Salmonella Virulence Plasmid: Modular Acquisition of the *spv* Virulence Region by an F-Plasmid in *Salmonella enterica* Subspecies I and Insertion Into the Chromosome of Subspecies II, IIIa, IV and VII Isolates

E. Fidelma Boyd and Daniel L. Hartl

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138

Manuscript received September 26, 1997
Accepted for publication April 1, 1998

ABSTRACT

The *spv* operon is common to all Salmonella virulence plasmids. DNA hybridization analysis indicates that the *spv* region is limited in distribution to serovars of *Salmonella enterica* subspecies I, II, IIIa, IV, and VII and is absent from *Salmonella bongori* isolates. Among strains of subspecies II, IIIa, and VII, all isolates examined contained sequences that hybridized with the *spv* region. However, among isolates of subspecies I, DNA sequences capable of hybridizing with the *spv* region were found in some isolates of certain serovars. Furthermore, in isolates of subspecies I, the virulence plasmid was found in the same set of isolates as an F-related plasmid, as determined by the presence of the *spv* region of the virulence plasmid and the *finO*, *traD*, and *repA* sequences of the F-plasmid. The concordance of the virulence plasmid and all three F-plasmid sequences in subspecies I serovar Choleraesuis, Paratyphi, and Typhimurium is most easily explained if the *spv* region is carried in an F-related plasmid in these isolates. In contrast, among *S. enterica* subspecies II, IIIa, IV, and VII, the isolates that contain *spv* sequences did not hybridize with an F-related plasmid or any other identifiable plasmid. With the use of pulse-field gel electrophoresis, the *spv* region in subspecies II, IIIa, and VII was found to be encoded on the chromosome. Analysis of the phylogenetic distribution of *spv* among Salmonella isolates and comparative nucleotide sequence analysis of *spvA* and *spvC* suggests that the *spv* region was acquired very recently, after speciation of the salmonellae.

SALMONELLA has been implicated in a wide variety of infections ranging from life-threatening typhoid to gastroenteritis and bacteremia. Salmonella is a facultative intracellular pathogen, typically colonizing reptiles, birds, and mammals, with some serovars showing remarkable host-adaptation; for example, serovars Typhi, Dublin, and Gallinarum infect humans, cattle, and birds, respectively, although by no means are these serovars restricted to these hosts (Falkow 1996). In view of the diversity of animal species infected and the complex mechanisms of infection, it is not surprising that the array of genes required for virulence depends on the particular host and the mode of infection.

In regard to *Salmonella enterica* subspecies I, the virulence properties of various serovars, Typhimurium, Choleraesuis, Dublin, and Enteriditis, depend on the presence of large plasmids 65–100 kb in size to cause systemic infection (Jones *et al.* 1982; Terakado *et al.* 1983; Nakamura *et al.* 1985; Gulig and Curtiss 1987). Analysis has shown that the virulence plasmids from different serovars share a related 7.8-kb virulence region, the *spv* gene cluster, which includes five open reading frames designated *spvR*, *spvA*, *spvB*, *spvC*, and *spvD* (Baird *et al.* 1985; Williamson *et al.* 1988; Norel *et al.* 1989; Taira and Rhen 1989; Krause *et al.* 1991; Gulig *et al.* 1993). The *spvR* gene encodes a regulatory protein of the LysR family that, together with the chromosomally encoded regulatory gene *rpoS*, regulates the *spvABCD* genes (Taira *et al.* 1991; Fang *et al.* 1992; Norel *et al.* 1992; Gulig *et al.* 1993). The function of the *spvABCD* gene products remains undetermined, but it has been shown that the presence of the virulence plasmid increases the growth rate of Salmonella in mice (Gulig and Doyle 1993).

The *spv* region of the Salmonella virulence plasmid represents only a small proportion of the plasmid-coding capability, and little is known about the coding capacity of the rest of the virulence plasmid. Tinge and Curtiss (1990) demonstrated that the virulence plasmid of serovar Typhimurium contained three replication (*rep*) regions *repA* (*par*), *repB*, and *repC*. The *repB* and *repC* regions hybridize weakly with IncFI plasmid F and IncFII plasmid R100, respectively; however, neither *repB* or *repC* conferred incompatibility with F or R100. The *rck* and *traT* genes also present in the serovar Typhimurium virulence plasmid encode outer membrane proteins whose expression confers survival in macrophages (Heffernan *et al.* 1992; Rhen *et al.* 1992). Recently, it has also been demonstrated that the virulence plasmid of Typhimurium encodes a fimbrial biosynthe-

Corresponding author: Daniel L. Hartl, 16 Divinity Avenue, Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138. E-mail: dhartl@oeb.harvard.edu

sis operon, *pef* (Friedrich *et al.* 1993). Examination of the virulence plasmid of Salmonella serovars Gallinarum, Enteritidis, and Typhimurium demonstrated the presence of F-like OriT, but this region was not found in the virulence plasmid of serovars Choleraesuis and Dublin (Ou *et al.* 1994). Further, DNA hybridization of the nonvirulence coding part of the virulence plasmid of serovar Dublin with *Escherichia coli* and Salmonella *spv*-negative plasmids indicates a common origin of this region of the plasmid (Aabo *et al.* 1995). Similarly, the large virulence plasmid of *Shigella flexneri* showed homology with part of the transfer region of F and with *vagC* and *vagD* of the Salmonella serovar Dublin virulence plasmid (Radnedge *et al.* 1997). Rodriguez-Pena and colleagues (1997) have recently shown the virulence plasmid from serovar Enteriditis could have arisen from the Typhimurium plasmid through deletions, and both plasmids share homology in the tra region from IncFII plasmids.

The genus Salmonella proper is divided into two species: *S. bongori* (formerly *S. enterica* subspecies V; Reeves *et al.* 1989) and *S. enterica*. The species *S. enterica* is further subdivided into seven subspecies, designated by the roman numerals I, II, IIIa, IIIb, IV, VI, and VII (Le Minor and Popoff 1987; Selander *et al.* 1994; Boyd *et al.* 1996). The majority of Salmonella human pathogens belong to subspecies I isolates, whereas subspecies IIIa, IIIb, II, IV, and VII and *S. bongori* are mainly associated with cold blooded vertebrates (Popoff and Le Minor 1992).

The true prevalence of the virulence plasmid among natural isolates of Salmonella is unknown. Previous studies have shown the occurrence of the virulence plasmid among a few serovars of *S. enterica* subspecies I, to which 99% of the medically important serovars belong (Williamson *et al.* 1988; Bäumler *et al.* 1997). Herein, we find that among the salmonellae, the *spv* region is found in all isolates of *S. enterica* subspecies II, IIIa, and VII isolates. Of the 72 subspecies I isolates examined of the Salmonella reference collection B (SARB), 15 strains showed homology with the *spv* probe, and further analysis showed these 15 isolates also hybridized with the F plasmid probes, *finO*, *traD*, and *repA*. Analysis of nucleotide and amino acid substitutions of the *spvA* gene suggests there has been rapid divergence, possibly related to niche specialization, among the subspecies. From the standpoint of population genetics, the most important result is that the presence of the Salmonella virulence plasmid is highly correlated with the presence of an F-related plasmid. Further, the finding of the presence of the *spv* region on the chromosome of subspecies II, IIIa, IV, and VII isolates is also significant. From the microbial pathogenesis perspective, the probability that the *spv* region is part of an F-related plasmid in isolates of subspecies I is significant in relation to the potential for frequent horizontal transmission among strains and species.

MATERIALS AND METHODS

Bacterial strains: We examined 72 isolates of the SARB collection, which includes 37 medically important serovars of subspecies I of *S. enterica* whose phylogenetic relationships are known (Boyd *et al.* 1993). In addition, we examined 80 natural isolates of the Salmonella reference collection C (SARC; Boyd *et al.* 1996), which represents most of the genetic diversity within the genus. SARC is composed of *S. bongori* (11 isolates) and all 7 subspecies of *S. enterica*: I (11 isolates), II (14 isolates), IIIa (4 isolates), IIIb (4 isolates), IV (9 isolates), VI (20 isolates), VII (4 isolates). Total genomic DNA was extracted with the G-Nome DNA isolation kit (Bio 101, La Jolla, CA).

PCR amplification: Primers for PCR and DNA sequencing were designed from published sequence of the *spv* virulence region (Krause *et al.* 1990). PCR products were purified using the Qiaquick PCR purification kit (Quigen Inc., La Jolla, CA).

DNA hybridization: Plasmid DNA was isolated from Salmonella strains. All isolates were examined by restriction digests with *Eco*RI and electrophoresed in 0.6% agarose gels and transferred to Hybond N^+ membranes. DNA fragments for use as fluoresin probes in Southern hybridizations were prepared from *S. enterica* serovar Typhimurium LT2. Three DNA probes were constructed, which encompassed the entire *spv* region. The *spv* fragments were amplified by long-range PCR and labeled to high-specify activity by the random-labeling method [enhanced chemilumence (ECL); Amersham, Arlington Heights, IL]. The F-plasmid probes were prepared as previously described (Boyd and Hartl 1997). Prehybridizations were performed in 10× SSC, 0.2% SDS, 10% dextran sulfate, and 5% ECL liquid block for 1 hr. Hybridization of probes was carried out overnight at 60° (55° for reduced stringency). Membrane washes were at 60° (55° for reduced stringency) in 1× SCC and 0.1% SDS. Hybridized fragments were detected with the ECL system.

Association of genes with plasmids: To determine whether the genes examined were plasmid encoded, total genomic DNA was separated on 0.6% agarose for an extended time to permit separation of plasmid DNA from the bacterial chromosomal band. DNA was transferred to Hybond N^+ membranes and DNA was cross-linked by ultraviolet exposure. Membranes were hybridized with plasmid probes overnight under both high (60°) and low (55°) stringency conditions.

Pulsed field gel electrophoresis (PFGE): To determine whether the *spv* region in isolates outside of subspecies I are encoded on the chromosome, PFGE was carried out. Agarose plugs were prepared as previously described (Bergthorsson and Ochman 1995). Agarose plugs were washed five times for 5 min each in 50 vol distilled H_2O and equilibrated in the appropriate restriction buffer. Agarose plugs were digested with I-*Ceu*I (NEB) as described by Bergthorsson and Ochman (1998). The restriction enzyme I-*Ceu*I cuts at rDNA operons only. Electrophoresis was performed with a CHEF-DR II apparatus (Bio-Rad Laboratories, Richmond, CA). Electrophoresis was carried out as previously described (Bergthorsson and Ochman 1998). DNA was transferred to a nylon membrane, and an *spv* DNA probe was used to detect homology.

Nucleotide sequencing: Two genes from the *spvABCD* operon were sequenced from representative isolates of Salmonella. DNA sequencing of PCR-amplified DNA was performed with a 370A DNA sequencer following manufacturers' instructions. Both dyeterminator and dyeprimer chemistries were used. All sequences have been deposited in GenBank accession numbers AF051816-AF051829.

Phylogenetic and statistical analysis: DNA sequence data were assembled and edited with sequencer programs (1991). Phylogenetic analysis was performed with the programs Molec-

TABLE 1

Genotypic characterization of *spv*-positive SARB strains and their hybridization with F-plasmid sequences

Strain[a]	Serovar	Probe			
		spv	*finO*	*traD*	*repA*
4	Choleraesuis	+	+	+	+
6	Choleraesuis	+	+	+	+
10	Derby	+	+	−	−
11	Derby	+	+	+	−
13	Dublin	+	+	+	−
16	Enteritidis	+	+	−	+
18	Enteritidis	+	+	+	−
48	Paratyphi C	+	+	+	+
49	Paratyphi C	+	+	+	+
51	Pullorum	+	+	+	−
52	Pullorum	+	+	+	−
65	Typhimurium	+	+	+	+
66	Typhimurium	+	+	+	+
67	Typhimurium	+	+	+	+
68	Typhimurium	+	+	+	+

[a] SARB strain number (Boyd *et al.* 1993).

ular Evolutionary Genetic Analysis, version 1.0 (Kumar *et al.* 1993) and Molecular Evolutionary Analysis (Etsuko Moriyama, Yale University).

RESULTS

DNA hybridization: Southern blot analysis yielded 15 strains with hybridization to the *spv* probe among the 72 strains of the SARB collection, which encompasses 37 serovars of subspecies I (Boyd *et al.* 1993). Homology to the three F-plasmid probes, *finO*, *traD*, and *repA* was found in the same set of subspecies I strains as the *spv*-positive isolates (Table 1). Of the *spv*-negative strains of SARB, one strain SARB 58 (serovar Sendai) gave positive hybridization signals with the *finO*, *traD*, and *repA* probes; SARB 57 was positive with the *finO* probe only; and SARB 64 was positive with the *repA* probe only. The SARC collection represents the range of genetic diversity found within the genus and is comprised of isolates of *S. bongori* (formerly subspecies V) and the seven subspecies of *S. enterica* I, II, IIIa, IIIb, IV, VI, and VII (Boyd *et al.* 1996). Of the 80 isolates in the SARC collection examined, 33 gave a positive hybridization signal with the *spv* probe, including the 7 *spv*-positive strains of subspecies I that overlap with isolates from the SARB collection. In contrast to subspecies I isolates, all strains of subspecies II, IIIa, and VII and two isolates of subspecies IV gave positive hybridization signals with the *spv* probe, but only one of these strains (subspecies VII) showed homology with any of the F-plasmid sequences; the *traD* and *repA* probes. Of the *spv*-negative SARC isolates, one strain gave positive hybridization signals with the *finO* and *traD* probes, one strain gave positive hybridization signals with the *traD* and *repA* probes, and eight strains gave a positive hybridization signal with only the *repA* probe.

Phylogenetic distribution of the *spv* region: The 15 SARB isolates that gave a positive signal with the *spv* probe represented seven serovars: Choleraesuis, Derby, Dublin, Enteritidis, Paratyphi C, Pullorum, and Typhimurium. Within each of these serovars, the *spv* region was absent from some isolates. Phylogenetically, the strains of SARB that have the *spv* region are clustered into two lineages of the SARB neighbor-joining tree inferred from multilocus enzyme electrophoresis (Boyd *et al.* 1993). The *spv*-positive serovars Choleraesuis, Paratyphi C, and Derby all cluster together on the SARB tree in lineage B, and the other *spv*-positive isolates are all found on three separate branches of lineage A. The Typhimurium strains cluster together, the Enteritidis and Pullorum isolates group together, and (on a third and separate branch) the *spv*-positive Dublin strains and one Enteritidis strain cluster.

Figure 1 reveals an unusual and unexpected pattern of occurrence of the *spv* region among the SARC collection. The spv region was sporadic in occurrence among subspecies I isolates; however, it was present in all isolates of subspecies II, IIIa, and VII, and in two isolates of subspecies IV and was absent from all isolates of *S. bongori* and *S. enterica* subspecies IIIb and VI (Figure 1).

Association of genes with plasmids: To determine whether the genes analyzed are located in plasmids, uncut genomic DNA was electrophoresed slowly to separate linear genomic DNA from the closed circular plasmid DNA. The *spv* probe was always associated with a large plasmid band moving more slowly than the genomic DNA in all the SARB strains that contained this sequence. Similarly, for the F-plasmid probes, hybridization was always with the same band as that for the *spv* probe. Among subspecies II, IIIa, IV, and VII isolates, the *spv* probe was not associated with any plasmid, but positive signals were obtained only from the total genomic DNA band. Whether this indicates comigration of plasmid bands with genomic DNA could not be resolved by these methods.

Pulsed field gel electrophoresis: PFGE was carried out to determine whether the *spv* region is chromosomally encoded in isolates where no association with a plasmid could be determined. Digested and undigested genomic DNA were electrophoresed side-by-side. Four isolates were examined: S2980 (subspecies IIIa), S2983 (subspecies IIIa), S3027 (subspecies IV), and S4194 (subspecies I; Figure 2). S4194 was used as a positive control, and both S4194 lanes gave a positive signal in the 100-kb region of the gel, indicating the Salmonella virulence plasmid, but there was no hybridization with chromosomal DNA (Figure 3). In isolates S2980, S2983, and S3027, where the *spv* region is proposed to be chromosomally encoded, the *spv* probe gave a strong positive signal in the upper region of the gel with both cut and

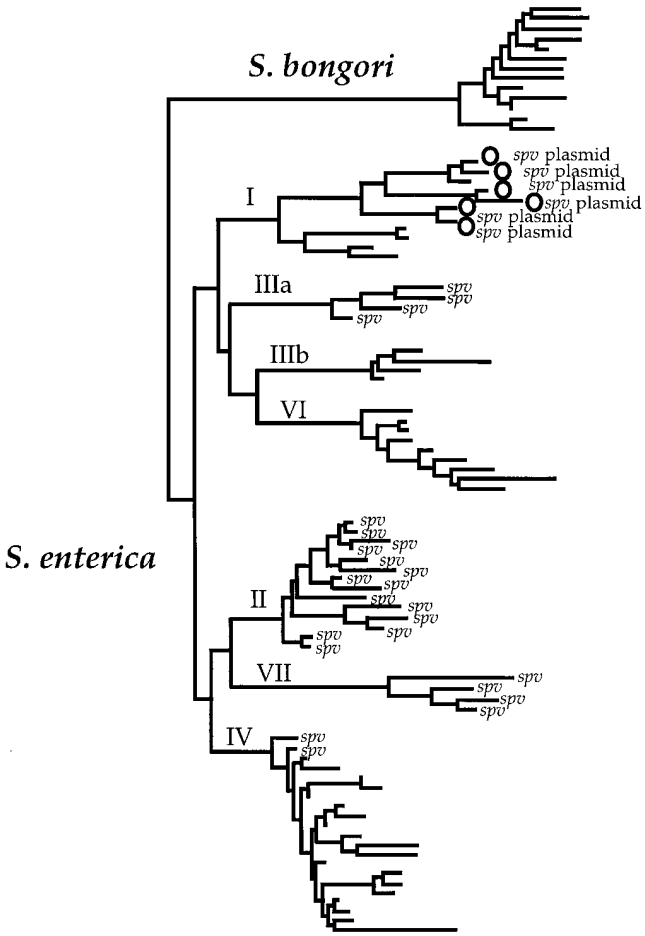

Figure 1.—Phylogenetic relationships of the Salmonella based on multilocus enzyme electrophoresis at 24 enzyme loci (Boyd *et al.* 1996). Roman numerals indicate subspecies of *S. enterica.* Circles represent the presence of the *spv* region in a plasmid. The subspecies I isolates in this phylogenetic tree are represented by 7 of the 15 *spv*-positive strains of the SARB collection.

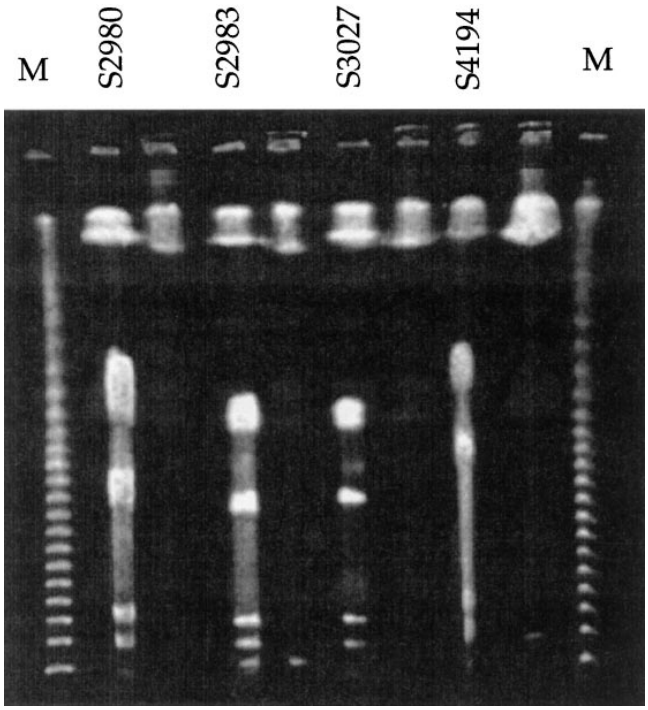

Figure 2.—Ethidium bromide-stained pulse field gel of four Salmonella isolates. Lanes 1 and 10 contain lambda marker. In lanes 2 to 9, even-numbered lanes represent genomic DNA cut with I-*Ceu*I, and odd-numbered lanes indicate uncut DNA. Isolates S2980, S2983, S3027, and S4194 are indicated over each of the wells.

uncut DNA, and no hybridization with a plasmid DNA was detected (Figure 3).

Nucleotide polymorphism: Two genes from the *spv* region were analyzed, *spvA* and *spvC.* We sequenced a 462-561-bp portion of the *spvA* gene from a sample of subspecies I (seven isolates), IIIa (four isolates), IV (two isolates), and VII (one isolate) strains to determine whether the phylogenetic distribution of *spv* reflects presence in the most recent common ancestor and widespread loss or, alternatively, recent horizontal transfer among lineages of Salmonella. There were both size and sequence polymorphisms in the *spvA* gene sequenced from the 14 Salmonella isolates. In total, there were 26 polymorphic nucleotide sites, 17 of which were replacement sites, and 15 of these replacement sites defined differences among subspecies (Figure 4). Of the seven subspecies I strains examined, there were six polymorphic nucleotide sites in the 462-bp region sequenced, three of which resulted in amino-acid substitutions. The host-adapted serovars Choleraesuis, Derby, Dublin, and Pullorum were virtually identical in *spvA* sequence (Figure 4), as were the Typhimurium and Enteriditis *spvA* sequences.

All subspecies IIIa isolates examined had an insertion of 99 bp after nucleotide position 387 from the start codon of the *spvA* gene. The two subspecies IV (SARC 60, SARC 61) and one subspecies VII (SARC 56) strains examined had a deletion of 219 bp from nucleotide position 168 from the start codon of *spvA* and an insertion of 60 bp identical to the last 60 bp of the insertion in subspecies IIIa isolates (Figure 5). The inserted region of *spvA* is flanked by a 10 nt direct repeat of CCGACCCTG found at nucleotide 388 in all isolates examined. This 10-bp sequence is also found at the beginning of the 99-bp insert in all strains of subspecies IIIa examined (Figure 5). The 99-bp insertion showed no similarity to any entries in the current sequence databases. Among the 530-bp region of *spvC* sequenced in seven SARC strains representing four subspecies (I, IIIa, IV, and VII) there were 11 polymorphic sites including five amino acid replacement sites.

We estimated the number of synonymous (silent) substitutions per synonymous sites (k_S) and the number of nonsynonymous (replacement) substitutions per nonsynonymous sites (k_N) for all pairwise comparisons, using the Jukes-Cantor correction for the proportion of

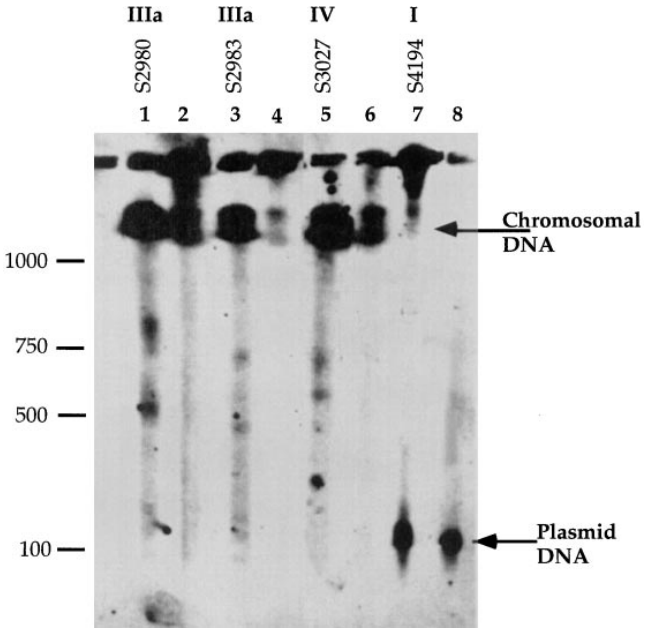

Figure 3.—Pulse field gel electrophoresis analysis of four *spv*-positive Salmonella isolates. In lanes 1 to 8, odd-numbered lanes represent genomic DNA cut with I-*Ceu*I, and even-numbered lanes indicate uncut DNA. Lanes 1 and 2, S2980 (subspecies IIIa); lanes 3 and 4, S2983 (subspecies IIIa); lanes 5 and 6, S3027 (subspecies IV); and lanes 7 and 8, S4194 (subspecies I). The lanes 1, 3, and 5 show *spv* probe hybridizing with large fragments of cut and partially cut genomic DNA. In lanes 2, 4, and 6 the *spv* probe hybridizes with uncut genomic DNA. No hybridization with plasmid DNA is detected. The last two lanes indicate *spv* probe hybridizing with Salmonella virulence plasmid DNA. Numbers along the side indicate sizes in kilobases.

differences. The level of synonymous site variation in the *spvA* (k_S, 0.019 ± 0.009) and *spvC* (k_S 0.028 ± 0.012) genes are unusually low relative to other chromosomal and plasmid genes previously analyzed in this species, which is typically about k_S, 0.20 (Boyd and Hartl 1997; Boyd *et al.* 1997). The k_N/k_S ratio, which is a measure of functional constraints on a protein if it is evolving in a neutral fashion, for the *spvA* and *spvC* genes was 0.55 and 0.21 respectively.

Evolutionary relationships among the *spv* region of *S. enterica*: The phylogenetic tree in Figure 6 is a neighbor-joining tree (Saitou and Nei 1987) constructed from a pairwise matrix of distances at silent-substitution sites (Jukes-Cantor 1969). The topology of nucleotide sequence variation among the subspecies examined is similar to that obtained for the consensus pattern for 12 chromosomal genes (Boyd *et al.* 1996). Representative isolates from each subspecies cluster together, indicating common ancestry (Figure 6). The single subspecies VII isolate examined clustered with the subspecies IV isolates, which have been found previously with a number of other virulence loci examined in this species (Boyd *et al.* 1996). One important feature of this evolutionary tree is the very short branch length when compared to other chromosomally encoded genes of Salmonella (Boyd *et al.* 1997).

DISCUSSION

Large virulence plasmids present in a few serovars of subspecies I of *S. enterica* are essential for the bacteria to produce systemic infection in both humans and other animals (Jones *et al.* 1982; Terakado *et al.* 1983; Nakamura *et al.* 1985; Gulig and Curtiss 1987). Outside of *S. enterica* subspecies I, the prevalence and distribution of the virulence plasmid is unknown. Analysis has shown that the virulence plasmids from different serovars share a related virulence region, the *spv* gene cluster (Baird *et al.* 1985; Williamson *et al.* 1988; Norel *et al.* 1989; Taira and Rhen 1989; Krause *et al.* 1991). Among a subset of 18 strains of the SARB collection (Boyd *et al.* 1993) examined, the *spv* region was found to be present in isolates of serovar Choleraesuis, Dublin, Enteriditis, Paratyphi, Pullorum, and Typhimurium (Bäumler *et al.* 1997).

A

```
                 111 112 222 222 233 344 455 666 77
                 677 790 135 567 806 699 918 115 04
                 958 883 746 992 232 403 844 174 61
I     SARB4      GGT GGA AAA TCC CTT AAC GCT CCC GT
I     SARB65     .A. ... ... ... T.. ... ... ... ..
I     SARB66     .A. ... ... ... ... .C. ... ... ..
I     SARB11     ... ... ... ... ... ... A.. ... ..
I     SARB13     ... ... ... ... ... ... A.. ... ..
I     SARB16     ... ... ... ..T ... ... ... ... ..
I     SARB51     ... .A. ... ... ... ... ... ... ..
IIIa  SARC25     ..A A.C CGG CTT .CA C.. ..C ... AG
IIIa  SARC26     A.A A.. CGG A.. .CA ... ..C .T. AG
IIIa  SARC27     A.A A.. CGG A.. .CA ... ..C T.. A.
IIIa  SARC28     A.A ... CGG A.. .CA ... ..C T.. A.
VII   SARC56     .-- --- --- --- --- -.T .T. ..T A.
IV    SARC60     .-- --- --- --- --- -.. .T. ..T A.
IV    SARC61     .-- --- --- --- --- -.. .T. ..T A.
```

B

```
                 1 1111222
          5566788992 2679003
          7908367011 2425466
I     SARB4      GAWQMTLTPV NTHISTA
I     SARB65     .T........ .......
I     SARB66     .T........ .P.....
I     SARB11     .......... .......
I     SARB13     .......... .......
I     SARB16     ........L. .......
I     SARB51     .......... .......
IIIa  SARC25     ..KPLA.ILD H..T..T
IIIa  SARC26     S.K.LAM..D ...T.IT
IIIa  SARC27     S.K.LAM..D ...TL.T
IIIa  SARC28     S.R.LAM..D ...TL.T
VII   SARC56     ---------- -.Y...T
IV    SARC60     ---------- -.Y...T
IV    SARC61     ---------- -.Y...T
```

Figure 4.—Polymorphic sites in the *spvA* gene among 14 *S. enterica* isolates. (A) The polymorphic nucleotide sites in the 561-bp region sequenced. (B) The polymorphic amino acid sites. Numbers across the top indicate the positions of polymorphic sites along the gene numbered from the start codon. Dots represent nucleotide and amino-acid identity; dashes indicate nucleotide and amino-acid deletions. Roman numerals indicate subspecies designation.

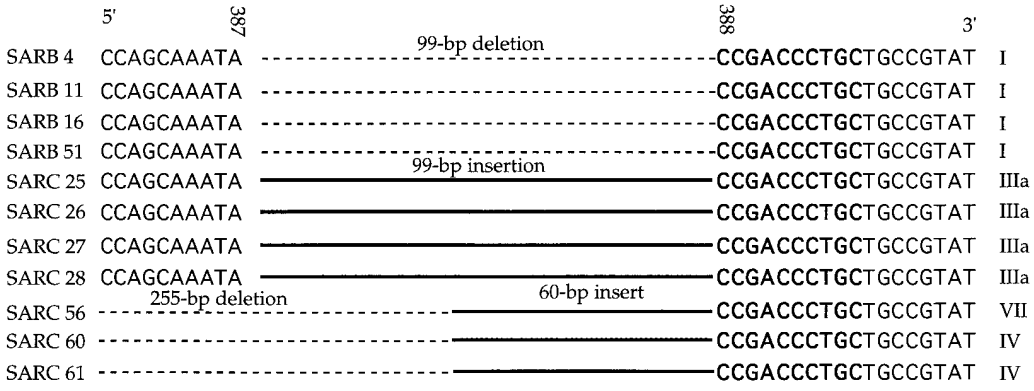

Figure 5.—Structure of the *spvA* from subspecies I, IIIa, IV, and VII isolates. The bold letters indicate the 10-bp repeat that marks the point insertion of the 99-bp and 60-bp inserts. The hatched lines represent inserts, and the broken line represents deleted regions in subspecies IV and VII isolates. The roman numerals indicate subspecies of *S. enterica*.

We show that the occurrence of the *spv* region of Salmonella is limited in distribution to isolates of subspecies I, II, IIIa, IV, and VII. Among the 72 subspecies I isolates of the SARB collection, 7 of 37 serovars examined have the virulence plasmid, but, within each serovar, not all isolates contain the *spv* region. The discordance may reflect the fact that similarity in serotype among strains of the salmonellae does not necessarily reflect close phylogenetic relationship (Boyd *et al.* 1993). Of the 15 *spv*-positive SARB isolates, all showed positive hybridization signals with the F-plasmid probes. Among the serovars of subspecies I that had the *spv* region, isolates of serovars Choleraesuis, Paratyphi, and Typhimurium all contained sequences homologous to all three F-plasmid probes, which suggests that the *spv* region, at least in these serovars, is carried on a similar F-related plasmid. Previously, others have shown homology of F-like sequence to the Salmonella virulence plasmid (Ou *et al.* 1994; Rodriguez-Pena *et al.* 1997). The other four *spv*-positive serovars, Derby, Dublin, Enteriditis, and Pullorum, all hybridized with at least two of the F-plasmid probes, confirming previous studies that show size variation among virulence plasmids among serovars, particularly serovar Dublin. Previously, we examined isolates of the Typhimurium complex for the presence of plasmids closely related to the *E. coli* F plasmid and found 19/72 Salmonella isolates with plasmid sequences homologous (and in some cases virtually identical) to those found in *E. coli* isolates (Boyd and Hartl 1997). Analysis of these 19 strains showed that 14 (all serovar Typhimurium) also hybridized with the *spv* probe. Studies have indicated the possible transfer of an epidemic R plasmid from Salmonella serovar Enteriditis to the normal human gut flora *E. coli*, which suggests that not only can *E. coli* act as a reservoir of resistant genes but also the factor that disseminates these and other important genes (Balis *et al.* 1996).

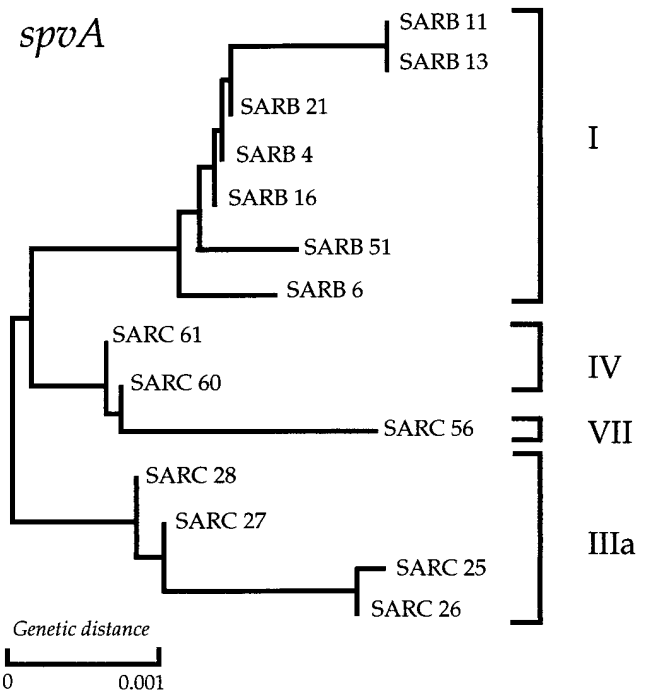

Figure 6.—Neighbor-joining dendrogram for the 14 *S. enterica* isolates, based on variation in the *spvA* sequences. The subspecies of *S. enterica* are designated by roman numerals; the SARC numbers are the same as those indicated in Figure 1.

Outside of subspecies I isolates, the *spv* region is found in all isolates of subspecies II, IIIa, and VII, as well as in two isolates of subspecies IV (Figure 1). The presence of the *spv* region in all isolates of subspecies II, IIIa, and VII is surprising and unusual given that this region is supposedly carried on a plasmid; therefore, sporadic loss is expected. The distribution found is consistent with a region carried on the chromosome. None of these *spv*-positive isolates from subspecies II, IIIa, and VII showed hybridization signals with all of the F-plasmid probes, but plasmid bands were found in some of these strains. Further, our analysis among strains outside of subspecies I could not demonstrate any association of the *spv* region with any type of plasmid. The possibility remains that the virulence plasmid may be unstable in these subspecies, but this seems unlikely given that Olsen *et al.* (1994) have shown that the virulence plasmid of serovar Enteriditis is stably maintained at a range of temperatures. Furthermore, the occurrence of the *spv* region in all isolates of subspecies IIIa, IV, and VII, along with the near sequence identity of *spvA* within

subspecies, strongly suggests that the *spv* region is a chromosomally encoded gene cluster in these subspecies. We carried out PFGE analysis to confirm the presence of the *spv* region on the chromosome of these isolates (Figure 3). The presence of the *spv* region on the chromosome of *S. enterica* subspecies II, IIIa, IV, and VII suggests that the *spv* region may confer some selective advantage for the survival of these lineages. This situation is reminiscent of that of the *E. coli* hemolysin determinants (*hly*), which are found on large transmissible plasmids in animal isolates of *E. coli* but are located within unique chromosomal insertions, termed pathogenicity islands, in human uropathogenic isolates (Muller *et al.* 1983; Welch *et al.* 1983).

Although the number of nucleotide substitutions in *spvA* was small (26 sites) relative to other Salmonella chromosomal genes involved in virulence (Boyd *et al.* 1997), many of these sites (17/27) were replacement sites. Furthermore, over half of these amino-acid replacement sites were found within a 90-bp region of the gene. Given that both the nucleotide and amino-acid polymorphisms in the *spvA* gene are mostly among and not within the subspecies (Figure 4), and, given that the lineages of Salmonella examined have existed for a very long time, the evidence suggests a case in which this region underwent a rapid period of strong positive selection and then clonal expansion of the new variant. It is apparent that this region was acquired after speciation of *S. bongori* from *S. enterica* and, within *S. enterica*, it was lost from subspecies IIIb and IV. The absence of the *spv* region from *S. bongori* may reflect the fact that this is the most divergent lineage of Salmonella as judged from an analysis of the nucleotide sequences of 12 genes and multilocus enzyme electrophoresis (Boyd *et al.* 1996) and also that this species is predominantly isolated from reptiles.

In addition to the concordance of the virulence-plasmid and F-plasmid sequences, a number of new insertion/deletion features of the *spvA* coding sequence were observed (Figure 5). The *spvA* sequence as previously reported is as shown for subspecies I in Figure 5, in which an A at nucleotide 387 (numbered relative to the A in the start codon) is adjacent to a C at nucleotide 388. As shown for subspecies IIIa, all four of the sequenced representatives have a 99-bp insertion relative to the canonical sequence that begins with the sequence CCGACCCTGC immediately after nucleotide 387; this sequence happens to be identical to nucleotides 388–397 in the canonical sequence. Considering the 10-bp repeat flanking the "insertion," it is possible that the subspecies IIIa sequence is ancestral and that the subspecies I sequence sustained a 99-bp deletion through unequal recombination or replication slippage involving the tandem duplication in the subspecies IIIa sequence. The corresponding region in the three sequences from subspecies IV and VII isolates differs from the regions in both of the other subspecies in exhibiting a 60-bp "insertion," which is identical to the last 60-bp of the "insertion" in the subspecies IIIa sequence. (This observation also supports the subspecies IIIa sequence as being ancestral.) Upstream of the 60-bp region in the subspecies IV and VII sequences, beginning at nucleotide 168 of *spvA*, there is a 219-bp deletion; this aberration results in a truncated SpvA protein. The subspecies IV and VII isolates examined are very similar in sequence for *spvA*, and this may reflect horizontal transfer of this region between the two groups, as has been previously hypothesized for another virulence locus, the chromosome encoded *inv/spa* invasion region (Boyd *et al.* 1997). Experiments by others on subcellular location of the Spv proteins indicated that the SpvA protein is located in the outer membrane, which may account for the selective pressure required for the rapid diversification seen at this locus. The SpvC protein was present in the cytoplasm (El-Gedaily *et al.* 1997). Studies by others have shown that the *spvA* gene from serovar Dublin is not essential for virulence in mice (Roudier *et al.* 1992) and this may account for the pattern of polymorphism and deletions we have found; *spvA* may be a nonessential gene and therefore not under any strong selective constraints.

The phylogenetically widespread distribution of the *spv* region but the limited sequence variation found in two genes, *spvA* and *spvC*, suggests that this region is evolutionarily new and has been acquired by lateral transfer much more recently than other virulence genes of Salmonella.

This work was supported by grant GM-40322 from the National Institutes of Health. We thank Ying Chen and Ulfar Bergthorsson for advice and technical assistance.

LITERATURE CITED

Aabo, S., J. E. Olsen, E. J. Threlfall and D. J. Brown, 1995 Characterization of non-virulence plasmids with homology to the virulence plasmid of *Salmonella dublin*. Res. Microbiol. **146:** 751–759.

Baird, G. D., E. J. Manning and P. W. Jones, 1985 Evidence for related virulence sequences in plasmids of *Salmonella dublin* and *Salmonella typhimurium*. J. Gen. Microbiol. **131:** 1815–1823.

Balis, E., A. C. Vatopoulos, M. Kanelopoulou, E. Mainas, G. Hatzoudis *et al.*, 1996 Indications of *in vivo* transfer of an epidemic R plasmid from *Salmonella enteriditis* to *Escherichia coli* of the normal human gut flora. J. Clin. Microbiol. **34:** 977–979.

Bäumler, A. J., R. M. Tsolis, P. J. Vatentine, T. A. Ficht and F. Heffron, 1997 Synergistic effect of mutations in *invA* and *lpfC* on the ability of *Salmonella* typhimurium to cause murine typhoid. J. Bacteriol. **65:** 2254–2259.

Bergthorsson, U., and H. Ochman, 1995 Heterogeneity of genome sizes among natural isolates of *Escherichia coli*. J. Bacteriol. **177:** 5784–5789.

Bergthorsson, U., and H. Ochman, 1998 Distribution of chromosome length variation in natural isolates of *Escherichia coli*. Mol. Biol. Evol. **15:** 6–16.

Boyd, E. F., and D. L. Hartl, 1997 Recent horizontal transmission of plasmids between natural populations of *Escherichia coli* and *Salmonella enterica*. J. Bacteriol. **179:** 1622–1627.

Boyd, E. F., F.-S. Wang, P. Beltran, S. A. Plock, K. Nelson *et al.*, 1993 *Salmonella* reference collection B (SARB): strains of 37 serovars of subspecies I. J. Gen. Microbiol. **139:** 1125–1132.

Boyd, E. F., F.-S. Wang, T. S. Whittam and R. K. Selander, 1996

Molecular genetic relationships of the salmonellae. Appl. Environ. Microbiol. **62:** 804–808.

Boyd, E. F., J. Li, H. Ochman and R. K. Selander, 1997 Comparative genetics of the *inv*/*spa* invasion gene complex of *Salmonella enterica*. J. Bacteriol. **179:** 1985–1991.

El-Gedaily, A., G. Paesold and M. Krause, 1997 Expression profile and subcellular location of the plasmid-encoded virulence (*spv*) proteins in wild-type *Salmonella dublin*. Infect. Immun. **65:** 3406–3411.

Falkow, S., 1996 The evolution of pathogenicity in *Escherichia coli*, *Shigella*, and *Salmonella*, pp. 2723–2729 in Escherichia coli *and* Salmonella: *Cellular and Molecular Biology*, edited by F. C. Neidhardt, J. Li, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter and H. E. Umbarger. American Society for Microbiology Press, Washington, DC.

Fang, F. C., S. J. Libby, N. A. Buchmeier, P. C. Loewen, J. Switala *et al.*, 1992 The alternative sigma factor *katF* (*rpoS*) regulates *Salmonella* virulence. Proc. Natl. Acad. Sci. USA **89:** 11978–11982.

Friedrich, M. J., N. E. Kinsey, J. Vila and R. J. Kadner, 1993 Nucleotide sequence of a 13.9-kb segment of the 90-kb virulence plasmid of *Salmonella typhimurium*: the presence of fimbrial biosynthetic genes. Mol. Microbiol. **8:** 543–558.

Gulig, P., and R. Curtiss, III, 1987 Plasmid-associated virulence of *Salmonella typhimurium*. Infect. Immun. **55:** 2891–2901.

Gulig, P.A., and T. J. Doyle, 1993 The *Salmonella typhimurium* virulence plasmid increases the growth rate of *Salmonellae* in mice. Infect. Immun. **61:** 504–511.

Gulig, P.A., H. Danbara, D. G. Guiney, A. J. Lax, F. Norel *et al.*, 1993 Molecular analysis of *spv* virulence genes of the *Salmonella* virulence plasmids. Mol. Microbiol. **7:** 825–830.

Heffernan, E. J., J. Harwood, J. Fierer and D. Guiney, 1992 The *Salmonella typhimurium* virulence plasmid complement resistance gene *rck* is including *pagC* and *ail*. J. Bacteriol. **174:** 84–91.

Jones, G. W., D. K. Rabert, D. M. Svinarich and H. J. Whitfield, 1982 Association of adhesive, invasive, and virulent phenotypes of *Salmonella typhimurium* with autonomous 60-megadalton plasmids. Infect. Immun. **38:** 476–486.

Jukes, T. H., and C. R. Cantor, 1969 *Evolution of Protein Molecules*. Academic Press, New York.

Krause, M., J. Fierer and D. Guiney, 1990 Homologous DNA sequences on the virulence plasmids of pathogenic *Yersinia* and *Salmonella dublin* lane. Mol. Microbiol. **4:** 905–911.

Krause, M., C. Roudier, J. Fierer, J. Harwood and D. Guiney, 1991 Molecular analysis of the virulence locus of the *Salmonella dublin* plasmid pSDL2. Mol. Microbiol. **5:** 307–316.

Kumar, S., K. Tamura and M. Nei, 1993 MEGA: Molecular Evolutionary Genetics Analysis, Version 1.0. The Pennsylvania State University, University Park, PA.

Le Minor, L., and M. Y. Popoff, 1987 Designation of *Salmonella enterica* sp. nov. rev., as the type and only species of the genus *Salmonella*. Int. J. Syst. Bacteriol. **37:** 465–468.

Muller, D., C. Hughes and W. Goebel, 1983 Relationship between plasmid and chromosomal hemolysin determinants of *Escherichia coli*. J. Bacteriol. **153:** 846–851.

Nakumura, M., S. Sato, T. Ohya, S. Suzuki and S. Ikeda, 1985 Possible relationship of a 36-megadalton *Salmonella enteriditis* plasmid to virulence in mice. Infect. Immun. **47:** 831–833.

Norel, F., C. Coynault, I. Miras, D. Hermant and M. Y. Popoff, 1989 Cloning and expression of plasmid DNA sequences involved in *Salmonella* serotype typhimurium virulence. Mol. Microbiol. **3:** 733–743.

Norel, F., V. Robbe-Saule, M. Y. Popoff and F. C. Coynault, 1992 The putative sigma factor KatF (RpoS) is required for the transcription of the *Salmonella typhimurium* virulence gene *spvB* in *Escherichia coli*. FEMS Microbiol. **78:** 271–276.

Olsen, J. E., D. J. Brown, D. L. Baggesen and M. Bisgaard, 1994 Stability of plasmids in five strains of *Salmonella* maintained in stab cultures at different temperatures. J. Appl. Bacteriol. **77:** 155–159.

Ou, J. T., M. Y. Lin and H. L. Chao, 1994 Presence of F-like OriT base-pair sequence on the virulence plasmids of *Salmonella* serovars Gallinarum, Enteriditis, and Typhimurium, but absent in those of Choleraesuis and Dublin. Microb. Pathog. **17:** 13–21

Popoff, M. Y., and L. Le Minor, 1992 Antigenic Formulas of the *Salmonella* Serovars. WHO Collaborating Center for Reference and Research on *Salmonella*. Institut Pasteur, Paris.

Radnedge, L., M. A. Davis, B. Youngren and S. J. Austin, 1997 Plasmid maintenance functions of the large virulence plasmid of *Shigella flexneri*. J. Bacteriol. **179:** 3670–3675.

Reeves, M. W., G. M. Evins, A. A. Heiba, B. D. Pliikaytis and J. J. Farmer, III, 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:** 311–320.

Rhen, M., C. D. O'Connor and S. Sukupolvi, 1992 The outer membrane permeability mutation of the virulence-associated plasmid of *Salmonella typhimurium* is located in a *traT*-like gene. FEMS Microbiol. Lett. **552:** 145–154.

Rodriguez-Pena, J. M., M. Bisna, M. Ibanez and R. Rotger, 1997 Genetic map of the virulence plasmid of *Salmonella enteritidis* and nucleotide sequence of its replicons. Gene **188:** 53–61.

Roudier, C., J. Fierer and D. G. Guiney, 1992 Characterization of translation termination mutations in the *spv* operon of the *Salmonella* virulence plasmid pSDL2. J. Bacteriol. **174:** 6318–6423.

Saitou, N., and M. Nei, 1987 The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Selander, R. K., J. Li, E. F. Boyd, F.-S. Wang and K. Nelson, 1994 DNA sequence analysis of the genetic structure of populations of *Salmonella enterica* and *Escherichia coli*, in Bacterial Diversity and Systematics, edited by F. G. Priest and B. J. Tindall. Plenum Press, New York.

Taira, S., and M. Rhen, 1989 Identification and genetic analysis of *mkaA*—a gene of the *Salmonella typhimurium* virulence plasmid necessary for intracellular growth. Microb. Pathog. **7:** 165–173.

Taira, S., P. Riikonen, H. Saarilahti, S. Sukupolvi and M. Rhen, 1991 The *mkaC* virulence gene of the *Salmonella* serovar typhimurium 96 kb plasmid encodes a transcriptional activator. Mol. Gen. Genet. **228:** 381–384.

Terakado, N., T. Sekizaki, K. Hashimoto and S. Naitoh, 1983 Correlation between the presence of a fifty-megadalton plasmid in *Salmonella dublin* and virulence in mice. Infect. Immun. **41:** 443–444.

Tinge, S. A., and R. Curtiss, III, 1990 Isolation of the replication and partitioning regions of the *Salmonella typhimurium* virulence plasmid and stabilization of heterologous replicons. J. Bacteriol. **172:** 5266–5277.

Williamson, M. C, G. D. Baird and E. J. Manning, 1988 A common virulence region on plasmids from eleven serotypes of *Salmonella*. J. Gen. Microbiol. **134:** 975–982.

Welch, R. A., R. Hull and S. Falkow, 1983 Molecular cloning and physical characterization of a chromosomal hemolysin from *Escherichia coli*. Infect. Immun. **42:** 178–186.

Communicating editor: P. L. Foster