3.10.10). The island has been linked with tissue tropism, as it is highly up-regulated in the porcine jejunum and less so in the colon, and contains genes which are essential for adhesion to, and invasion of, IPEC-J2 monolayers (refer to chapter 3). We could posit here that as *S*. Derby lineage 2 isolates D8 and D12 invade to a significantly lesser degree than SPI-23 harbouring, lineage 1 isolates, D1 and D2, and lack the nucleotide sequence between *gooN* and *genE,* which encodes 23 genes, that these genes may play a more significant role in invasion of IPEC-J2 monolayers than the rest of the island. This hypothesis is strengthened as this region contains five putative type III effector proteins *sanA, kayT, janE, chlE* and *yauM*, as well as the previously characterised *pilV*-like gene, *potR,* which was shown to be essential for invasion of IPEC-J2 monolayers by *S*. Derby D1 (refer to section 3.4.4). The absence of the same SPI-23 genes in D8 and D12 would suggest that they were lost prior to divergence of *S*. Derby lineage 2 from D8 and that the remainder of SPI-23 was likely lost from *S*. Derby isolates D4 and D14 recently, over a time period in which no SNPs were formed at the MLST loci. It would also appear that SPI-1 regions 1 and 2 were lost since the divergence of *S*. Derby isolate D8 from lineage 2. It is worth noting here that the numbers of *S*. Derby D1 and D2 cells that invaded the monolayer were not significantly different from *S*. Mbandaka M4 and M8, which both lack the entire SPI-23 sequence.

The development of a simple PCR for genes between *gooN* and *genE* of SPI-23 could be used as an epidemiological marker for distinguishing between lineages 1 and 2 of *S*. Derby.

However, as we have not tested either lineage against a turkey cell line we cannot conclude if the absence of this island signifies a more turkey pathogenic strain but with such a PCR tool assigning lineage with host of isolation would be an interesting avenue to explore in future surveillance.

### 6.5.3 *S.* Mbandaka is adapted towards environmental persistence

All *S.* Mbandaka isolates with the exception of M9 formed biofilms at 25$^{o}$C. *S.* Mbandaka M1 and M2 were both previously shown to contain the gene *sciN*, which is required by *E.coli* to form a biofilm; these isolates also showed the potential to experience increased gene dosage of the curli operon, an operon associated with biofilm formation (Aschtgen et al. 2008; Hayward et al. 2013) (refer to section 2.3.2). Biofilm formation by *S.* Derby lineage 2 isolates at 25$^{o}$C appears to have been acquired after the split between *S.* Derby isolates D8 and D9. The sequence of *S.* Derby isolate D4 should shed some light on differences between *S.* Derby lineage 1 and 2 with regards to biofilm formation.

*S.* Mbandaka isolates underwent a second period of respiration on soybean homogenates at 25$^{o}$C which did not occur with *S.* Derby isolates. This could reflect catabolite repression, and hence the ability of *S.* Mbandaka to compete with *S.* Derby for the same set of metabolites, until they are exhausted before utilising a second set of metabolites. We showed in previous work that *S.* Mbandaka M1, and not *S.* Derby D1, was able to respire at 25$^{o}$C on components of the soybean metabolome: D-saccharic acid, succinic acid, D-trehalose, mellibiose, methyl succinate and fumaric acid (refer section 4.4.1). It would appear that *S.* Mbandaka initially uses the same components as *S.* Derby, until they are

exhausted, at which point they begin to utilise other components potentially including the above mentioned metabolites.

Biofilm formation and more proficient use of soybean metabolites, suggest that *S.* Mbandaka is better adapted to survival in the farm environment, persisting in biofilms, utilising a greater number of soybean metabolites, a common ingredient in animal feed in the UK (Csáky & Fekete 2004; Food Standards Agency 2013). This would also explain the reduced level of antibiotic resistance amongst *S.* Mbandaka strains as they can persist in the, antibiotic free, external farm environment. This could present a model for niche partitioning where one population is better adapted to colonising and persisting in the host and the other adapted to persisting and growing in the external environment: an intriguing hypothesis worthy of further study.

### 6.5.4 Tetracycline resistance an adaptation to surviving in farmed pigs

With the exception of D6 and D11, *S.* Derby isolates were resistant to tetracycline. Whereas only five isolates of *S.* Mbandaka were resistant to tetracycline. In the past tetracycline has been used as a growth stimulant in pigs (Horvath & Noot 1954). The compound was prohibited for this use in the UK in 1971(Smith 1975), yet remains, in the UK, the most used antibiotic in livestock (pigs, cattle and poultry) accounting for 32% of antibiotics used in 2011. Pigs account for the highest amount of antibiotic usage in the UK amongst all livestock species (VMD 2012). This may explain why 60-66% of *S. enterica* isolates obtained from pigs in the UK in 2011 were resistant to tetracycline (AHVLA 2011). The high usage of all antibiotics in pigs in the UK may explain why all isolates except D6 of *S.* Derby were multi-drug resistant. This suggests that tetracycline resistance may define, in

part, the niche occupied by *S*. Derby. We could posit that antibiotics are a barrier to *S*. Mbandaka colonisation of pigs. Those isolates of *S*. Mbandaka that did display resistance to antimicrobials, were resistant to sulphonamide compounds, as well as three lineage 2 isolates of *S*. Derby. In 2011, 86% of *S. enterica* isolations made from turkeys were resistant to sulphonamide compounds (VMD 2012). The high number of *S*. Mbandaka isolates with no resistance suggests that the serovar largely occupies an environment free, in whole or in part, of antibiotic stress.

# Chapter 7:
# **Discussion**

## 7.1 An approach for discovering host adaptations

In this thesis, I set about trying to answer the question "why do *S. enterica* serovars Derby and Mbandaka have distinct host associations?" by identifying and characterising those genetic and phenotypic features which distinguish the serovars, and therefore may act as adaptations to different hosts and their surrounding environments. The serovars studied were identified through comparison of isolation statistics, published annually by the AHVLA, for the period 2000 to 2010. *S.* Derby was found to be isolated predominantly from pigs and turkeys, whereas *S.* Mbandaka was isolated predominantly from cattle and chickens. This was corroborated in part by a meta-analysis of the global literature, the majority of which identified pig and poultry as the host of isolation of *S.* Derby and *S.* Mbandaka respectively. I was careful to select serovars that were consistently isolated from different, defined subset of host species to avoid the possibility of short term bias such as new farming practices, point source contamination and so forth, which could confound and obscure host adapted serovars. The avenues pursued and experiments performed were designed in response to differences discovered through comparative functional genomics of two isolates of each serovar. In parallel to this approach, metabolic phenome profiling was used to study differences in metabolite utilisation under different conditions, reflecting possible temperatures and oxygen compositions encountered in a porcine host and in the environment. Genome-scale metabolic models were generated and curated for isolates *S.* Derby D1 and *S.* Mbandaka M1. These allowed the simulation of the interactions between all the identified metabolic genes in the genomes when challenged with the different metabolite compositions of the porcine jejunum and colon. Genotypic and phenotypic differences discovered between the four sequenced isolates were studied against 14 isolates of each serovar isolated from across the UK between 2000 and 2010, for which the phylogenetic relationship between isolates was determined. This was done to see which phenotypes and genotypes were representative of the whole serovar, or correlated with

certain branches of the phylogeny, to see which of these features most likely influenced the isolation statistics in the UK. A question that arises when considering the approach utilised in this study is "can this approach be applied more widely to not just other pair-wise comparisons but as a general systematic approach to all serovars?" One intriguing issue would be how this approach might be modified to distinguish convergent host adaptation. A caveat for any pair-wise analysis is that it may not translate more widely to other serovars and it should be kept in mind that quite discrete gene regions, although unrelated in structure and sequence, may confer the same phenotype or effect. With these caveats in mind, this discussion will focus on the pairwise analyses undertaken, and draw careful abstractions when considering the wider picture of differences in the complement of host adaptations between divergent lineages of *Salmonella enterica*.

### 7.2 Structure of thesis discussion

As each chapter has its own extensive discussion, I will give a brief summary of the main results, and the significance of the results with regards to identifying potential phenotypic differences that could influence host distribution. I will also discuss the strengths and limitations of the overall study. Where possible I will incorporate results from multiple chapters and suggest experiments to be performed in the future that could potentially strengthen the conclusions drawn.

### 7.3 Using comparative functional genomics to discover host adaptations

Genome sequences were generated for *S*. Derby D1 and D2 isolated from pigs in 2008 and *S*. Mbandaka M1 and M2 isolated from cattle in 2008 and 2009 respectively. The sequences were annotated using the RAST annotation system. The annotations were compared in

SEED genome viewer by function; this identified genes that differed between the strains which could potentially convey a unique phenotype. One assumption that must be addressed here is that the comparison performed by SEED relied on the correct identification of homologous function by RAST from the pfam database. As the database was updated over a one year period, the putative functional differences between the serovars did not change, this suggests that during this period of database curation the functions attributed to the gene sequences did not change, or at least not in a way that further distinguished the strains. Yet as the proficiency of the annotation system is likely to increase we could potentially find genes with distinct functions amongst the 251 unique hypothetical proteins which distinguished the serovars; posing new targets for studying differences in host adaptations between these two serovars.

In this study I have assumed that highly similar genes have the same functions in both genetic backgrounds and that major phenotypic differences are to be identified from the functional prediction of genes that are not shared between the serovars. This is one potential draw-back of using RAST for comparative functional genomics, as it is function orientated, it does not consider the genetic background in which the genes function was identified. One scenario in which this could be a problem is where similar genes may be expressed differentially as a result of variation in the genetic background, and hence result in a very different phenotype. The method presented in this study has proven very useful when it comes to identifying potential phenotypic differences, though one of the major draw-backs in comparing the static genome sequence, is the lack of information regarding the conditions under which the phenotypes are manifest.

Scaffolding of the genome sequence provided information regarding contig orientation, and facilitated the discovery of a large sequence inversion in the chromosome sequence of *S.* Mbandaka, which may influence the expression level of genes pertaining to environmental adaptation. The inversion is particularly interesting when considering the discovery of host adaptations and the mechanisms that lead to their evolution, as it is a mechanism independent of horizontal gene acquisition, which is vertically transmitted. It is also a mechanism that may be missed when comparing genomes that have been assembled from contigs, using a reference genome as a scaffold.

In section 6.4.1 I generated a phylogenetic reconstruction of 14 *S.* Derby and 14 *S.* Mbandaka isolates obtained during background monitoring across the UK between 2000 and 2010. I showed that for the seven MLST loci studied, *S.* Mbandaka and the sequenced isolates of *S.* Derby were clonal. Comparison of the genome sequences and the functional annotations of these clonal strains in chapter 2 showed that there was genetic variation between isolates which was missed through MLST and serological typing. These results suggest that there is a finer resolution of genetic diversity, potentially functional, between strains which is not accounted for when using the sequences of MLST loci to perform phylogenetic reconstructions. Phenotypic variation was also observed between clonal isolates, in chapter 3 *S.* Derby D2 associated in significantly greater numbers than *S.* Derby D1 in both the colon and jejunum IVOC assays. In addition to this, in chapter 4, I showed that there was phenotypic variation in metabolite utilisation for individual strains under different conditions. This suggests that caution must be taken when inferring phenotypic properties from functional genomics, as the environmental conditions may in some cases influence greater phenotypic variation than subtle differences in gene content.

Another potential mechanism of host adaptation discussed in section 2.3.12, is the adaptive immune response encoded by the CRISPR operons. There was little homology in spacer sequences between serovars. *S.* Derby contained a complete complement of *cas* genes required for implementing the CRISPR spacer sequences, whereas *S.* Mbandaka lacked these (refer to section 2.3.6). This could indicate that *S.* Mbandaka no longer has a functioning CRIPSR operon, potentially leaving itself open to bacteriophage and other foreign DNA elements. The potential significance of this is a higher rate of integration of foreign DNA, which could result in greater genome plasticity. The spacer sequences may not necessarily reflect different foreign DNA elements, and may reflect different regions of the same element. This adaptive immune response could define the habitable regions for different CRISPR spacer genovars of *S.* Derby. This potential host adaptation will be further studied during future analysis and comparison of the *S.* Derby lineage 2 isolate, D4 with D1 and D2. Assuming some relationship between specific bacteriophage and particular hosts, it is possible that host range could be defined in part by phage resistance (Duerkop et al. 2012; Moreno Switt et al. 2013).

### 7.4 Pathogenesis, host adaptation and tissue tropism

To establish if *S.* Derby D1 and D2 were more proficient at associating and invading porcine cells than *S.* Mbandaka M1 and M2, association and invasion assays were performed with a porcine jejunum derived cell monolayer, IPEC-J2 (refer to section 3.4.1). Over a 60 minute incubation period *S.* Derby D1 and D2 associated with, and invaded, the monolayer at a faster rate than *S.* Mbandaka M1 and M2. These results suggested that there was a difference in proficiency of pathogenesis between the serovars, which may reflect an adaptation of *S.* Derby isolates to the porcine host. Association and invasion assays were also performed in section 6.4.5 to include a larger number of isolates, and to see if

pathogenicity varied across lineages. The isolates were tested under the same conditions, but for just one incubation period, 30 minutes. This incubation period was selected based on the results of the comparison between *S*. Derby D1 and *S*. Mbandaka M1, presented in section 3.4.1, which were performed as preliminary experiments to select incubation periods. Further experimentation, presented in section 3.4.1, showed that *S*. Derby D2 and *S*. Mbandaka M2 were not significantly different at this time point. It may be desirable to repeat the population association assays, in section 6.4.5, for an hour incubation period as at this time point *S*. Derby D1 and D2 and *S*. Mbandaka M1 and M2 isolates were significantly different. However, the results presented in section 6.4.5 showed that the trend in association and invasion between the serovars did not hold for the comparison between all *S*. Derby lineage 1 isolates (D1, D2 and D9), those with SPI-23, and all *S*. Mbandaka isolates (M1, M2, M4, M8), which lack SPI-23. Whereas, *S*. Derby lineage 1, isolates with SPI-23, invaded monolayers in significantly greater numbers than lineage 2 isolates, which lack SPI-23. This may suggest that *S*. Derby lineage 2 is less pathogenic than *S*. Derby lineage 1 and *S*. Mbandaka isolates, though this would require further experimentation, including IVOC and animal dosing studies. It is also worth noting that in some studies tissue culture models for host colonisation have compared poorly to inoculation studies; in some instances showing the opposite trend. This may suggest that there are additional host barriers that must be over come by the pathogen in addition to the invasion of the host cell. A notable example of non-concordance between *in vivo* and *in vitro* models is presented in a study by Watson et al (Watson et al. 2000). In this study the virulence of serovars *S*. Choleraesuis, *S*. Dublin and *S*. Typhimurium was investigated in both the tissues of a porcine host post inoculation and as a result of *in vitro* infection of porcine alveolar macrophage. The study found that *S*. Typhimurium and *S*. Dublin were less virulent in a porcine host then *S*. Choleraesuis, whereas in the *in vitro,* alveolar macrophage experiment the opposite trend was observed. The virulence of *S*. Dublin in the porcine alveolar

macrophage was much higher than *S*. Choleraesuis, yet was found to be avirulent *in vivo*. This observation, suggests that care must be taken when interpreting the results of cell culture experiments. In relation to the topic of this thesis the higher invasion rate of *S*. Derby in to IPEC-J2 cells compared to *S*. Mbandaka may not manifest as higher virulence once within the porcine intestine. In the absence of inoculation studies it can not be concluded with certainty that pathogenesis plays a direct role in host-specificity.

The above mentioned serovar *S*. Choleraesuis is considered host restricted, being predominantly isolated from a porcine host. Due to the high prevalence in pigs it may be expected that *S*. Choleraesuis would be virulent when experimentally ingested by a porcine host and avirulent in other common *Salmonella* hosts. Intestingly, studies have shown that *S*. Choleraesuis is highly virulent in a porcine host, as would be expected, but is also highly virulent in a bovine host (Watson 2000, Paulin 2002). This suggests that cattle are not regularly in contact with sources of *S*. Choleraesuis, unlike pigs, and that other environmental factors, *ex vivo*, must influence the prevalence of this serovar on cattle farms. These results signify that care must be taken when combining the results of meta-analyses, host trends observed in isolation statistics and the results of cell culture experiments. Whereby, a higher degree of invasion of a monolayer may be identified as an adaptation to cellular invasion, but not nessicarily an adaptation to increased virulence when ingested by the host or as a determinant of frequency of isolation from a particular host.

Pathogenesis of *S*. Derby D1 and D2 was further studied for signs of tissue tropism by performing IVOC association assays with sections of porcine jejunum and colon. The

numbers of *S*. Derby D2 cells associated to IVOC porcine jejunum and colon sections after 30 minutes were significantly greater than *S*. Derby D1 cells. Both isolates showed tropism to porcine jejunum over colon. As I identified potential pathogenicity related genes in SPI-23, it may be hypothesised that the island would be regulated in a tissue specific manner, reflecting the tropism of the strain to porcine jejunum. When the expression levels of ten genes from SPI-23 of *S*. Derby D1 were determined after 30 minutes incubation with porcine jejunum and colon IVOC, and compared against a no tissue control, it was found that the island was up-regulated when exposed to porcine tissues. The genes *potR*, *genE*, *sadZ, tinY* and *docB* were up-regulated to significantly greater degree in the porcine jejunum when compared to the porcine colon. This identified these genes as potentially playing a role in the observed tissue tropism. It would be interesting to see how the island is regulated when the pathogen is confronted with turkey, cattle or chicken, cells lines or IVOC, to see if the regulation of these genes could also play a role in host determination as well as tissue tropism. Further experimentation is required to identify if *potR*, *genE*, *sadZ, tinY* and *docB* are regulated in response to the same environmental stimuli in both the colon and jejunum, and if up-regulation is concentration dependent or due to the presence of unique environmental components. Expanding the expression studies to include the other 32 genes found on SPI-23 may identify the number of operons spread across the island. It would also be interesting to test if SPI-23 acts in a modular fashion, whereby its acquisition by *S*. Mbandaka or *S*. Derby lineage 2 isolates conveys increased pathogenicity. This could be studied using plasmid capture of the island from the chromosome of *S*. Derby lineage 1, followed by transfection, on an expression vector, in to *S*. Mbandaka and *S*. Derby lineage 2 isolates. These transfected strains could then be tested for increased rates of invasion and association to IPEC-J2 monolayers. If the regulatory machinery of SPI-23 is located exclusively within the SPI-23, expression would be as anticipated in the new background. However, if there are dependencies for expression such as uptake of metabolic signals or

the need to integrate expression in wider regulatory networks, then this experiment may not yield the anticipated increased rates of expression. This leads to interesting questions regarding recent acquisition of new genetic elements and what intra-genetic adaptations are needed for appropriate co-ordinated regulation?

The knock-out mutant *S*. Derby D1 Δ*potR* displayed a strong agglutinating phenotype when left static at room temperature, which on several occasions, prevented sub-culturing. Negative stain microscopy showed the mutant cells to be covered in pili-like appendages, which were largely lacking from the parent strain. *S*. Derby D1 Δ*potR* displayed a similar phenotype to that seen in *E. coli* when the gene subunit *pilVA'* is knocked-out of the shufflon; expression of *pilV* causes the production of detachable pili on the cell-surface which reduce the degree of agglutination between bacteria (Yoshida et al. 1998; Tam et al. 2006). Differential staining confocal microscopy showed that the mutant strain associated to an IPEC-J2 monolayer in significantly lower numbers than the parent strain, and unlike the parent strain did not invade. The agglutinating phenotype of the mutant strain and the significantly greater up-regulation of *potR* in the jejunum compared to the colon, suggest a mechanism which may explain the tissue tropism observed from the IVOC association assays. Whereby the gene *potR* is up-regulated in the porcine jejunum, increasing the expression of detachable pili from the cell surface, these detachable pili allow the bacteria to attain a planktonic state, in which they can spread out and attach to a greater surface area of the porcine jejunum. Once the bacteria migrate out of the jejunum in to the colon the expression level of *potR* drops significantly, producing fixed pili as opposed to detachable pili, potentially causing the bacteria to agglutinate prior to exiting the body. To address this hypothetical model the correlation between the expression level of *potR* and the proportion of agglutinated to planktonic bacteria would need to be determined. The expression level of

*potR* when *in vivo*, passing through several distinct gastrointestinal environments before reaching the colon and jejunum, would need to be tested to see if expression levels are similar to those measured, *in vitro*, using IVOC methods. Another interesting experiment might be to place *potR* under a different, inducible, promoter to test whether tissue tropism can be experimentally influenced.

This project has focused largely on pigs, one of the preferred hosts of *S*. Derby, identified from the AHVLA and CDC statistics and the meta-analysis presented in section 1.4; as a result of this selection it has not been possible to comment on the potential determinants of the full host range of both of these serovars. As was mentioned above with regards to host association identified from isolation statistics, it is not possible to comment on the role virulence may play in range determination (Paulin 2002, AHVLA 2011). It is possible that host barriers, metabolite compositions or farming practices may determine the associations gleaned from the isolation statistics and meta-analysis.

The porcine host was chosen, as a well characterised cell line was available, in addition to the facilities, expertise and protocols to perform IVOC with porcine tissue. Further experimentation, with regards to pathogenesis of both of these serovars, across their full host range, including invasion and association assays on livestock and human intestinal cells, livestock IVOC and inoculation studies, would be incredibly informative, and may shed light on the importance of pathogenesis in host association.

## 7.5 Potential niche partitioning as a result of differences in proficiency of metabolite utilisation

In chapter 4 I showed how temperature and oxygen influence the proficiency of *S.* Derby D1 and *S.* Mbandaka M1 to utilise different metabolites in the wells of Biolog phenotypic microassays PM1-4 and PM6. The experimental set up allowed the discovery of three properties of metabolite utilisation, these being, oxygen dependent metabolite utilisation, temperature dependent utilisation and a significantly more proficient use of the metabolite by one of the two serovars. Proficiency was defined as a faster start of utilisation post inoculation, a steeper log-phase respiratory gradient, which was shown in section 5.4.1 to correlate with growth, and a greater maximum dye intensity, reflecting a higher production of NADH in the well. For the final parameter there is the limitation that if the dye has been fully reduced, then no further respiration will be measured, making it possible to miss differences in maximum respiration that are beyond this limit.

One of the most interesting observations made during this study was that the differences between isolates in proficiency of metabolite utilisation observed under one condition were not observed under other conditions. At an ambient temperature of 25$^{o}$C with oxygen, *S.* Mbandaka M1 was found to be more proficient at utilising 10 metabolites than *S.* Derby D1. These metabolites were shown to be constituents of the soybean metabolome. This study focused on the soybean metabolite as it has previously been shown that soybean feed is a route by which *S.* Mbandaka has previously entered the farm environment (Wierup & Häggblom 2010); though it is possible that these same metabolites may also be found in other plants or plant based feed (Torres et al. 2011). In particular rapeseed oil which *S.* Mbandaka has also been shown to be a contaminant of (Morita et al. 2007). I also showed in section 6.4.8 that *S.* Mbandaka experienced a second period of respiration after 24 hours

which *S*. Derby lineage 1 and 2 did not, when cultured in soybean homogenate at 25°C. These dynamics may reflect aspects of catabolite repression, whereby the pathogen utilises a particular set of metabolites first, competing with *S*. Derby until they are depleted before moving on to other carbon sources. The same dynamics may reflect the utilisation of those metabolites that had longer lag times. In section 2.3.8, regulatory, transport and metabolic genes were shown to be present in the chromosome of *S*. Mbandaka M1 and not *S*. Derby D1 for the utilisation of the sugar D-galactonolactone, the Biolog results suggest that these genes are regulated in a temperature dependent fashion, and used at 25°C and not 37°C. Expression studies, like those performed in section 3.3.4 for SPI-23, would need to be performed to confirm this relationship. We can also not infer the threshold temperature at which *S*. Mbandaka begins and stops utilising D-galactonolactone, but these results suggest there is a difference in utilisation between ambient and body temperature.

Once the incubation temperature of the Biolog PMs was raised, from 25°C to 37°C, the metabolic advantages displayed by *S.* Mbandaka M1 disappeared, and *S*. Derby D1 became more proficient at utilising four metabolites. Once oxygen was removed these advantages again disappeared and *S.* Mbandaka M1 became more proficient at utilising four metabolites. Of these, glycine which is used to supplement animal feed, as a pH buffer and flavouring, and the dipeptide form, glycyline-alanine, both were utilised sooner by *S*. Derby at 37°C in an aerobic environment (glycine is sold as a feed additive by: Azelis, Europe; United Foods Corporation, USA; FoodChemChina, China). Glycine was not utilised at ambient temperatures by either strain. This may reflect an adaptation to animal feed once it is within the host. Of course one significant confounder of this 'dissecting' approach is that each metabolite is looked at in isolation from the real and often exceedingly complex environment and we have as yet little understanding, other than catabolite repression, of

metabolite use when many are in abundance without a preferred energy source such as glucose available. That said, the approach used in this study gives clues as to strain-metabolite competencies and preferences which surely contribute to the adaptive response of these organisms. This is an example of where the models presented in section 5.4.2 go further towards helping us to dissect these complex interactions, and assess the significance of individual metabolites on the proficiency of biomass incorporation.

Soybean feed is used in the USA and Europe and frequently fed to cattle, cows, pigs, turkeys and other livestock species. As was mentioned in section 2.3.6, *S*. Derby D1 and D2 possess the *mer* operon, which conveys mercury resistance. Fish meal, a common feed additive, is high in mercury. It may be posited that fish meal irrespective of the presence of soybean meal, a combined feed commonly given to pigs could inhibit *S*. Mbandaka but not *S*. Derby. A temporary EU-wide ban on the use of fishmeal in ruminant feed has been in place since 2001 (EC 2000/766 and 2001/9). In 2003, the temporary ban was transferred into permanent measures within the TSE regulations (EC 999/2001) (Andreoletti et al. 2007). This means by law in the EU that pig, chicken and turkey feed can contain fishmeal but not cattle, the feeding regimes of different livestock species could significantly influence the habitable niche of the two serovars. Though it must be noted that fish meal is used in one of *S*. Mbandaka preferred hosts, poultry, and therefore it is unlikely that mercury resistance poses such a strong selection pressure that it excludes *S*. Mbandaka from pigs and turkeys.

In section 5.4.5 I showed that *S*. Derby D1 out competes *S*. Mbandaka M1 on both porcine jejunum and colon homogenates. This was shown not to be influenced by cellular exudates

of either serovar, determined through spent media assays. *S*. Derby also respired to a significantly greater extent on porcine jejunum compared to the colon. This may suggest that the tissue tropism displayed by *S*. Derby to porcine jejunum with regards to association is also accompanied by a more proficient ability to utilise the metabolite available in that environment compared to those found in the colon. It was also shown that *S*. Derby D1 grew to a greater extent in the porcine colon homogenate than *S*. Mbandaka M1. These results suggest that when in mixed culture *S*. Derby D1 is able to utilise the available metabolites at a faster rate than *S*. Mbandaka M1. This could lead to competitive exclusion of *S*. Mbandaka when in the colon or jejunum of a porcine host. But as was shown in sections 4.5.1 and 6.4.8, *S*. Mbandaka grows more proficiently on soybean based feed at ambient temperatures. The relative fitness of these two serovars is potentially dependent on the phase in pathogenesis in which they are encountered and the relative availability of nutrients from the feed. It can be assumed that the prevalence of soy will reduce along the length of the gut in which case Mbandaka should be nutritionally favoured in the small intestine and less so in the large intestine. To confirm this, mixed culture competition on soybean homogenate should be performed. It was shown in section 5.4.5 that *S*. Derby D1 out-competes *S*. Mbandaka M1 on porcine jejunum and colon homogenates. It could therefore be posited that the two serovars may have a greater relative fitness in different phases of pathogenesis, *S*. Derby within a porcine host and *S*. Mbandaka in soybean based feed. It should be noted here that increased growth rate of *S*. Typhimurium in porcine intestinal tissue has been found to be linked with increased levels of proinflammatory cytokines. This was demonstrated in a study by Paulin et al. as a potential limiting factor, keeping *S*. Typhimurium infection largely localised to the mucosal wall of the intestine, and hence causing fewer systemic and potentially fatal infections. Whereas in the same study, the pig restricted serovar *S*. Choleraesuis was found to undergo a period of slow growth and a higher degree of persistence in the mesenteric lymphnodes leading to a systemic

infection. Paulin et al. suggested this may allow the serovar to go unnoticed by the host's immune system (Paulin et al 2007). This suggests that *S*. Derby's faster replication rate on porcine intestinal homogenates could actually hinder the persistence of the pathogen and development of systemic infection by triggering an immune response. To determine if this is indeed the case, inoculation studies considering persistence of the infection and cytokine levels would need to be performed.

### 7.6 Identifying potential differences in the proficiency of metabolite utilisation in the porcine intestine using genome-scale metabolic reconstructions

In section 2.3.9 I showed that there were differences in metabolic pathways between *S*. Derby D1 and D2 and *S*. Mbandaka M1 and M2. In section 4.4.3 I showed that *S*. Derby D1 and *S*. Mbandaka M1 differed in their proficiency to utilise several metabolites at 37$^{o}$C under aerobic conditions. In section 3.4.2 I showed that *S*. Derby D1 and D2 display tropism towards the porcine jejunum over the porcine colon. In section 5.4.8 I attempted to merge the results from section 2.3.9 with those of section 4.4.3 to see if it was possible to explain the tropism observed in section 3.4.2, and potentially shed further light on the role of metabolic composition, and competence of utilisation, on defining the distinct host distributions of these two serovars, and comment on the possible implications on the proposed model of niche partitioning.

In section 2.5.3, transport reactions were automatically added to the metabolic models by modelSEED, and were used as indicators of metabolites that the organisms could potentially use. Where the transport reactions, porcine jejunum and colon, mucus and tissue metabolomes had a corresponding transport reaction, I assumed that the metabolite could be utilised by the pathogen and therefore included it in a virtual medium which was used to

challenge the models. Future curation will allow the flux of metabolites in to the model to reflect the optimal growth rate on the metabolite tested in isolation in the well of a Biolog PM at 37$^{o}$C with oxygen. I also plan to produce a separate set of models which will be curated with the flux rates for the Biolog PM performed at 25$^{o}$C and will be challenged with the soybean metabolome mentioned in section 4.5.1 (Joshi et al. 2010). These models will identify which secondary metabolic reactions influence the optimal growth rate when exposed to environments with different metabolite compositions. These results could potentially show if metabolite composition contributes to the observed host specificity and tissue tropisms. The models can also be used to investigate the significance of single metabolites in a complex mix of metabolites, by performing flux-balance analysis and seeing which metabolites most significantly affect the growth parameter when removed from the virtual media.

The models presented in chapter 5 need additional curation as the automated models produced from the RAST annotation did not account for pseudogenes. As a consequence of this limitation the number of pathways in the metabolic networks may have been over estimated. Since the majority of metabolic genes in *Salmonella* are not essential and therefore are susceptible to decay when not maintained under positive selection, it may be reasonable to expect the model predictions to change significantly (Becker et al. 2006). One of the most dramatic examples of gene decay influencing metabolic potential is found when comparing *Mycobacterium leprae* with *Mycobacterium bovis* where over a thousand genes have decayed to become pseudogenes (Vissa and Brennan 2001). In addition to producing more accurate models, the identification of pseudogenes could potentially identify the degree of selection on different metabolic genes and pathways. In addition to the models presented here it may also be interesting to see if other serovars associated with similar

hosts to *S*. Derby and *S*. Mbandaka have accumulated non-synonymous SNPs in over lapping sets of metabolic genes, potentially identifying genes which are no longer under purifying selection and therefore used in environments foreign to the host adapted serovars.

I showed in chapter 4 that temperature and oxygen are potentially significant influences defining the different niches of *S*. Derby and *S*. Mbandaka. The concept of metabolic niche, in potentially defining host range and tissue tropism, is elaborated on in section 5.5 where the metabolite composition of porcine jejunum and colon, mucus and tissue, is determined. Perhaps a more profound issue is that the potential exists for differential expression of the same metabolic pathway in different serovars. Is it possible that the metabolic pathway is essential, and therefore is retained through selection, but that its regulation is a target for more rapid evolution so that similar organisms sharing similar metabolic capabilities become adapted to specific niches. It can be argued that over-reliance on genetic data alone to understand evolution is potentially dangerous as it is the phenotype within the many environments in which an organism is challenged is more prescient: genotype reflects 'potential' competence whereas the phenotype reflects 'actual' competence. This limitation has been demonstrated in a study by Becker et al. where it was experimentally found that genes which had been previously identified through genome-scale metabolic reconstruction to be essential to the growth of *S*. Typhi *in vivo* were in fact redundant when experimentally knocked-out. Based on the models alone it was predicted that there would be many more targets for antimicrobial development then were found experimentally. In addition to this, several of the genes found to be essential to *S*. Typhi were also essential for human metabolism and therefore not candidates for clinical antimicrobials (Becker et al. 2006).

It is becoming increasingly common to integrate data from RNAseq and Transposon Site Hybridization (TraSH) with genome wide metabolic reconstructions (Chindelevitch et al. 2012, Lee, D. et al. 2012, Puniva et al. 2013). Through these methods it is now possible to consider focal points in metabolic networks that receive a high level of flux, and are also regulated in responses to environmental perturbations (Soons et al. 2013). Within the confines of my current study it has been shown that *S.* Derby and *S.* Mbandaka differ considerably with regards to metabolic gene compoisiton (chapter 2), metabolic potential (chapter 4) and metabolic pathways (chapter 5), yet with limited concordance between these different elements. The addition of gene essentiality and regulatory data may help explain the cryptic phenotypes seen under varying conditions in the Biolog system (chapter 4). Further parametratisation of the metabolic reconstructions may also shed light on the mechanisms by which these metabolic networks have diverged and potentially become adapted towards different host environments, comprising of different metabolites under different conditions.

The porcine metabolomes presented in section 5.4.6 could be used outside of this study with metabolic models generated for other enteric pathogens of pigs. The pig is also being used increasingly as a surrogate model for the human digestive tract and, therefore, this information could be useful to others interested in the composition of the porcine jejunum and colon (Merrifield et al. 2011). The blunt end integration of the host metabolome and Biolog data with the metabolic reconstructions allows the modelling of host and pathogen interactions at the metabolic level; yet more complex relationships such as the requirement of tetrathionate for anaerobic respiration of ethanolamine in the gut is not modelled here. This is an example of a two step interaction whereby *Salmonella* induces the release of the

electron acceptor tetrathionate from the host before it can utilise host derived ethanolamine (Thienimitr et al 2011).

### 7.7 Determination of the population structure to relate phenotypes to serovar level host associations

The statistics that formed the justification for studying *S. enterica* serovars Derby and Mbandaka, were based on the typing of two surface antigens, previous work has shown that serovars may be polyphyletic and comprise of very distinct genovars (Achtman et al. 2012). Therefore it is not possible to assume that all of the genotypes and phenotypes are shared by all isolates. There is also the possibility that a single broad range serovar could represent several host restricted genovars, one associated with each host species. Therefore to relate the proposed host adaptations, discovered from the sequenced isolates, to the host distributions inferred from isolation statistics, it is essential to understand the phylogenetic relationship of the isolates that make up these statistics.

The phylogenetic reconstruction showed that the *S*. Mbandaka isolates were clonal. As was shown in section 2.3.4, the clonality of strains for MLST loci does not necessarily reflect the same genome composition, but did signify a much lower degree of variance than observed between isolates of different serovars and MLST types. *S*. Derby was more diverse and consisted of five distinct genovars, forming two lineages, with two intermediary isolates. Here we have found that the serovar level of classification can represent several genovars or a single clonal expansion. With regards to *S*. Mbandaka we can discuss the possibility of a host adapted serovar, as it is formed of one clonal lineage, whereas with *S*. Derby we find that lineage 2 (D4, D12 and D14) is only isolated from turkeys, whereas lineage 1 (D1, D2, D3, D5, D6, D10, D11 and D13), separated by 39

SNPs across 3336bp, is isolated from both pigs and turkeys, therefore we potentially have identified a host restricted genovar (lineage 2) and a host adapted genovar (lineage 1). This hypothesis is further strengthened by the results of genotypic and phenotypic assays. Though further MLST typing of *S*. Derby isolates from pigs and turkeys in the UK is needed to give a better perspective on how prevalent these two lineages are, and if lineage 2 is exclusively isolated from turkeys.

It was shown in section 6.4.2, using PCR, that SPI-23 was completely absent from lineage 2 isolates D4 and D14, whereas the lineage 2 isolate D12 and the non-clustering isolate D8 lacked an estimated 24 genes, found at the 5' end of the island. All of these isolates associated to IPEC-J2 monolayers in significantly lower numbers, with the exception of isolate D4, than lineage 1 isolates D1, D2 and D9. Lineage 1 isolates D1 and D2, but not D9 invaded the monolayers in greater numbers than all lineage 2 isolates and *S*. Mbandaka isolates M1 and M2. This strengthens the hypothesis that *S*. Derby in the UK consists of two distinct lineages, lineage 1 is better at invading porcine cells then lineage 2. As we have seen throughout this thesis, the proficiency of invasion is unlikely to be the sole determinant of host range, and several other hypotheses have been constructed that posit a role for metabolism and environmental persistence. Also, it is clear that the numbers of isolates used in the study was modest; but the isolates were selected to give the greatest possible breadth of coverage by host, geography and time. As such the data presented and deductions made do require further interrogation by wider population studies. However, this work has produced readily testable hypotheses regarding likely population structures in these two serovars.

The chromosome sequence of a lineage 2 isolate has not been presented in this thesis, and therefore it is not possible to comment on the chromosome sequence outside of SPI-23 or SPI-1 gap 1 and 2. To address this I have acquired a full genome sequence for the isolate D4. This isolate lacks the SPI-23 locus, and the SPI-1 additional regions described in sections 2.3.10.10 and 2.3.10.2 respectively. In addition to the lower degree of invasion of IPEC-J2 monolayers, isolate D4, was shown to possess the ability to form a biofilm at 25$^{o}$C which is absent from the sequenced *S*. Derby strains. The new sequences will undergo the same analysis as described in chapter 2, and will be compared with *S*. Derby D1 and D2 genome sequences. This will potentially lead to the identification of new host adaptations within the polyphyletic serovar *S*. Derby which was shown to consist of at least two distinct genovars. Due to time constraints this work is beyond the scope of this already very extensive PhD thesis.

There are several interesting questions that arise from the polyphyletic nature of *S*. Derby. Some of these for instance are "would infection of a pig with a lineage 2 isolate be symptomatic?" and "would a lineage 2 isolate lacking SPI-23 colonise or just pass through?" also "what is the significance of the partial SPI-23 in some isolates?" as well as "could regulation on the shared pathways differ between *S*. Derby lineages effecting what may be inferred from the genome sequences as potential core phenotypes?" to name just a few.

In chapters 4 and 5 the partitioned niche was defined by metabolic proficiency of the serovars on single and complex mixes of metabolites. In addition to this I have also shown the greater level of invasion and association of porcine cells by *S*. Derby lineage 1 isolates

compared to *S*. Mbandaka isolates in section 3.4.1 and 6.4.5. The hypothesis of *S*. Mbandaka being adapted to feed in the external environment, is developed further in chapter 6, to suggest general adaptation to the external environment. This is suggested through the observation that with the exception of one isolate, *S*. Mbandaka, at the serovar level of classification, formed a biofilm at 25$^{o}$C, whereas none of the *S*. Derby lineage 1 isolates did. This structure has been shown to increase resistance to desiccation and antimicrobials, including disinfectants (Joseph et al. 2001). In addition to biofilm formation, *S*. Mbandaka isolates also displayed a greater degree of sensitivity to antimicrobial compounds when compared to *S*. Derby isolates, which could reflect a weaker selection pressure presented by antimicrobials in the host, as they are fitter in the secondary niche of the external environment.

The majority of *S*. Derby isolates were resistant to tetracycline which could reflect adaptation to pathogenesis within a porcine host where there is a high usage of this particular antimicrobial. The lineage 1 isolates also respired to a significantly greater degree on porcine colon then either lineage 2 or *S*. Mbandaka isolates. The higher invasion rate of *S*. Derby lineage 1 may reflect an adaptation that has developed in unison with the ability to utilise a distinct set of metabolites from the host intestine, while resisting high levels of tetracycline, and mercury found in pig feed.

## 7.8 Conclusion

The data generated clearly defines differences between *S*. Derby and *S.* Mbandaka at a number of levels that can be interpreted, quite compellingly, when considering the limited niches that have been considered. These interpretations need to be challenged but, within the confines of a PhD study and without access to unlimited resources, proper experimental challenge has not been possible. Rather I have established some compelling hypotheses that need exploration. For example, I have not undertaken animal challenge or environmental survival studies with representatives of the two serovars and three genovars that I describe. By way of extension, it would be interesting to study the *in vivo* behaviour and gene response of the three serovars in several species at two sites in the gut at least. IVOC transcription studies are not completely representative of the host and very expensive to perform. Given the limitations of the studies and perhaps their selectivity, I suggest this overall approach can still be used to explore *S. enterica* and its many variants as a means of exploring how divergence has been influenced by environment. What is intriguing is that serotyping has stood the test of time and has been a useful tool in starting a description of difference. To have such useful markers is helpful, but to understand why an organism is found to have particular associations/niches requires deeper functional genomics, and here I argue also phenomic approaches. Indeed, for organisms that lack these robust measures of difference, such as *Campylobacter*, such comprehensive approaches are the only way forward. MLST has gone some way to suggesting *Campylobacter* has clusters that have preferential host associations. The different conditions defining the habitable niche need to be explored; this thesis has gone some way to indicating the type of questions and approaches that do yield data leading to convincing interpretation regarding host association.

# References

AbuOun, M. et al., 2009. Genome scale reconstruction of a *Salmonella* metabolic model: comparison of similarity and differences with a commensal *Escherichia coli* strain. *The Journal of biological chemistry*, 284(43), pp.29480–29488.

Achtman, M. et al., 2012. Multilocus sequence typing as a replacement for serotyping in *Salmonella enterica*. *PLoS pathogens*, 8(6), p.e1002776.

Aguilera, A. & Benitez, T., 1988. Relationship between growth, fermentation, and respiration rates in *Saccharomyces cerevisiae*: a study based on the analysis of the yield Ypx. *Biotechnology and bioengineering*, 32, pp.240–244.

Ahmer, B.M., Tran, M. & Heffron, F., 1999. The virulence plasmid of *Salmonella* typhimurium is self-transmissible. *Journal of bacteriology*, 181(4), pp.1364–1368.

AHVLA, D., 2011. *Salmonella in livestock production, annual report*, Available at: http://vla.defra.gov.uk/reports/rep_salm_rep11.htm.

Alexander, J. et al., 2008. Mercury as undesirable substance in animal feed 1 scientific opinion of the panel on contaminants in the food chain adopted on 20 February 2008. *The EFSA journal*, 654, pp.1–76.

Altschul, S.F. et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), pp.3389–3402.

Amavisit, P. et al., 2003. Variation between Pathogenic Serovars within *Salmonella* Pathogenicity Islands. *The journal of bacteriology*, 185(12), pp.3624–3635.

Andreoletti, O. et al., 2007. Health risks of feeding of ruminants with fishmeal in relation to the risk to TSE. *The EFSA*, 443, pp.1–26.

Anjum, M.F. et al., 2005. Identification of core and variable components of the *Salmonella enterica* subspecies I genome by microarray. *Infection and immunity*, 73(12), pp.7894–7905.

Aragon, V., 2011. India - Pork production representing only 7% of the country's animal protein sources. *pig333*. Available at: http://www.pig333.com/latest_swine_news/india-pork-production-representing-only-7-of-the-country's-animal-p_4250/.

Arricau, N. et al., 1998. The RcsB-RcsC regulatory system of *Salmonella* Typhi differentially modulates the expression of invasion proteins, flagellin and Vi antigen in response to osmolarity. *Molecular microbiology*, 29(3), pp.835–850.

Aschtgen, M.-S.M.-S. et al., 2008. SciN is an outer membrane lipoprotein required for type VI secretion in enteroaggregative *Escherichia coli*. *Journal of bacteriology*, 190(22), pp.7523–7531.

Athale, A., Sujatha, S. & Himatsin, A., 2010. Consumption of chicken set to double by 2014. *The Economic Times*. Available at: http://articles.economictimes.indiatimes.com/2010-06-08/news/27598670_1_poultry-industry-chicken-consumption-b-soundararajan.

Azelis, 2013. Azelis animal nutrition. Available at: http://www.azelis.com/en/markets/animal-nutrition/products/feed-additives/feed-additives-nutritional/.

Aziz, R.K. et al., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9, pp.75–90.

Barbe, V. et al., 2004. Unique features revealed by the genome sequence of *Acinetobacter sp*. ADP1, a versatile and naturally transformation competent bacterium. *Nucleic acids research*, 32(19), pp.5766–5779.

Barkay, T., Miller, S.M. & Summers, A.O., 2003. Bacterial mercury resistance from atoms to ecosystems. *FEMS Microbiology Reviews*, 27, pp.355–384.

Barrick, J.E. et al., 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461(7268), pp.1243–1247.

Baumler, A.J. et al., 1998. Evolution of host adaptation in *Salmonella enterica*. *Infection and immunity*, 66(10), pp.4579–4587.

Becker D. 2006. Robust *Salmonella* metabolism limits possibilities for new antimicrobials. *Nature*. 440(7082), pp.303-307.

Bellanti, J.A., 1963. The development of the immune response: studies of the agglutinin response to *Salmonella* flagerllar antigens in the newborn rabbit. *Journal of experimental medicine*, 117(3), pp.479–496.

Betancor, L. et al., 2010. Prevalence of *Salmonella enterica* in poultry and eggs in Uruguay during an epidemic due to *Salmonella enterica* serovar Enteritidis. *Journal of clinical microbiology*, 48(7), pp.2413–2423.

Biolog, 2011. Omnilog Temperature Stability. *Biolog Manual*.

Bishop, A. et al., 2008. Interaction of *Salmonella enterica* serovar Typhi with cultured epithelial cells: roles of surface structures in adhesion and invasion. *Microbiology*, 154(7), pp.1914–1926.

Blanc-Potard, A.-B. et al., 1999. The SPI-3 pathogenicity island of *Salmonella enterica*. *The journal of bacteriology*, 181(3), pp.998–1004.

Blanc-Potard, A.B. & Groisman, E.A., 1997. The *Salmonella selC* locus contains a pathogenicity island mediating intramacrophage survival. *The EMBO journal*, 16(17), pp.5376–5385.

Blaser, M.J. & Newman, L.S., 1982. A review of human Salmonellosis: I. infective dose. *Reviews of infectious diseases*, 4(6), pp.1096–1106.

Blondel, C. et al., 2009. Comparative genomic analysis uncovers 3 novel loci encoding type six secretion systems differentially distributed in *Salmonella* serotypes. *BMC Genomics*, 10(1), pp.354–371.

Bochner, B.R., 2009. Global phenotypic characterization of bacteria. *FEMS microbiology reviews*, 33(1), pp.191–205.

Bochner, B.R., Gadzinski, P. & Panomitros, E., 2001. Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. *Genome research*, 11(7), pp.1246–1255.

Bohez, L. et al., 2006. *Salmonella enterica* serovar Enteritidis colonization of the chicken caecum requires the HilA regulatory protein. *Veterinary microbiology*, 116, pp.202–210.

Boyd, E.F., Carpenter, M.R. & Chowdhury, N., 2012. Mobile effector proteins on phage genomes. *Bacteriophage*, 2(3), pp.139–148.

Brombacher, E. et al., 2006. Gene expression regulation by the curli Activator CsgD protein: modulation of cellulose biosynthesis and control of negative determinants for microbial adhesion. *Journal of bacteriology*, 188(6), pp.2027–2037.

Brown, N.F. et al., 2011. *Salmonella* phage ST64B encodes a member of the SseK/NleB effector family. *PloS one*, 6(3), p.e17824.

Buchwald, D.S. & Blaser, M.J., 1984. A review of human salmonellosis: II. duration of excretion following infection with non-Typhi *Salmonella*. *Reviews of infectious diseases*, 6(3), pp.345–356.

Buddruhs, N., Fiebig, A. & Goeker, M.M., 2012. Package “ opm .”

Caiazza, N.C. et al., 2007. Inverse regulation of biofilm formation and swarming motility by *Pseudomonas aeruginosa* PA14. *Journal of bacteriology*, 189(9), pp.3603–3612.

Canchaya, C. et al., 2003. Prophage genomics. *Microbiology and molecular biology reviews*, 67(2), pp.238–276.

Carlson, S.A. et al., 2000. Identification of diminished tissue culture invasiveness among multiple antibiotic resistant *Salmonella* Typhimurium DT104. *Microbial pathogenesis*, 28(1), pp.37–44.

Carver, T.J. et al., 2005. ACT: the Artemis Comparison Tool. *Bioinformatics (Oxford, England)*, 21(16), pp.3422–3423.

CDC PHLIS, 2009. *Salmonella annual survey*, Available at: http://www.cdc.gov/ncidod/dbmd/phlisdata/*Salmonella*.htm.

CDC, 2011. National enteric disease surveillance : *Salmonella* surveillance overview

Chan, K. et al., 2003. Genomic comparison of *Salmonella enterica* serovars and *Salmonella bongori* by use of an *S. enterica* serovar Typhimurium DNA microarray. *Journal of bacteriology*, 185(2), pp.553–563.

Chang, D.-E., Smalley, D.J. & Conway, T., 2002. Gene expression profiling of *Escherichia coli* growth transitions: an expanded stringent response model. *Molecular microbiology*, 45(2), pp.289–306.

Charif, D. & Lobry, J., 2007. SeqinR 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. *Structural Approaches to Sequence Evolution*. Springer Berlin Heidelberg, pp. 207–232.

Chindelevitch, L. et al., 2012. MetaMerge: scaling up genome-scale metabolic reconstructions with applications to *Mycobacterium tuberculosis. Genome Biology*, 13, pp.1-13.

Chiu, C.H. et al., 2005. The genome sequence of *Salmonella enterica* serovar Choleraesuis, a highly invasive and resistant zoonotic pathogen. *Nucleic acids research*, 33(5), pp.1690–1698.

Clokie, M.R. et al., 2011. Phages in nature. *Bacteriophage*, 1(1), pp.31–45.

Collins, J.W. et al., 2010. Response of porcine intestinal *in vitro* organ culture tissues following exposure to *Lactobacillus plantarum* JC1 and *Salmonella enterica* serovar Typhimurium SL1344. *Applied and environmental microbiology*, 76(19), pp.6645–6657.

Collinson, S.K. et al., 1993. Thin, aggregative fimbriae mediate binding of *Salmonella* Enteritidis to fibronectin. *Journal of bacteriology*, 175(1), pp.12–18.

Corpet, D.E., 1986. Ecological factors influencing the transfer of plasmids *in vitro* and *in vivo*. *Journal of antimicrobial chemotherapy*, 18, pp.127–132.

Couturier, E. & Rocha, E.P.C., 2006. Replication-associated gene dosage effects shape the genomes of fast-growing bacteria but only for transcription and translation genes. *Molecular microbiology*, 59(5), pp.1506–1518.

Csáky, I. & Fekete, S., 2004. Soybean: feed quality and safety. Part 1: biologically active components. A review. *Acta veterinaria Hungarica*, 52(3), pp.299–313.

Darling, A.C.E. et al., 2004. Mauve: multiple alignment of conserved genomic sequence With rearrangements. *Genome research*, 14(7), pp.1394–1403.

Davis, M., 2003. Feedstuffs as a vehicle of cattle exposure to *Escherichia coli* O157:H7 and *Salmonella enterica*. *Veterinary microbiology*, 95(3), pp.199–210.

Dean, P., 2011. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS microbiology reviews*, 35(6), pp.1100–1125.

Deiwick, J. et al., 1998. Mutations in *Salmonella* pathogenicity island 2 (SPI2) genes affecting transcription of SPI1 genes and resistance to antimicrobial agents. *The journal of bacteriology*, 180(18), pp.4775–4780.

Delcher, A.L. et al., 2007. Identifying bacterial genes and endosymbiont DNA with Glimmer . *Bioinformatics*, 23(6), p.673-679.

Duerkop, B.A. et al., 2012. A composite bacteriophage alters colonization by an intestinal commensal bacterium. *PNAS*, 109(43), pp.17621–17626.

Edwards, J.S., Ibarra, R.U. & Palsson, B.O., 2001. *In silico* predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data . *Nature biotechnology*, 19(2), pp.125–130.

Elasri, M.O. & Miller, R. V., 1999. Study of the response of a biofilm bacterial community to uv radiation. *Applied envionmental microbiology*, 65(5), pp.2025–2031.

Elsinghorst, E.A., Baron, L.S. & Kopecko, D.J., 1989. Penetration of human intestinal epithelial cells by *Salmonella*: molecular cloning and expression of *Salmonella* Typhi invasion determinants in *Escherichia coli*. *PNAS*, 86(13), pp.5173–5177.

ESRI 2011. ArcGIS Desktop: Release 10. Redlands, CA: Environmental Systems Research Institute.

Eswarappa, S.M. et al., 2009. Host-specificity of *Salmonella enterica* serovar Gallinarum: insights from comparative genomics. *Infection, genetics and evolution*, 9(4), pp.468–473.

Figueroa-Bossi, N. et al., 2001. Variable assortment of prophages provides a transferable repertoire of pathogenic determinants in *Salmonella*. *Molecular microbiology*, 39(2), pp.260–272.

Folkesson, a et al., 1999. Multiple insertions of fimbrial operons correlate with the evolution of *Salmonella* serovars responsible for human disease. *Molecular microbiology*, 33(3), pp.612–622.

Food Standards Agency, 2013. What farm animals eat. Available at: http://www.food.gov.uk/business-industry/farmingfood/animalfeed/what-farm-animals-eat/#.UT5mCBwUx2M [Accessed March 10, 2013].

Fookes, M. et al., 2011. *Salmonella bongori* provides insights into the evolution of the *Salmonella*e. *PLoS pathogens*, 7(8), p.e1002191.

Fordtran, J.S. & Locklear, T.W., 1966. Ionic constituents and osmolality of gastric and small-intestinal fluids after eating . *The American journal of digestive diseases*, 11(7), p.503-521.

Foster, J.T. et al., 2009. Whole-genome-based phylogeny and divergence of the genus *Brucella*. *Journal of bacteriology*, 191(8), pp.2864–2870.

Frenzen, P. et al., 1999. *Salmonella* cost estimate updated using FoodNet data. *Food safety*, 22(2), p.10.

Fricke, W.F. et al., 2011. Comparative genomics of 28 *Salmonella enterica* isolates: evidence for CRISPR-mediated adaptive sublineage evolution. *Journal of bacteriology*, 193(14), pp.3556–3568.

Fuentes, J.A. et al., 2009. RpoS- and Crp-dependent transcriptional control of *Salmonella* Typhi *taiA* and *hlyE* genes: role of environmental conditions. *Research in microbiology*, 160(10), pp.800–808.

Fukushima, M., Kakinuma, K. & Kawaguchi, R., 2002. Phylogenetic analysis of *Salmonella*, *Shigella*, and *Escherichia coli* strains on the basis of the *gyrB* gene sequence. *Journal of Clinical Microbiology*, 40(8), pp.2779–2785.

Galyov, E.E. 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(5), pp.903–912.

Gevorgyan, A. et al., 2011. SurreyFBA: a command line tool and graphics user interface for constraint-based modeling of genome-scale metabolic reaction networks. *Bioinformatics (Oxford, England)*, 27(3), pp.433–434.

Ghosh, S. et al., 2011. An adhesion protein of *Salmonella enterica* serovar Typhi is required for pathogenesis and potential target for vaccine development. *PNAS*, 108(8), pp.3348–3353.

Gilbert, P., Collier, P.J. & Brown, M.R.W., 1990. Minireview: Influence of growth rate on susceptibility to antimicrobial agents : biofilms , cell cycle , dormancy, and stringent response. *Antimicrobial agents and chemotherapy*, 34(10), pp.1865-1868.

Giuffra, E. et al., 2000. The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, 154(4), pp.1785–1791.

Goldstein, F.W. et al., 1986. Plasmid-mediated resistance to multiple antibiotics in *Salmonella* Typhi. *The Journal of infectious diseases*, 153(2), pp.261–266.

Goodall, E.D. & Kay, R.N., 1965. Digestion and absorption in the large intestine of the sheep. *The journal of physiology*, 176, pp.12–23.

Google, 2013. Google Scholar. Available at: http://scholar.google.co.uk/schhp?as_sdt=0,5.

Green, B.T. & Brown, D.R., 2006. Differential effects of clathrin and actin inhibitors on internalization of *Escherichia coli* and *Salmonella* Choleraesuis in porcine jejunal Peyer's patches. *Veterinary microbiology*, 113(1-2), pp.117–122.

Grimont, P., Weill, F. & Grimont and Weill, 2007. *Antigenic formulae of the Salmonella serovars*, Available at: http://www.scacm.org/free/Antigenic Formulae of the *Salmonella* Serovars 2007 9th edition.pdf

Grissa, I., Vergnaud, G. & Pourcel, C., 2007a. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research*, 35, pp.52–57.

Grissa, I., Vergnaud, G. & Pourcel, C., 2007b. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC bioinformatics*, 8(1), pp.172–182.

Groisman, E. a & Ochman, H., 1997. How *Salmonella* became a pathogen. *Trends in microbiology*, 5(9), pp.343–349.

Groisman, E.A. et al., 1993. Molecular, functional, and evolutionary analysis of sequences specific to *Salmonella*. *PNAS*, 90(3), pp.1033–1037.

Guilloteau, P. et al., 1997. Gut regulatory peptides in young cattle and sheep. *Journal of veterinary medicine series A*, 44(1-10), pp.1–23.

Gunn, J.S. et al., 1995. Characterization of the *Salmonella* Typhimurium *pagC*/*pagD* chromosomal region. *The Journal of bacteriology*, 177(17), pp.5040–5047.

Hald, T. & Wingstrand, A., 2006. Human health impact of *Salmonella* contamination in imported soybean products: a semiquantitative risk assessment. *Foodborne pathogens & disease*, 3(4), pp.422–431.

Hammer, B.K. & Bassler, B.L., 2003. Quorum sensing controls biofilm formation in *Vibrio cholerae*. *Molecular microbiology*, 50(1), pp.101–104.

Han, X. et al., 2010. *Escherichia coli* genes that reduce the lethal effects of stress. *BMC microbiology*, 10(1), pp.35–44.

Hao, A.J. et al., 2012. Package “ batman .”

Hapfelmeier, S. et al., 2004. Role of the *Salmonella* Pathogenicity Island 1 effector proteins SipA , SopB , SopE , and SopE2 in *Salmonella enterica* subspecies 1 serovar Typhimurium colitis in streptomycin-pretreated mice. *Immunity and infection*, 72(2), pp.795–809.

Harris, J.B. et al., 2006. Identification of i*n vivo*-induced bacterial protein antigens during human infection with *Salmonella enterica* serovar Typhi. *Infection and immunity*, 74(9), pp.5161–5168.

Hayashi, F. et al., 2001. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*, 410(6832), pp.1099–1103.

Hayward, M.R., Jansen, V.A. & Woodward, M.J., 2013. Comparative genomics of *Salmonella enterica* serovars Derby and Mbandaka, two prevalent serovars associated with different livestock species in the UK. *BMC Genomics*, 14(1), p.365-384.

Henry, C.S. et al., 2010. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature biotechnology*, 28(9), pp.977–982.

Hensel, M., 2004. Evolution of pathogenicity islands of *Salmonella enterica*. *International journal of medical microbiology* , 294, pp.95–102.

Hibbing, M.E. et al., 2010. Bacterial competition: surviving and thriving in the microbial jungle. *Nature reviews microbiology*, 8(1), pp.15–25.

Hill, N.S. et al., 2012. Cell size and the initiation of DNA replication in bacteria. *PLoS genetics*, 8(3), p.e1002549.

Hoben, D. a, Ashton, D.H. & Peterson, a C., 1973. Some observations on the incorporation of novobiocin into Hektoen enteric agar for improved *Salmonella* isolation. *Applied microbiology*, 26(1), pp.126–127.

Hone, D. M. et al. 1988. A *galE via* (Vi antigen-negative) mutant of *Salmonella* Typhi Ty2 retains virulence in humans. *Infection and Immunity*. 56, pp.1326-1333.

Horvath, D. & Noot, G. Vander, 1954. Effect of three levels of a new antibiotic, tetracycline, in a swine ration. *Journal of animal science*, 13, pp.899–903.

HPA., 2004. DoubleACT. Available at: http://www.hpa-bioinfotools.org.uk/pise/double_act.html.

Hu, Q. et al., 2008. *Salmonella enterica* serovar Senftenberg human clinical isolates lacking SPI-1. *Journal of clinical microbiology*, 46(4), pp.1330–1336.

Hurme, R. et al., 1997. A proteinaceous gene regulatory thermometer in *Salmonella*. *Cell*, 90(1), pp.55–64.

Iino, T. & Enomoto, M., 1966. Genetical studies of non-flagellate mutants of *Salmonella*. *Journal of general microbiology*, 43(3), pp.315–327.

Izumiya, H. et al., 2010. Whole-genome analysis of *Salmonella enterica* serovar Typhimurium T000240 reveals the acquisition of a genomic island involved in multidrug resistance via IS1 derivatives on the chromosome. *Antimicrobial agents and chemotherapy,* 55(2), pp.623-630.

Jarvik, T. et al., 2010. Short-term signatures of evolutionary change in the *Salmonella enterica* serovar Typhimurium 14028 genome. *Journal of bacteriology*, 192(2), pp.560–567.

Johnston, C. et al., 1996. Transcriptional activation of *Salmonella* Typhimurium invasion genes by a member of the phosphorylated response-regulator superfamily . *Molecular microbiology*, 22(4), pp.715–727.

Jones, B.D., 1994. *Salmonella* Typhimurium initiates murine infection by penetrating and destroying the specialized epithelial M cells of the Peyer's patches . *Journal of experimental medicine*, 180(1), p.15-23.

Jones, M.A. et al., 2001. *Salmonella enterica* serovar Gallinarum requires the *Salmonella* pathogenicity island 2 type III secretion system but not the *Salmonella* pathogenicity island 1 type III secretion system for virulence in chickens *Salmonella enterica* serovar Gallinarum . *Infection and immunity*, 69(9), 5471-5476.

Jones, P.W. et al., 1982. Transmission of *Salmonella* Mbandaka to cattle from contaminated feed. *The journal of hygiene*, 88(2), pp.255–263.

Jormakka, M., Byrne, B. & Iwata, S., 2003. Formate dehydrogenase – a versatile enzyme in changing environments. *Current opinion in structural biology*, 13(4), pp.418–423.

Joseph, B. et al., 2001. Biofilm formation by *Salmonella* spp. on food contact surfaces and their sensitivity to sanitizers. *International journal of food microbiology*, 64(3), pp.367–372.

Joshi, T. et al., 2010. SoyMetDB: the soybean metabolome database. In *In Proceedings of IEEE International Conference on Bioinformatics & Biomedicine*. Hong Kong, pp. 203–208.

Kado, C.I. & Liu, S.T., 1981. Rapid procedure for detection and isolation of large and small plasmids. *Journal of bacteriology*, 145(3), pp.1365–1373.

Kahm, M.M., 2012. Package " grofit ."

Kingsley, R. & Bäumler, J., 2000. Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. *Molecular microbiology*, 36(5), pp.1006–1014.

Kingsley, R.A. et al., 2003. Molecular and phenotypic analysis of the CS54 island of *Salmonella enterica* serotype Typhimurium: identification of intestinal colonization and persistence determinants. *Infection and immunity*, 71(2), pp.629–640.

Klumpp, J. & Fuchs, T.M., 2007. Identification of novel genes in genomic islands that contribute to *Salmonella* Typhimurium replication in macrophages. *Microbiology*, 153(4), pp.1207–1220.

Knodler, L.A. et al., 2002. *Salmonella* effectors within a single pathogenicity island are differentially expressed and translocated by separate type III secretion systems. *Molecular microbiology*, 43(5), pp.1089–1103.

Kraatz, M., 2011. Isolation of lactic acid-related bacteria from the pig mucosal proximal gastrointestinal tract, including *Olsenella Umbonata Sp. Nov.* and *Veillonella Magna Sp. Nov.* Berlin: Logos Verlag Berlin GmbH.

Krzywinski, M. et al., 2009. Circos: an information aesthetic for comparative genomics. *Genome research*, 19(9), pp.1639–1645.

Lee, C.A. & Falkow, S., 1990. The ability of *Salmonella* to enter mammalian cells is affected by bacterial growth state. *PNAS*, 87(11), pp.4304–4308.

Lee, D, et al. 2012. Improving metabolic flux predictions using absolute gene expression data. *BMC Systems Biology*, 6, 73.

Lenski, R.E., Winkworth, C.L. & Riley, M.A., 2003. Rates of DNA sequence evolution in experimental populations of *Escherichia coli* during 20,000 generations. *Journal of molecular evolution*, 56(4), pp.498–508.

Lianou, A. & Koutsoumanis, K.P., 2012. Strain variability of the biofilm-forming ability of *Salmonella enterica* under various environmental conditions. *International journal of food microbiology*, 160(2), pp.171–178.

Librado, P. & Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics (Oxford, England)*, 25(11), pp.1451–1452.

Liu, S.L. & Sanderson, K.E., 1996. Highly plastic chromosomal organization in *Salmonella* Typhi. *PNAS*, 93(19), pp.10303–10308.

Looney, RJ and Steigbigel, RT. 1986. Role of the Vi antigen of *Salmonella* Typhi in resistance to host defense *in vitro*. *Journal of Laboratory and Clinical Medicine*. 108(5), pp.506-516.

López-Garrido, J. & Casadesús, J., 2010. Regulation of *Salmonella enterica* pathogenicity island 1 by DNA adenine methylation. *Genetics*, 184(3), pp.637–649.

Lu, S., Killoran, P.B. & Riley, L.W., 2003. Association of *Salmonella enterica* serovar Enteritidis YafD with resistance to chicken egg albumen. *Infection and immunity*, 71(12), pp.6734–6741.

Manning, P.A. et al., 1995. Putative O-antigen transport genes within the *rfb* region of *Vibrio cholerae* O1 are homologous to those for capsule transport. *Gene*, 158(1), pp.1–7.

Mappley, L.J. et al., 2012. Comparative genomics of *Brachyspira pilosicoli* strains: genome rearrangements, reductions and correlation of genetic complement with phenotypic diversity. *BMC genomics*, 13, p.454-472.

Marcus, S.L. et al., 2000. *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes and infection*, 2(2), pp.145–156.

Matsushita, S. et al., 1996. Serovar-distribution and drug-resistance of *Salmonella* strains isolated from domestic and imported cases in 1990-1994 in Tokyo . *Kansenshogaku zasshi.The Journal of the Japanese Association for Infectious Diseases*, 70(1), pp.42–50.

McClelland, M. et al., 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature*, 413(6858), pp.852–856.

McCelland M, et al. 2004. Comparison of genome degradation in Paratyphi A and Typhi, human-restricted serovars of *Salmonella enterica* that cause typhoid. *Nature Genetics.* 36, pp.1268–1274.

McCullough, N.B. & Eisele, C.W., 1951. Experimental human salmonellosis. III. Pathogenicity of strains of *Salmonella* Newport, *Salmonella* Derby, and *Salmonella* Bareilly obtained from spray-dried whole egg. *The journal of infectious diseases*, 89(3), pp.209–213.

McLennan, A.G., 2006. The Nudix hydrolase superfamily. *Cellular and molecular life sciences*, 63(2), pp.123–143.

McMillan, D., 2013. China's soaring demand for beef imports creates opportunities. *The Western Producer*. Available at: http://www.producer.com/2013/06/chinas-soaring-demand-for-beef-imports-creates-opportunities/.

McQuiston, J.R. et al., 2008. Do *Salmonella* carry spare tyres? *Trends in microbiology*, 16(4), pp.142–148.

Mennerat, A. et al., 2010. Intensive farming: evolutionary implications for parasites and pathogens. *Evolutionary biology*, 37, pp.59–67.

Merrifield, C. a et al., 2011. A metabolic system-wide characterisation of the pig: a model for human physiology. *Molecular biosystems*, 7(9), pp.2577–2588.

Merritt, M.E. & Donaldson, J.R., 2009. Effect of bile salts on the DNA and membrane integrity of enteric bacteria. *Journal of medical microbiology*, 58, pp.1533–1541.

Messens, W. et al., 2004. Growth of *Salmonella* serovars in hens' egg albumen as affected by storage prior to inoculation. *Food microbiology*, 21(1), pp.25–32.

Moreno Switt, A.I. et al., 2013. *Salmonella* bacteriophage diversity reflects host diversity on dairy farms. *Food Microbiology,* 36(2), 275-285.

Morgan, E. et al., 2004. Identification of host-specific colonization factors of *Salmonella enterica* serovar Typhimurium. *Molecular microbiology*, 54(4), pp.994–1010.

Morgan, M.C. et al., 2009. Comparison of the Biolog OmniLog Identification System and 16S ribosomal RNA gene sequencing for accuracy in identification of atypical bacteria of clinical origin. *Journal of microbiological methods*, 79(3), pp.336–343.

Morita, T. et al., 2007. Risk of *Salmonella* that invade oilmeal manufacturing plants via imported oilseed. *Japanese journal of animal hygiene*, 32(2), pp.79-85.

Morris, C. et al., 2003. The shufflon of *Salmonella enterica* serovar Typhi regulates type IVB pilus-mediated bacterial self-association. *Infection and immunity*, 71(3), pp.1141–1146.

Mount, L., 1959. The metabolic rate of the new-born pig in relation to environmental temperature and to age. *The journal of physiology*, 147, pp.333–345.

Norris, F.A. et al., 1998. SopB, a protein required for virulence of *Salmonella* Dublin, is an inositol phosphate phosphatase. *PNAS*, 95(24), pp.14057–14059.

O'Callaghan, D. & Charbit, A., 1990. High efficiency transformation of *Salmonella* Typhimurium and *Salmonella* Typhi by electroporation . *MGG Molecular & General Genetics*, 223(1), pp.156-158.

Ochman, H. & Groisman, E.A., 1996. Distribution of pathogenicity islands in *Salmonella* spp. *Infection and immunity*, 64(12), pp.5410–5412.

Ochman, H., Elwyn, S. & Moran, N.A., 1999. Calibrating bacterial evolution. *PNAS*, 96(22), pp.12638–12643.

Ogata, H. et al., 1999. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic acids research*, 27(1), pp.29–34.

Ong, S.Y. et al., 2010. Analysis and construction of pathogenicity island regulatory pathways in *Salmonella enterica* serovar Typhi. *Journal of integrative bioinformatics*, 7(1), pp.145–179.

Oti, M., Huynen, M. a & Brunner, H.G., 2008. Phenome connections. *Trends in genetics*, 24(3), pp.103–106.

Overbeek, R. et al., 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic acids research*, 33(17), pp.5691–5702.

Palsson, B., 2006. *Systems biology: properties of reconstructed networks* 1st ed., Cambridge: Cambridge University Press.

Pan, Z. et al., 2009. Identification of genetic and phenotypic differences associated with prevalent and non-prevalent *Salmonella* Enteritidis phage types: analysis of variation in amino acid transport. *Microbiology (Reading, England)*, 155, pp.3200–13.

Papadopoulou, C. et al., 2009. Retrospective analysis of *Salmonella* isolates recovered from animal feed in Great Britain. *The veterinary record*, 165(23), pp.681–688.

Parkhill, J. et al., 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature*, 413(6858), pp.848–852.

Paulin, SM et al. 2002. Analysis of *Salmonella enterica* serotype-host specificity in calves: avirulence of *S. enterica* serotype gallinarum correlates with bacterial dissemination from mesenteric lymph nodes and persistence *in vivo*. *Infection and Immunity*. 70, pp.6788–6797.

Paulin, SM. et al. 2007. Net replication of *Salmonella enteric* serovars Typhimurium and Choleraesuis in porcine intestinal mucosa and nodes is associated with their differential virulence. *Infection and Immunity*, 75, pp.3950-3960.

Pettersson, D. & Åman, P., 1986. Enzyme supplementation of a poultry diet containing rye and wheat. *British journal of nutrition*, 62, pp.139–149.

Phaniraja, K.L. & Panchasara, H.H., 2009. Indian draught animals power. *Veterinary world*, 2(10), pp.404–407.

Pickard, D. et al., 2003. Composition, acquisition, and distribution of the Vi exopolysaccharide-encoding *Salmonella enterica* pathogenicity island SPI-7. *Journal of bacteriology*, 185(17), pp.5055–5065.

Poon, W.W. et al., 2000. Identification of *Escherichia coli ubiB*, a gene required for the first monooxygenase step in ubiquinone biosynthesis. *The journal of bacteriology*, 182(18), pp.5139–5146.

Porwollik, S., Wong, R.M.-Y. & McClelland, M., 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *PNAS*, 99(13), pp.8956–8961.

Pulkkinen, K. et al., 2010. Intensive fish farming and the evolution of pathogen virulence: the case of columnaris disease in Finland. *Proceedings of the royal society*, 277(1681), pp.593–600.

Puniva, BL. et al. 2013. Integrated gene co-exression network analysis in the growth phase of *Mycobacterium tuberculosis* reveals new potential drug targets. *Molecular BioSystems*, 11, pp. 2798-2815.

Purvis, G.M. et al., 2005. Persistence of *Salmonella* Typhimurium DT120 in abattoir paddocks holding sheep. *Veterinary record*, 157(6), pp.165–167.

Qu, A. et al., 2008. Comparative metagenomics reveals host specific metavirulomes and horizontal gene transfer elements in the chicken cecum microbiome. *PloS one*, 3(8), p.e2945.

R Core Team, 2012. R: A language and environment for statistical computing. Available at: http://www.r-project.org.

Rabsch, W. et al., 2002. *Salmonella enterica* serotype Typhimurium and its host-adapted variants. *Infection and immunity*, 70(5), pp.2249–2255.

Raffatellu, M. et al. 2005. The Vi capsular antigen of *Salmonella enterica* serotype Typhi reduces Toll-like receptor-dependent interleukin-8 expression in the intestinal mucosa. *Infection and Immunity*. 73: pp.3367–3374.

Raghunathan, A. et al., 2009. Constraint-based analysis of metabolic capacity of *Salmonella* Typhimurium during host-pathogen interaction. *BMC systems biology*, 3, p.38-54.

Raina, S. et al., 1993. Identification and transcriptional analysis of the *Escherichia coli htrE* operon which is homologous to pap and related pilin operons. *Journal of bacteriology*, 175(16), pp.5009–5021.

Reid, E., Cole, J. & Eaves, D.J., 2001. The *Escherichia coli* CcmG protein fulfils a specific role in cytochrome c assembly. *Biochemical journal*, 58, pp.51–58.

Ren, Q., Ahuja, U. & Thöny-Meyer, L., 2002. A bacterial cytochrome c heme lyase. CcmF forms a complex with the heme chaperone CcmE and CcmH but not with apocytochrome c. *The Journal of biological chemistry*, 277(10), pp.7657–7663.

Reuven, N.B., Tomer, G. & Livneh, Z., 1998. The mutagenesis proteins UmuD and UmuC prevent lethal frameshifts while increasing base substitution mutations. *Molecular cell*, 2(2), pp.191–199.

RGM, 2012. fragment of *E. coli* chromosome. Available at: http://rgm3.lab.nig.ac.jp/RGM/r_function?p=seqinr&f=m16j.

Rohmer, L., Hocquet, D. & Miller, S.I., 2011. Are pathogenic bacteria just looking for food? Metabolism and microbial pathogenesis. *Trends in microbiology*, 19(7), pp.341–348.

Rolfe, M.D. et al., 2012. Lag phase is a distinct growth phase that prepares bacteria for exponential growth and involves transient metal accumulation. *Journal of bacteriology*, 194(3), pp.686–701.

Röling, W.F.M., Ferrer, M. & Golyshin, P.N., 2010. Systems approaches to microbial communities and their functioning. *Current opinion in biotechnology*, 21(4), pp.532–538.

Ronquist, F. & Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, 19(12), pp.1572–1574.

Rotger, R. & Casadesús, J., 1999. The virulence plasmids of *Salmonella*. *International microbiology*, 2(3), pp.177–184.

Rubin, R.H. & Weinstein, L., 1977. *Salmonellosis : microbiologic, pathologic, and clinical features*, New York: Stratton Intercontinental Medical Book Corp.

Rutherford, K. et al., 2000. Artemis: sequence visualization and annotation. *Bioinformatics*, 16(10), pp.944–945.

Rychlik, I. & Barrow, P. a, 2005. *Salmonella* stress management and its relevance to behaviour during intestinal colonisation and infection. *FEMS microbiology reviews*, 29(5), pp.1021–1040.

Rychlik, I. et al., 2009. Virulence potential of five major pathogenicity islands (SPI-1 to SPI-5) of *Salmonella enterica* serovar Enteritidis for chickens. *BMC microbiology*, 9, p.268-277.

Sambrook, J., Fritsch, E.F. & Maniatis, T., 2002. *Molecular Cloning: A Laboratory Manual* C. S. H. L. Press, ed., New York: Cold Spring Harbor.

Samudrala, R., Heffron, F. & McDermott, J.E., 2009. Accurate prediction of secreted substrates and identification of a conserved putative secretion signal for type III secretion systems. *PLoS pathogens*, 5(4), p.e1000375.

Sanders, C. et al., 2010. Cytochrome c biogenesis: the Ccm system. *Trends in microbiology*, 18(6), pp.266–274.

Santos, R.L. et al., 2009. Life in the inflamed intestine, *Salmonella* style. *Trends in microbiology*, 17(11), pp.498–506.

Saroj, S.D. et al., 2008. Distribution of *Salmonella* pathogenicity island (SPI)-8 and SPI-10 among different serotypes of *Salmonella*. *Journal of medical microbiology*, 57, pp.424–427.

Sbrogio-Almeida, M. & Mosca, T., 2004. Host and bacterial factors affecting induction of immune responses to flagellin expressed by attenuated *Salmonella* vaccine strains. *Infection and immunity*, 72(5), pp.2546–2555.

Schierack, P. et al., 2006. Characterization of a porcine intestinal epithelial cell line for *in vitro* studies of microbial pathogenesis in swine. *Histochemistry and cell biology*, 125(3), pp.293–305.

Schikora, A. et al., 2008. The dark side of the salad: *Salmonella* Typhimurium overcomes the innate immune response of *Arabidopsis thaliana* and shows an endopathogenic lifestyle. *PloS one*, 3(5), p.e2279.

Schikora, A., Garcia, A. V & Hirt, H., 2012. Plants as alternative hosts for *Salmonella*. *Trends in plant science*, 17(5), pp.245–249.

Schmid, M.B. & Roth, J.R., 1987. Gene location affects expression level in *Salmonella* Typhimurium. *Journal of bacteriology*, 169(6), pp.2872–2875.

Searle, L.E.J. et al., 2009. A mixture containing galactooligosaccharide, produced by the enzymic activity of *Bifidobacterium bifidum*, reduces *Salmonella enterica* serovar Typhimurium infection in mice. *Journal of medical microbiology*, 58, pp.37–48.

Seth-Smith, H.M.B., 2008. SPI-7: *Salmonella*'s Vi-Encoding Pathogenicity Island. *The journal of infection in developing countries*, 2(4), pp.267–271.

Shabala, L. et al., 2009. Ion transport and osmotic adjustment in *Escherichia coli* in response to ionic and non-ionic osmotica. *Environmental microbiology*, 11(1), pp.137–148.

Shah, D.H. et al., 2005. Identification of *Salmonella* Gallinarum virulence genes in a chicken infection model using PCR-based signature-tagged mutagenesis. *Microbiology (Reading, England)*, 151, pp.3957–3968.

Shao, F. et al., 2002. A *Yersinia* effector and a *Pseudomonas* avirulence protein define a family of cysteine proteases functioning in bacterial pathogenesis. *Cell*, 109(5), pp.575–588.

Shea, J.E. et al., 1999. Influence of the *Salmonella* Typhimurium Pathogenicity Island 2 Type III Secretion System on Bacterial Growth in the Mouse. *Infection and immunity*, 67(1), pp.213–219.

Sinha, N. & Ferguson, S.J., 1998. An *Escherichia coli ccm* (cytochrome c maturation) deletion strain substantially expresses *Hydrogenobacter thermophilus* cytochrome c552 in the cytoplasm: availability of haem influences cytochrome c552 maturation. *FEMS microbiology letters*, 161(1), pp.1–6.

Smith, H., 1975. Persistence of tetracycline resistance in pig *E. coli*. *Nature*, 258(258), pp.628–629.

Soons, ZTA. et al. 2013. Identification of metabolic engineering targets through analysis of optimal and sub-optimal routes. *PlosOne*, 8, e61648.

Sorek, R., Kunin, V. & Hugenholtz, P., 2008. CRISPR--a widespread system that provides acquired resistance against phages in bacteria and archaea. *Nature reviews.microbiology*, 6(3), pp.181–186.

Spano, S. & Galan, JE. 2011. A Rab32-dependent pathway contributes to *S.* Typi host restriction. *Science*, 338, pp.960-963.

Spano, S. Liu, X. & Galan, JE. 2011. Proteolytic targeting of Rab29 by an effector protein distinguishes the intracellular compartments of human-adapted and broad-host *Salmonella*. *PNAS,* 108, pp.18418-18423.

Steinbach, G., Lauterbach, L. & Methner, U., 2000. Studies of the phenomenon of host adaptation in *Salmonella*. *Journal of veterinary medicine, series B*, 47(9), pp.707–719.

Stevenson, G. et al., 1994. Structure of the O-Antigen of *Escherichia coli* K-12 and the Sequence of It's *rIb* Gene Cluster. *Journal of bacteriology*, 176(13), pp.4144–4156.

Suar, M., Jantsch, J. & Hapfelmeier, S., 2006. Virulence of broad-and narrow-host-range *Salmonella enterica* serovars in the streptomycin-pretreated mouse model. *Infection and immunity*, 74(1), pp.632–644.

Tam, C.K.P., Morris, C. & Hackett, J., 2006. The *Salmonella enterica* serovar Typhi type IVB self-association pili are detached from the bacterial cell by the PilV minor pilus proteins. *Infection and immunity*, 74(9), pp.5414–5418.

Tenor, J.L. et al., 2004. *Caenorhabditis elegans*-based screen identifies *Salmonella* virulence factors required for conserved host-pathogen interactions. *Current biology*, 14(11), pp.1018–1024.

Teplitski, M., Goodier, R.I. & Ahmer, B.M.M., 2003. Pathways leading from BarA/SirA to motility and virulence gene expression in *Salmonella*. *Journal of bacteriology*, 185(24), pp.7257–7265.

The Document Foundation, 2012. OfficeLibre Calc.

Thienimitr, P, et al. 2011. Intestinal inflammation allows *Salmonella* to use ethanolamine to compete with the microbiota. *PNAS*. 108(42), pp.17480-17485.

Thomson, N. et al., 2004. The role of prophage-like elements in the diversity of *Salmonella enterica* serovars. *Journal of molecular biology*, 339(2), pp.279–300.

Thomson, N.R. et al., 2008. Comparative genome analysis of *Salmonella* Enteritidis PT4 and *Salmonella* Gallinarum 287/91 provides insights into evolutionary and host adaptation pathways. *Genome research*, 18(10), pp.1624–1637.

Toguchi, A. et al., 2000. Genetics of swarming motility in *Salmonella enterica* serovar Typhimurium: critical role for lipopolysaccharide. *Journal Of bacteriology*, 182(22), pp.6308–6321.

Torres, G.J. et al., 2011. The prevalence of *Salmonella enterica* in Spanish feed mills and potential feed-related risk factors for contamination. *Preventive veterinary medicine*, 98, pp.81–87.

Tsolis, R.M. et al., 1999. Of mice, calves, and men. Comparison of the mouse typhoid model with other *Salmonella* infections . *Advances in experimental medicine and biology*, 473, pp.261–274.

Uzzau, S. et al., 2000. Host adapted serotypes of *Salmonella enterica*. *Epidemiology and infection*, 125(2), pp.229–255.

Van Cuyck, H. et al., 2011. MLVA polymorphism of *Salmonella enterica* subspecies isolated from humans, animals, and food in Cambodia. *BMC research notes*, 4(1), pp.306–314.

Vernikos, G.S. & Parkhill, J., 2006. Interpolated variable order motifs for identification of horizontally acquired DNA: revisiting the *Salmonella* pathogenicity islands. *Bioinformatics*, 22(18), pp.2196–2203.

Vijayaraghavan, A., 2012. How China's dietary habits are changing world agriculture. *TriplePundit*. Available at: http://www.triplepundit.com/2012/05/feeding-worlds-largest-population-chinas-dietary-habits-changing-world-agriculture/.

Virgilio, R. & Cordano, A.M., 1981. Naturally occurring prototrophic strains of *Salmonella* Typhi . *Canadian journal of microbiology*, 27(12), pp.1272–1275.

Vissa, VD and Brennan, PJ. 2001. The genome of Mycobacterium leprae: a minimal mycobacterial gene set. *Genome Biology*. 2(8), 1-8.

Vladoianu, I.-R., Chang, H.R. & Pechère, J.-C., 1990. Expression of host resistance to *Salmonella* typhi and *Salmonella* Typhimurium: bacterial survival within macrophages of murine and human origin. *Microbial pathogenesis*, 8(2), pp.83–90.

VMD, 2012. *Sales of antimicrobial products auhtorised for use as veterinary medicines in the UK in 2011*. Available at: www.vmd.defra.gov.uk/pdf/salesanti10.pdf

Waterfield, N.R., Daborn, P.J. & Ffrench-Constant, R.H., 2002. Genomic islands in *Photorhabdus*. *Trends in microbiology*, 10(12), pp.541–545.

Waterman, S.R. & Holden, D.W., 2003. Functions and effectors of the *Salmonella* pathogenicity island 2 type III secretion system . *Cellular microbiology*, 5(8), pp.501–511.

Watson, PR. 2000. Interaction of *Salmonella* serotypes with porcine macrophages*in vitro* does not correlate with virulence. *Microbiology*. 146, pp.1639-1649

Westerhoff, H. V et al., 2009. Systems biology: the elements and principles of life. *FEBS letters*, 583(24), pp.3882–3890.

White, a P. et al., 2008. Aggregation via the red, dry, and rough morphotype is not a virulence adaptation in *Salmonella enterica* serovar Typhimurium. *Infection and immunity*, 76(3), pp.1048–1058.

White, A.P. et al., 2006. Thin aggregative fimbriae and cellulose enhance long-term survival and persistence of *Salmonella*. *Journal of bacteriology*, 188(9), pp.3219–3227.

White, A.P. et al., 2010. A global metabolic shift is linked to *Salmonella* multicellular development. *PloS one*, 5(7), p.e11814.

Wierup, M. & Häggblom, P., 2010. An assessment of soybeans and other vegetable proteins as source of *Salmonella* contamination in pig production. *Acta veterinaria Scandinavica*, 52, pp.15-24.

Wildschutte, H. & Lawrence, J.G., 2007. Differential *Salmonella* survival against communities of intestinal amoebae. *Microbiology (Reading, England)*, 153, pp.1781–1789.

Wildschutte, H. et al., 2004. Protozoan predation, diversifying selection, and the evolution of antigenic diversity in *Salmonella*. *PNAS*, 101(29), pp.10644–10649.

Wood, R.L., Pospischil, A. & Rose, R., 1989. Distribution of persistent *Salmonella* typhimurium infection in internal organs of swine. *American journal of veterinary research*, 50(7), pp.1015–1021.

Woodward, M.J. et al., 2000. The role of SEF14 and SEF17 fimbriae in the adherence of *Salmonella enterica* serotype Enteritidis to inanimate surfaces. *Journal of medical microbiology*, 49(5), pp.481–487.

Xiang, S.H., Haase, a M. & Reeves, P.R., 1993. Variation of the *rfb* gene clusters in *Salmonella enterica*. *Journal of bacteriology*, 175(15), pp.4877–4884.

Xu, Z. et al., 2011. Mutation of a novel virulence-related gene *mltD* in *Vibrio anguillarum* enhances lethality in zebra fish. *Research in microbiology*, 162(2), pp.144–150.

Yoon, S.H. et al., 2007. Towards pathogenomics: a web-based resource for pathogenicity islands. *Nucleic acids research*, 35, pp.395–400.

Yoshida, T. et al., 1998. Purification and characterization of thin pili of IncI1 plasmids ColIb-P9 and R64: formation of PilV-specific cell aggregates by type IV pili. *Journal of Bacteriology*, 180(11), pp.2842–2848.

Yoshimura, Y. et al., 2006. Effects of age, egg-laying activity, and *Salmonella*-inoculation on the expressions of gallinacin mRNA in the vagina of the hen oviduct . *Journal of reproduction and development*, 52(2), p.211-218.

Zhou, Y. et al., 2011. PHAST: a fast phage search tool. *Nucleic acids research*, 39, pp.347–352.

**Theodosius Dobzhansky Quote: 1973, NABT, Vol 35, No 3, pp. 125-129.**