1 Introduction

1.1 The genus *Salmonella*

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

1.1.1 Classification and nomenclature

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

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

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

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

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

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

[b] Examples of serotype designations are Salmonella serotype (ser.) Typhimurium, Salmonella II 50:b:z_6, Salmonella IIIb 60:k:z, and Salmonella ser. Marina (IV 48:g,z_{51}:−).

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

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

1.2 Disease syndromes caused by *Salmonella*

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

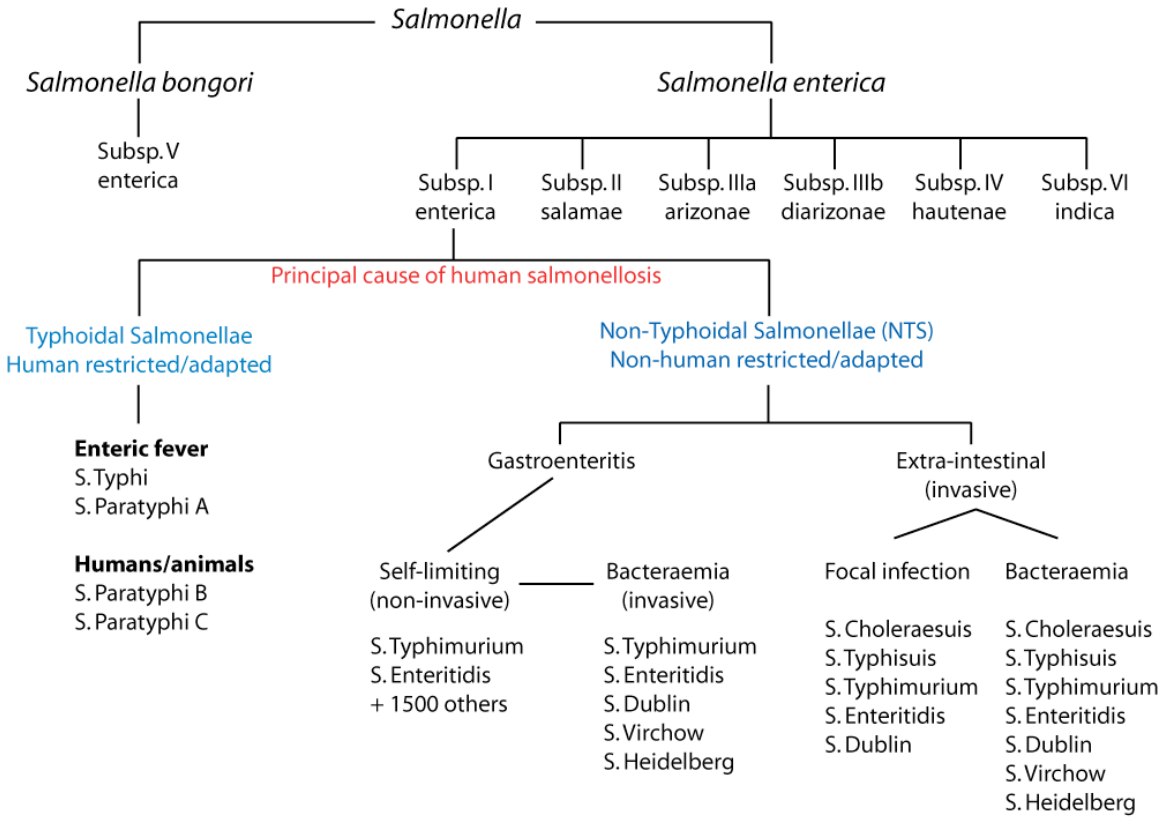


Figure 1.1 Disease syndromes and subspecies [7].

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

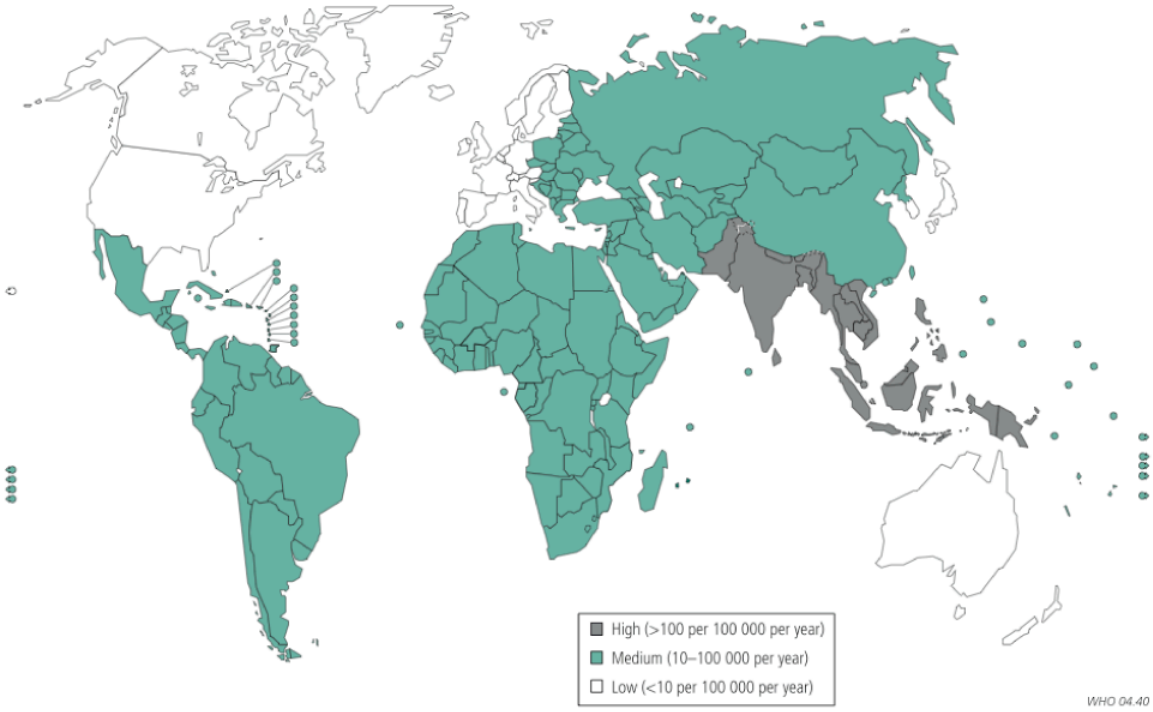


Figure 1.2 WHO estimates of the global burden of typhoid fever and its geographical distribution [8]

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

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

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

1.3 Salmonella pathogenicity

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

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

1.3.1 The biology of *Salmonella* infection

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

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

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

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

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

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

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

1.3.2 Typhoid vaccines

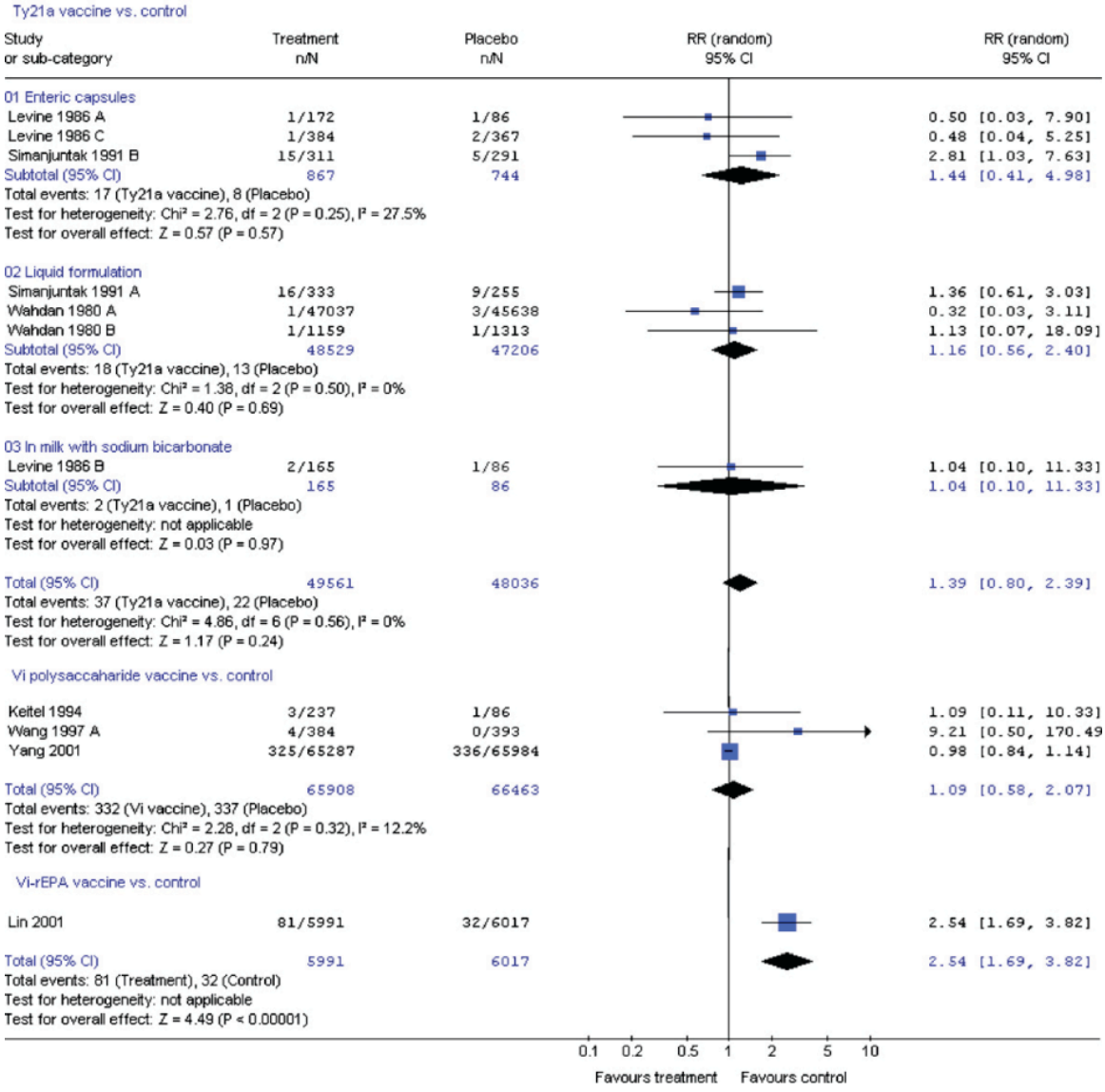
The first typhoid vaccine, introduced in 1896, was based on a whole-cell inactivated *S.* Typhi preparation. The efficacy of the vaccine has been questioned but during controlled trials in 1960 the efficacy after two parenterally administered doses was 73% over a 3-year period [60]. However, 10% of recipients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

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

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

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

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



1.4 The genomics of invasive *Salmonella*

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

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

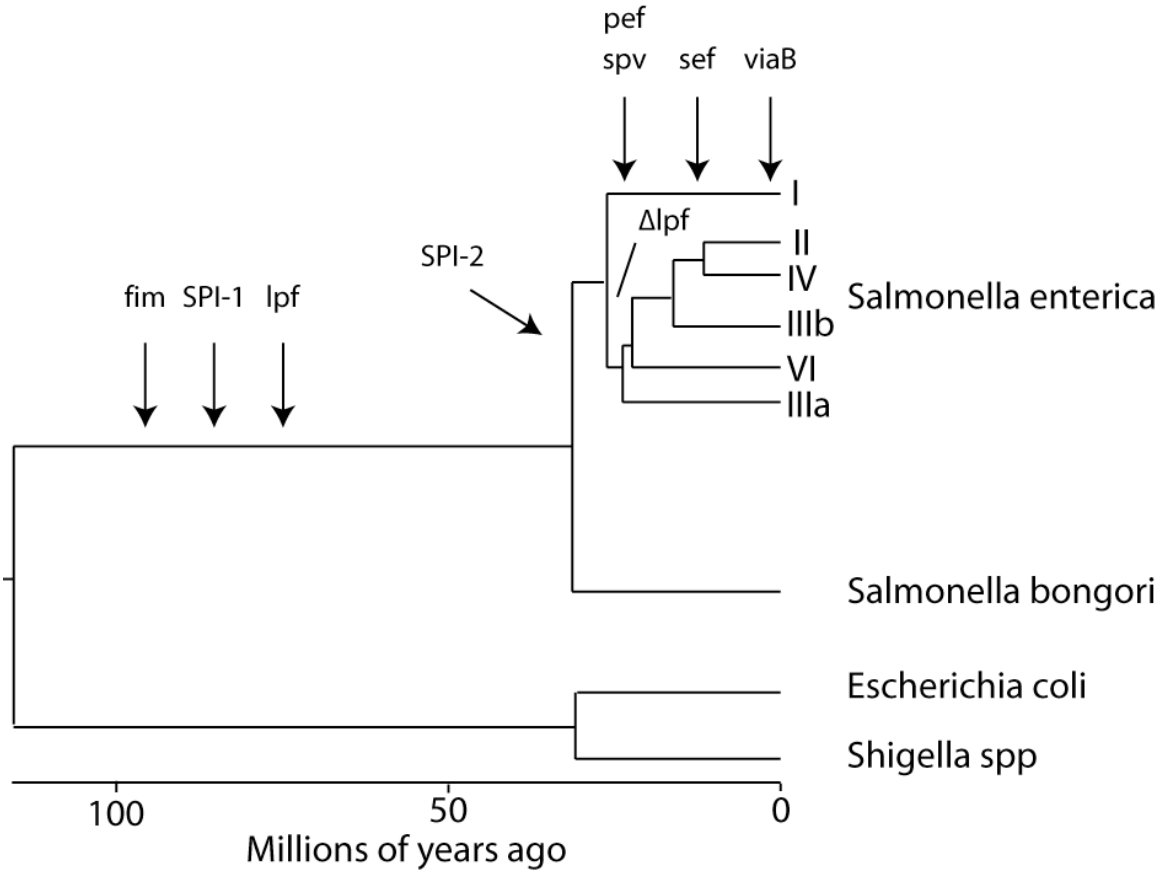


Figure 1.3 The horizontal transfer of virulence-associated genes within the genus *Salmonella* [78].

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

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

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

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

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

Table 1.3 Genes unique to pairs of genomes [82]

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

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

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

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

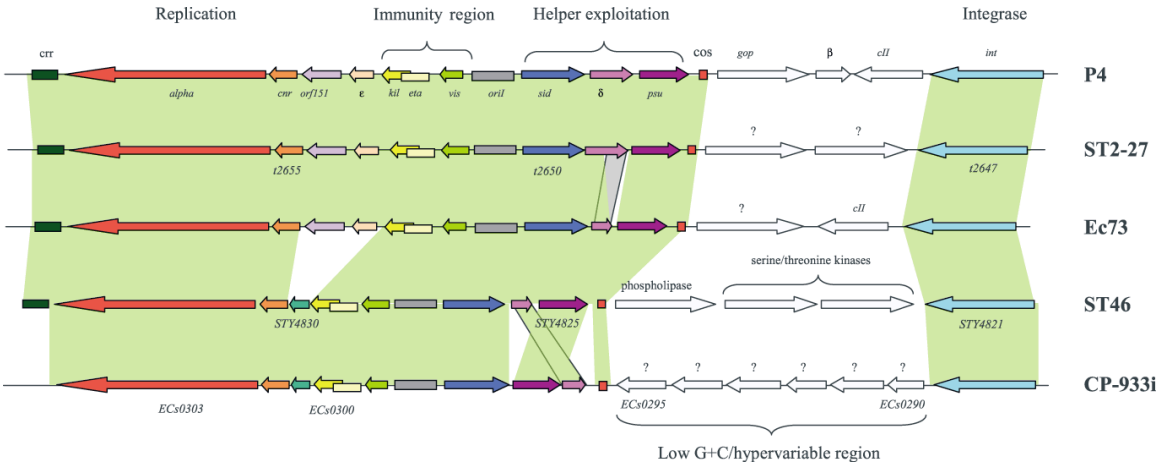


Figure 1.4 Analysis of the P4-like bacteriophage ST46 [89].

Alignment of the bacteriophage ST46 with other P4 family bacteriophages including ST2-27 (Ty2) (NC_00463; this study) Retron 73 (*E. coli* M64113), CP-933i (*E. coli* 0157:H7) (NC_002655) and P4 (X51522). Regions of significant amino acid homology are indicated via a shaded connection. Orthologous genes have the same colour. Genes in the low GC region are coloured white. A question mark indicates that the gene product has no known function.

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

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

1.5 Genetics of *Salmonella* disease

1.5.1 TTSSs

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

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

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

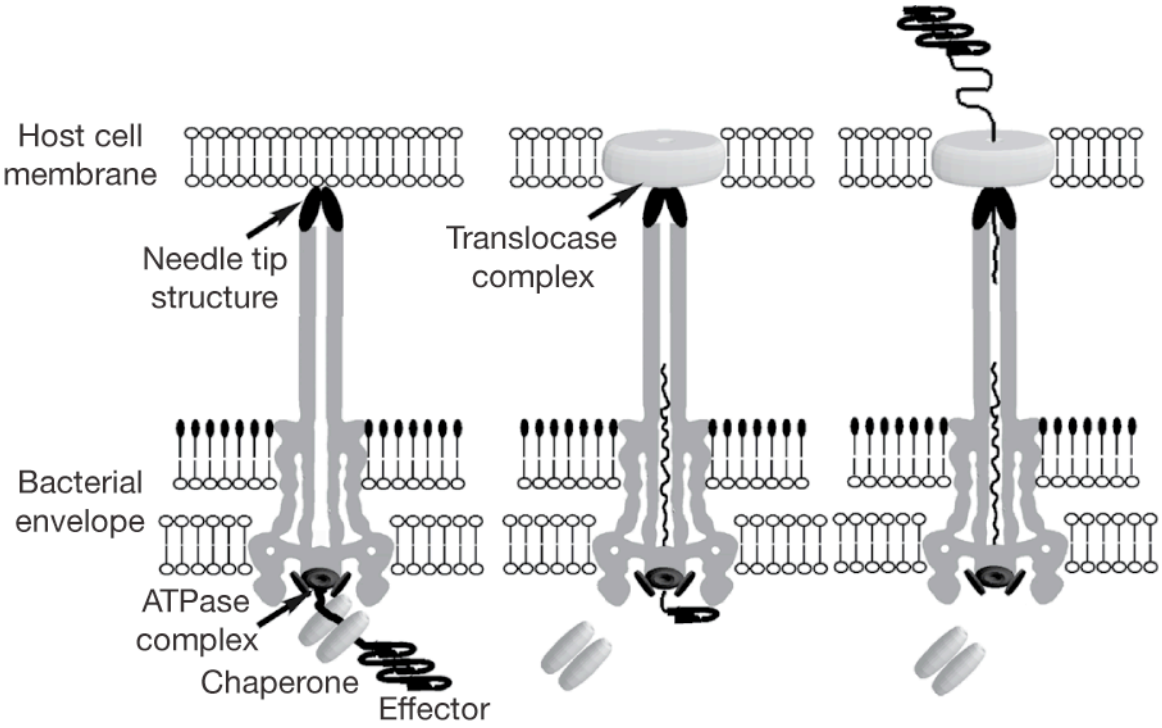


Figure 1.5 Model for substrate recognition and delivery of proteins by TTSS machines [96].

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

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

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

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

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

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

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

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

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

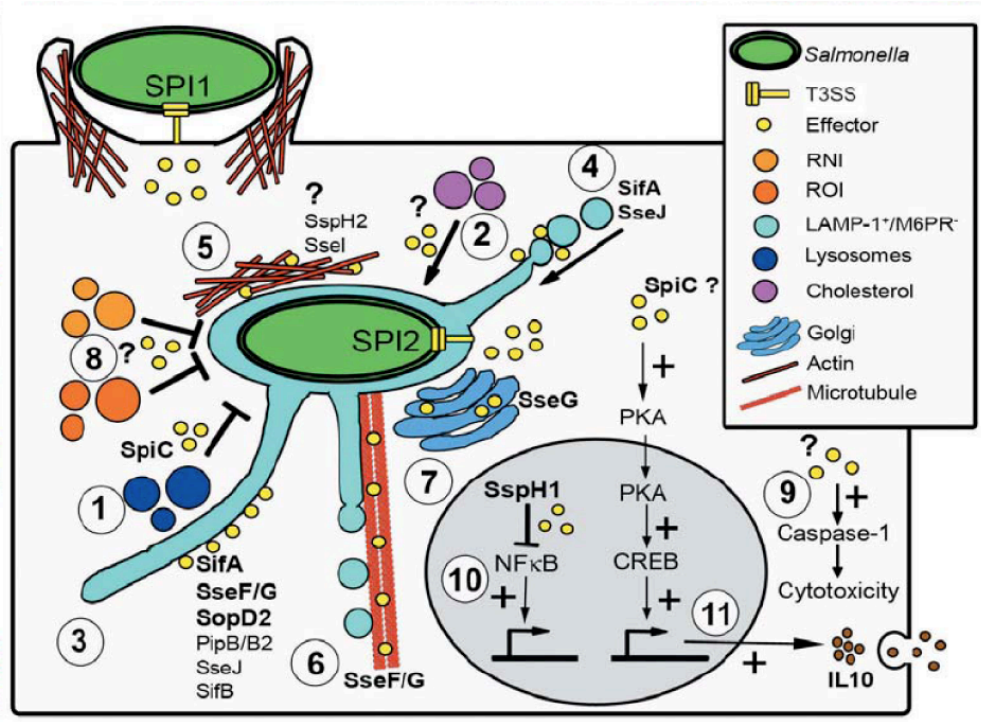


Figure 1.6 Cellular phenotypes associated with the function of SPI-2 [113].

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

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

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

1.5.2 The virulence (Vi) antigen

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

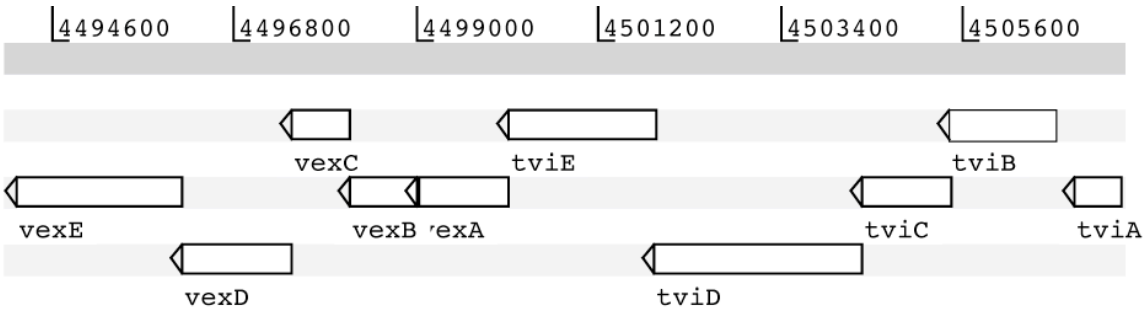


Figure 1.7 Genetic structure of the viaB locus in *S.* Typhi Ty2.

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

1.5.3 Horizontally acquired DNA

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

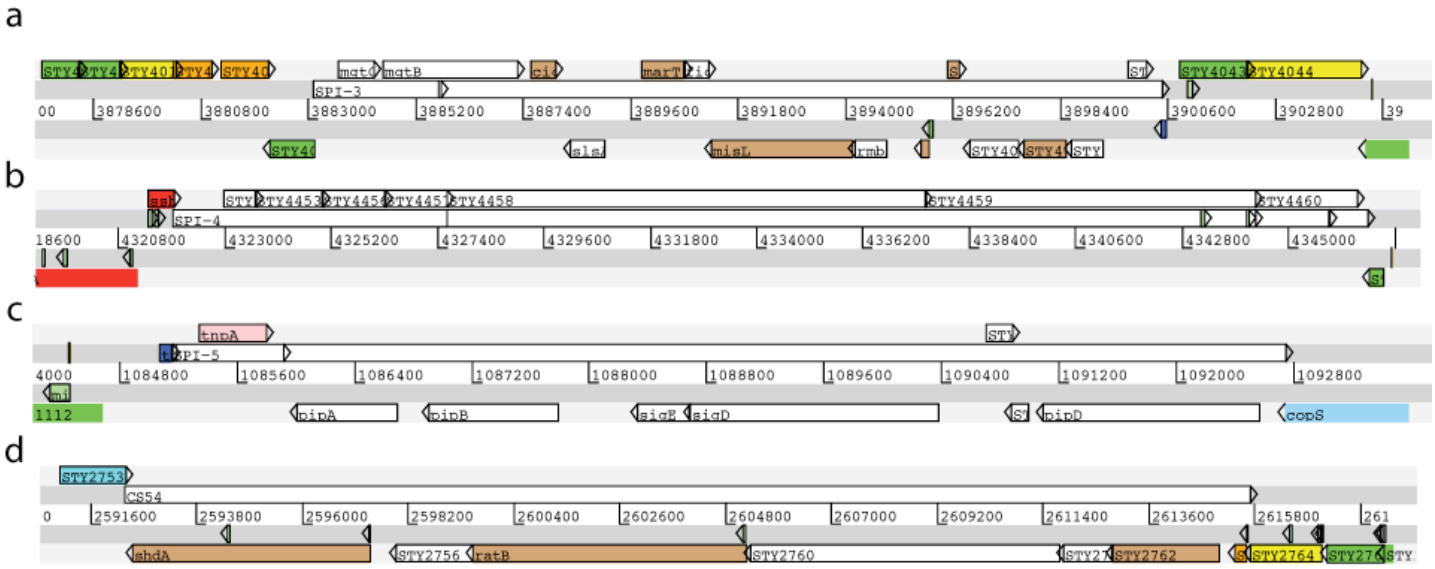


Figure 1.8 Genetic content of horizontally acquired regions in S. Typhi CT18.

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

1.5.4 Adherence

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

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

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

1.5.5 Motility

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

1.5.6 Two-component regulators and virulence

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

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

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

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

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

1.6 Regulation of bacterial gene expression

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

1.6.1 Transcriptional activation

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

Table 1.5 Major families of bacterial transcription factors [164]

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

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

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

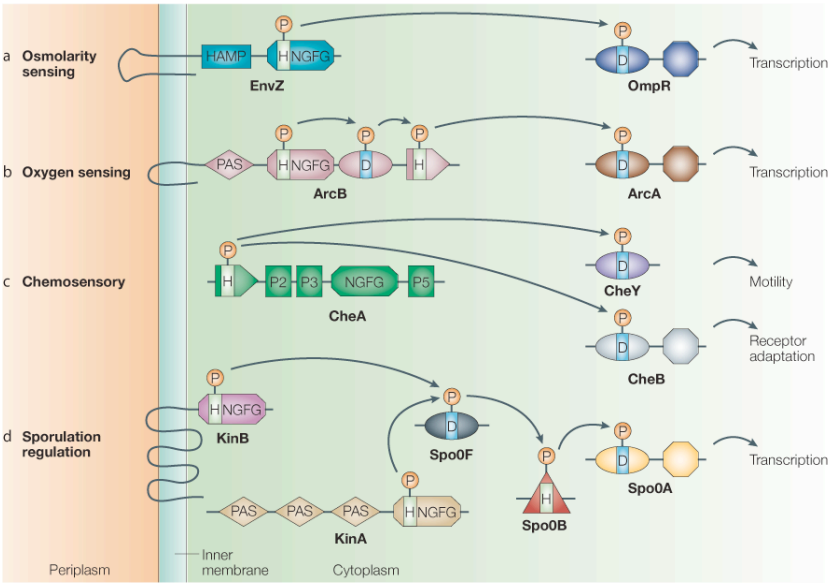


Figure 1.9 Some of the different combinations of histidine protein kinase (HPK) and aspartate response regulator (RR) domains in histidine–aspartate phosphorelay (HAP) systems [166].

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

1.6.2 Post-transcriptional control

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

1.7 DNA sequencing technologies

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

1.7.1 DNA microarrays

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

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

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

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

1.7.2 High-capacity DNA sequencing

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

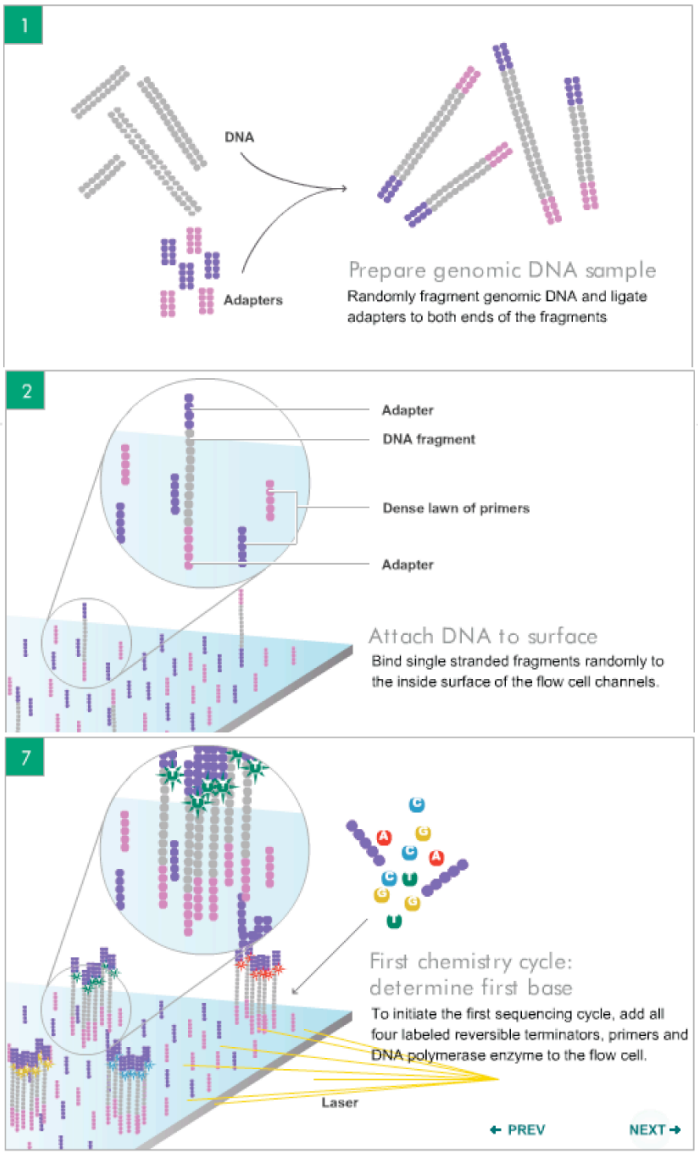


Figure 1.10 Schematic of Illumina sequencing technology (http://www.illumina.com/).

1.8 *In silico* - software and tools

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

1.8.1 Artemis and Artemis Comparison Tool (ACT)

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

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

1.8.2 BLAST and other packages

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

1.9 Aims of the Thesis

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