Identification of a pathogenicity island required for *Salmonella* survival in host cells

HOWARD OCHMAN*, FERNANDO C. SONCINI†, FELIX SOLOMON†, AND EDUARDO A. GROISMAN†‡

*Department of Biology, University of Rochester, Rochester, NY 14627; and †Department of Molecular Microbiology, Washington University School of Medicine, Campus Box 8230, St. Louis, MO 63110

ABSTRACT   We have identified a region unique to the *Salmonella typhimurium* chromosome that is essential for virulence in mice. This region harbors at least three genes: two (*spiA* and *spiB*) encode products that are similar to proteins found in type III secretion systems, and a third (*spiR*) encodes a putative regulator. A strain with a mutation in *spiA* was unable to survive within macrophages but displayed wild-type levels of epithelial cell invasion. The culture supernatants of the *spi* mutants lacked a modified form of flagellin, which was present in the supernatant of the wild-type strain. This suggests that the Spi secretory apparatus exports a protease, or a protein that can alter the activity of a secreted protease. The "pathogenicity island" harboring the *spi* genes may encode the virulence determinants that set *Salmonella* apart from other enteric pathogens.

Although virulence genes map to numerous chromosomal locations (1), several large regions—termed "pathogenicity islands"—often define virulence characteristics in enteric bacteria. Pathogenicity islands are found at several positions in the chromosome, and the genes within a given island often determine the specific disease condition that results from infection. For example, uropathogenic (2) and enteropathogenic (3) strains of *Escherichia coli* harbor different pathogenicity islands at the same chromosomal location (the *selC* locus). Although the pathogenicity island specific to uropathogenic strains encodes a hemolysin (2), the island in enteropathogenic *E. coli* encodes a type III secretion system that exports proteins responsible for the attachment and effacing lesions of intestinal cells (4). Therefore, the specific virulence properties of two types of pathogenic *E. coli* are largely determined by the "cassette" present at the *selC* locus.

Through the systematic screening of a *Salmonella typhimurium* DNA library, Fitts (5) recovered numerous clones containing sequences that were apparently unique to the salmonellae. Three of these clones appeared to have been acquired by horizontal gene transfer since they had base compositions much lower than the overall G+C content of 52–54% of the *Salmonella* genome (6). The restricted phylogenetic distributions of these sequences suggested that they encode biochemical or cellular functions that set *Salmonella* apart from other enteric species (7). One of these three clones contained a gene cluster—designated *spa*—that enables *Salmonella* to invade epithelial cells (8). The *spa* gene cluster is part of a 40-kb pathogenicity island that encodes a variety of determinants that mediate the entry of *Salmonella* into nonphagocytic cells (9). These determinants include a type III secretion system and its substrates, which are homologous to an antigen export apparatus on the virulence plasmid of *Shigella* (10). Another of these clones specified a member of the LysR family of transcriptional regulators—SinR—and has no role in virulence (6).

In a survey of enteric species, only one of the clones—RF333—was strictly confined to *Salmonella* (6). Here we report the molecular genetic characterization of this clone and the virulence properties of strains harboring mutations in the corresponding region of the chromosome. These analyses define a new pathogenicity island (*spi*) within *Salmonella* encoding a regulator and a type III secretion system essential for virulence in mice and survival in macrophages. The *spi* island is distinct from the invasion region containing the *spa* locus, suggesting that multiple events of horizontal transfer have been responsible for the acquisition of pathogenic properties in *Salmonella*.

MATERIALS AND METHODS

Bacterial Strains, Plasmids, and Growth Conditions. All *S. typhimurium* strains used in this study were derived from the mouse-virulent wild-type strain 14028s. Strains with mutations in the RF333 chromosomal region were constructed from the pMS333 derivatives pEG7186 and pEG7200, which harbor *kan* insertions in *spiA* and *spiR*, respectively (6). The resulting strains—EG5793 and EG5799—contain the *Sma*I 1.3-kb *kan* fragment from plasmid pUC4-KIXX (Pharmacia) inserted at the *Pme*I and *Nae*I sites of RF333, respectively (Fig. 1). The *kan* gene is in the same transcriptional orientation as *spiA* in EG5793 and as *spiR* in EG5799, and the *kan* promoter is predicted to transcribe the genes located downstream of *spiA* and *spiR*. [Similar mutations within the *spa* region generated by inserting the same *kan* cassette were not polar on downstream genes (8)]. The structure of the *spi* locus in the mutant strains was verified by PCR with RF333-specific primers and by Southern hybridization analysis with both the 5.7-kb *Bam*HI fragment of RF333 and *kan*-specific probes. Plasmid pMS333 is a pUC19 derivative with the *Bam*HI fragment from RF333 inserted at the *Bam*HI site. Ampicillin was used at 50 μg/ml and kanamycin at 40 μg/ml.

DNA Sequencing, Analysis, and Other Molecular Biological Manipulations. Sequencing was carried out on both strands of plasmid pMS333 by the dideoxy chain termination method using the Sequenase kit (United States Biochemical) with ³⁵S-labeled dATP. Additional primers were synthesized as the partial sequences were obtained. Sequence analyses were performed with both GeneWorks (IntelliGenetics) and the GCG package (University of Wisconsin). Restriction endonucleases and phage T4 ligase were purchased from Bethesda Research Laboratories, Boehringer Mannheim, or New England Biolabs, and used according to the supplier's specifications. Other protocols were taken from Maniatis *et al.* (11).

Virulence Assays and Preparation of Culture Supernatants. Macrophage survival assays were conducted with the macrophage-like cell line J774 as described (12). Adhesion to and invasion of human intestinal Henle-407 cells were investigated

Genetics: Ochman *et al.*
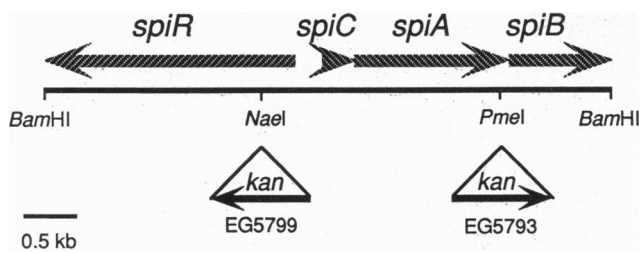
Proc. Natl. Acad. Sci. USA 93 (1996) 7801

FIG. 1. Genetic and physical maps of the *S. typhimurium* RF333 (*spi*) region. The arrows indicate the size and direction of transcription of the four largest open reading frames (ORF) contained within RF333. The positions of the *kan* gene insertions in the mutants EG5793 and EG5799 are indicated.

as described (13) except that dilutions of bacterial cells were performed in PBS instead of Luria–Bertani broth. Virulence assays were performed with 7–8-week-old female BALB/c mice inoculated intraperitoneally with 100 μl of bacteria diluted in PBS. Viability was recorded for at least 30 days in 10 mice per mutant at each dose. Culture supernatants were prepared and analyzed as described (14).

RESULTS

Molecular Genetic Analysis of a *Salmonella*-Specific Chromosome Segment. We determined the nucleotide sequence of the 5.7-kb *Bam*HI insert of clone RF333 and identified four ORFs, which were designated as *spiA*, *spiB*, *spiC*, and *spiR* (*spi* for *Salmonella* pathogenicity island). The *spiA*, *spiB*, and *spiC* genes are organized in an operon that is divergently transcribed from *spiR* (Fig. 1). The *spiA* gene encodes a protein of 497 amino acids, and the *spiB* ORF is >323 amino acids and extends beyond the *Bam*HI site at the end of the insert. The *spiA* gene is preceded by the *spiC* gene, which encodes a protein of 127 or 133 amino acids, depending on the translational start site. (Presumably, the 127-amino acid ORF is translated because the sequence of its putative ribosome binding site—AGGAG—corresponds to the 3' end of the 16s rRNA.) The *spiA* initiation codon is 1 bp downstream of the stop codon for *spiC*, which suggests that these proteins are translationally coupled. The initiation codon of *spiC* is separated by 400 bp from that of the *spiR* gene, which encodes a protein of >822 amino acids.

SpiR Is Similar to Proteins of the Two-Component Family. Two-component systems generally consist of a sensor protein that, in response to environmental cues, modifies the ability of a regulatory protein to affect transcription of particular genes (15). Several such systems have been implicated in the regulation of virulence functions in the salmonellae (16). The deduced amino acid sequence of the SpiR protein exhibits homology to the subgroup of two-component systems that have both conserved domains—the histidine kinase domain of sensors and the receiver domain of response regulators—in a single molecule (Fig. 2A). The SpiR protein sequence is related to: (*i*) *E. coli* BarA, which is encoded by a gene that was isolated as a multicopy number suppressor of an *envZ* mutation (17); (*ii*) RscC, which controls capsule synthesis in *E. coli* (18); (*iii*) BvgS, a major regulator of virulence determinants in *Bordetella* (19); as well as two-component systems of plants and yeast (20).

SpiA and SpiB Are Similar to Proteins of Type III Secretion Systems. Type III secretion systems export proteins by a mechanism distinct from either the classical *sec*-dependent pathway or that responsible for secretion of hemolysin in *E. coli* (21). These systems mediate secretion of virulence proteins in both animal and plant pathogens: Homologs include the *Salmonella* Inv/Spa and the *Shigella* Mxi/Spa complexes, which are necessary for host cell invasion (22); the *Yersinia* Ysc/Lcr proteins, which secrete a tyrosine phosphatase and a

cytotoxin (23); and proteins in the plant pathogens *Erwinia*, *Pseudomonas*, and *Xanthomonas*, which are involved in the hypersensitive response (24, 25).

As shown in Fig. 2B, SpiA displays the highest degree of sequence similarity to SepC of enteropathogenic *E. coli* (4), YscC of *Yersinia* (26), and InvG of *Salmonella* (27); and SpiB exhibited a low level of similarity to *Yersinia* YscD (26) and to an unreported ORF adjacent to *eaeA* in enteropathogenic *E. coli* (4). The gene order of *spiA* and *spiB* on the *Salmonella* chromosome is the same as that of *yscC* and *yscD* on the virulence plasmid of *Yersinia* (26). But *spiC* has no homologs in the sequence databases and apparently is a component specific to the *spi* locus.

Virulence Properties of *spi* Mutants. To determine the function of the genes present in RF333, we constructed mutations in the *spiA* and *spiR* ORFs (Fig. 1; see *Materials and Methods*). When these mutations were transferred back to the *S. typhimurium* chromosome, viable colonies were obtained, indicating that the inactivated genes are not essential for growth in complex laboratory media. These mutants behaved like the wild-type parent in their ability to grow at different temperatures and in defined media.

Given the prevalence of type III secretion systems among animal and plant pathogens (21, 22), we hypothesized that strains with mutations in *spi* genes might be unable to export certain virulence determinants and, consequently, would be attenuated for virulence. Indeed, neither *spiR* nor *spiA* mutants killed BALB/c mice when inoculated intraperitoneally at >10,000 times the median lethal dose of the wild-type parent. This finding is in contrast to the phenotypes of the *inv/spa* mutants of *S. typhimurium* that are attenuated only when the bacteria are administered orally (28).

To determine the particular stage of infection in which the RF333-encoded determinants are required, we tested the ability of the *spiA* mutant to invade cultured epithelial cells and to survive within macrophages *in vitro*. The *spiA* mutant displayed wild-type levels of epithelial cell invasion but was defective for intramacrophage survival (Fig. 3), a result that further distinguishes the role of the *spi* genes from that of the genes in the *inv/spa* locus. Taken together with the mice virulence data, these experiments establish that the Spi proteins are essential for stages of infection beyond the initial interaction with intestinal cells.

Identification of Proteins Exported by the Spi Secretion System. Because *spi* apparently encodes components of a type III secretion system, we investigated whether *spi* mutants were defective in the export of proteins to the growth media. Culture supernatants of wild-type, *spiA*, and *spiR* mutant strains were prepared and their protein profiles were compared. The supernatant of the wild-type strain had a 45-kDa protein that was absent from the supernatant of the mutant strains; instead, the mutant supernatant harbored a 47-kDa protein that was absent from the wild-type cultures (Fig. 4). We isolated both the 45- and 47-kDa proteins and determined the amino acid sequences of their first 15 residues. The two sequences were identical to one another and to the *S. typhimurium* flagellin, and it is likely that the 45-kDa protein is a processed form of the 47-kDa protein. The profile of secreted proteins for the *spiR* mutant was identical to that displayed by the *spiA* mutant strain, suggesting that SpiR controls transcription of the *spiCAB* operon.

DISCUSSION

The virulence phenotypes of several enteric pathogens have often been attributed to the presence of DNA segments that are absent from the genomes of nonpathogenic strains. These pathogenicity islands are largely responsible for the virulence properties of enteropathogenic and uropathogenic strains of *E. coli* (2, 3). In *Salmonella*, many of the invasion determinants

A

```
                                                                              *
SpiR   371..LE... .......NKV A......... ..ERTQALNE AKKRAERANK RKSIHLTVIS HELRTPMNGV LGAIE..LLQ TTPLNIEQQG LADTARNCTL  442
BarA   247..LAAYH EEMQHNIDQA TSDLRETLEQ MEIQNVELDL AKKRAQEAAR IKSEFLANMS HELRTPLNGV IGFTR..LTL KTELTPTQRD HLNTIERSAN  339
RcsC   408..LQISF VHSRYRNENV AICVLVDVSS RVKMEESLQE MAQAAEQASQ SKSMFLATVS HELRTPLYGI IGNLD..LLQ TKELPKGVDR LVTAMNNSSS  500
LemA   229..LQNAQ EELQLSIDQA TEDVRQNLET IEIQNIELDL ARKEALEASR IKSEFLANMS HEIRTPLNGI LGFTH..ILQ KSELTPRQFD YLGTIEKSAD  321
BvgS   677..IPYGD SLGELKGIIG GWIDITERAE LLRE...LHD AKESADAANR AKTTFLATMS HEIRTPMNAI IGMLELALLR PTDQEPD.RQ SIQVAYDSAR  767
Etrl   307..LSHAA ILEE...... ...SMRARDL LMEQNVALDL ARREAETAIR ARNDFLAVMN HEMRTPMHAI IAL..SSLLQ ETELTPEQRL MVETILKSSN  390

SpiR   SLLAIINNLL DFSRIESGHF TLHMEETALL PLLDQAMQTI QGPAQSKKLS LRTFVGQHVP LYFHTDSIRL RQILVNLLGN AVKFTETGGI RLTV.....K  537
BarA   NLLAIINDVL DFSKLEAGKL ILESIPPFLR STLDEVVTLL AHSSHDKGLE LTLNIKSDVP DNVIGDPLRL QQIITNLVGN AIKFTENGNI DILVEKRALS  439
RcsC   LLLKIISDIL DFSKIESEQL KIEPREFSPR EVMNHITANY LPLVVRKQLG LYCFIEPDVP VALNGDPMRL QQVISNLLSN AIKFTDTGCI VLHV.....R  595
LemA   NLLSIINEIL DFSKIEAGKL VLDNIPFNLK DLLQDTLTIL APAAHAKQLE LVSLVYRDTP LALSGDPLRL RQILTNLVSN AIKFTREGTI VARAMLEDET  421
BvgS   SLLELIGDIL DIAKIEAGKF DLAPVRTALR VLPEGAIRVF DGLARQKGIE LVLKTDIVGV DDVLIDPLRM KQVLSNLVGN AIKFTTEGQV .VLAVTARPD  866
Etrl   LLATLMNDVL DLSRLEDGSL QLELGTFNLH TLFREVLNLI KPIAVVKKLP ITLNLAPDLP EFVVGDEKRL MQIILNIVGN AVKFSKQGSI SVTALVTKSD  490

SpiR   RHEEQLIFLV SDSGKGIEIQ QQSQIFTAFY QADT....NS Q.GTGIGLTI ASSLAKMMGG NLTLKSVP.. GVGTCVSLVL PLQEYQPPQP IK.GTLSAPF  629
BarA   NTKVQIEVQI RDTGIGIPER DQSRLFQAFR QADASISRRH G.GTGLGLVI TQKLVNEMGG DISFHSQPNR GSTFWFHINL DLNPNIIIEG PSTQCLAGKR  538
RcsC   ADGDYLSIRV RDTGVGIPAK EVVRLFDPFF QVGTGVQRNF Q.GTGLGLAI CEKLISMMDG DISVDSEP.. GMGSQFTVRI PLYGAQYPQK KGVEGLSGKR  692
LemA   EEHAQLRISV QDTGIGLSSQ DVRALFQAFS QADNSLSRQP G.GTGLGLVI SKRLIEQMGG EIGVDSTPGE GSEFWISLKL PKAREDKEES LNIP.LGGLR  519
BvgS   GDAAHVQFSV SDTGCGISEA DQRQLFKPFS QVGGSAEAGP APGTGLGLSI SRRLVELMGG TLVMRSAP.. GVGTTVSVDL RLT....... ..........  947
Etrl   TRA......V KDSGAGINPQ DIPKIFTKFA QTQSLATRSS G.GSGLGLAI SKRFVNLMEG NIWIESDG.L GKGCTAIFDV KLG....... ..........  565

                                                        *
SpiR   689..LQILL VDDADINRDI IGKMLVSLGQ HVTIAASSNE ALTLSQQQRF DLVLIDIRMP EIDGIE...C VRLWHDEPNN LDPDCMFVAL SASVATEDIH  780
BarA   668..MTVMA VDDNPANLKL IGALLEDMVQ HVELCDSGHQ AVERAKQMPF DLILMDIQMP DMDGIRACEL IHQL...PHQ QQTP..VIAV TAHAMAGQKE  757
RcsC   809..MMILV VDDHPINRRL LADQLGSLGY QCKTANDGVD ALNVLSKNHI DIVLSDVNMP NMDGYR.... ...LTQRIRQ LGLTLPVIGV TANALAEEKQ  896
LemA   657..PRVLC VDDNPANLLL VQTLLEDMGA EVVAVEGGYA AVNAVQQEAF DLVLMDVQMP GMDGRQATEA IRAWEAERNQ SSLP..IVAL TAHAMANEKR  749
BvgS   973..LRVLV VDDHKPNLML LRQQLDYLGQ RVIAADSGEA ALALWREHAF DVVITDCNMP GISGYELARR IRAAEAAPGY GRTRCILFGF TASAQMDEAQ 1067
Etrl   593..LKVLV MDENGVSRMV TKGLLVHLGC EVTTVSSNEE CLRV.VSHEH KVVFMDVCMP GVENYQIALR IH..EKFTKQ RHQRPLLVAL SGNTDKSTKE  684

              *
SpiR   RCKKNGIHHY ITKPVTLATL ARYISIAA//      808
BarA   KLLGAGMSDY LAKPIEEERL HNLLLRYK//      785
RcsC   RCLESGMDSC LSKPLTLDVI KQSLTLYA//      924
LemA   SLLQSGMDDY LTKPISERQL AQVVLKWT//      777
BvgS   ACRAAGMDDC LFKPIGVDAL RQRLNEAV//     1095
Etrl   KCMSFGLDGV LLKPVSLDNI RDVLSDLL//      712
```

B

```
SpiA        MVVN KRLILILLFI ...LNTAKSD ELSWKGNDFT LYARQMPLAE VLHLLSENYD TAITISPLIT AT.FSGKIPP GPPVDILNNL AAQYDLLTWF  90
SepC        MKKI SFFIFTALFC CSAQAAPSSL EKRLGKNEYF IITKSSPVRA ILNDFAANYS IPVFISSSVN DD.FSGEIKN EKPVKVLEKL SKLYHLTWYY  93
InvG   MKTHILL ARVLACAALV LVTPGYSSEK IPVTGSG... FVAKDDSLRT FFDAMALQLK EPVIVSKMAA RKKITGNFEF HDPNALLEKL SLQLGLIWYF  94
MxiD   MKKFNIK SLTLLIVLLP LIVNANNIDS HLLEQNDIAK YVAQSDTVGS FFERFSALLN YPIVVSKQAA KKRISGEFDL SNPEEMLEKL TLLVGLIWYK  97
YscC   MAFPLHSFFK RVLTGTLLLL SSY...SWAQ ELDWLPIPYV YVAKGESLRD LLTDFGANYD ATVVVSDKIN DK.VSGQFEH DNPQDFLQHI ASLYNLVWYY  96

SpiA   DGSMLYVYPA SLLKHQVITF NILSTGRFIH YLRSQNILSS PGCEVKEITG TKAVEVSGVP SCLTRISQLA SVLDNALI.. KRKDSAVSVS IYTLKYATAM  188
SepC   DENILYIYKT NEISRSIITP TYLDIDSLLK YLSDTISVNK NSCNVRKITT FNSIEVRGVP ECIKYITSLS ESLDKEAQ.. SKAKNKDVVK VFKLNYASAT  191
InvG   DGQAIYIYDA SEMRNAVVSL RNVSLNEFNN FLKR.SGLYN KNYPLRGDNR KGTFYVSGPP VVVNAAIVOA TMMDKQND.. GIELGRQKIG VMRLNNTFVG  191
MxiD   DGNALYIYDS GELISKVILL ENISLNYLIQ YLKD.ANLYD HRYPIRGNIS DCTFYISGPP ALVELVANTA TLLDKQVS.. SIGTDKVNFG VIKLKNTFVS  194
YscC   DGNVLYIFKN SEVASRLIRL QESEAAELKQ AL.QRSGIWE PRFGWRPDAS NRLVYVSGPP RYLELVEQTA AALEQQTQIR SEKTGALAIE IFPLKYASAS  195

SpiA   DTQYQYRDQS VVVPGVVSVL REM.SKTSVP ..TSSTNNGS PAT....... .......... .........Q AL...PMFAA DPRQNAVIVR DYAANMAGYR  256
SepC   DITYKYRDQN VVVPGVVSIL KTMASNGSLP ..STGKGAVE RSG....... .......... .........N LFDNSVTISA DPRLNAVVVK DREITMDIYQ  263
InvG   DRTYNLRDQK MVIPGIATAI ERLLQGEEQP LGNIVSSEPP AMPAFSANGE KGKAANYAGG MSLQEALKQN AAAGNIKIVA YPDTNSLLVK GTAEQVHFIE  291
MxiD   DRTYNMRGED IVIPGVATVV ERLLNN..... .GKALSNRQA QNDPMPPFNI TQKVSEDSND FSF.SSVTNS SILEDVSILA YPETNSILVK GNDQQIQIIR  288
YscC   DRTIHYRDDE VAAPGVATIL QRVLSDATIQ QVTVDNQRIP QAA....... .......... .........T RASAQARVEA DPSLNAIIVR DSPERMPMYQ  269

SpiA   KLITELDQRQ QMIEISVKII DVNAGDINQL GIDWGTAVSL GGK.KIAFNT GLNDGGASGF STVIS..... DTSNFMVRLN ALEKSSQAYV LSQPSVVTLN  350
SepC   QLISELDIEQ RQIEISVSII DVDANDLQQL GVNWSGTLNA GQG.TIAFNS STAQVNIS.. SSVIS..... NASNFMIRVN ALQQNSKAKI LSQPSIITLN  355
InvG   MLVKALDVAK RHVELSLWIV DLNKSDLERL GTSWSGTITI GDK....LGV SLNQS...SI STL....... DGSRFIAAVN ALEEKKQATV VSRPVLLTQE  377
MxiD   DIITQLDVAK RHIELSLWIV DIDKSELNNL GVNWQGTASF GDS....FGA SFNMSSSASI STL....... DGNKFIASVM ALNQKKKANV VSRPVILTQE  377
YscC   RLIHALDKPS ARIEVALSIV DINADQLTEL GVDWRVGIRT GNNHQVVIKT TGDQSNIASN GALGSLVDAR GLDYLLARVN LLENEGSAQV VSRPTLLTQE  369

SpiA   NIQAVLDKNI TFYTKLQGEK VAKLESITTG SLLRVTPRLL NDNGT..... ..QKIMLNLN IQDGQQSDTQ ....SETD... ..PLPEVQNS EIASQATLLA  435
SepC   NMQAILDKNV TFYTKVSGEK VASLRSITSG TLLRVTPRIL DDSSNSLTGK RRERVRLLLD IQDGNQSTNQ ...SNAQDA SSTLPEVQNS EMTTEATLSA  451
InvG   NVPAIFDNNR TFYTKLIGER NVALEHVTYG TMIRVLPRFS ADG....... ...QIEMSLD IEDGNDKTPQ ....SDTTTS VDALPEVGRT LISTIARVPH  463
MxiD   NIPAIFDNNR TFYVSLVGER NSSLEHVTYG TLINVIPRFS SRG....... ...QIEMSLT IEDGTGNSQS NYNYNNENTS V..LPEVGRT KISTIARVPQ  465
YscC   NAQAVIDHSE TYYVKVTGKE VAELKGITYG TMLRMTPRVL TQGDKS.... ...EISLNLH IEDGNQKPNS ....SGIE.. ..GIPTISRT VVDTVARVGH  454

SpiA   GQSLLLGGFK QGKQIHSQNK IPLLGDIPVV GHLFRNDTTQ VHSVIRLFLI KASVVNNGIS HG                                         498
SepC   GKVVYLADLS KIKKVSPKDG IPLLSDIPVI GSLFSSTVKQ KATPIKSASS E                                                  512
InvG   GKSLLVGGYT RDANTDTVQS IPFLGKLPLI GSLFRYSSKN KSNVVRVFMI EPKEIVDPLT PDASE..... SVNNILKQSG AWSGDDKL.. .QKWVRV///  552
MxiD   GKSLLIGGYT HETNSNEIIS IPFLSSIPVI GNVFKYKTSN ISNIVRVFLI QPREIKESSY YNTAEYKSLI SEREIQKTTQ IIPSETTLLE DEKSLVS///  562
YscC   GQSLIIGGIY RDELSVALSK VPLLGDIPYI GALFRRKSEL TRRTVRLFII EPRIIDEGIA HHLALGNGQD LRTGILTVDE ISNQSTTLNK LLGGSQC///  551
```
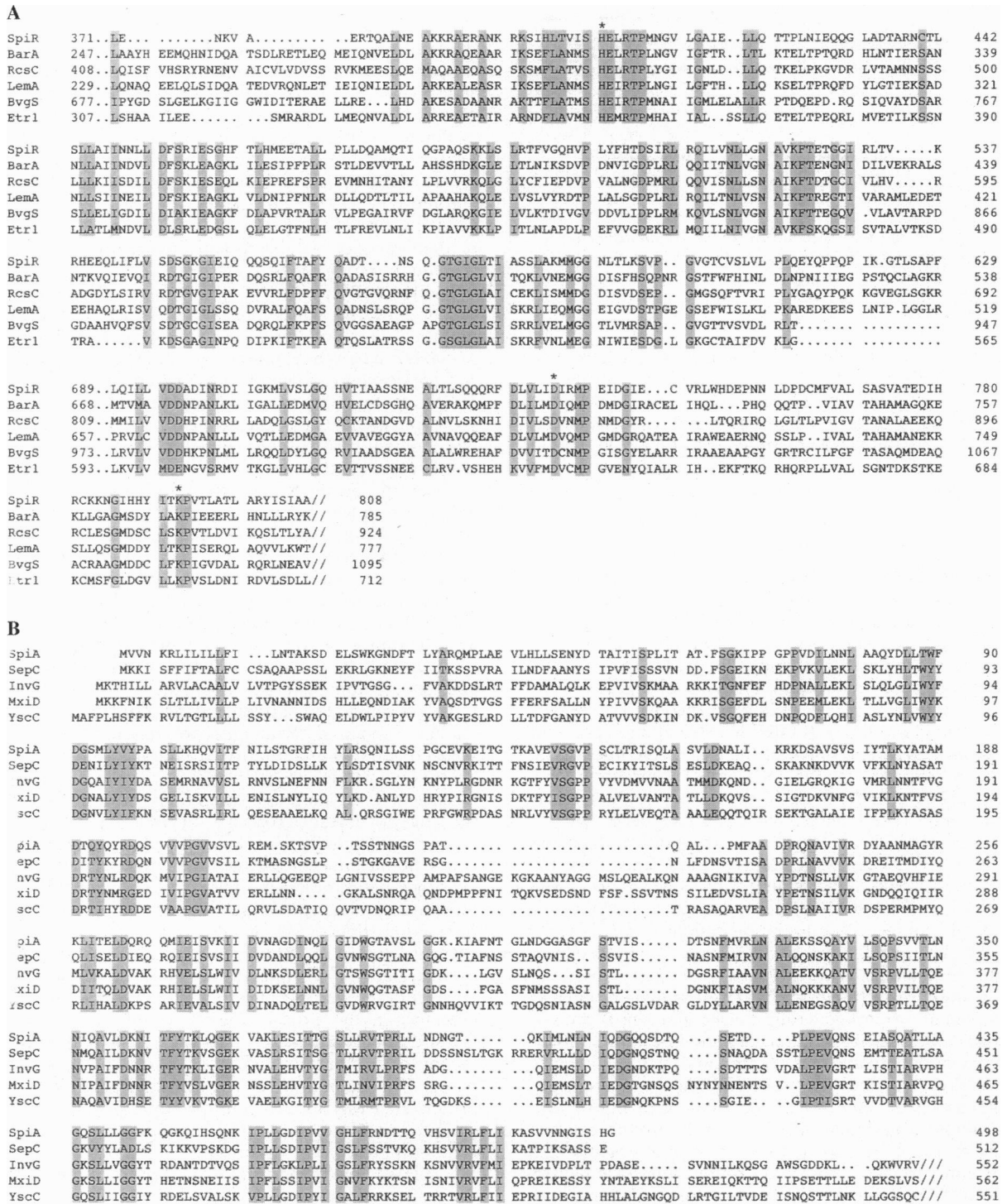
FIG. 2. Comparison of *Salmonella* Spi with related proteins. The deduced amino acid sequences of proteins encoded by the *S. typhimurium spi* island are shown in the single-letter code. (*A*) Comparison of SpiR to members of the two-component family. (*B*) Comparison of SpiA to components of type III secretion systems. Alignment was conducted using the PILEUP program (GCG). Asterisks indicate residues conserved in all members of this family. For any given comparison, residues present in the majority of related proteins are highlighted.

have been localized to a 40-kb segment of the chromosome specifying at least 25 genes encoding a secretion system, its effectors, and the regulators controlling their expression (9). Because these genes are ancestral to all eight subgroups of *Salmonella enterica*, it has been hypothesized that the acquisition of this segment was an essential step in the evolution of invasion by *Salmonella* (14). However, the presence of structurally and functionally equivalent genes in the *Shigella flexneri* virulence plasmid (22) indicates that this invasion island does

not confer the specific properties that distinguish *Salmonella* from other enteric pathogens.

We have identified a new gene cluster essential for *Salmonella* virulence that includes the *spiCAB* operon, encoding components of a putative type III secretion apparatus, and a regulatory gene, *spiR*. Because the SpiA and SpiB proteins were similar to components of secretion apparatuses, we examined the profiles of proteins exported by *spi* and wild-type strains. The most conspicuous difference was the presence of
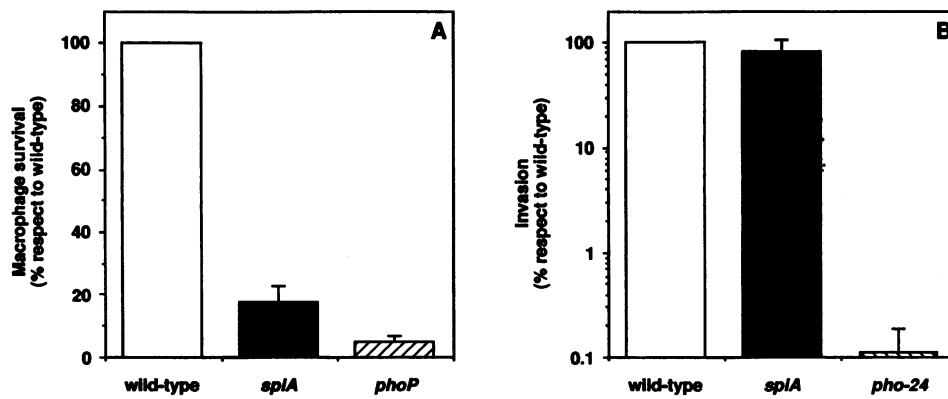
Genetics: Ochman *et al.*

Proc. Natl. Acad. Sci. USA 93 (1996)   7803

FIG. 3. Intramacrophage survival (*A*) and invasion (*B*) properties of an *spiA* mutant of *S. typhimurium*. Values represent the mean of triplicate samples ± SD. Properties of *phoP* and *pho-24* strains are shown for comparative purposes.

two distinct forms of flagellin: 45 kDa in wild-type and 47 kDa in the *spi* mutants (Fig. 4). Thus, Spi may export a protease or a protein that modifies the activity of a secreted protease. However, flagellin is probably not the physiologically relevant substrate of this protease because flagellin is not essential for virulence and *spi* mutants are motile.

Although the Spi secretion system is reminiscent of that encoded within the *spa* invasion island of *Salmonella*, strains with mutations in the *spi* genes displayed wild-type levels of invasion (Fig. 3). Unlike *invA* mutants, which are only attenuated upon oral inoculation (28), the *spi* mutants failed to cause lethal infections in mice even when inoculated intraperitoneally. The *spi* genes were required for intramacrophage survival—a trait fundamental to *Salmonella* pathogenesis (29, 30)—and not surprisingly, homologous sequences have not been detected in other enteric species (6). Thus, Spi constitutes a novel type III secretion system that exports proteins that permit the survival of *Salmonella* within phagocytic cells, perhaps by modifying host factors required for phagosome–lysosome fusion (31) or phagosome acidification (32).

The *spi* gene cluster has an anomalous base composition of only 42.1% G+C, which is much lower than the overall G+C content of 52–54% of the *Salmonella* genome (6), and suggests that it was acquired by horizontal transfer. Regions acquired through horizontal processes often harbor the genes encoding both structural and regulatory elements in a contiguous DNA segment (9, 33); and within the 5.7-kb region analyzed in this study, we identified genes encoding products similar to components of a secretion apparatus (*spiA* and *spiB*) and a transcriptional regulator (*spiR*). SpiR exhibits broad-scale similarity to proteins of the two-component regulatory family and is likely to control the expression of adjacent loci in the *spi* pathogenicity island. Indeed, the *spiR* mutant displayed a
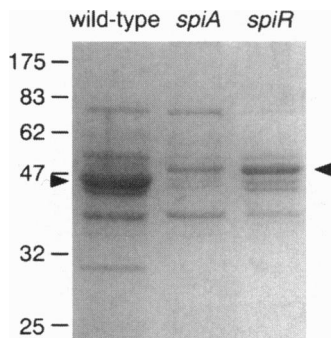
profile of exported proteins identical to that of the *spiA* strain (Fig. 4), implying that SpiR regulates the *spiCAB* operon and/or genes encoding the substrates of the Spi secretion system.

The *spi* locus is situated between 30.5′ and 33.5′, a region of the *S. typhimurium* chromosome that is relatively devoid of mapped genes (34). Only two other *Salmonella*-specific genes have been mapped to this region: *sly* and *ompD*. Despite the role of *sly* in virulence (35), it is not likely to be contained within the *spi* pathogenicity island due to its base composition—50.1% vs. 42.1% G+C for the *spi* locus—and its presence in *Shigella* and enteropathogenic *E. coli*. The *spi* genes do not correspond to any previously described macrophage-survival loci (36). However, it is presently unknown whether *spiA* is allelic with a virulence gene that has been reported to be homologous to *invG* because neither the DNA sequence nor the map position of this gene was provided (37).

What is the mechanism by which horizontally acquired DNA segments are incorporated into the bacterial genome? In *E. coli*, a site adjacent to the *selC* tRNA gene is the target of insertion of two different pathogenicity islands (2, 3) and of a retronphage (38). Moreover, the second pathogenicity island present in certain uropathogenic strains of *E. coli* is located next to *leuX*, which, like *selC*, encodes a tRNA gene for rarely used codons (2). On the other hand, a 100-kb virulence region in the *Yersinia pestis* chromosome is flanked by repeated sequences (39). Thus, different mechanisms may mediate the incorporation of foreign DNA sequences. Future analysis of the entire *spi* island should reveal the manner in which the *spi* region was obtained by ancestral *Salmonella*.

The enteric pathogens *Salmonella* and *Shigella* employ similar secretion systems to promote entry into host cells (22). The differences in disease pathology and host range displayed by these pathogens most likely result from genes that are species-specific. The *spi* pathogenicity island may harbor such genes because: (*i*) it is present in all *Salmonella* serotypes investigated (5); (*ii*) *spi*-hybridizing sequences have not been detected in the genomes of other enteric species (6); and (*iii*) this region is essential for virulence functions beyond host cell invasion.

FIG. 4. Protein profile of supernatants prepared from wild-type, *spiA*, and *spiR* strains of *S. typhimurium*.

1. Groisman, E. A. & Ochman, H. (1994) *Trends Microbiol.* **2**, 289–294.

2. Blum, G., Ott, M., Lischewski, A., Ritter, A., Imrich, H., Tschäpe, H. & Hacker, J. (1994) *Infect. Immun.* **62**, 606–614.
3. McDaniel, T. K., Jarvis, K. G., Donnenberg, M. S. & Kaper, J. B. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 1664–1668.
4. Jarvis, K. G., Girón, J. A., Jerse, A. E., McDaniel, T. K., Donnenberg, M. S. & Kaper, J. B. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7996–8000.
5. Fitts, R. (1985) *Food Technol.* **39**, 95–102.
6. Groisman, E. A., Sturmoski, M. A., Solomon, F. R., Lin, R. & Ochman, H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 1033–1037.
7. Ochman, H. & Groisman, E. A. (1994) in *The Origin and Evolution of Species Differences in Escherichia coli and Salmonella typhimurium*, eds. Schierwater, B., Streit, B., Wagner, G. P. & DeSalle, R. (Birkhauser, New York), pp. 479–493.
8. Groisman, E. A. & Ochman, H. (1993) *EMBO J.* **12**, 3779–3787.
9. Mills, D. M., Bajaj, V. & Lee, C. A. (1995) *Mol. Microbiol.* **15**, 749–759.
10. Parsot, C. (1994) *Curr. Top. Microbiol. Immunol.* **192**, 217–241.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab. Press, Plainview, NY).
12. Buchmeier, N. A. & Heffron, F. (1989) *Infect. Immun.* **57**, 1–7.
13. Lee, C. A., Jones, B. D. & Falkow, S. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 1847–1851.
14. Li, J., Ochman, H., Groisman, E. A., Boyd, E. F., Solomon, F., Nelson, K. & Selander, R. K. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 7252–7256.
15. Stock, J. B., Surette, M. G., Levit, M. & Park, P. (1995) in *Two-Component Signal Transduction Systems: Structure-Function Relationships and Mechanisms of Catalysis*, eds. Hoch, J. A. & Silhavy, T. J. (Am. Soc. Microbiol., Washington, DC), pp. 25–51.
16. Groisman, E. A. & Heffron, F. (1995) in *Regulation of Salmonella Virulence by Two-Component Regulatory Systems*, eds. Hoch, J. A. & Silhavy, T. J. (Am. Soc. Microbiol., Washington, DC), pp. 319–332.
17. Nagasawa, S., Tokishita, S., Aiba, H. & Mizuno, T. (1992) *Mol. Microbiol.* **6**, 799–807.
18. Gottesman, S. (1995) in *Regulation of Capsule Synthesis: Modification of the Two-Component Paradigm by an Accessory Unstable Regulator*, eds. Hoch, J. A. & Silhavy, T. J. (Am. Soc. Microbiol., Washington, DC), pp. 253–262.
19. Uhl, A. A. & Miller, J. F. (1995) in *Bordetella pertussis BvgAS Virulence Control System*, eds. Hoch, J. A. & Silhavy, T. J. (Am. Soc. Microbiol., Washington, DC), pp. 333–349.
20. Chang, C. & Meyerowitz, E. M. (1994) *Res. Microbiol.* **145**, 481–486.
21. Salmond, G. P. C. & Reeves, P. J. (1993) *Trends Biochem. Sci.* **18**, 7–12.
22. Ochman, H. & Groisman, E. A. (1995) *Can. J. Microbiol.* **41**, 555–561.
23. Cornelis, G. R. (1994) *Curr. Top. Microbiol. Immunol.* **192**, 243–263.
24. Van Gijsegem, F., Gough, C., Zischek, C., Niqueux, E., Arlat, M., Genin, S., Barberis, P., German, S., Castello, P. & Boucher, C. (1995) *Mol. Microbiol.* **15**, 1095–1114.
25. Wei, Z.-M. & Beer, S. V. (1993) *J. Bacteriol.* **175**, 7958–7967.
26. Michiels, T., Vanooteghem, J.-C., Lambert de Rouvroit, C., China, B., Gustin, A., Boudry, P. & Cornelis, G. R. (1991) *J. Bacteriol.* **173**, 4994–5009.
27. Kaniga, K., Bossio, J. C. & Galán, J. E. (1994) *Mol. Microbiol.* **13**, 555–568.
28. Galán, J. & Curtiss, R. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6383–6387.
29. Finlay, B. B. (1994) *Curr. Top. Microbiol. Immunol.* **192**, 163–185.
30. Fields, P. I., Swanson, R. V., Haidaris, C. G. & Heffron, F. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5189–5193.
31. García-del Portillo, F. & Finlay, B. B. (1995) *J. Cell Biol.* **129**, 81–97.
32. Alpuche-Aranda, C. M., Swanson, J. A., Loomis, W. P. & Miller, S. I. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 10079–10083.
33. Roth, J. R., Lawrence, J. G., Rubenfield, M., Kieffer-Higgins, S. & Church, G. M. (1993) *J. Bacteriol.* **175**, 3303–3316.
34. Sanderson, K. E., Hessel, A. & Rudd, K. E. (1995) *Microbiol. Rev.* **59**, 241–303.
35. Libby, S. J., Goebel, W., Ludwig, A., Buchmeier, N., Bowe, F., Fang, F. C., Guiney, D. G., Songer, J. G. & Heffron, F. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 489–493.
36. Bäumler, A. J., Kusters, J. G., Stojiljkovic, I. & Heffron, F. (1994) *Infect. Immun.* **62**, 1623–1630.
37. Hensel, M., Shea, J. E., Gleeson, C., Jones, M. D., Dalton, E. & Holden, D. W. (1995) *Science* **269**, 400–403.
38. Inouye, S., Sunshine, M. G., Six, E. W. & Inouye, M. (1991) *Science* **252**, 969–971.
39. Fetherston, J. D., Schuetze, P. & Perry, R. D. (1992) *Mol. Microbiol.* **6**, 2693–2704.