THE ROYAL SOCIETY

*doi* 10.1098/rstb.2001.0891

# Acquisition of virulence-associated factors by the enteric pathogens *Escherichia coli* and *Salmonella enterica*

**John Wain*, Deborah House, Derek Pickard, Gordon Dougan and Gad Frankel**

*Centre for Molecular Microbiology and Infection, Imperial College of Science, Technology and Medicine, South Kensington, London W7 2AZ, UK*

In this review we summarize recent genomic studies that shed light on the mechanism through which pathogenic *Escherichia coli* and *Salmonella enterica* have evolved. We show how acquisition of DNA at specific sites on the chromosome has contributed to increased genetic variation and virulence of these two genera of the Enterobacteriaceae.

**Keywords:** *Salmonella*; enteropathogenic *Escherichia coli*; horizontal gene transfer; pathogenicity islands; microbial pathogenesis

## 1. INTRODUCTION

Commensal organisms have adapted to coexist with the human host without causing overt damage, although they may cause opportunistic infections under certain circumstances, for example in the immunocompromised host. Pathogenic bacteria, on the other hand, are capable of colonizing the human host and have acquired the ability to cause clinically significant pathology, either by causing localized damage to host mucosal cells or by breaching cellular barriers and causing systemic disease. These organisms are also able to transfer from one host to another and are therefore able to respond to and adapt to multiple changes in environmental conditions, both inside and outside the host.

*Escherichia coli* and *Salmonella enterica* (Euzéby 1999) belong to the Enterobacteriaceae and share a common ancestor that diverged approximately 150 million years ago. The genomes of the two species share extensive regions of homology and are essentially superimposable. However, variation, in terms of DNA inversions, deletions and insertions, between the two genomes does exist and it appears that the chromosomes of both *E. coli* and the *S. enterica* have preferred sites where novel DNA sequences can be integrated (figure 1 and table 1) and tolerated without significantly affecting fitness (Achtman & Pluschke 1986). The term pathogenicity island is used when these sets of inserted genes are demonstrated to contribute to an increase in virulence potential (Blum *et al.* 1991; Hacker *et al.* 1997). These genetic islands often have a guanine and cytosine (G + C) content different from that of the rest of the chromosome and so have almost certainly been acquired by horizontal transfer from a different bacterial genus. Some have been introduced into particular strains via horizontal transfer involving vectors such as bacteriophage (Acheson *et al.* 1998) and individual *E. coli* and *S. enterica* isolates often harbour either functional or defective integrated bacteriophage in different chromosomal locations. Bacteriophages may use transfer RNA (tRNA) genes as their insertion sites and so association of horizontally acquired DNA with a tRNA gene is suggestive of bacteriophage origin.

*Author for correspondence (j.wain@ic.ac.uk).

## 2. PATHOGENICITY ISLANDS OF *E. COLI*

There are multiple serotypes of *E. coli*, the classification of which is based on the O or lipopolysaccharide (LPS), H or flagella and capsular antigens. Many of the serotypes are not normally associated with human disease, and the majority of the non-anaerobic, commensal gut flora of man are commensal strains of *E. coli*. Some strains of *E. coli* have, however, acquired the ability to cause severe disease in humans, including neonatal meningitis, urinary tract infections, and gastro-intestinal infections.

Phenotypic and genetic differences between commensal and pathogenic strains of *E. coli* from a range of disease syndromes have been defined, and it appears the evolution of the pathogenic strains was dependent to a large extent on the ordered acquisition and retention of virulence-associated genes (Reid *et al.* 2000). *E. coli* K12 is a non-pathogenic isolate which has been fully sequenced. The genome revealed a single chromosome encoding over 4280 genes (Blattner *et al.* 1997). The chromosome consists of a relatively stable framework of genes, many of which are conserved in other *E. coli* and enteric bacteria, and has a complete repertoire of flagella genes similar to those defined in *Salmonella*. Perhaps the first studies linking gene acquisition to increased virulence in *E. coli* came from the studies of Smith & Huggins (1971) showing plasmid-encoded toxin and adhesion genes in enterotoxigenic *E. coli*. Early genetic proof that pathogenic *E. coli* encode novel chromosomal genes that facilitate adaptation to host colonization came from studies on *E. coli* associated with urinary tract infections (Knapp *et al.*

 

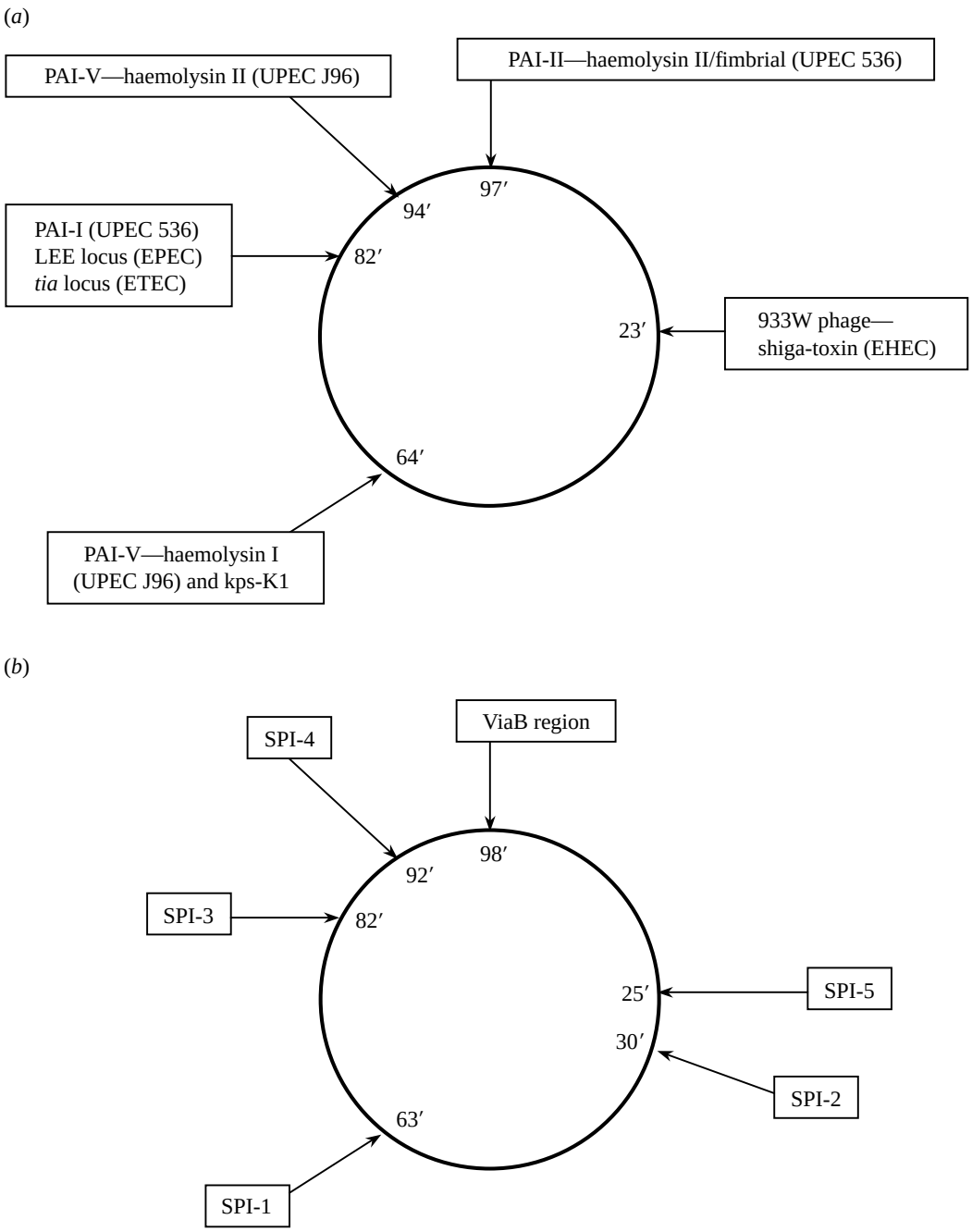


Figure 1. Chromosomal location of insertion points of pathogenicity islands and virulence-associated genes in *E. coli* (*a*) and *S. enterica* (*b*). EPEC, enteropathogenic *E. coli*; EHEC, enterohaemorrhagic *E. coli*; ETEC, enterotoxigenic *E. coli*; UPEC, uropathogenic *E. coli*; kps, k-polysaccharide synthesis

1986). These strains were found to encode novel fimbriae, named P-fimbriae, which facilitated attachment to epithelial cells. It was subsequently shown that the genes required for the biosynthesis of a haemolysin were also encoded in this same locus, bringing forward the concept of a pathogenicity island. Pathogenic *E. coli* have acquired a number of different pathogenicity islands not present in *E. coli* K12 that facilitate subversion of host-cell function and which benefit the bacterium. Particularly good examples of this phenomenon are the interactions of enteropathogenic *E. coli* (EPEC) and enterohaemorrhagic *E. coli* (EHEC) with mammalian intestinal epithelium. EPEC, an established aetiological agent of human diarrhoea, remains an important cause of mortality among young infants in developing countries. EHEC is an emerging cause of acute gastro-enteritis and haemorrhagic colitis and is often associated with severe or fatal renal and neurological complications (Nataro & Kaper 1998). Subversion of intestinal epithelial cell function by EPEC and EHEC leads to the formation of distinctive 'attaching and effacing' (A/E) lesions, which are characterized by localized destruction (effacement) of brush border microvilli, intimate attachment of the bacillus to the host-cell membrane and the formation of an actin-rich underlying pedestal-like structure in the host cell (Frankel *et al.* 1998) (figure 2). The genes encoding the A/E phenotype are encoded on a pathogenicity island termed the locus of enterocyte effacement (LEE) (McDaniel *et al.* 1995). The G + C content of the LEE (38.3%) is significantly lower than that of the *E. coli* chromosome (50.8%) and both the

Table 1. Chromosomal location of insertion points of pathogenicity islands (PAI) and virulence-associated genes in *E. coli* and *S. enterica*.

| insertion site (minutes) | insertion site (gene) | insertion |
|---|---|---|
| *E. coli* | | |
| 23 | *wrbA* | 933W phage |
| 64 | tRNA*pheV* | PAI-IV |
| 82 | tRNA*selC* | PAI-I |
| 94 | tRNA*pheR* | PAI-V |
| 97 | tRNA*leuX* | PAI-III |
| *S. enterica* | | |
| 25 | tRNA*aspV-yafV* | SPI-5 |
| 30 | *serT-copS/copR* | SPI-2 |
| 63 | tRNA*val* | SPI-1 |
| 82 | tRNA*selC* | SPI-3 |
| 92 | *ccb-soxSR* | SPI-4 |
| 98 | tRNA*pheU* | ViaB region |

EPEC and EHEC LEEs contain 41 open reading frames (ORFs). The majority of the genes are organized in five polycistronic operons (*LEE1*, *LEE2*, *LEE3*, *tir* and *LEE4*). *LEE1*, *LEE2* and *LEE3* encode components of a type III secretion system (TTSS), *LEE4* encodes proteins secreted by the TTSS, termed *E. coli* secreted proteins (ESPs), and the *tir* operon encodes for the outer membrane adhesion molecule intimin, the translocated intimin receptor (Tir), and CesT (the Tir chaperon) (Zhu *et al.* 2001; Frankel *et al.* 1998). In both EPEC and EHEC, transcription of the *LEE2*, *LEE3* and *tir* operons is positively regulated by Ler (Mellies *et al.* 1999), the product of the first ORF in the *LEE1* operon which belongs to the H-NS family of transcriptional activators. In EPEC only, a plasmid-encoded regulator, Per, activates transcription of *ler* and the *LEE1* operon (Gomez-Duarte & Kaper 1995). Expression of *ler* is also dependent upon the integration host factor (Friedberg *et al.* 1999) and quorum sensing (Sperandio *et al.* 1999).

Population genetic surveys have shown that EPEC and EHEC strains can be divided into two related clones, designated EPEC clones 1 and 2 and EHEC clones 1 and 2 (Whittam & McGraw 1996). The insertion site of the LEE in the *E. coli* chromosome varies according to the clonal phylogeny of the strain. EPEC or EHEC that are clonally grouped together have an identical LEE insertion site (*selC* or *pheU*) (Wieler *et al.* 1997). This suggests that the LEE has inserted at multiple times during the evolution of the EPEC and EHEC family. These *E. coli* derivatives subsequently acquired additional virulence factors such as verocytotoxin encoded on bacteriophage (EHEC only) and on large (*ca.* 90 kb) plasmids found in EPEC and EHEC. Genetic analysis of the A/E gene (*eae*) from EHEC strains supports these data, as the level of DNA-sequence variation is above that expected from the natural accumulation of mutations (Boerlin *et al.* 1998). The different clonal types of EHEC harbouring *eae*, however, cluster on phylogenetic analysis within closely related lineages (Boerlin *et al.* 1998) and there is very little sequence variation in the genes from the LEE of EHEC and EPEC not directly associated with interaction with the host (Perna *et al.* 1998). This suggests that only closely related *E. coli* maintain the LEE, probably because of a dependence on genetic background. Transfer of the LEE therefore seems to be an event which has taken place several times but only a small subset of the *E. coli* population is capable of maintaining these genes.

## 3. *SALMONELLA* PATHOGENICITY ISLANDS

The genus *Salmonella* consists of two species, *enterica* and *bongori*. The species *enterica* is divided into seven subspecies (groups I, II, IIIa, IIIb, IV, VI and VII) containing over 2300 different serotypes (Popoff *et al.* 1998), while the species *bongori* (group V) (Reeves *et al.* 1989) contains 17 serotypes. As with *E. coli*, many of the *Salmonellae* are not known to be pathogenic for man, with the majority of the serotypes being isolated from cold-blooded animals. However, several of the group I serotypes can cause disease in humans, including *S. enterica* serotype Enteritidis (*S.* Enteritidis), *S.* Typhimurium, and *S.* Dublin, which cause gastro-enteritis, and *S.* Typhi and *S.* Paratyphi A, which cause enteric fever. Some serotypes that cause gastro-enteritis in man are known to be host-adapted for birds or mammals other than man. Disease in humans is the result of contact with infected animals or the inadvertent introduction of the bacteria into the food chain. On the other hand, *S.* Typhi and *S.* Paratyphi A serotypes are host-adapted to man and have the ability to breach the gut mucosal barriers and cause a severe systemic infection.

As with *E. coli*, *S. enterica* has acquired single genes and large pathogenicity islands by horizontal transfer. The pathogenicity islands of *Salmonella* are generally characterized as having a G + C content different from that of the remainder of the chromosome (45%) and are usually, but not always, flanked by tRNA genes or repetitive elements. To date, at least five putative pathogenicity islands have been identified in *S.* Typhimurium: SPI-1, -2, -3, -4 and -5 (Marcus *et al.* 2000). SPI-1 and SPI-2 are the best characterized and are located at 63′ and 30′, respectively, on the *S.* Typhimurium chromosome (figure 1*b*). SPI-1 is a 40 kb pathogenicity island that is inserted not in a tRNA gene but between *fhlA* and *mutS*, two genes that are adjacent on the *E. coli* chromosome (Mills *et al.* 1995). SPI-2, which is also a 40 kb DNA region, is inserted adjacent to a tRNA gene (Hensel 2000). Both SPI-1 and SPI-2 contain genes that code for TTSSs in addition to other genes, including those for two-component regulatory systems (Hensel 2000). The genes encoding the secreted proteins, or substrates, of the SPI-1 and SPI-2 TTSSs are found both within the pathogenicity islands themselves as well as on other regions of the chromosome (Hensel 2000; Marcus *et al.* 2000).

The genes within SPI-1 encode proteins that facilitate entry of *Salmonellae* into epithelial cells and appear to have been acquired between 50 and 100 million years ago (Mills *et al.* 1995). SPI-1 was probably acquired before the divergence into the two *Salmonella* species because the *spa* genes from this region are present in all serotypes of *Salmonella*, including those of *S. bongori* (Ochman & Groisman 1996). Furthermore, dendrograms for the subspecies of *Salmonella* based on the *inv/spa* gene sequences, the products of which are components of the

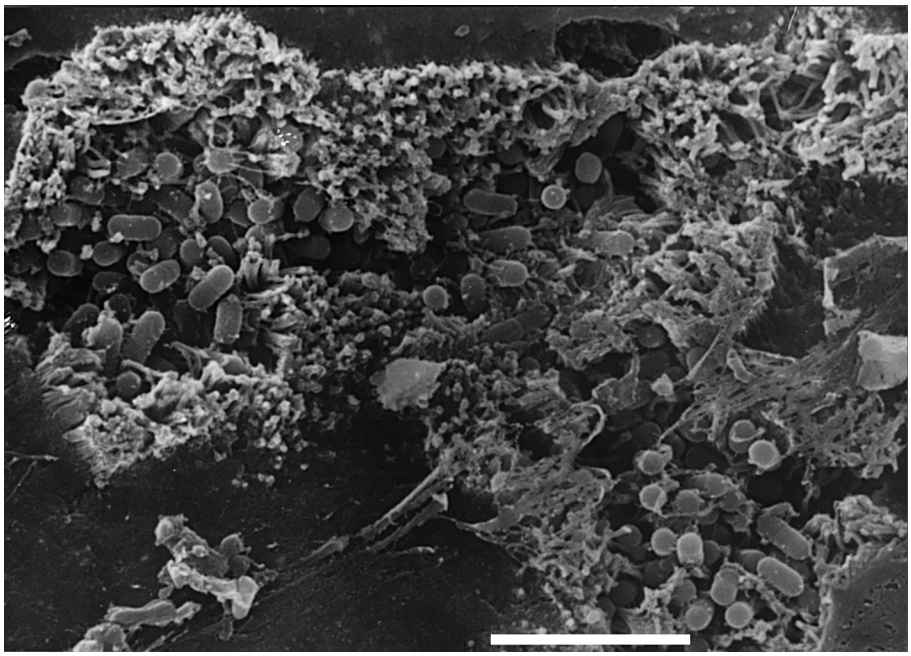

Figure 2. Localized colony of EPEC O111:MN (B171) with A/E lesion formation on human intestinal ileal explant following 8 h organ culture. Bar = 5 μm.

SPI-1 TTSS, are very similar to dendrograms based on the sequences of five housekeeping genes with one exception (Boyd *et al.* 1996). The *inv*/*spa* genes of group IV and group VII are very similar, despite differences in the housekeeping genes, suggesting that the entire island has been transferred as a single entity between these two subspecies (Boyd *et al.* 1996). SPI-1 itself has elements of a mosaic structure, with the nucleotide composition of a putative iron transport system encoded within the region being different from that of the rest of the pathogenicity island (Zhou *et al.* 1999). This suggests that the final organization of SPI-1 may have been the result of more than one genetic event but because the *inv*/*spa* genes vary in the same way as the housekeeping genes these genetic events most probably occurred before the *Salmonellae* split into subspecies.

Although the apparatus encoded within SPI-1 facilitates invasion of epithelial cells, the initial interaction between *Salmonella* and the cell is facilitated by the expression of adhesins or fimbriae, including those encoded within two operons, *fim* (type I fimbriae) and *lpf* (long polar fimbriae). As with SPI-1, it appears that these operons were acquired before the division of the *Salmonella* into two species, as both operons are present in *S. bongori* and *S. enterica* subspecies I. Although the *fim* operon is present in all *Salmonella*e, it appears that *lpf* has been lost by a deletion event from the genomes of *S. enterica* subspecies II, IIIa, IIIb, IV and VI (Baumler 1997). SPI-1, *fim*, and *lpf* map to different regions on the *Salmonella* chromosome, suggesting that a series of transfer events occurred leading to the acquisition of all the virulence gene clusters required for adherence to and invasion of epithelial cells (Baumler 1997). Preliminary analysis of the completed *S.* Typhi genome sequence indicates the presence of multiple fimbriae operons, indicating that acquisition of fimbriae (or gene duplication events) may be relatively common.

A second pathogenicity island, SPI-2, encodes the necessary machinery for *Salmonellae* to survive within macrophages and thus establish a systemic infection (Hensel 2000). Early studies suggested that, unlike SPI-1, the virulence genes encoded by SPI-2 were acquired by *S. enterica* following the split from *S. bongori* but prior to the diversification into the various subspecies (Ochman & Groisman 1996; Hensel *et al.* 1997). Further studies have shown that SPI-2 is composed of two distinct genetic elements that may have been acquired independently: a 25.3 kb region located between the $tRNA^{ValV}$ gene at 31′ and *ssrB* of SPI-2, and a smaller 14.5 kb region at 30.5′. The 25.3 kb region contains virulence-associated genes (e.g. those for the TTSS), while the smaller 14.5 kb region harbours a cluster of five *ttr* genes involved in anaerobic tetrathionate reduction and seven additional ORFs (Hensel *et al.* 1999). Interestingly, the smaller 14.5 kb region is present in *S. bongori* and it is possible that this region was acquired before the divergence into the two species of *Salmonella*, with the larger virulence-associated region being inserted later.

A smaller pathogenicity island, SPI-3, is located at 82′ on the chromosome immediately adjacent to *selC*, a tRNA locus that is the insertion site for distinct pathogenicity islands in EPEC and uropathogenic *E. coli* (UPEC; Blanc-Potard *et al.* 1999). SPI-3 is a 17 kb DNA region containing ten ORFs organized into six transcriptional units, including the *mgtCB* operon encoding the macrophage survival protein MgtC and the magnesium transporter MgtB. The distribution of SPI-3 sequences varies among the *Salmonellae*. The right end of the island, which harbours the virulence gene *mgtC*, is present in all eight subspecies, while a four-gene cluster at the centre of SPI-3 is found only in subspecies I, II and IV. This cluster has been shown to be bracketed by remnants of insertion sequences, suggesting a multistep process in the evolution of SPI-3 sequences (Blanc-Potard *et al.* 1999).

The acquisition of potential virulence-associated genes is only one step towards increasing pathogenic potential. The newly acquired genes must be expressed at the appropriate stages of the pathogenic process in a manner that does not reduce overall fitness. The pathogenic process for invasive *Salmonella* is highly complex, involving several major environmental changes in pH, oxygen tension, osmolarity and temperature. The expression of SPI-1 and SPI-2 virulence-associated genes is dependent upon the environmental conditions and, at least for SPI-2, only occurs in specific host intracellular compartments (Deiwick *et al.* 1999; Hensel 2000; Lucas & Lee 2000). The regulation of the genes encoded within the SPIs is highly complex and involves global regulators, such as Phop-PhoQ and EnvZ-OmpR (Lucas & Lee 2000), as well as regulatory systems encoded within the pathogenicity islands themselves. The SirA/HilA regulon associated with SPI-1 is expressed primarily when *Salmonella* are extracellular. SirA, a possible two-component regulator, is required for the expression of *hilA*, and the product HilA positively activates the *spa*, *inv*, and *prg* operons, which encode the proteins that form the type III secretion apparatus of SPI-1 (Ahmer *et al.* 1999; Eichelberg & Galan 1999). Similarly SsrA–SsrB, a two-component regulatory system encoded within SPI-2, is required for SPI-2 gene expression (Cirillo *et al.* 1998; Hensel *et al.* 1998). The regulator *ssrAB* and the structural gene *ssaH* are transcribed after *Salmonella* enters host cells. The early transcription of *ssrAB* itself is regulated by EnvZ-OmpR (Lee *et al.* 2000). It is possible that *ssr*AB is a global regulon controlling several sets of horizontally acquired genes. At least ten genes are regulated by SsrB within epithelial cells and macrophages (Worley *et al.* 2000). It has been suggested that SPI-1 and SPI-2 are inversely regulated, and mutations in SPI-2 have been shown to affect the expression of SPI-1-encoded genes (Deiwick *et al.* 1998).

## 4. CAPSULAR GENES

The interaction between the bacterial surface and host cells is an important phase in the infection and can be influenced by surface-associated polysaccharides produced by *E. coli* and *S. enterica*, in addition to lipopolysaccharide. The genes for many of these products have been defined in *E. coli* (Kotloff *et al.* 1992). Although the exact function of the capsule in *E. coli* is not clear, many *E. coli* express one of the 80 serotypes of K capsular antigens. The capsules of *E. coli* have been recently reviewed (Whitfield & Roberts 1999) and here we will concentrate on the capsule of *S.* Typhi.

Several serotypes of *S. enterica* group I are host-adapted, that is, they are pathogens that are able to circulate and cause systemic disease within a particular host population (Kingsley & Baumler 2000). An example of such a serotype is *S.* Typhi, which causes human typhoid fever. Man and higher primates are the only known hosts for *S.* Typhi and the bacterium is considered to be host-adapted as well as host-restricted to man because it can cause systemic disease in humans and can circulate within the human population. Although other serotypes such as *S.* Choleraesuis can cause systemic disease in humans, these serotypes do not normally circulate within the human population and are usually acquired following contact with infected animals. The factor(s) responsible for the host-specificity of *S.* Typhi are not known. It is possible that either gene inactivation or the acquisition of additional genetic material after the divergence of the species *S. enterica* into the various subspecies may have contributed to the host-specificity of group I pathogenic *Salmonella* (Baumler 1997), however, there has been no specific genotype or phenotype proven to be associated as a determinant for host-specificity. One region that is present in the *S.* Typhi chromosome but which is absent from the *S.* Typhimurium and *E. coli* genomes is the ViaB locus. This locus is located on a 118 kb loop of DNA at position 98′ in the *S.* Typhi chromosome (figure 1*b*) (Liu & Sanderson 1995) and contains the genes responsible for the synthesis and polymerization of the Vi capsular antigen as well as the genes required for its transport to the surface of the bacterium (Hashimoto *et al.* 1993). The Vi antigen is a homopolymer of *O*-acetylated $\alpha 1 \rightarrow 4$, 2-deoxy-2-*N*-acetyl galacturonic acid and was first described by Felix & Pitt (1934). It is expressed by almost all clinical isolates of *S.* Typhi (Lesmana *et al.* 1980) but has never been reported to be expressed by *S.* Typhimurium. Despite the reports of spontaneous Vi-negative mutants, the presence of the antigen on the vast majority of clinical isolates of *S.* Typhi suggests that *S.* Typhi expressing Vi have a selective advantage. Vi has been shown to be a virulence factor in a mouse model of typhoid fever using *S.* Typhi inoculated in hog mucin (Felix & Pitt 1934) and, in humans, the infectivity of Vi-expressing strains of *S.* Typhi is greater than that of Vi-negative spontaneous mutants (Hornick *et al.* 1970). However, the antigen is not essential for virulence as Vi-negative strains of *S.* Typhi are able to cause infection and disease in humans (Hornick *et al.* 1970). Furthermore, occasional isolates of other bacteria, including some serotypes of *Salmonellae* which do not commonly cause invasive disease in humans, may express the Vi antigen, e.g. *S.* Dublin and *Citrobacter freundii*. The role of Vi in the pathogenic process is not clear, although there is some evidence to suggest that it is antiopsonic and antiphagocytic (Looney & Steigbigel 1986). The 118 kb loop of DNA containing the ViaB locus is located inside a larger 135 kb island (figure 3) and it is possible that other genes within the locus are important virulence determinants.

The Vi capsule can act as a physical barrier to the export of flagella protein and the invasion-promoting proteins of SPI-1, and the expression of Vi capsule and SPI-1 is coordinated (Arricau *et al.* 1998). For example, the expression of Vi is down regulated and SPI-1 up-regulated under conditions of high osmolarity, as would be encountered in the gut. Indeed, osmolarity is an important environmental signal for the expression of many bacterial proteins, and triggers an adherent and invasive phenotype in *S.* Typhi (Tartera & Metcalf 1993). One locus involved in the regulation of Vi expression is ViaA (Snellings *et al.* 1981), a locus that is present in several bacterial species (Baron *et al.* 1982). Mutations in ViaA that lead to the absence of Vi expression are invariably within the *rcs*B gene (Houng *et al.* 1992), which regulates Vi expression via the two component system RcsA–RcsB (Arricau *et al.* 1998). The RcsA–RcsB system is a common mechanism in the activation of capsule

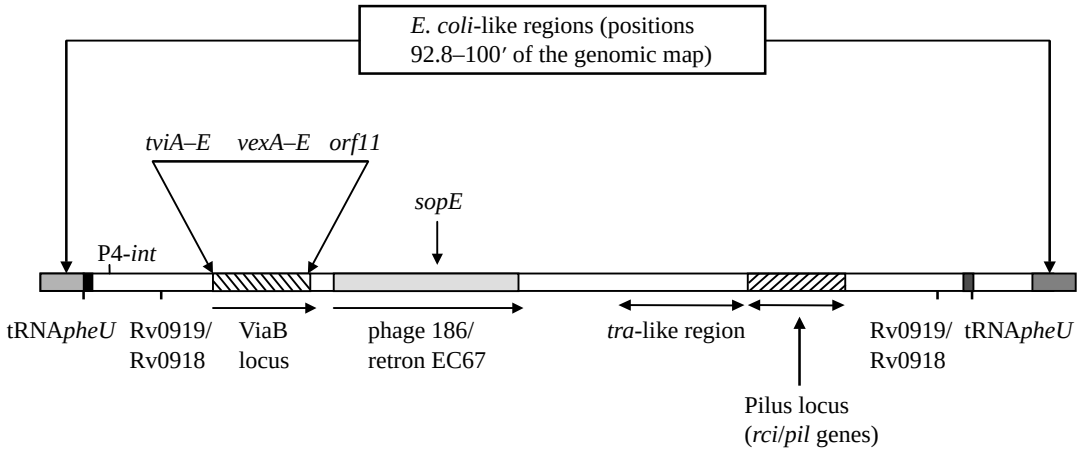


Figure 3. The ViaB region of *S.* Typhi showing the ViaB locus, which encodes the genes involved in Vi capsule biosynthesis (*tviB–E*) and expression (*vexA–E*), and other gene clusters encoding for a bacteriophage and pilin subunits.

biosynthesis in many enteric bacteria (Wehland & Bernhard 2000). Transfer of the ViaB region alone into *E. coli* allows the expression of a Vi capsule (Houng *et al.* 1992), showing that the ViaA locus present in *E. coli* is functional with respect to Vi regulation. The first gene within the ViaB locus, *tviA*, is involved in the RcsB regulation of Vi expression (Virlogeux *et al.* 1996). The mechanism of the interaction between TviA (from ViaB) and RcsB (from ViaA) and the process by which the two genes have come together in a single pathogen is still not understood. There are certain similarities between this and the *E. coli* serotype K30 capsule. In K30, expression is positively regulated by the Rcs system and by the first ORF of the gene cluster *orf*X. (Whitfield & Roberts 1999).

## 5. GENOMIC INITIATIVES

Currently several whole-genome sequencing programmes for specific *E. coli* isolates are underway, in particular EHEC and *E. coli* O157. This and similar DNA-sequencing programmes will help us define more clearly the differences between commensal and pathogenic *E. coli* and help us understand more about the evolution of this micro-organism. *E. coli* K12 is genetically the most thoroughly characterized micro-organism and so it was an *E. coli* K12 derivative, MG1655, that was the first *E. coli* genome to be completely sequenced (Blattner *et al.* 1997). Many of the ORFs identified through the sequencing project had not been discovered during the previous 50 years of intensive conventional genetic investigation. A significant proportion of the ORFs encode proteins for which no obvious gene function can be assigned. Some of these genes of unknown function have homologues in other bacteria, particularly those within the Enterobacteriaceae but there are no clear clusters of genes that might contribute to virulence traits such as adhesins or toxins.

The sequencing of the *S.* Typhi and *S.* Typhimurium genomes is nearing completion and information is available at http://www.sanger.ac.uk/Projects/S_typhi. The *S.* Typhi strain being sequenced at the Sanger Centre (Cambridge, UK) is a multidrug-resistant strain, CT18, that harbours two large plasmids in addition to the chromosome. One of these plasmids, pHCM1, encodes multiple antibiotic resistance determinants. Plasmid pHCM1 is related over *ca.* 80% of the sequence to R27 and it is possible to readily identify additional sequences present on pHCM1 against the background of shared sequence. Detailed comparisons of *S. enterica* and *E. coli* genome organization await the completion of these sequencing projects.

As sequencing programmes progress, the comparison between species of Enterobacteriaceae will allow a more general understanding of the evolution of this broad group of bacteria. However, of more direct relevance to the emergence of pathogens from within a species are the genetic differences between the various clinical groups of *E. coli* and between the serotypes of *S. enterica.*

## 6. CONCLUSIONS

Modern molecular studies are providing tremendous insights into the functional organization and evolutionary relationships of *E. coli* and *S. enterica*. Both species appear to encode a related framework of common genes, which determine the characteristics of these species. Variation is built on this framework through a number of different processes, some involving gene acquisition at specific sites on the chromosome or extra-chromosomal elements. For the first time we are able to accurately define virulence-associated traits and distinguish factors that contribute to commensalism or pathogenicity. This information will help us predict potential future evolutionary developments in these species as we identify mechanisms that have contributed to past diversity. Increasing knowledge of the area will also yield information on how best to intervene in diseases precipitated by these micro-organisms through the identification of novel antigens and potential drug targets. Genome-sequencing programmes are already impacting in this area and completion of ongoing programmes will eventually provide new blocks of information to guide future work. The ability to genetically manipulate the

species will facilitate experimental approaches, making this an area for expansion and discovery in the next decade.

We thank Dr Alan Phillips from the Royal Free Hospital, London, for contributing figure 2. This work was supported by grants from The UK Wellcome Trust. Much of the *S.* Typhi work was carried out at The Wellcome Trust Clinical Research Unit and The Centre for Tropical Diseases, Ben Ham Tu Street, Ho Chi Minh City, Vietnam.

## REFERENCES

Acheson, D. W., Reidl, J., Zhang, X., Keusch, G. T., Mekalanos, J. J. & Waldor, M. K. 1998 *In vivo* transduction with shiga toxin 1-encoding phage. *Infect. Immun.* **66**, 4496–4498.

Achtman, M. & Pluschke, G. 1986 Clonal analysis of descent and virulence among selected *Escherichia coli*. *A. Rev. Microbiol.* **40**, 185–210.

Ahmer, B. M., Van Reeuwijk, J., Watson, P. R., Wallis, T. S. & Heffron, F. 1999 *Salmonella* SirA is a global regulator of genes mediating enteropathogenesis. *Mol. Microbiol.* **31**, 971–982.

Arricau, N., Hermant, D., Waxin, H., Ecobichon, C., Duffey, P. S. & Popoff, M. Y. 1998 The RcsB–RcsC regulatory system of *Salmonella typhi* differentially modulates the expression of invasion proteins, flagellin and Vi antigen in response to osmolarity. *Mol. Microbiol.* **29**, 835–850.

Baron, L. S., Kopecko, D. J., McCowen, S. M., Snellings, N. J., Johnson, E. M. & Reid, W. C. 1982 Genetic and molecular studies of the regulation of atypical citrate utilization and variable Vi antigen expression in enteric bacteria. *Basic Life Sci.* **19**, 175–194.

Baumler, A. J. 1997 The record of horizontal gene transfer in *Salmonella*. *Trends Microbiol.* **5**, 318–322.

Blanc-Potard, A. B., Solomon, F., Kayser, J. & Groisman, E. A. 1999 The SPI-3 pathogenicity island of *Salmonella enterica*. *J. Bacteriol.* **181**, 998–1004.

Blattner, F. R. (and 16 others) 1997 The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1474.

Blum, G., Ott, M., Cross, A. & Hacker, J. 1991 Virulence determinants of *Escherichia coli* O6 extraintestinal isolates analysed by Southern hybridizations and DNA long range mapping techniques. *Microb. Pathogen.* **10**, 127–136.

Boerlin, P., Chen, S., Colbourne, J. K., Johnson, R., De Grandis, S. & Gyles, C. 1998 Evolution of enterohemorrhagic *Escherichia coli* hemolysin plasmids and the locus for enterocyte effacement in shiga toxin-producing *E. coli*. *Infect. Immun.* **66**, 2553–2561.

Boyd, E. F., Wang, F. S., Whittam, T. S. & Selander, R. K. 1996 Molecular genetic relationships of the *Salmonellae*. *Appl. Environ. Microbiol.* **62**, 804–808.

Cirillo, D. M., Valdivia, R. H., Monack, D. M. & Falkow, S. 1998 Macrophage-dependent induction of the *Salmonella* pathogenicity island 2 type III secretion system and its role in intracellular survival. *Mol. Microbiol.* **30**, 175–188.

Deiwick, J., Nikolaus, T., Shea, J. E., Gleeson, C., Holden, D. W. & Hensel, M. 1998 Mutations in *Salmonella* pathogenicity island 2 (SPI2) genes affecting transcription of SPI1 genes and resistance to antimicrobial agents. *J. Bacteriol.* **180**, 4775–4780.

Deiwick, J., Nikolaus, T., Erdogan, S. & Hensel, M. 1999 Environmental regulation of *Salmonella* pathogenicity island 2 gene expression. *Mol. Microbiol.* **31**, 1759–1773.

Eichelberg, K. & Galan, J. E. 1999 Differential regulation of *Salmonella typhimurium* type III secreted proteins by pathogenicity island 1 (SPI-1)-encoded transcriptional activators InvF and hilA. *Infect. Immun.* **67**, 4099–4105.

Euzéby, J. P. 1999 Revised *Salmonella* nomenclature: designation of *Salmonella enterica* (ex Kauffmann and Edwards 1952) Le Minor and Popoff 1987 sp. nov., nom. rev. as the neotype species of the genus *Salmonella* Lignieres 1900 (approved lists 1980), rejection of the name *Salmonella choleraesuis* (Smith 1894) Weldin 1927 (approved lists 1980), and conservation of the name *Salmonella typhi* (Schroeter 1886) Warren and Scott 1930 (approved lists 1980). Request for an opinion. *Int. J. System. Bacteriol.* **49**, 927–930.

Felix, A. & Pitt, R. 1934 A new antigen of *B. typhosus*. *The Lancet* **227** (ii), 186–191.

Frankel, G., Phillips, A. D., Rosenshine, I., Dougan, G., Kaper, J. B. & Knutton, S. 1998 Enteropathogenic and enterohaemorrhagic *Escherichia coli*: more subversive elements. *Mol. Microbiol.* **30**, 911–921.

Friedberg, D., Umanski, T., Fang, Y. & Rosenshine, I. 1999 Hierarchy in the expression of the locus of enterocyte effacement genes of enteropathogenic *Escherichia coli*. *Mol. Microbiol.* **34**, 941–952.

Gomez-Duarte, O. G. & Kaper, J. B. 1995 A plasmid-encoded regulatory region activates chromosomal *eaeA* expression in enteropathogenic *Escherichia coli*. *Infect. Immun.* **63**, 1767–1776.

Hacker, J., Blum-Oehler, G., Muhldorfer, I. & Tschape, H. 1997 Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. *Mol. Microbiol.* **23**, 1089–1097.

Hashimoto, Y., Li, N., Yokoyama, H. & Ezaki, T. 1993 Complete nucleotide sequence and molecular characterization of *ViaB* region encoding Vi antigen in *Salmonella typhi*. *J. Bacteriol.* **175**, 4456–4465.

Hensel, M. 2000 MicroReview: *Salmonella* pathogenicity island 2. *Mol. Microbiol.* **36**, 1015–1023.

Hensel, M., Shea, J. E., Baumler, A. J., Gleeson, C., Blattner, F. & Holden, D. W. 1997 Analysis of the boundaries of *Salmonella* pathogenicity island 2 and the corresponding chromosomal region of *Escherichia coli* K-12. *J. Bacteriol.* **179**, 1105–1111.

Hensel, M., Shea, J. E., Waterman, S. R., Mundy, R., Nikolaus, T., Banks, G., Vazquez-Torres, A., Gleeson, C., Fang, F. C. & Holden, D. W. 1998 Genes encoding putative effector proteins of the type III secretion system of *Salmonella* pathogenicity island 2 are required for bacterial virulence and proliferation in macrophages. *Mol. Microbiol.* **30**, 163–174.

Hensel, M., Nikolaus, T. & Egelseer, C. 1999 Molecular and functional analysis indicates a mosaic structure of *Salmonella* pathogenicity island 2. *Mol. Microbiol.* **31**, 489–498.

Hornick, R. B., Greisman, S. E., Woodward, T. E., DuPont, H. L., Dawkins, A. T. & Snyder, M. J. 1970 Typhoid fever: pathogenesis and immunologic control. *New Engl. J. Med.* **283**, 686–691.

Houng, H. S., Noon, K. F., Ou, J. T. & Baron, L. S. 1992 Expression of Vi antigen in *Escherichia coli* K-12: characterization of *ViaB* from *Citrobacter freundii* and identity of *ViaA* with RcsB. *J. Bacteriol.* **174**, 5910–5915.

Kingsley, R. A. & Baumler, A. J. 2000 Host adaptation and the emergence of infectious disease: the *Salmonella* paradigm. *Mol. Microbiol.* **36**, 1006–1014.

Knapp, S., Hacker, J., Jarchau, T. & Goebel, W. 1986 Large, unstable inserts in the chromosome affect virulence properties of uropathogenic I O6 strain 536. *J. Bacteriol.* **168**, 22–30.

Kotloff, K. L., Herrington, D. A., Hale, T. L., Newland, J. W., Van De Verg, L., Cogan, J. P., Snoy, P. J., Sadoff, J. C., Formal, S. B. & Levine, M. M. 1992 Safety, immunogenicity, and efficacy in monkeys and humans of invasive *Escherichia coli* K-12 hybrid vaccine candidates expressing *Shigella flexneri* 2a somatic antigen. *Infect. Immun.* **60**, 2218–2224.

Lee, A. K., Detweiler, C. S. & Falkow, S. 2000 OmpR regulates the two-component system SsrA–SsrB in *Salmonella* pathogenicity island 2. *J. Bacteriol.* **182**, 771–781.

Lesmana, M., Rockhill, R. C. & Sanborn, W. R. 1980 A coagglutination method for presumptive identification of

*Salmonella* typhi. *Southeast Asian J. Trop. Med. Public Health* **11**, 302–307.

Liu, S. L. & Sanderson, K. E. 1995 Genomic cleavage map of *Salmonella typhi* Ty2. *J. Bacteriol.* **177**, 5099–5107.

Looney, R. J. & Steigbigel, R. T. 1986 Role of the Vi antigen of *Salmonella typhi* in resistance to host defence *in vitro*. *J. Lab. Clin. Med.* **108**, 506–516.

Lucas, R. L. & Lee, C. A. 2000 Unravelling the mysteries of virulence gene regulation in *Salmonella typhimurium*. *Mol. Microbiol.* **36**, 1024–1033.

Marcus, S. L., Brumell, J. H., Pfeifer, C. G. & Finlay, B. B. 2000 *Salmonella* pathogenicity islands: big virulence in small packages. *Microbes Infect.* **2**, 145–156.

McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S. & Kaper, J. B. 1995 A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. *Proc. Natl Acad. Sci. USA* **92**, 1664–1668.

Mellies, J. L., Elliott, S. J., Sperandio, V., Donnenberg, M. S. & Kaper, J. B. 1999 The Per regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). *Mol. Microbiol.* **33**, 296–306.

Mills, D. M., Bajaj, V. & Lee, C. A. 1995 A 40 kb chromosomal fragment encoding *Salmonella* typhimurium invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. *Mol. Microbiol.* **15**, 749–759.

Nataro, J. P. & Kaper, J. B. 1998 Diarrheagenic *Escherichia coli*. *Clin. Microbiol. Rev.* **11**, 142–201.

Ochman, H. & Groisman, E. A. 1996 Distribution of pathogenicity islands in *Salmonella* spp. *Infect. Immun.* **64**, 5410–5412.

Perna, N. T., Mayhew, G. F., Posfai, G., Elliott, S., Donnenberg, M. S., Kaper, J. B. & Blattner, F. R. 1998 Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. *Infect. Immun.* **66**, 3810–3817.

Popoff, M., Bockemuhl, J. & Brenner, F. W. 1998 Supplement 1997 (no. 41) to the Kaufman–White scheme. *Res. Microbiol.* **149**, 601–604.

Reeves, M. W., Evins, G. M., Heiba, A. A., Plikaytis, B. D. & Farmer, III, J. J. 1989 Clonal nature of *Salmonella typhi* and its genetic relatedness to other salmonellae as shown by multilocus enzyme electrophoresis, and proposal of *Salmonella bongori* comb. nov. *J. Clin. Microbiol.* **27**, 313–320.

Reid, S. D., Herbelin, C. J., Bumbaugh, A. C., Selander, R. K. & Whittam, T. S. 2000 Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**, 64–67.

Smith, H. W. & Huggins, M. B. 1971 The influence of plasmid determined and other characteristics of enteropathogenic *Escherichia coli* on their ability to proliferate in the alimentary tract of piglets, calves and lambs. *J. Med. Microbiol.* **11**, 471–492.

Snellings, N. J., Johnson, E. M., Kopecko, D. J., Collins, H. H. & Baron, L. S. 1981 Genetic regulation of variable Vi antigen expression in a strain of *Citrobacter freundii*. *J. Bacteriol.* **145**, 1010–1017.

Sperandio, V., Mellies, J. L., Nguyen, W., Shin, S. & Kaper, J. B. 1999 Quorum sensing controls expression of the type III secretion gene transcription and protein secretion in enterohemorrhagic and enteropathogenic *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **96**, 15 196–15 201.

Tartera, C. & Metcalf, E. S. 1993 Osmolarity and growth phase overlap in regulation of *Salmonella typhi* adherence to and invasion of human intestinal cells. *Infect. Immun.* **61**, 3084–3089.

Virlogeux, I., Waxin, H., Ecobichon, C., Lee, J. O. & Popoff, M. Y. 1996 Characterization of the rcsA and rcsB genes from *Salmonella* typhi: rcsB through tviA is involved in regulation of Vi antigen synthesis. *J. Bacteriol.* **178**, 1691–1698.

Wehland, M. & Bernhard, F. 2000 The RcsAB box. Characterization of a new operator essential for the regulation of exopolysaccharide biosynthesis in enteric bacteria. *J. Biol. Chem.* **275**, 7013–7020.

Whitfield, C. & Roberts, I. S. 1999 Structure, assembly and regulation of expression of capsules in *Escherichia coli*. *Mol. Microbiol.* **31**, 1307–1319.

Whittam, T. & McGraw, E. 1996 Clonal analysis of EPEC. *Rev. Microbiol. Sao Paulo* **27** (Suppl. 1), 7–16.

Wieler, L. H., McDaniel, T. K., Whittam, T. S. & Kaper, J. B. 1997 Insertion site of the locus of enterocyte effacement in enteropathogenic and enterohemorrhagic *Escherichia coli* differs in relation to the clonal phylogeny of the strains. *FEMS Microbiol. Lett.* **156**, 49–53.

Worley, M. J., Ching, K. H. & Heffron, F. 2000 *Salmonella* SsrB activates a global regulon of horizontally acquired genes. *Mol. Microbiol.* **36**, 749–761.

Zhou, D., Hardt, W. D. & Galan, J. E. 1999 *Salmonella typhimurium* encodes a putative iron transport system within the centisome 63 pathogenicity island. *Infect. Immun.* **67**, 1974–1981.

Zhu, C., Agin, T. S., Elliott, S. J., Johnson, L. A., Thate, T. E., Kaper, J. B. & Boedeker, E. C. 2001 Complete nucleotide sequence and analysis of the locus of enterocyte Effacement from rabbit diarrheogenic *Escherichia coli* RDEC-1. *Infect. Immun.* **69**, 2107–2115.