The EMBO Journal vol.12 no.10 pp.3779-3787, 1993

Cognate gene clusters govern invasion of host epithelial cells by *Salmonella typhimurium* and *Shigella flexneri*

Eduardo A.Groisman[1] and Howard Ochman[2]

[1]Department of Molecular Microbiology, Washington University School of Medicine, Box 8230, St Louis, MO 63110 and [2]Department of Biology, University of Rochester, Rochester, NY 14627, USA

Communicated by S.Normark

The enteric pathogens *Salmonella typhimurium* and *Shigella flexneri* differ in most virulence attributes including infectivity, pathology and host range. We have identified a new assemblage of genes responsible for invasion properties of *Salmonella* which is remarkably similar in order, arrangement and sequence to the gene cluster controlling the presentation of surface antigens (*spa*) on the virulence plasmid of *Shigella*. In *Salmonella*, this chromosomally encoded complex consists of over 12 genes, mutations in which abolish bacterial entry into epithelial cells. Although these genera use distinct invasion antigens, a non-invasive *spa* mutant of *Salmonella* could be rescued by the corresponding *Shigella* homolog. While *spa* promotes equivalent functions in *Shigella* and *Salmonella*, this constellation of genes has been acquired independently by each genus and displays motifs used by diverse antigen export systems including those required for flagellar assembly and protein secretion.

Key words: gene transfer/invasion/*Salmonella*/*Shigella*/virulence

Introduction

Enteric bacteria including *Salmonella* and *Shigella* are among the principal etiologic agents of diarrheal diseases in mammalian hosts. The initial step in the pathogenesis of shigellosis and salmonellosis consists of bacterial attachment and entry into the epithelial cells that line the upper intestinal tract. Following adhesion and invasion, *Shigella* infections progress by lysis of the surrounding vacuole and spreading to neighboring cells using projections consisting of polymerized actin (Sansonetti, 1991). In contrast, *Salmonella* remains in membrane-bound compartments during transit through epithelial cells to deeper tissues (Finlay *et al.*, 1988a).

Enteric species have adopted distinct strategies for entry into eukaryotic cells (Falkow, 1991; Isberg, 1991). The best characterized system is that mediated by the temperature-regulated Inv protein of *Yersinia* (Isberg *et al.*, 1987). This protein interacts with members of the integrin family of cell adhesion molecules to mediate uptake of Inv-containing microorganisms by non-phagocytic cells (Isberg and Leong, 1990). Yersiniae harbor two additional loci which can direct entry into epithelial cells: the chromosomally encoded *ail* (Miller and Falkow, 1988) and one residing in the 70 kb virulence plasmid (Isberg, 1988). While three independent loci confer invasion properties to *Yersinia*, the genetic determinants necessary for entry of epithelial cells by *Shigella* encompass a 37 kb region from the 220 kb virulence plasmid (Maurelli *et al.*, 1985). In contrast, *Salmonella* invasion has been ascribed to numerous loci distributed around the chromosome suggesting a complex and multifactorial process (Finlay *et al.*, 1988b; Galán and Curtiss, 1989; Betts and Finlay, 1992; Lee *et al.*, 1992; Stone *et al.*, 1992). Moreover, the invasion process is regulated by temperature in *Shigella* (Maurelli *et al.*, 1984, 1985; Small *et al.*, 1987) but by oxygen tension in *Salmonella* (Ernst *et al.*, 1990; Lee and Falkow, 1990; Schiemann and Shope, 1991). Despite these apparent genetic differences, it has been hypothesized that entry into epithelial cells by *Salmonella* and *Shigella* proceeds by a similar mechanism (Francis *et al.*, 1992). For example, both genera promote their phagocytosis into the host cell by inducing actin polymerization in a process that requires the function of microfilaments but not microtubules (Clerc and Sansonetti, 1987; Finlay and Falkow, 1988).

The search for invasion characters in *Salmonella* has identified several determinants including loci that control flagellar and chemotactic activity (Liu *et al.*, 1988; Khoramian-Falsafi *et al.*, 1990; Betts and Finlay, 1992; Jones *et al.*, 1992), cell envelope properties and the production of lipopolysaccharides (Finlay *et al.*, 1988b; McCormick *et al.*, 1988). Several loci, comprising the *inv* complex, were isolated based on their ability to complement a non-invasive strain of *Salmonella typhimurium* (Galán and Curtiss, 1989) and were subsequently mapped to 59 min on the chromosome (Galán *et al.*, 1992) near the *hil* locus, which was identified by mutations allowing invasion under non-physiological conditions (Lee *et al.*, 1992). Although the *inv* genes were originally reported as being unique to *Salmonella* (Galán and Curtiss, 1991), two gene products—InvA and InvE—exhibit some sequence similarities to proteins from *Yersinia* (Galán *et al.*, 1992; Ginocchio *et al.*, 1992). InvA and InvE act to alter both the distribution of polymerized actin microfilaments and the normal architecture of microvilli of infected cells (Galán *et al.*, 1992; Ginocchio *et al.*, 1992). In addition, a locus was recovered from *Salmonella typhi* by its ability to confer invasion properties to *Escherichia coli*, but the corresponding 33 kb region from *S.typhimurium* was ineffective (Elsinghorst *et al.*, 1989).

We have recently characterized a set of clones that were originally described as being confined to the salmonellae (Groisman *et al.*, 1993). One of these clones, RF319, showed hybridization with DNA from plasmid-linked sequences of *Shigella flexneri* and mapped between 57 and 60 min in the *Salmonella* chromosome. These findings raised the possibility that the region included within RF319 encoded determinants required for similar steps in pathogenesis by these two microbes. Although the *inv* complex had also been mapped within this interval (Galán *et al.*, 1992), differences in restriction maps (Fitts, 1985; Galán and Curtiss, 1989) and phylogenetic distribution (Galán and Curtiss, 1991;

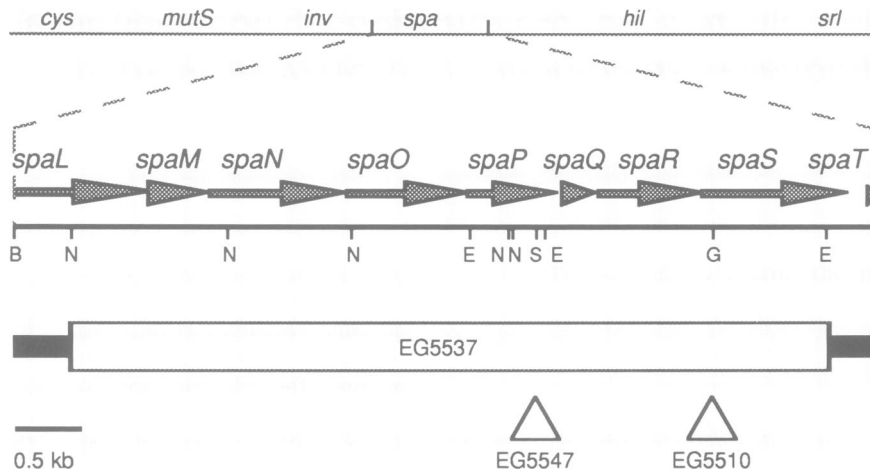

Fig. 1. Genetic and physical maps of the *S.typhimurium* RF319 (*spa*) region. The arrows indicate the size and direction of transcription of the nine ORFs contained within RF319. **Top**: genetic map of the 59–60 min region illustrating the relative positions of the *spa* gene complex and genetic markers in the region. The map is based on our data and those reported by Galán *et al.* (1992) and Lee *et al.* (1992). **Bottom**: restriction map of RF319 and position of *kan* gene insertions and insertion/deletion in three mutant strains. B, *Bam*HI; N, *Nru*I; E, *Eco*RV; S, *Sal*I and G, *Bgl*II.

Groisman *et al.*, 1993) indicated that RF319 and *inv* were quite distinct. In this report, we conducted a molecular and functional analysis of RF319 and established the presence of a suite of genes required for *Salmonella* invasion. In all, this gene assemblage comprises some dozen genes homologous to the region of the *Shigella* virulence plasmid required for the presentation of invasion proteins. It also indicates that antigen export pathways, including those responsible for flagellar assembly, utilize a common protein localization strategy.

Results

Mapping of RF319

The 6.4 kb clone, RF319, was mapped to the 57–60 min region by hybridization to an ordered array of segments from the *S.typhimurium* chromosome (Groisman *et al.*, 1993). To localize RF319 further, we conducted phage P22 transductions with a donor strain harboring a kanamycin resistance marker within the RF319 chromosomal region and a set of recipient strains with markers mapping between 57 and 60 min. RF319::*kan* was 24% linked to *mutS* (59/250) and 4% linked to *hil* (17/404) in the 59.5 min region but not to *proU* (0/252), *cysC* (0/246), *srl* (0/240) or *zfi-2009*::Tn*10* (0/252). To establish if RF319 mapped clockwise or counterclockwise to *mutS* (Sanderson and Roth, 1988), we used Tn*10* to isolate adjacent deletions from a *mutS*::Tn*10* RF319::*kan* strain. We selected for fusaric acid resistance, and tetracycline-sensitive derivatives were screened for kanamycin susceptibility and cysteine prototrophy: all kanamycin-sensitive derivatives were cysteine prototrophs and all cysteine auxotrophs were kanamycin-resistant, suggesting that the gene order is *srl*, *RF319*, *mutS*, *cysJIHDC* (Figure 1). The mapping placed RF319 close to *invA*, which also shows linkage to *mutS* (Galán *et al.*, 1992), suggesting that both RF319 and *inv* may be part of a single complex involved in related activities.

Mutations in the RF319 chromosomal region abolish epithelial cell invasion

To evaluate the function of any genes within the 6.4 kb region contained in RF319, we constructed a deletion that removed ~5.6 kb of *Salmonella* DNA and replaced it by a kanamycin resistance cassette (Figure 1). The mutation was recombined into the chromosome and the recovery of colonies indicated that the deleted segment did not encode essential genes. The resulting strain, EG5537, was prototrophic and behaved like the wild-type parent in its ability to grow at temperatures ranging from 25°C to 42°C, and under anaerobic conditions.

An initial step in *Salmonella* pathogenesis is the entry into the epithelial cells lining the walls of the small intestine. The possibility that RF319 harbored genes responsible for invasion was tested due to (i) its proximity to known invasion loci and (ii) the presence of hybridizing sequences in the virulence plasmid of *S.flexneri* (Groisman *et al.*, 1993). The RF319 deletion mutant exhibited reduced invasion: >250-fold for human intestinal Henle-407 cells and 20-fold for the human laryngeal epithelium HEp-2 cell line (Figure 2A). This experiment defined a new set of genes required for the invasion properties of *Salmonella*. The mutant strain adhered to these cell lines to the same extent as wild-type, confirming previous observations that attachment and invasion are genetically distinct processes in *Salmonella* (Galán and Curtiss, 1989).

Conservation in the gene order and structure of the Salmonella and Shigella invasion regions

The nucleotide sequence of the 6.4 kb insert of RF319 was analyzed in order to identify the determinants contributing to the invasion properties of *Salmonella*. There were nine open reading frames (ORFs) on one strand, eight of which had structural features—such as the presence of overlapping stop and start codons and absence of long intergenic regions—consistent with being part of an operon whose products participate in the same biochemical or assembly pathway. All ORFs except for the first, which is clearly the 3′ end of a gene, contained candidate ribosome binding sites located at the appropriate distance from their start codons. The last ORF is separated by 140 bp and could be part of a distinct transcription unit (the nucleotide sequences have been deposited in the EMBL Data Library under accession number X73525).

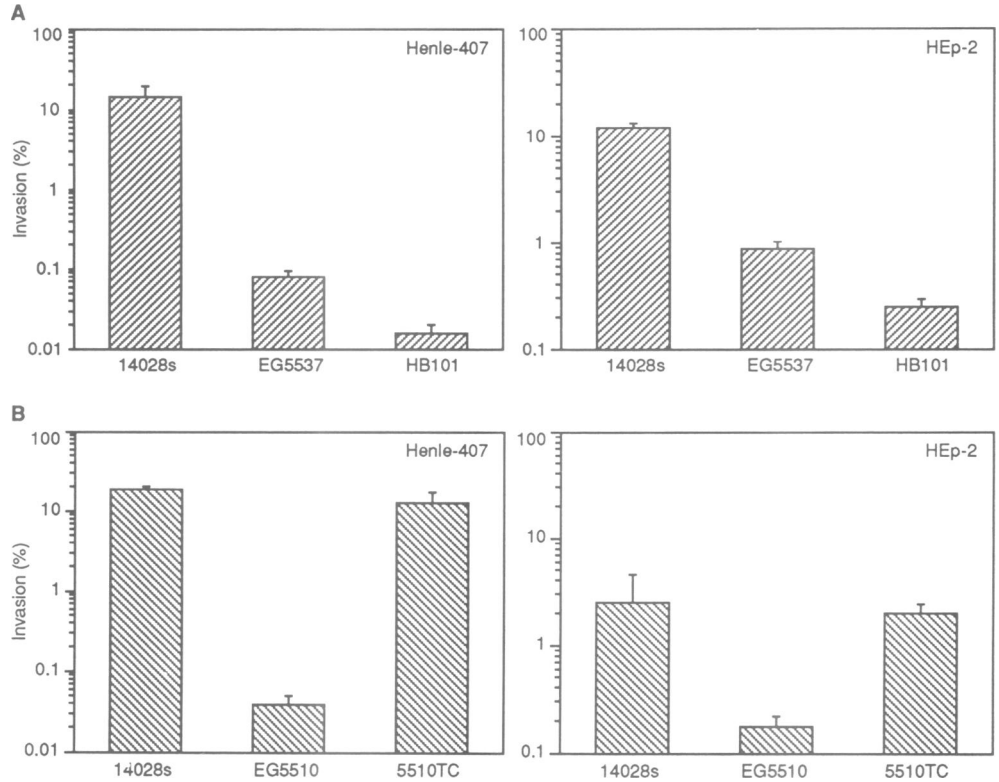

Fig. 2. Invasion properties of (**A**) *S.typhimurium* strains 14028s (wild-type) and EG5537 (Δ*spaLMOPQR5*::*kan*), and non-invasive *E.coli* HB101 and of (**B**) *S.typhimurium* strains 14028s (wild-type), EG5510 (*spaS*::*kan*) and EG5510TC (*spaS*$^+$ derivative of EG5510). Invasion assays were performed as described in Materials and methods. Values represent the mean of triplicate samples ± SD.

The gene products encoded by RF319 are highly similar to the surface presentation antigens of the *S.flexneri* virulence plasmid (Venkatesan *et al.*, 1992). This similarity extends beyond the primary sequence and predicted secondary structures of the individual proteins: the arrangement of the corresponding genes in *Shigella* and *Salmonella* is conserved with respect to size, gene order and orientation. Based on these characteristics, the products are likely to have similar functions, so the sequenced genes were named *spa* (for *s*urface *p*resentation of *a*ntigens), in analogy to their counterparts in *Shigella*. These genes are also similar to the flagellar assembly loci of *Bacillus subtilis* and enteric bacteria in both primary sequence and arrangement (Albertini *et al.*, 1991; Macnab and Parkinson, 1991; Bischoff *et al.*, 1992).

Mosaic patterns of genetic relationships within the spa region

We searched the GenBank database (release 73) for nucleotide or peptide sequences bearing similarity to the *spa* genes and their predicted products. The first sequence, corresponding to SpaL, had the highest degree of sequence identity, 59% (74% similarity introducing a gap of only one amino acid in length), with the Spa47 protein of *Shigella* in the 326 amino acid region of overlap (Figure 3). SpaL was also 43–44% identical to the FliI proteins of *S.typhimurium* and *Bacillus subtilis* which are required for the assembly of flagella (Albertini *et al.*, 1991; Vogler *et al.*, 1991), and to a lesser extent (28–33% identity) with several other ATP-utilizing proteins of both prokaryotic and eukaryotic origin (Kibak *et al.*, 1992). The regions of similarity with these ATPases were limited to two clusters believed to constitute the nucleotide binding sites (Walker *et al.*, 1982). In contrast, the sequence identity of SpaL with the *Shigella* Spa47 protein, which is required for the delivery of invasion antigens to the correct subcellular location (Venkatesan *et al.*, 1992), extended across the entire protein. SpaP was 61% identical to the Spa24 protein of *Shigella*, 38% identical to a protein in *Xanthomonas* required for plant pathogenicity (Hwang *et al.*, 1992), and 32% identical to the FliP protein of the *B.subtilis* flagellar export apparatus (Figure 3). Although no genes have been identified immediately downstream of *spa24* in the *S.flexneri* plasmid, the hydrophobic protein SpaQ exhibited 78% identity with the sequenced portion of an unreported ORF adjacent to *spa24*. The 86 amino acid SpaQ was 35% identical to the 89 amino acid *B.subtilis* FliQ (Figure 3), which is encoded by a gene located downstream of the *spaP* homolog *fliP* (Bischoff *et al.*, 1992).

Between *spaL* and *spaP* there are three additional ORFs—*spaM*, *spaN* and *spaO*—which are coincident with the *Shigella spa* and the *Bacillus* and *Salmonella fli* loci. Despite the remarkable similarity exhibited by *spaL* and *spaP*, the extent of amino acid identity decreased dramatically within this region: SpaO, although of corresponding length to Spa33 of *Shigella*, had only 24% identity; SpaN exhibited just 19% identity with Spa32 even after the introduction of nine gaps into the alignment; SpaM showed 22% identity with a previously unrecognized ORF located between *spa47* and *spa32* of Shigella (Figure 3). The 147 amino acid SpaM is identical in length to the *Salmonella* FliJ protein but displays only 18% identity. The C-terminus of SpaO exhibited 32% sequence identity (57% similarity with no gaps) with the last 80 amino acids of FliN of *S.typhimurium*, which is about half its size. As noted, sequence information from the

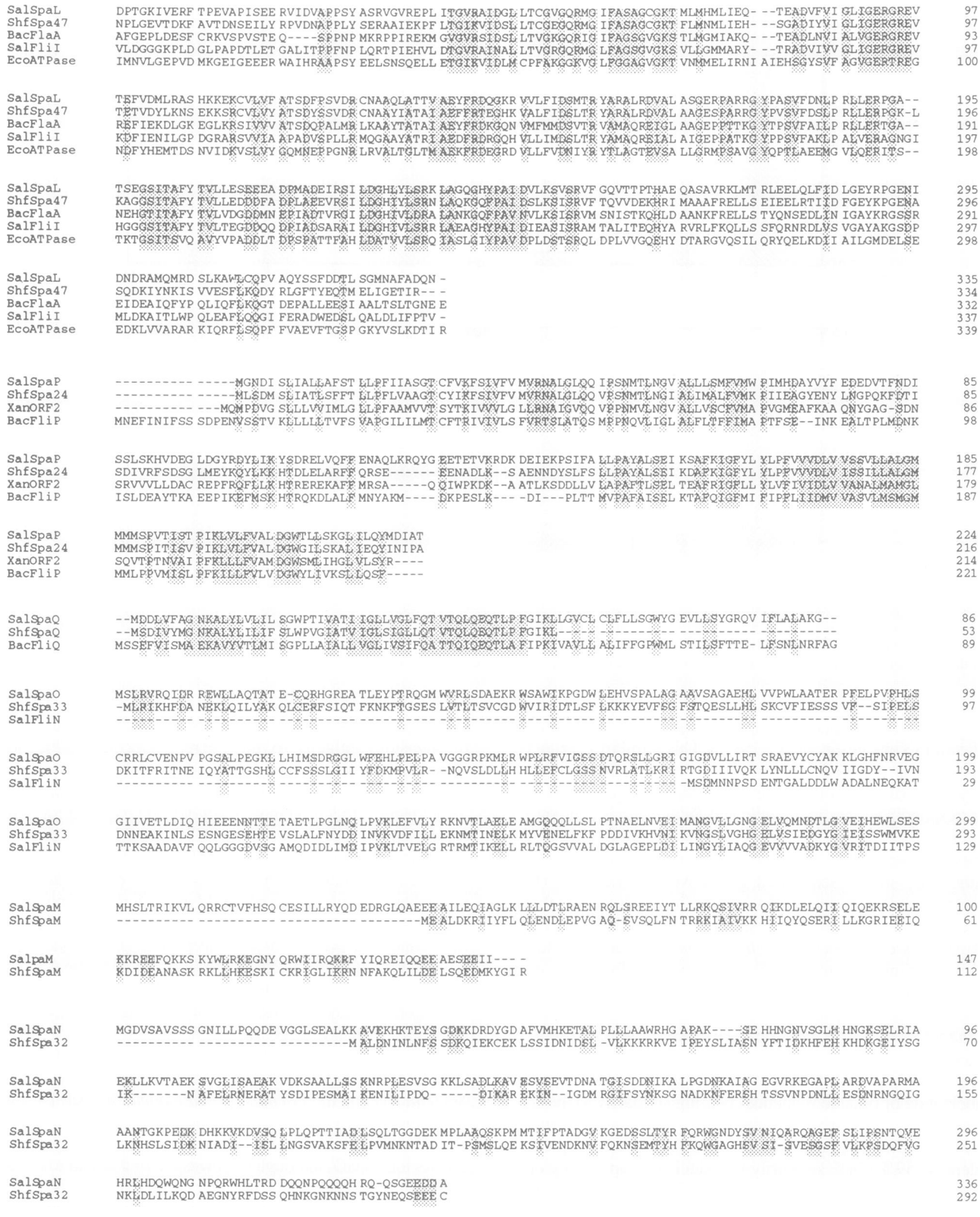

Fig. 3. Comparison of *Salmonella* Spa with proteins of antigen export systems. The deduced amino acid sequences of six proteins of the *S.typhimurium* Spa complex are shown in the single letter code. For any given comparison, residues present in the majority of the related proteins are highlighted. We used the GeneWorks alignment algorithm with a cost of 1 (this algorithm scores mismatches and gaps rather than matches, with lower cost signifying better alignment; identities have a cost of 0). Sequences were obtained from GenBank (release 73).

Shigella virulence plasmid ends within the *spaQ* coding region and the three additional genes encoded by RF319—*spaR*, *spaS* and *spaT*—had no homologs in the database.

SpaS is a novel protein required for the invasion of epithelial cells

The N-terminus of SpaS contained a motif reminiscent of energy transport proteins from mitochondria such as the

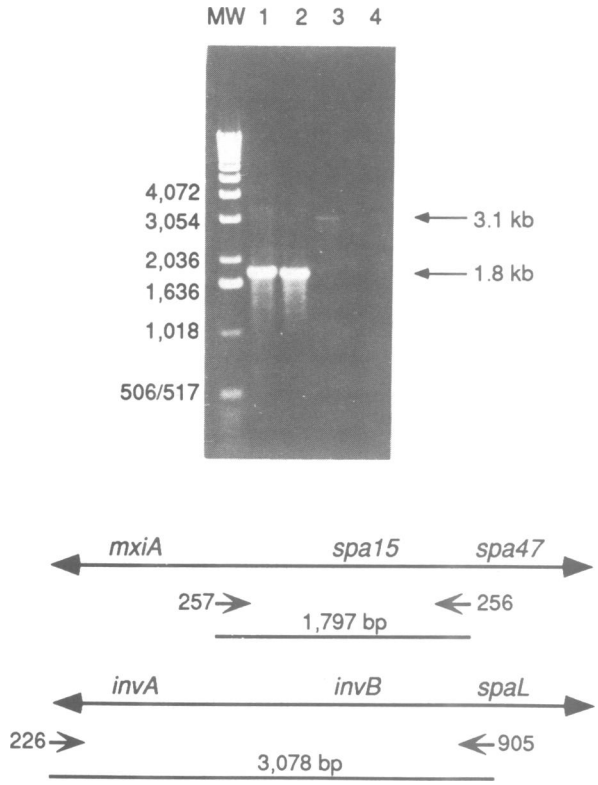

Fig. 4. Determination of the physical distance between the *inv* and *spa* invasion regions. **Top**: ethidium bromide-stained agarose gel of PCR products obtained by amplification of DNA sequences from the *S.typhimurium* 14028s chromosome (lanes 1 and 3) and the enteroinvasive plasmid pSF204 (lanes 2 and 4). Products obtained using primers 256 and 257 were loaded in lanes 1 and 2, and those obtained with primers 905 and 226 were loaded in lanes 3 and 4. **Bottom**: distances between the two sets of primers as deduced from the DNA sequences of the *mxi/spa* regions (Andrews and Maurelli, 1992; Venkatesan *et al.*, 1992).

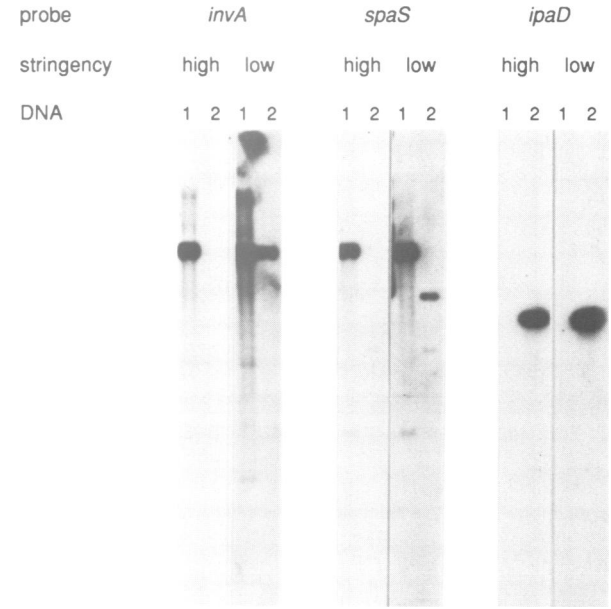

Fig. 5. Presence or absence of invasion loci in enteric bacteria. Southern hybridization of *Pst*I-digested DNA from *S.typhimurium* 14028s (lane 1) and of purified enteroinvasive plasmid pSF204 (lane 2). Gene-specific probes were generated by amplification of the coding regions of each locus by the PCR. High and low stringency conditions for hybridizations are as described in Materials and methods.

ADP/ATP translocase and the phosphate carrier protein (Klingenberg, 1990). It was of interest to establish whether SpaS was required for invasion because of potential interactions with other proteins within this cluster, particularly SpaL, which contains ATP binding sites. Moreover, the SpaL homolog in *S.flexneri*—Spa47—is required for entry of epithelial cells (Venkatesan *et al.*, 1992). We found that a *spaS*::*kan* mutant was defective for invasion and that this defect could be rescued following transduction with the wild-type allele using a linked Tn*10* insertion (Figure 2B). Thus *spaS* (or a gene downstream of *spaS*) is an invasion determinant not previously described in *Salmonella* or in other invasive pathogens.

The similarity between Salmonella and Shigella extends beyond the spa regions

In *Shigella*, the *spa* genes are located immediately downstream of the *mxi* loci which have also been implicated in the membrane export of antigens and invasive properties of this pathogen (Allaoui *et al.*, 1992b, 1993; Andrews and Maurelli, 1992). *Salmonella* InvA and InvE have sequence identity with the *Shigella* MxiA and MxiC proteins, respectively, and are encoded by genes which, like *spa*, are closely linked to *mutS*. To examine whether the organization of genes outside of the *spa* region was conserved in both *Salmonella* and *Shigella*, we determined the physical distance between *invA* and *spaL* using the polymerase chain reaction (PCR). Employing primers designed to anneal to conserved regions of the corresponding genes, we obtained amplification products of ~1.8 kb for both *Salmonella* and *Shigella* (Figure 4), as predicted from the published sequences of this region in the *Shigella* plasmid (Venkatesan *et al.*, 1992). Using primers to non-conserved regions, we obtained a 3.1 kb product for *Salmonella* corresponding to the expected size of this region in *Shigella*. The physical distance between *invA* and *spaL* determined by this PCR cannot accommodate a gene encoding the 64 kDa InvB protein as proposed by Galán and Curtiss (1989), and it is likely that *invB* encodes a product of ~15 kDa, like its *Shigella* counterpart, *spa15* (Venkatesan *et al.*, 1992). Moreover, based on its reported size and that of Spa47, InvC is likely to correspond to SpaL.

To define the boundaries of the region conserved between *Salmonella* and *Shigella*, we performed a series of hybridization experiments using sequences that had been described only in one of the two genera. As expected from the high degree of nucleotide sequence identity (63%), we detected homologous sequences in the invasion plasmid pSF204 using the *S.typhimurium invA* gene as a probe (Figure 5). This experiment was necessary to establish the proper hybridization conditions for sequences of this degree of identity since Galán and Curtiss (1991) could not detect *invA* homologs in other invasive enteric species. Using a probe corresponding to the coding region of *Salmonella spaS* we detected homologs in the enteroinvasive-derived plasmid pSF204 (Figure 5). In contrast, we found no sequences in *S.typhimurium* that would cross-hybridize with *ipaD* (Figure 5) which encodes one of the recognized invasion antigens in *S.flexneri* (Venkatesan *et al.*, 1988b). Cumulatively, a complex of at least 12 contiguous genes is

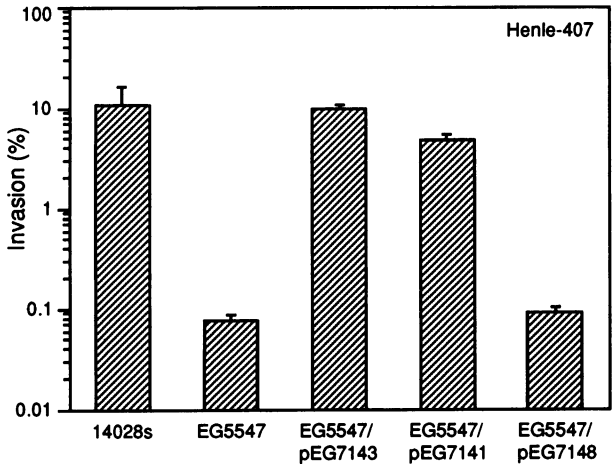

Fig. 6. Rescue of invasion-deficient *Salmonella* with *Shigella* invasion determinant. Invasion properties of *S.typhimurium* strains 14028s (wild-type), EG5547 (*spaP*::*kan*), and EG5547 derivatives harboring either plasmid pEG7143 (*Salmonella spaP*), pEG7141 (*Shigella spa24*) or pEG7148 (*Shigella spa24*::*kan*). Invasion assays were performed as described in the legend to Figure 2.

conserved between the plasmid-borne invasion locus of *Shigella* and the chromosomal *inv/spa* cluster of *Salmonella*. The absence of IpaD—and perhaps of other neighboring Ipa export antigens—from *Salmonella* could account for some of the observed differences in the invasion process by these two genera.

Complementation of an invasion-deficient Salmonella with the spa homolog from Shigella

While *Salmonella* and *Shigella* have distinct antigens to promote phagocytosis by epithelial cells, their invasion systems appear to be very similar. To investigate the functional equivalence of the *spa* genes from these genera, we tested the ability of *spa24* from *Shigella* to complement a non-invasive *spaP* strain of *Salmonella*. *spaP* was disrupted by inserting a 1.3 kb *kan* cassette within its coding region and the resulting strain was unable to invade Henle-407 epithelial cells. By introducing a plasmid containing the *Shigella spa24* gene into the *spaP* mutant we could rescue the invasion phenotype to the same levels obtained with a plasmid harboring the wild-type *S.typhimurium spaP* gene (Figure 6). As expected, a derivative of the complementing plasmid containing a *kan* insertion within *spa24* could not complement the *Salmonella spaP*. The complementing plasmid harbored the coding region of *spa24*, and no additional *Shigella* DNA establishing a role for *spaP* in invasion and demonstrating that Spa24 of *Shigella* could function with the *Salmonella* invasion machinery.

Discussion

The invasion mechanisms of enteric bacteria

Similar host cell morphological changes accompany entry of *Shigella* and *Salmonella* into epithelial cells leading Falkow and co-workers to propose that these organisms use a similar mechanism for invasion (Francis *et al.*, 1992). However, *Shigella* and *Salmonella* follow distinct intracellular pathways upon invasion of epithelial cells (Sansonetti, 1991) and the different routes adopted by these microbes could be determined by the entry step. For example, the course of entry employed by *Toxoplasma gondii* into Fc receptor-containing fibroblasts determines whether this parasite will reside in vacuoles that are capable of acidification and fusion with other organelles (Joiner *et al.*, 1990).

Despite the close genetic relatedness of *Salmonella* and *Shigella*, the discovery of such an extensive set of cognate genes was unforeseen given the known differences in the patterns of infection and the genetic location of invasion loci in each genus (Finlay and Falkow, 1989; Hale, 1991). After entry into host tissues, *Shigella* lyses the membrane of the phagocytic vacuole (Sansonetti *et al.*, 1986) and spreads through the cytoplasm to infect adjoining cells (Allaoui *et al.*, 1992b). In contrast, Salmonellae are enclosed within the phagocytic vacuole and transported through the epithelial cells before infecting other tissues (Finlay and Falkow, 1988; Finlay *et al.*, 1988a). The underlying differences between these pathogens go beyond the processes seen within epithelial cells: *Shigella* spp. have a very restricted host range, infecting only humans and other primates, and most of the identified genes required for virulence are plasmid-borne (Hale, 1991), whereas *Salmonella* spp. have been recovered from a wide range of animal species (Brenner, 1992) and their virulence determinants have been mapped principally to chromosomal locations (Curtiss and Kelly, 1987; Miller *et al.*, 1989; Groisman *et al.*, 1992; Stone *et al.*, 1992). Although the *spa* gene complex is present in both *Shigella* and *Salmonella*, the complement of genes responsible for invasion is not fully identical. In *Shigella*, the Spa proteins serve to export the invasion plasmid antigens—known as Ipa proteins—to the proper subcellular location (Venkatesan *et al.*, 1992) but homologs to the *ipa* genes could not be detected in *S.typhimurium* (Figure 5; Venkatesan *et al.*, 1988a). While the present study demonstrates that both pathogens use an equivalent set of genes to deliver the relevant invasion proteins, the antigens required for entry by *Salmonella* remain to be identified.

Spa exemplifies a distinct secretory pathway

Many of the Spa proteins display features, such as being extremely hydrophobic but devoid of signal sequences, recognized in other products that function in protein transport including those involved in flagellar assembly, plant pathogenicity and in secretion of virulence proteins in *Yersinia*. This family of proteins constitute a third secretory pathway in bacteria, distinct from both the classical *sec*-dependent pathway—which includes a periplasmic intermediate—and the Hly and Cva systems responsible for the secretion of hemolysin and colicin V in *E.coli* (Salmond and Reeves, 1993). While complementation has been observed between the hemolysin and colicin secretion systems (Fath *et al.*, 1991), Hly of *E.coli* could not rescue the export defect of *Shigella* strains with mutations in the *mxi/spa* region (Andrews *et al.*, 1991). On the other hand, we have shown the restoration of invasion in a *Salmonella* mutant by a *Shigella* Spa protein (Figure 6) demonstrating complementation between members of this secretory pathway (Salmond and Reeves, 1993).

Structural and functional diversity of invasion determinants

The *inv/spa* complex comprises at least 12 overlapping or adjoining genes suggesting a single transcription unit. While

the overall sizes and organization of genes within the *spa* region are the same in *Salmonella* and *Shigella* (Figure 7), there are sharp discontinuities in the extent of sequence divergence among the Spa homologs. SpaL has a high degree of sequence similarity with Spa47 but it is followed by a region encoding proteins (SpaM, SpaN and SpaO) that exhibit only about 20% identity with their *Shigella* counterparts. Downstream of these sequences, SpaP exhibits a high level of sequence identity (61%) with its *Shigella* homolog. The mosaic pattern of sequence identities, with very divergent sequences bordered by highly conserved sequences, results from differential selective constraints on invasion determinants encoded within the *spa* complex. That selection could be acting to conserve the functions of SpaL and SpaP is supported by the ability of *spa24* from *Shigella* to replace the homologous gene in *Salmonella*. Moreover, both genes utilize a very limited set of synonymous codons indicative of constraints at the nucleotide level. On the other hand, the internal genes (*spaM*, *spaN* and *spaO*) could be accumulating variation at an enhanced rate due to selection for diversification, as observed for other antigen-associated loci such as *fljB* in *Salmonella* (Smith and Selander, 1991) and the histocompatibility complex in mammals (Hughes and Nei, 1988), and provide specificity to the export apparatus, perhaps by interacting with the invasion antigens themselves, to determine the invasion properties of each genus.

The fact that chloroform- or heat-killed *S.typhimurium* cells can adhere to, but do not enter, epithelial cells (Francis *et al.*, 1992) suggests that invasion is an energy-dependent process. Two of the proteins encoded by the *spa* complex contain nucleotide-binding motifs: SpaL has the two ATP-binding domains typical of ATPases of prokaryotic and eukaryotic origin (Kibak *et al.*, 1992), and SpaS, which has been shown to be involved in invasion (Figure 2), contains a motif that defines a group of mitochondrial proteins responsible for energy transfer. While these mitochondrial proteins harbor the motif twice (Klingenberg, 1990), we have found that SpaS, like other bacterial proteins that interact with nucleotides (such as translation initiation factors and DNA polymerase III), contain a single copy of this motif. SpaO, one of the least conserved proteins of the Spa complex, exhibited partial identity with *Salmonella* FliN, a protein half its size which was originally thought to be part of the flagellar motor switch and later found to be also necessary for assembly/export of this organelle (Irikura *et al.*, 1993). The region of identity is the same as that exhibited by FliY, a larger *Bacillus* protein which can functionally complement the motility—but not the chemotaxis—defect of a *S.typhimurium fliN* mutant (Bischoff and Ordal, 1992).

Origin of invasion gene complexes

Both the phylogenetic distribution and anomalous base composition of the *spa* region indicate that salmonellae acquired these genes by horizontal gene transfer. The nucleotide content of *spa* genes was 47% G+C, much lower than the 51−53% G+C estimated for the entire *Salmonella* chromosome (Fasman, 1976). Because the majority of species typed to *Salmonella* contain sequences hybridizing to *spa* (Fitts, 1985) it is likely that the entire gene complex is ancestral to the genus and that the characteristics of contemporary *spa* genes reflect, in part, an amelioration in nucleotide content and codon usage to that more typical of *Salmonella* genes.

The order and structure of the *spa* genes are conserved in *Shigella*, where the genes are plasmid-borne and have a G+C content of only 35%, much lower than the average of 52% for the *Shigella* chromosome. To date, all of the sequenced genes from the *Shigella* virulence plasmid are very AT-rich (only 25−35% G+C), which has been interpreted as evidence that virulence determinants were acquired from a distantly related organism (Hale, 1991). *Shigella* spp., which are known to be derived from *E.coli* (Ochman *et al.*, 1983), have a relatively recent origin and could not be the source of the *inv/spa* complex in salmonellae. Conversely, the extremely AT-rich *mxi* and *spa* genes of *S.flexneri* could not have been derived from their moderately AT-rich homologs in *Salmonella*.

The similarity between certain Inv proteins of *Salmonella* (Galán *et al.*, 1992; Ginocchio *et al.*, 1992) and Mxi proteins of *Shigella* (Allaoui *et al.*, 1992b; Andrews and Maurelli, 1992) to sequences involved in *Yersinia* virulence has led others to suggest that some of the genes required for invasion were acquired from this genus. Although certain surface antigens of *Yersinia* (Yop) are broadly similar to the Ipa proteins of *Shigella* (Håkansson *et al.*, 1993) and the G+C content of Yersinia is 46% (Fasman, 1976), the chromosomal organization within the *spa* complex makes this genus a rather unlikely source for the invasion genes. In sum, the *spa* gene cluster is not ancestral to enteric bacteria but, rather, was acquired independently by *Shigella* and *Salmonella* [making these genes 'xenologous' rather than 'homologous' using the terminology proposed by Gray and Fitch (1983)], and this assemblage of genes has been adopted for the invasion of host epithelial cells by each genus.

Salmonella-specific sequences and the evolution of complex pathways

One of the remarkable features of the *inv/spa* complexes of *Shigella* and *Salmonella* is that the sizes and positions of the genes are mantained even in regions of low sequence identity. While several authors have noted the similarity of individual invasion determinants of *Shigella* and *Salmonella* with certain flagellar genes from enteric bacteria (Galán *et al.*, 1992; Venkatesan *et al.*, 1992), our data showed the structural integrity of large genetic regions controlling protein export systems (Figure 7) and indicated that certain gene complexes can be adapted by divergent taxa for related functions.

The present study has shown that in addition to point-mutational evolution, bacterial chromosomes can acquire large assemblages of genes capable of conferring novel and complex phenotypes to the organism. Molecular analyses of the *Salmonella* chromosome have already provided evidence that such processes have a significant role in molding the bacterial genome. For example, large regions of the *Salmonella* chromosome, such as those responsible for vitamin B_{12} biosynthesis (J.G.Lawrence and J.R.Roth, personal communication) and for the production of the multiple forms of surface polysaccharides (Reeves, 1993) have been acquired by horizontal gene transfer. The incongruities in the broadly conserved genomes of *E.coli* and *S.typhimurium* (Riley and Sanderson, 1990) reflect the incorporation of large DNA segments which help each species prosper in unique environmental niches. Since it

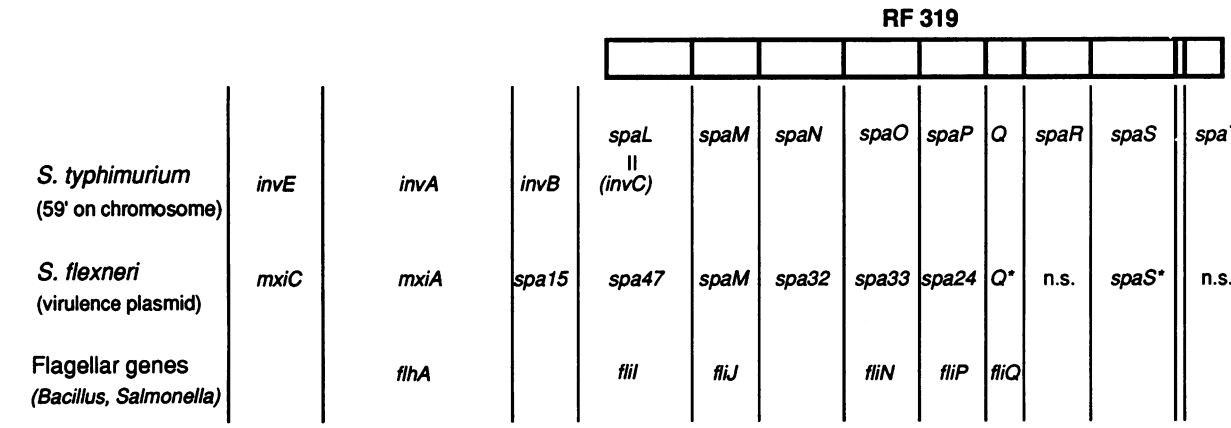

Fig. 7. Organization of gene complexes encoded by protein export and assembly systems. Relative sizes of genes are shown. Asterisks denote unidentified genes detected by hybridization or sequence comparisons. N.S.: no sequence information available in GenBank (release 73).

diverged from *E.coli* some 100 million years ago, *Salmonella* has acquired several gene complexes, including the *inv/spa* cluster, which has enabled the differential exploitation of animal hosts.

Materials and methods

Bacterial strains, plasmids and growth conditions

All *S.typhimurium* strains were derived from the mouse-virulent, invasion-competent 14028s (lab collection). Strain genotypes are as follows: EG5510, *spaS*::*kan*; EG5537, Δ*spaLMNOPQRS*::*kan*; and EG5547, *spaP*::*kan*. Plasmids RF319 (Fitts, 1985) and pSF204 (Small and Falkow, 1988) were obtained from the Agricultural Research Culture collection (Peoria, IL) and Patrick Bavoil (University of Rochester), respectively. Plasmids used are as follows. pMS319: pUC19 with *Salmonella Bam*HI fragment from RF319; pEG7143: pUC19 with *spaP*; pEG7141: pUC19 with *spa24*; and pEG7148: pUC19 with *spa24*::*kan*. The *S.typhimurium* 14028s- and *S.flexneri* M90T-derived sequences in plasmids pEG7143 and pEG7141, respectively, were obtained by the PCR. Plasmid pEG7148 harbored the *Bam*HI *kan* fragment from plasmid pUC4-K inserted into the *Bgl*II site of plasmid pEG7141, within the *spa24* coding region. LB and M63 minimal media have been described (Miller, 1972). Ampicillin was used at 50 μg/ml, chloramphenicol at 25 μg/ml, kanamycin at 40 μg/ml and tetracycline at 10 μg/ml. HEp-2 cells, an epidermal carcinoma cell line derived from human larynx (obtained from Joe St Geme), were grown in Dulbecco's modified Eagle's medium (DMEM; Gibco) + 10% fetal calf serum (Gibco) supplemented with D-glucose and L-glutamine. Henle-407 cells (obtained from Roy Curtiss III) were grown in MEM (Gibco) supplemented with Earle's salts, L-glutamine, 10% fetal calf serum and 1% non-essential amino acids.

Construction of S.typhimurium strains with mutations in the RF319 region

Strains harboring mutations in the RF319 chromosomal region were constructed as described in Groisman *et al.* (1993) using pMS319 derivatives harboring insertions of either of two genes that confer resistance to kanamycin. Strain EG5510 harbors the *Bam*HI 1.3 kb *kan* fragment from plasmid pUC4-K (Pharmacia) inserted at the *Bgl*II site in RF319 in the opposite orientation relative to the *spa* genes (Figure 1). Strain EG5537 contains the 1.3 *Hin*cII fragment of pUC4-K inserted between the outermost *Nru*I and *Eco*RV sites with the *kan* fragment oriented in the same orientation as *spa* (Figure 1). Strain EG5547 has the *Sal*I 1.3 kb *kan* fragment from plasmid pUC4-KIXX (Pharmacia) inserted at the unique *Sal*I site in RF319. This *kan* gene is present in the same orientation as the *spa* operon and its promoter transcribes the genes located downstream of *spaP*. The structure of the genes in the mutant strains was verified by Southern hybridization analysis using both the 6.4 kb *Bam*HI fragment of RF319 and *kan*-specific probes. Transformation of *S.typhimurium* with plasmid DNA was carried out by electroporation with a GenPulser apparatus according to the manufacturer's recommendations (Bio-Rad). Insertions of mini-Mu elements into plasmid pMS319 were performed as described in Groisman (1991).

Adhesion and invasion assays

Assays were performed in 24-well tissue culture plates (Corning) following the protocol described in Lee *et al.* (1992) with the following modifications: infected monolayers were incubated for 3 h rather than 1 h and dilutions of bacterial cells were performed in PBS instead of LB broth. The percentage of cell-associated bacteria was calculated by dividing the number of bacteria that remained after the initial 3 h of incubation by the number present in the original inoculum, and multiplied by 100. The percentage of intracellular bacteria was calculated by dividing the number of gentamicin-resistant bacteria by the number present in the original inoculum, and multiplied by 100.

DNA sequencing, analysis and other molecular biological techniques

Double-stranded DNA sequencing was carried out on both strands of plasmid pMS319 and derivatives harboring mini-Mu insertions, by the dideoxy chain-termination method using the Sequenase kit (United States Biochemicals) with ^{35}S-labeled dATP. Primers were complementary to the ends of Mu or synthesized as the partial sequences were obtained. Sequence analyses were performed using both GeneWorks (Intelligenetics) and the GCG package (University of Wisconsin). Southern hybridizations were performed as described by Maniatis *et al.* (1982) using as probes PCR-generated DNA fragments corresponding to the coding regions of each of the genes. High stringency conditions were as described by Groisman *et al.* (1993) except that the washes were carried out for 20 min. Hybridizations under low stringency conditions were carried out for 18 h at 54°C (in the same buffer as for high stringency) and followed by one wash at 25°C for 15 min and two washes at 54°C for 20 min each in the following buffer: 1 × standard saline citrate/0.1% SDS. The PCR experiments described in Figure 4 were carried out as follows: denaturation at 94°C for 1 min, annealing at 46°C for 1 min and extension at 72°C for 3 min using the following primers: 256, 5′-GCAAAAATGCCCATTCGCTG-3′; 257, 5′-GTCAGTTCTTT-ATTGATTATGG-3′; 226, 5′-AGTGCTGCTTTCTCTAC-3′; and 905: 5′-CAACATATCCACGAATTCAGTG-3′. The nucleotide sequences reported in this paper have been deposited in the EMBL Data Library under accession number X73525).

Acknowledgements

We thank Michael Baer, Felix Solomon, Mark Sturmoski and Robert Lin for technical assistance, and Ken Sanderson, Cathy Lee and Patrick Bavoil for strains and plasmids. Joe St Geme provided assistance with invasion assays. This work was supported in part by NIH grants RO1 AI29554 (EAG) and GM40995 (HO) and by NSF grant DCB-8916403 (EAG).

References

Albertini,A.M., Caramori,T., Crabb,W.D., Scoffone,F. and Galizzi,A. (1991) *J. Bacteriol.*, **173**, 3573–3579.

Allaoui,A., Mounier,J., Prévost,M.-C., Sansonetti,P.J. and Parsot,C. (1992a) *Mol. Microbiol.*, **6**, 1605–1616.

Allaoui,A., Sansonetti,P.J. and Parsot,C. (1992b) *J. Bacteriol.*, **174**, 7661–7669.

Allaoui,A., Sansonetti,P.J. and Parsot,C. (1993) *Mol. Microbiol.*, **7**, 59–68.

Andrews,G.P. and Maurelli,A.T. (1992) *Infect. Immun.*, **60**, 3287–3295.

Andrews,G.P., Hromockyj,A.E. and Coker,C. (1991) *Infect. Immun.*, **59**, 1997–2005.

Betts,J. and Finlay,B.B. (1992) *Can. J. Microbiol.*, **38**, 852–857.
Bischoff,D.S. and Ordal,G.W. (1992) *Mol. Microbiol.*, **6**, 2715–2723.
Bischoff,D.S., Weinreich,M.D. and Ordal,G.W. (1992) *J. Bacteriol.*, **174**, 4017–4025.
Brenner,D.J. (1992) In Balows,H., Trüper,H.G., Dworkin,M., Harder,W. and Schliefer,K.-H. (eds), *The Prokaryotes*. Springer, New York, Vol. 3, pp. 2673–2696.
Clerc,P. and Sansonetti,P.J. (1987) *Infect. Immun.*, **55**, 2681–2688.
Curtiss,R.,III and Kelly,S.M. (1987) *Infect. Immun.*, **55**, 3035–3043.
Elsinghorst,E.A., Baron,L.S. and Kopecko,D.J. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5173–5177.
Ernst,R.K., Dombroski,D.M. and Merrick,J.M. (1990) *Infect. Immun.*, **58**, 2014–2016.
Falkow,S. (1991) *Cell*, **65**, 1099–1102.
Fasman,G.D. (1976) *CRC Handbook of Biochemistry and Molecular Biology*. CRC Press, Cleveland.
Fath,M.J., Skvirsky,R.C. and Kolter,R. (1991) *J. Bacteriol.*, **173**, 7549–7556.
Finlay,B.B. and Falkow,S. (1988) *Biochimie*, **70**, 1089–1099.
Finlay,B.B. and Falkow,S. (1989) *Mol. Microbiol.*, **3**, 1833–1841.
Finlay,B.B., Gumbiner,B. and Falkow,S. (1988a) *J. Cell Biol.*, **107**, 221–230.
Finlay,B.B., Starnbach,M.N., Francis,C.L., Stocker,B.A.D., Chatfield,S., Dougan,G. and Falkow,S. (1988b) *Mol. Microbiol.*, **2**, 757–766.
Fitts,R. (1985) *Food Technol.*, **39**, 95–102.
Francis,C.L., Starnbach,M.N. and Falkow,S. (1992) *Mol. Microbiol.*, **6**, 3077–3087.
Galán,J. and Curtiss,R. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 6383–6387.
Galán,J. and Curtiss,R. (1991) *Infect. Immun.*, **59**, 2901–2908.
Galán,J., Ginocchio,C. and Costeas,P. (1992) *J. Bacteriol.*, **174**, 4338–4349.
Ginocchio,C., Pace,J. and Galán,J. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 5976–5980.
Gray,G.S. and Fitch,W.M. (1983) *Mol. Biol. Evol.*, **1**, 57–66.
Groisman,E.A. (1991) *Methods Enzymol.*, **204**, 180–212.
Groisman,E.A., Parra-Lopez,C., Salcedo,M., Lipps,C.J. and Heffron,F. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 11939–11943.
Groisman,E.A., Sturmoski,M.A., Solomon,F.R., Lin,R. and Ochman,H. (1993) *Proc. Natl Acad. Sci. USA*, **90**, 1033–1037.
Håkansson,S., Bergman,T., Vanooteghem,J.-C., Cornelis,G. and Wolf-Watz,H. (1993) *Infect. Immun.*, **61**, 71–80.
Hale,T.L. (1991) *Microbiol. Rev.*, **55**, 206–224.
Hughes,A. and Nei,M. (1988) *Nature*, **355**, 167–170.
Hwang,I., Lim,S. and Shaw,P. (1992) *J. Bacteriol.*, **174**, 1923–1931.
Irikura,V., Kihara,M., Yamaguchi,S., Sockett,H. and Macnab,R. (1993) *J. Bacteriol.*, **175**, 802–810.
Isberg,R.R. (1988) *Infect. Immun.*, **57**, 1998–2005.
Isberg,R.R. (1991) *Science*, **252**, 934–938.
Isberg,R.R. and Leong,J.M. (1990) *Cell*, **60**, 861–871.
Isberg,R.R., Voorhis,D.L. and Falkow,S. (1987) *Cell*, **50**, 769–778.
Joiner,K.A., Fuhrman,S.A., Miettinen,H.M., Kasper,L.H. and Mellman,I. (1990) *Science*, **249**, 641–646.
Jones,B., Lee,C. and Falkow,S. (1992) *Infect. Immun.*, **60**, 2475–2480.
Khoramian-Falsafi,T., Harayama,S., Kutsukake,K. and Pechère,J.C. (1990) *Microb. Pathog.*, **9**, 47–53.
Kibak,H., Taiz,L., Starke,T., Bernasconi,P. and Gogarten,J.P. (1992) *J. Bioenerg. Biomembr.*, **24**, 415–424.
Klingenberg,M. (1990) *Trends Biochem. Sci.*, **15**, 108–112.
Lee,C. and Falkow,S. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 4304–4308.
Lee,C., Jones,B. and Falkow,S. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 1847–1851.
Liu,S.-L., Ezaki,T., Miura,H., Matsui,K. and Yabuuchi,E. (1988) *Infect. Immun.*, **56**, 1967–1973.
Macnab,R. and Parkinson,J.S. (1991) *Trends Genet.*, **7**, 196–200.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Maurelli,A.T., Blackmon,B. and Curtiss,R.I. (1984) *Infect. Immun.*, **43**, 195–201.
Maurelli,A., Baudry,B., D'Hauteville,H., Hale,T. and Sansonetti,P.J. (1985) *Infect. Immun.*, **49**, 164–171.
McCormick,B.A., Stocker,B.A.D., Laux,D.C. and Cohen,P.S. (1988) *Infect. Immun.*, **56**, 2209–2217.
Miller,J.H. (1972) *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Miller,S.I., Kukral,A.M. and Mekalanos,J.J. (1989) *Proc. Natl Acad. Sci. USA*, **86**, 5054–8.
Miller,V.L. and Falkow,S. (1988) *Infect. Immun.*, **56**, 1242–1248.
Ochman,H., Whittam,T.S., Caugant,D.A. and Selander,R.K. (1983) *J. Gen. Microbiol.*, **129**, 2715–2726.
Reeves,P. (1993) *Trends Genet.*, **9**, 17–22.
Riley,M. and Sanderson,K.E. (1990) In Drlica,K. and Riley,M. (eds), *The Bacterial Chromosome*. American Society for Microbiology, Washington, DC, pp. 85–95.
Salmond,G.P.C. and Reeves,P.J. (1993) *Trends Biochem. Sci.*, **18**, 7–12.
Sanderson,K.E. and Roth,J.R. (1988) *Microbiol. Rev.*, **52**, 485–532.
Sansonetti,P.J. (1991) *Rev. Infect. Dis.*, **13** (Suppl. 4), S285–292.
Sansonetti,P.J., Ryter,A., Clerc,P., Maurelli,A.T. and Mournier,J. (1986) *Infect. Immun.*, **1**, 461–469.
Schiemann,D.A. and Shope,S.R. (1991) *Infect. Immun.*, **59**, 437–440.
Small,P.L.C. and Falkow,S. (1988) *J. Bacteriol.*, **56**, 225–229.
Small,P.L.C., Isberg,R.R. and Falkow,S. (1987) *Infect. Immun.*, **55**, 1674–1679.
Smith,N.H. and Selander,R.K. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 956–960.
Stone,B.J., Garcia,C.M., Badger,J.L., Hassett,T., Smith,R.I.F. and Miller,V.L. (1992) *J. Bacteriol.*, **174**, 3945–3952.
Venkatesan,M., Buysse,J.M., Vandendries,E. and Kopecko,D.J. (1988a) *J. Clin. Microbiol.*, **26**, 261–266.
Venkatesan,M.M., Buysse,J.M. and Kopecko,D.J. (1988b) *Proc. Natl Acad. Sci. USA*, **85**, 9317–9321.
Venkatesan,M.M., Buysse,J.M. and Oaks,E.V. (1992) *J. Bacteriol.*, **174**, 1990–2001.
Vogler,A.P., Homma,M., Irikura,V.M. and MacNab,R.M. (1991) *J. Bacteriol.*, **173**, 3564–3572.
Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J.*, **1**, 945–951.

Received on May 14, 1993; revised on June 29, 1993

Note added in proof

Sasakawa *et al.* [(1993) *J. Bacteriol.*, **175**, 2334–2346] have recently analysed the *Shigella flexneri spa* locus beyond the region reported by Venkatesan *et al.* (1992). In agreement with the data presented in our paper, which demonstrated that the homology between the *spa* regions of *Shigella flexneri* and *Salmonella typhimurium* extended several kilobases downstream of the genes originally characterized by Venkatesan *et al.* (1992), the recent report by Sasakawa *et al.* (1993) identified several genes downstream of *spa24* and recognized the ORF located between *spa47* and *spa32* identified in our paper but ignored by Venkatesan *et al.* (1992). Thus, the *Shigella flexneri spa13*, *spa9*, *spa29* and *spa40* correspond to the *Salmonella typhimurium spaM*, *spaQ*, *spaR* and *spaS*. This demonstrates that *spa* regions of both species contain at least 11 genes of similar sizes and organization.