JOURNAL OF BACTERIOLOGY, Sept. 2003, p. 5055–5065 Vol. 185, No. 17
0021-9193/03/$08.00+0 DOI: 10.1128/JB.185.17.5055–5065.2003

Composition, Acquisition, and Distribution of the Vi Exopolysaccharide-Encoding *Salmonella enterica* Pathogenicity Island SPI-7

Derek Pickard,[1]* John Wain,[2] Stephen Baker,[1] Alexandra Line,[3] Sonia Chohan,[3] Maria Fookes,[3] Andrew Barron,[3] Peadar Ó Gaora,[1] José A. Chabalgoity,[4] Niren Thanky,[1] Christoph Scholes,[1] Nicholas Thomson,[3] Michael Quail,[3] Julian Parkhill,[3] and Gordon Dougan[1]

Centre for Molecular Microbiology and Infection, Department of Biological Sciences,[1] and Department of Infectious Diseases and Microbiology,[2] Imperial College, London, and The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge,[3] United Kingdom, and Laboratory for Vaccine, Research, Department of Biotechnology, Instituto de Higiene, Facultad de Medicina, Montevideo, Uruguay[4]

Received 10 March 2003/Accepted 28 May 2003

Vi capsular polysaccharide production is encoded by the *viaB* locus, which has a limited distribution in *Salmonella enterica* serovars. In *S. enterica* serovar Typhi, *viaB* is encoded on a 134-kb pathogenicity island known as SPI-7 that is located between partially duplicated tRNA^{pheU} sites. Functional and bioinformatic analysis suggests that SPI-7 has a mosaic structure and may have evolved as a consequence of several independent insertion events. Analysis of *viaB*-associated DNA in Vi-positive *S. enterica* serovar Paratyphi C and *S. enterica* serovar Dublin isolates revealed the presence of similar SPI-7 islands. In *S. enterica* serovars Paratyphi C and Dublin, the SopE bacteriophage and a 15-kb fragment adjacent to the intact tRNA^{pheU} site were absent. In *S. enterica* serovar Paratyphi C only, a region encoding a type IV pilus involved in the adherence of *S. enterica* serovar Typhi to host cells was missing. The remainder of the SPI-7 islands investigated exhibited over 99% DNA sequence identity in the three serovars. Of 30 other *Salmonella* serovars examined, 24 contained no insertions at the equivalent tRNA^{pheU} site, 2 had a 3.7-kb insertion, and 4 showed sequence variation at the tRNA^{pheU}-*phoN* junction, which was not analyzed further. Sequence analysis of the SPI-7 region from *S. enterica* serovar Typhi strain CT18 revealed significant synteny with clusters of genes from a variety of saprophytic bacteria and phytobacteria, including *Pseudomonas aeruginosa* and *Xanthomonas axonopodis* pv. citri. This analysis suggested that SPI-7 may be a mobile element, such as a conjugative transposon or an integrated plasmid remnant.

Salmonella enterica subspecies I serovar Typhi is a host-adapted, human-restricted pathogen that causes typhoid fever (38). In addition to some isolates of *S. enterica* serovar Paratyphi C (11), *S. enterica* serovar Dublin, and *Citrobacter freundii* (35), most clinical isolates of *S. enterica* serovar Typhi express the Vi exopolysaccharide (23, 41). In contrast to the human host-restricted taxon *S. enterica* serovar Typhi, *S. enterica* serovar Paratyphi C is pathogenic for both humans and other animal species (30). *S. enterica* serovar Dublin Vi-positive isolates have been found mainly associated with cattle, while *C. freundii* is rarely associated with pathogenicity and the role of the Vi antigen in this species is equivocal. Vi expression is associated with a cluster of 10 genes located at position 4409519 on the *S. enterica* serovar Typhi chromosome, known as the *viaB* operon (29, 37), that comprises both Vi antigen biosynthetic genes (*tviB* to *tviE*) and export genes (*vexA* to *vexE*). Vi expression is under control of the *rcsB-rcsC* (2, 24) and *ompR-envZ* (39) two-component regulator systems that lie outside the pathogenicity island. The regulators in turn interact with the first gene of the *viaB* gene cluster, *tviA* (48), and regions upstream of the *tviA* promoter. The DNA sequences of *viaB* in *S. enterica* serovar Typhi strains CT18 (37) and Ty2 (21), as well as a Vi-positive isolate of *C. freundii* (accession number AF316551), have been determined. The *viaB* operon in *S. enterica* serovar Typhi resides on a 134-kb pathogenicity island that is located between the partially duplicated copies of the *pheU* tRNA gene, which has been designated *Salmonella* pathogenicity island 7 (SPI-7) (37).

There are multiple genes encoded within SPI-7 that are not directly associated with expression of the Vi capsule, but none of them are present in the genome of the completely sequenced strain *S. enterica* serovar Typhimurium strain LT2 (20, 31). However, little else is known about the conservation of SPI-7 in Vi-positive serovars or indeed about possible related or unrelated DNA insertions at the tRNA^{pheU} site in Vi-negative serovars. It is known that tRNA genes are common sites for insertion of horizontally acquired DNA in *S. enterica,* but little is known about how variable such tRNA-associated DNA insertions are in other *S. enterica* serovars. The abilities of different *S. enterica* serovars or isolates to cause distinct disease syndromes are likely to be reflected in their genetic makeups, and so variability, especially in horizontally acquired regions, may have relevance for pathogenicity and host specificity. We present here a detailed comparison of the SPI-7 regions from representative isolates of several Vi-positive *S. enterica* strains

* Corresponding author. Mailing address: Centre for Molecular Microbiology and Infection, Department of Biological Sciences, Imperial College of Science, Technology and Medicine, Armstrong Road, London, SW7 2AZ, United Kingdom. Phone: 44 207 594 3070. Fax: 44 207 594 3069. E-mail: d.pickard@imperial.ac.uk.

TABLE 1. Bacterial isolates

Tazon	Strain	Source
C. freundii	ID 7821 (Vi positive)	M. Popoff, Institut Pasteur, Paris, France
S. enterica serovar Agona	SARB 1	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Anatum	SARB 2	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Brandenburg	SARB 3	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Cholerasuis	SARB 4	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Decataur	SARB 8	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Derby	SARB 9	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Dublin	SARB 12	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Dublin	ID 1662K (Vi positive)	M. Popoff, Institut Pasteur, Paris, France
S. enterica serovar Duisburg	SARB 15	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Emek	SARB 20	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Enteritidis	SARB 16	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar S. Gallinarum	SARB 21	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Haifa	SARB 22	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Heidelberg	SARB 23	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Indiana	SARB 25	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Infantis	SARB 26	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Miami	SARB 28	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Montevideo	SARB 30	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Muenchen	SARB 32	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Newport	SARB 36	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Panama	SARB 39	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Paratyphi A	SARB 42	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Paratyphi C	SARB 48	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Paratyphi C	ID 32K (Vi positive)	M. Popoff, Institut Pasteur, Paris, France
S. enterica serovar Pullorum	SARB 51	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Reading	SARB 53	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Rubislaw	SARB 54	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Saintpaul	SARB 56	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Senftenberg	SARB 59	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Stanley	SARB 60	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Stanleyville	SARB 61	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Typhi	SARB 63	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Typhi	SARB 64	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Typhi	TY2	Imperial College Collection, London, United Kingdom
S. enterica serovar Typhi	CT18	Imperial College Collection, London, United Kingdom
S. enterica serovar Typhimurium	SARC1	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Typhimurium	SL1344	Imperial College Collection, London, United Kingdom
S. enterica serovar Typhimurium	LT2	Imperial College Collection, London, United Kingdom
S. enterica serovar Typhisuis	SARB 70	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Wien	SARB 71	Salmonella Genetic Stock Centre, Calgary, Canada
S. enterica serovar Wien	SARB 72	Salmonella Genetic Stock Centre, Calgary, Canada

belonging to different serovars. We also describe the distribution of DNA insertions in the tRNA^{pheU} region of the *S. enterica* chromosome from selected isolates from *Salmonella* Reference Collection B (SARB) (8) and a bioinformatic analysis which indicated that specific genomic regions from a variety of soil and plant bacteria, including a recently sequenced gene island in *Pseudomonas aeruginosa,* share a number of gene clusters with SPI-7.

MATERIALS AND METHODS

Bacterial strains and culture. The *S. enterica* strains used in this study are listed in Table 1. The majority of the *S. enterica* strains were from SARB (8). Bacteria were routinely cultured in Luria-Bertani broth or on Luria-Bertani agar. *S. enterica* serotypes were determined by using typing sera from Murex (Dartford, United Kingdom), and biochemical traits were analyzed by using Analytical Profile Index (API 20E) strips obtained from BioMerieux (Marcy-l'Etoile, France).

Oligonucleotides and PCR conditions. Table 2 shows the PCR oligonucleotides designed to investigate the structure of SPI-7 from various serovars of *S. enterica*. PCR was performed by using an Expand high-fidelity PCR kit (Roche, St. Albans, United Kingdom) for fragments smaller than 10 kb and DNA polymerase XL PCR kits (Perkin-Elmer, Branchburg, N.J.) for PCR fragments larger than 10 kb. The locations of the oligonucleotides in SPI-7 of *S. enterica* serovar Typhi strain CT18 are shown in Table 2. Where possible, the PCR primers used to amplify the PCR fragments were chosen to give some degree of overlap between the fragments generated. This allowed production of contiguous segments covering much of the SPI-7 region for the *Salmonella* serovars examined. Oligonucleotides pairs were designed for PCR by using the MacVector 7.1 program (Accelrys Ltd., Cambridge, United Kingdom). These oligonucleotides were based on DNA sequences within genes when possible. Genome sequences from other *S. enterica* serovars were obtained from Washington University, St. Louis, Mo. (http://genome.wustl.edu/projects/bacterial/).

Sequencing, DNA alignment, BLASTP analysis, and annotation. Genomic DNA of *S. enterica* serovar Paratyphi C and *S. enterica* serovar Dublin for use in PCRs were prepared by using the method of Hull et al. (26). DNA from the various *Salmonella* serovars examined were routinely prepared by using the cetyltrimethylammonium bromide method for chromosomal DNA (3). PCR-generated DNA fragments obtained by using oligonucleotides based on SPI-7 sequences from different DNA templates were randomly fragmented by sonication, and DNA that was approximately 1.5 kb long was gel purified and cloned into pUC18. The cloned DNA was sequenced in both directions by using pUC18-based primers. The coverage was at least threefold for all sequences. The sequences were assembled by using GAP4 (6) to produce contiguous sequences for each PCR product. These PCR-generated fragments were eventually aligned to cover most of the SPI-7 regions by using MacVector 7.1 alignment tools (Accelrys Ltd.). Analysis of SPI-7 sequences was performed with ARTEMIS (43) and

TABLE 2. Oligonucleotides used to obtain and analyze the regions of SPI-7 from *S. enterica* serovars Dublin and Paratyphi C by using PCR fragments

Target DNA covered by PCR oligonucleotide pairs	Primer[a]	Primer sequence (5′-3′)
phoN-cut3A	SB001 (F)	GCATATCACTCTGCTTTGCCCGAACTC
	SB002 (R)	GCCGATACTACGCTGGAAATGGTGGC
tRNA^{pheU} region-*cut3A*	DE101 (F)	CGCAAAGTGTGCTGGGTCATTCAAAC
	DE102 (R)	AGGCAGATGGAACTGGTGCTGGAAGG
tviA-tRNA^{pheU}	DE032 (F)	CCAATTCTTTTCTCCAGCGATACATAGTC
	DE033 (R)	GCTCAGTCGGTAGAGCAGGGGATT
samB-vexE	DE076 (F)	GCCCTCAAACTGCCGCCATTCCTTAGAC
	DE079 (R)	GCGACTGTTGCTGACTTAGCCGAG
traC-samA	CS100 (F)	CCACGCCTCTTTATTGCCGAAGCC
	CS101 (R)	CCATAGCATCCATTCCTGAAGCAC
traG-traC	DE063 (F)	GAGACCTATGCCGAGAAGTTACTGAC
	DE064 (R)	CTCAAGCTCACTGCACTGATACTCTC
traG region	DE105 (F)	TCAGTGCCCGTTATAATGCAGTCG
	DE106 (R)	CCGCAGAAGCAACTTCAGAATCTG
pilNa-traG	DE065 (F)	CAGATGAACCGCCAGGTCAGGCTC
	DE066 (R)	AGCGTTATTTCCTTCTCCCCCAAG
phoN-pilNa	CS104 (F)	TTGCCGAAGCAGTAGCGTAGTAGC
	CS105 (R)	GCTGAATCCAGAATAGATACCGACG
rci-pilR (primers designed to confirm *S. enterica* serovar Paratyphi C deletion in this region)	CS106 (F)	GATGTGAACGGAAAAAAACGGGACGCAC
	CS107 (R)	ATCTCTGATGGTCAGTGGGGACGGGGC
cut3A/tRNA^{pheU}-*int2* (primers used to analyze *S. enterica* serovars Typhi and Dublin/Paratyphi C in this region)	CS108 (F)	TTTTATCCCGACCCAACCCATCCTATTC
	CS109 (R)	ATCAGTCCTCATCGTTTCAGACACACGC
viaB operon region 1-*yjhP* (STY4650)-*tviD*	DE028 (F)	GCCACACTATTTTCGCCCCTGCCAGGA
	DE029 (R)	GCTCATCCTGAGAACCCGCCAGACTG
viaB operon region 2-*tviD-helD* (STY4664)	DE026 (F)	GCCATGAGTCTGAAGCCAGGAGGAATT
	DE027 (R)	GCTGGAACCGTCATTCTATCCCGTAGT
samA-samB (primer pair designed to confirm the presence or absence of the SopE phage)	DE078 (F)	TCCCATAGCATCCATTCCTGAAGCACT
	DE079 (R)	GCCTCAAACTGCCGCCATTCCTTAGA

[a] F, forward; R, reverse.

the BLASTP database search program (1). For some of the work described here we used preliminary sequence data obtained from the DOE Joint Genome Institute (http://www.jgi.doe.gov/JGI_microbial/html/index.html).

Nucleotide sequence accession number. The nucleotide sequence of the *C. freundii recJ-tviA* region has been deposited in the GenBank database under accession number AY282413.

RESULTS

Characterization of the SPI-7 region of *S. enterica* serovar Typhi. Vi polysaccharide expression in *S. enterica* serovar Typhi is associated with the *viaB* operon encoded on a pathogenicity island known as SPI-7 (37). SPI-7 is a 134-kb DNA insertion in duplicated tRNA^{pheU} sequences between positions 4409574 and 4543073 on the *S. enterica* serovar Typhi strain CT18 chromosome (20). The intact tRNA^{pheU} sequence (positions 4543073 to 4543148) is 75 bp long, while the truncated tRNA^{pheU} sequence (positions 4409519 to 4409574) lacks the first 20 nucleotides. The *viaB* operons from two Vi-positive *S. enterica* strains, *S. enterica* serovar Typhi strain CT18 (37) and *S. enterica* serovar Typhi strain Ty2 (21), and a Vi-positive isolate of *C. freundii* (accession number AF316551) have been sequenced previously.

The *phoN* gene, STY4519, defines the left boundary of SPI-7 and is adjacent to the truncated tRNA^{pheU} and *oriT* remnant. The 12-kb DNA segment upstream of this truncated tRNA^{pheU} (including Δ*oriT* and the region between STY4503 and STY4519) probably represents a region horizontally transferred earlier, as originally proposed by Groisman et al. (19), since this small gene cluster is present in all *Salmonella* subspecies 1 strains tested by Porwollik et al. in microarrays, while it is absent from a variety of enteric bacteria examined, including *Escherichia coli* K-12 and O157:H7 (40).

The gene content of SPI-7 can be divided into distinct regions. Lying between the truncated tRNA^{pheU} and STY4536 (*ssb*) (Fig. 1A) are a number of genes that are likely to encode functions related to conjugal transfer of DNA, DNA replication, or transposition. For example, STY4536 (*ssb*) encodes a putative single-stranded DNA binding protein, and STY4533 (*topB*) encodes a putative topoisomerase B. From STY4537 to STY4553 (*pilK*) (Fig. 1B) are several previously characterized genes encoding a type IVB pilus system with a role in attachment to eukaryotic cells (50). The type IVB pilus system may originally have served as a mating pair formation cluster for a conjugative plasmid or conjugative transposon and is similar to the related gene cluster of the R64 plasmid (49).

The region from STY4553 (*pilK*) to STY4598 (*samB*) (Fig. 1C) is predicted to encode at least 24 hypothetical proteins, as well as a number of genes with significant matches as determined by BLASTP analysis with DNA transfer genes, such as STY4554 (*traE*), STY4562 (*traG*), and STY4573 (*traC*). Finally, STY4592 (*ardC*) exhibits similarity to a protein with a putative role in the conjugative process itself as a DNA escorting protein in the *incW* plasmid, pSa (4). A gene encoding an ArdC homologue has also been found in the symbiosis island

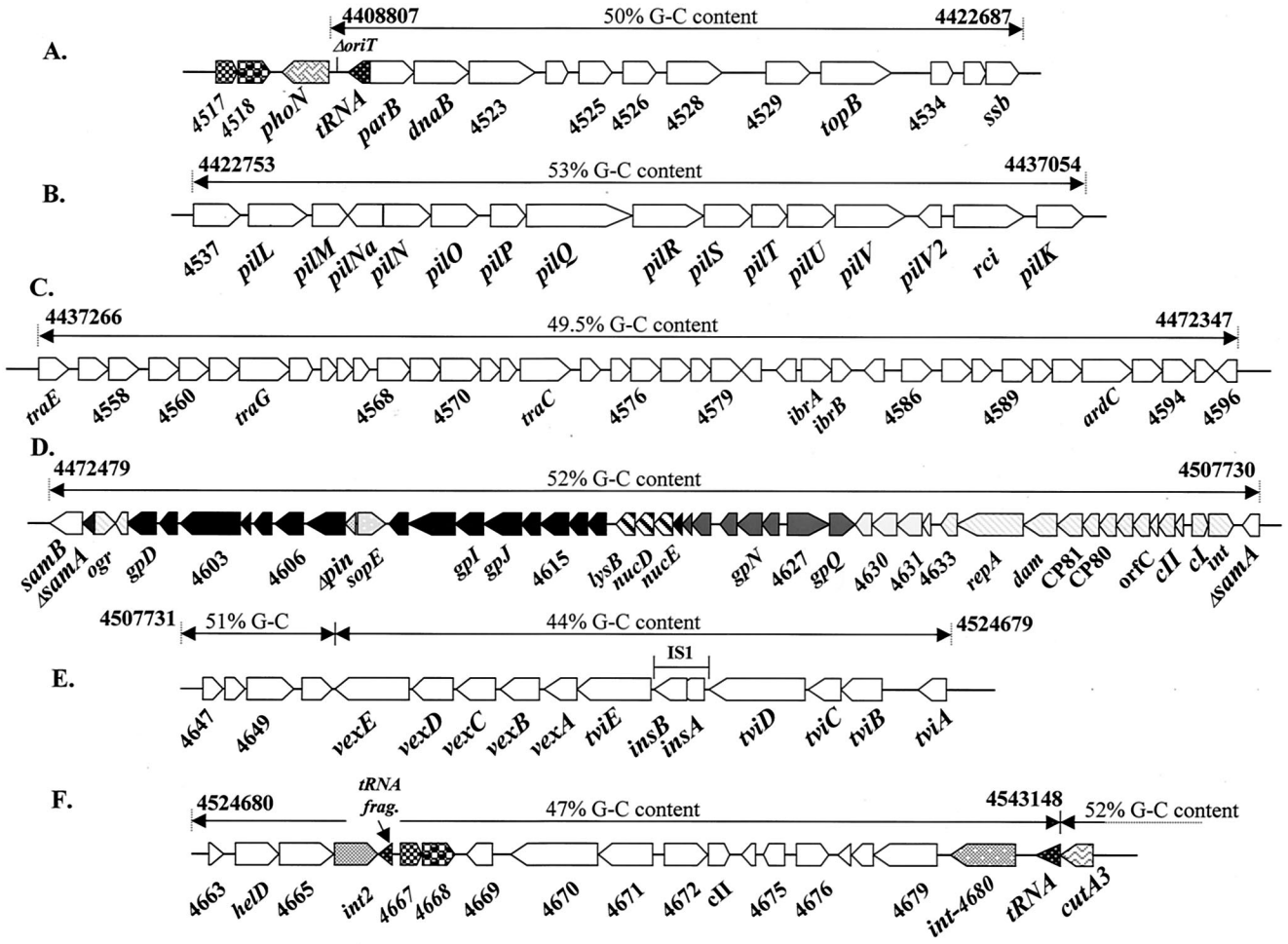

FIG. 1. Detailed annotation of six regions of SPI-7 from *S. enterica* serovar Typhi strain CT18. (A) *phoN-ssb* region, including the truncated tRNA and the adjacent *phoN* gene along with STY4517 and STY4518, which show homology to genes present in SPI-7. (B) *pil* locus, which encodes a type IVB pilus system and extends from the *pilL* gene to the *pilK* gene. (C) STY4553 to STY4596. This region harbors a number of genes that show identity to related genes involved in DNA transfer. (D) SopE bacteriophage, a P2-type bacteriophage located within the *samA* gene. The CDSs of this phage are shaded to highlight the region. The genes represented by black arrows represent the tail region; the genes represented by arrows with black stripes encode the lysis proteins of the phage; the genes represented by dark grey arrows encode capsid or head functions; the genes represented by light grey arrows (STY4629 to STY4632) are genes of foreign origin and have a lower G+C content than the surrounding genes; the genes represented by arrows with light grey stripes encode the regulatory functions of the phage. The *cos* site is located between STY4532 and STY4533 at coordinates 4499001 to 4499019 and has the sequence GGCGTGGCGGGGATACGAG. (E) *viaB* operon encoding the genes required for production and export of the Vi polysaccharide. Note the presence of the IS*1* element. (F) STY4663-*cutA3* region encoding putative integrases (shaded), a 16-bp duplication of a section of $tRNA^{pheU}$, and the $tRNA^{pheU}$-*cut3A* SPI-7 junction. Note that the STY4667 and STY4668 genes are highly homologous to the STY4517 and STY4518 genes, respectively, located near *phoN* and are highlighted to emphasize this. The numbers associated with genes are the STY numbers used in the Sanger Institute annotation of the *S. enterica* serovar Typhi strain CT18 genome (accession number AL62783). The numbers above the regions indicate locations (in base pairs) on the *S. enterica* serovar Typhi strain CT18 chromosome, as indicated by the Sanger Institute annotation.

of *Mesorhizobium. loti* (46). This region, along with the *pil* locus and the region located between STY4539 (*pilL*) and the truncated $tRNA^{pheU}$, hint that there is an integrated functional role for the entire locus in conjugation and DNA transfer. This region in SPI-7 is possibly derived from a conjugative plasmid with similarity to R64.

A bacteriophage encoding the SPI-1 effector protein, SopE (Fig. 1D), is inserted in the *samA* gene (32). Analysis of the SopE phage DNA sequence showed significant DNA sequence and coding sequence (CDS) similarity to the available sequence of phage Retron Ec67. That Retron Ec67 phage belongs to the P2/186 family of bacteriophages (13) and was originally isolated from *E. coli* strain CL-1 (25). Detailed analysis of the SopE phage sequence has revealed that many of the promoter sites regulating lysogenic and lytic functions are still present (14), which indicates that the phage may still be capable of excision from SPI-7. Additionally, the attachment site for the SopE phage has been determined to be duplications on either side of the phage insertion (*attR* at coordinates 4507385 to 4507393 and *attL* at coordinates 4473831 to 4473839) (34). A *lexA* binding site, characteristic of phage 186, has also been found in the SopE phage at coordinates 4499689 to 4499708.

This site plays a role in 186 prophage induction (27). Four CDSs with low G+C contents are located adjacent to the *cos* site (cohesive ends) of the SopE phage, a site where foreign DNA is commonly inserted into phages of the P2 family (33). After the *viaB* operon (Fig. 1E) and between STY4663 and STY4680 are many genes whose functions are unknown. The genes encoding two candidate integrases, STY4666 and STY4680, were identified within this region of SPI-7 (Fig. 1F). The integrase encoded adjacent to the 3′ $tRNA^{pheU}$, STY4680 (*int*), shows similarity to integrases encoded by the P4 family of bacteriophages (54% identity over 416 amino acids to a P4-like integrase identified in *E. coli* CFT073, for example). The gene encoding the other candidate integrase, STY4666 (*int2*), which is nearer the *viaB* locus, is more similar to the gene encoding an integrase found in *Actinobacillus actinomycetemcomitans*. A 16-bp duplication of the last bases of $tRNA^{pheU}$, alongside the STY4666 gene, is indicative of further recombination events in *S. enterica* serovar Typhi strain CT18.

Two genes adjacent to STY4666 (*int2*), STY4667 and STY4668, are also worthy of note since they are highly similar to genes found just outside the truncated 5′ $tRNA^{pheU}$ (Fig. 1A and Table 3), STY4517 and STY4518, respectively. Additionally, in *Shigella flexneri* plasmid pWR100 (9), genes with identity to STY4667 and STY4668 are present as similar duplications on either side of a known *Shigella* pathogenicity region (between the gene duplications of *orf46*/*47* and *orf85a*/*85b* in pWR100) (16, 47). Genes similar to STY4667 and STY4668 have also been found to be associated with other well-characterized horizontally transferred DNA, including SPI-1 (36).

Comparison of the *viaB*-associated DNA found in Vi-positive *S. enterica* serovar Dublin and *S. enterica* serovar Paratyphi C and identification of the *viaB* insertion site in the *C. freundii* genome. Relatively few *S. enterica* serovars or isolates express Vi polysaccharide, and for those that do it is not known whether the Vi locus is also harbored on an SPI-7-related element. To address this question, the regions flanking the *viaB* operon were obtained by PCR from *S. enterica* serovar Paratyphi C Vi-positive strain 32K and *S. enterica* serovar Dublin Vi-positive strain 1622K by using PCR primers based on the SPI-7 region of *S. enterica* serovar Typhi (Table 2). PCR fragments generated from both *S. enterica* serovar Paratyphi C and *S. enterica* serovar Dublin by using these primers were subjected to DNA sequencing. Both *S. enterica* serovar Paratyphi C strain 32K and *S. enterica* serovar Dublin strain 1622K were found to encode the *viaB* operon associated with significant stretches of DNA showing similarity to SPI-7 (Fig. 2). Additionally, the SPI-7 regions of both *S. enterica* serovar Dublin and *S. enterica* serovar Paratyphi C, like that in *S. enterica* serovar Typhi, were inserted in duplicated $tRNA^{pheU}$ sites and were over 99.5% identical to the equivalent regions of *S. enterica* serovar Typhi strain CT18. *S. enterica* serovar Typhi strain CT18 contains an additional 15 kb of DNA located in the region bordered by the STY4678 integrase gene (*int*) near $tRNA^{pheU}$ and the gene encoding the other integrase STY4666 (*int2*) compared to the sequences of both *S. enterica* serovar Dublin and *S. enterica* serovar Paratyphi C. Consequently, the corresponding PCR product generated from *S. enterica* serovar Paratyphi C or *S. enterica* serovar Dublin was smaller (5.8 kb) than the 14-kb PCR product amplified from *S. enterica* serovar Typhi (Fig. 1). PCR analysis also demonstrated that SPI-7 of *S. enterica* serovar Paratyphi C and SPI-7 of *S. enterica* serovar Dublin lack the SopE phage that disrupts the *samAB* locus of SPI-7 of *S. enterica* serovar Typhi. The DNA sequence encoding the type IVB pilus cluster is intact in *S. enterica* serovar Dublin, but a deletion is present in *S. enterica* serovar Paratyphi C strain 32K between the STY4552 *rci* and STY4547 *pilS* genes (Fig. 2). Loss of these genes would result in possibly deleterious effects on pilus assembly. The remaining DNA sequences showing similarity to the *S. enterica* serovar Typhi SPI-7 sequence are virtually identical (over 99%) to each other in the three serovars.

In *C. freundii*, the *viaB* operon is inserted at a different tRNA locus, $tRNA^{glyU}$, based on our analysis of the flanking DNA (GenBank accession number AY282413). The gene order 4 kb upstream of this $tRNA^{glyU}$ site resembles that in a number of previously sequenced enteric bacteria, including *S. enterica* serovar Typhi strain CT18. The presence of the *C. freundii viaB* gene adjacent to $tRNA^{glyU}$ is of interest as this is the site where the *P. aeruginosa* SG17M genetic island is inserted, in contrast to the sequenced strain *P. aeruginosa* PAO1 (28); additionally, significant parts of this island show remarkable synteny with gene clusters in SPI-7, as described in detail below and shown in Fig. 3.

Analysis of SPI-7 indicates the presence of regions of synteny with bacteria of soil and plant origin, including *Xanthomonas axonopodis* pv. citri, *Burkholderia fungorum*, and *P. aeruginosa*. In the original annotation, a number of regions of *S. enterica* serovar Typhi strain CT18 SPI-7 yielded few matches in database searches (37). These regions were reexamined by using the ARTEMIS program to discover if any new data would shed light on their origin and function. The presence of duplicated $tRNA^{pheU}$ sites on either side of SPI-7 hinted that this pathogenicity island may have originally been a mobile element similar to 100-kb conjugative transposon CTnscr94, identified in *S. enterica* serovar Senftenberg strain 5494-57, which may also be located within $tRNA^{pheU}$ (22). From database searches, it became evident that a large number of genes from both saprophytic bacteria and phytobacteria showed homology and were syntenic with specific regions in SPI-7 (Fig. 3 and Table 3).

A large number of these matches were to genes of unknown function in the region between the STY4597 (*samB*) and STY4552 (*rci*) genes of SPI-7. The close synteny extended to the region lying between STY4539 (*pilL*) and STY4521, the first gene downstream of the truncated $tRNA^{pheU}$-*phoN* junction (Fig. 1A). The soil bacteria and phytobacteria which have DNA sequences which exhibit synteny with regions of SPI-7 include the plant-adapted organism *X. axonopodis* pv. citri (12) and, to a lesser extent, *Xylella fastidosa* (15, 45), as well as *Pseudomonas fluorescens*, *B. fungorum*, and *Ralstonia metallidurans* (also called *Ralstonia eutrophus*). Significantly, gene islands recently identified in two isolates of *P. aeruginosa* (28) also showed matching synteny with these same regions in SPI-7. These isolates were obtained from a cystic fibrosis patient (*P. aeruginosa* strain C) and an aquatic environment (*P. aeruginosa* strain SG17M) (42). Similar synteny with SPI-7 was also observed with *Haemophilus somnus* strain 129PT (a bacterium found in the respiratory or genitourinary tract of cattle).

Figure 3 shows the extensive synteny of these regions from

TABLE 3. Selected genes showing synteny between SPI-7 and *X. axonopodis* pv. citri (accession number AE011859) or *R. metallidurans* (accession number NZ_AAA101000352)

Gene(s)	Location (coordinates)	Product size (amino acids)	Function	% Amino acid identity(ies) (range)[a]
tRNA^{pheU} (truncated)	4409519-4409574			
STY4521	4409652-4410686	344	*parB*-like partition protein	41 (286 aa to XAC2205), 39 (289 aa to Reut4191)
STY4523	4412040-4413839	599	Hypothetical	31 (580 aa to XAC2206), 30 (600 aa to Reut4193)
STY4526	4415107-4415664	185	Hypothetical	32 (173 aa to Reut4194), 28 (165 aa to XAC2207)
STY4528	4415909-4417252	447	Hypothetical	32 (405 aa to Reut4195), 29 (399 aa to XAC2208)
STY4529	4417838-4418617	259	Hypothetical	30 (163 aa to XAC2209), 29 (227 aa to Reut4196)
STY4530	4418728-4420722	664	*topB* topoisomerase B	50 (675 aa to XAC2212), 49 (677 aa to Reut4199)
STY4534	4421390-4421839	149	Hypothetical	38 (136 aa to Reut4221), 33 (128 aa to XAC2236)
STY4535	4421920-4422138	72	Hypothetical	56 (65 aa to XAC2217)
STY4536	4422151-4422687	178	*ssb* single-stranded DNA-binding protein	60 (174 aa to pWWO Ssb), 68 (162 aa to P1 phage Ssb-p1)
STY4539	4423868-4425112	414	*pilL* membrane-located open reading frame	46 (126 aa to XAC2253), 38 (165 aa to Reut4232)
STY4554	4437266-4438141	291	*traE* (putative)	23 (219 aa to R64 plasmid open reading frame TraE)
STY4557	4438648-4439562	304	Hypothetical membrane protein	25 (170 aa to Reut4233)
STY4558	4439581-4440321	246	Hypothetical	40 (227 aa to XAC2255), 37 (237 aa to Reut4234)
STY4559	4440435-4440986	183	Hypothetical	44 (141 aa to XAC2256), 42 (148 aa to Reut4235)
STY4560	4441022-4441522	166	Hypothetical	41 (160 aa to XAC2257), 39 (153 aa to Reut4236)
STY4561	4441532-4442101	189	Hypothetical	51 (213 aa to NP_511181 [R46 plasmid])
STY4562	4442121-4444211	696	*traG*-like (DNA transfer region-associated gene)	65 (541 aa to XAC2259), 61 (709 aa to Reut4237), 26 (630 aa to TraG-like open reading frame in R27 plasmid)
STY4563	4444208-4444966	252	Hypothetical	42 (234 aa to XAC2260), 42 (228 aa to Reut4238)
STY4564	4445185-4445457	90	Hypothetical	42 (82 aa to Reut4247), 39 (81 aa to XAC2269)
STY4565	4445457-4445696	79	Hypothetical membrane-spanning domains	29 (77 aa to Reut4248)
STY4566	4445726-4446088	120	Hypothetical membrane-spanning domains	37 (107 aa to XAC2270), 35 (89 aa to Reut4249)
STY4567	4446098-4446472	124	Hypothetical	30 (90 aa to XAC2271)
STY4568	4446469-4447122	217	Hypothetical exported protein	54 (219 aa to Reut4250)
STY4569	4447122-4448021	299	Hypothetical	50 (255 aa to Reut4252), 44 (292 aa to XAC2272)
STY4570	4448011-4449489	492	Hypothetical exported protein	38 (513 aa to XAC2273)
STY4571	4449482-4449925	147	Hypothetical lipoprotein	47 (127 aa to Reut4254)
STY4572	4449922-4450341	139	Hypothetical	50 (70 aa to XAC2274), 45 (108 aa to Reut4255 [all N-terminal end only])
STY4573	4450338-4452761	807	*traC*-like (putative Dtr protein)	52 (817 aa to XAC2274), 51 (815 aa to Reut4255), 20 (354 aa to Virb4 Dtr protein from *Helicobacter pylori*)
STY4575	4453457-4453849	392	Hypothetical	36 (113 aa to XAC2282), 36 (125 aa to Reut4258)
STY4576	4453846-4454814	323	Hypothetical	49 (320 aa to XAC2283), 48 (320 aa to Reut4259)
STY4577	4454829-4456259	476	Hypothetical	41 (402 aa to XAC2284), 40 (439 aa to Reut4260)
STY4579	4456592-4458097	501	Hypothetical membrane protein	45 (499 aa to XAC2286), 42 (510 aa to Reut4262)
STY4583	4459393-4460640	415	*ibrA* regulatory protein (previously *ybdN*)	85 (40 aa to IbrA from EcoR-9), 84 (415 aa to IbrA [Z1203 and Z1643] from O157:H7 EDL933)
STY4584	4460652-4461266	204	*IbrB* regulatory protein (previously designated *ybdM*)	78 (203 aa to IbrB from EcoR-9), 78 (204 aa to IbrB [Z1204 and Z1644] from O157:H7 EDL933)
STY4585	4461477-4462019	180	Hypothetical	74 (178 aa to STY4858), 74 (178 aa to STM4502)
STY4592	4467675-4469624	649	*ardC* (DNA escorting protein anti-restriction)	34 (282 aa to ArdC from *Mesorhizobium loti* symbiosis island)
STY4593	4469696-4470604	302	Hypothetical	33 (279 aa to STY4594)
STY4594	4470678-4471577	299	Hypothetical	33 (279 aa to STY4593)
STY4595	4471619-4471978	119	Hypothetical	41 (105 aa to XAC2221), 36 (74 aa to Reut4219), 38 (76 aa to Orf23 of plasmid pCTX-M3)
STY4599-STY4645	4473830-4507389		SopE phage (S. *enterica* serovar typhi only)	
STY4630	4496774-4497838	354	Hypothetical within A-T-rich region of SopE phage	35 (363 aa to S060 of SXT element from *Vibrio cholerae*)
STY4631	4497835-4498899	354	Hypothetical within A-T-rich region of SopE phage	40 (356 aa to S061 of SXT element from *Vibrio cholerae*)
STY4649	4508645-4509535	296	Putative 3-methyltransferase	75 (288 aa to STY4856)
STY4651-STY4662	4510583-4524679		ViaB operon	
STY4664	4525533-4527044	503	*helD* DNA helicase	39 (472 aa to Orf193 of plasmid Rts1)
STY4665	4527028-4528617	529	Hypothetical	29 (451 aa to XAC2196), 28 (449 aa to Reut2852)
STY4666	4528781-4529794	337	*int2* integrase	35 (265 aa to integrase from *Actinobacillus actinomycetemcomitans*)
Truncated tRNA^{pheU}	4529973-4529988			
STY4667	4530221-4530514	97	Similar open reading frames often associated with transposon elements	98 (97 aa to STY4517), 98 (97 aa to STM4317), 26 (96 aa to Rv0918 of *Mycobacterium tuberculosis*)
STY4668	4530511-4530999	162	As for STY4667	92 (162 aa to STY4518), 92 (162 aa to STM4318), 50 (156 aa to Rv0919)
STY4673	4537312-4537533	73	*ner*-like Mu DNA binding	66 (68 aa to Ner open reading frame of Mu phage)
STY4680	4541654-4542913 (complement)	419	*int* integrase	54 (416 aa to P4-like prophage integrase from *Escherichia coli* CFT073)
tRNA^{pheU} (full length, 75 bp)	4543148-4543073			

[a] aa, amino acids. The designations of the CDSs of *X. axonopodis* pv. citri begin with XAC, and the designations of the CDSs of *R. metallidurans* begin with Reut.

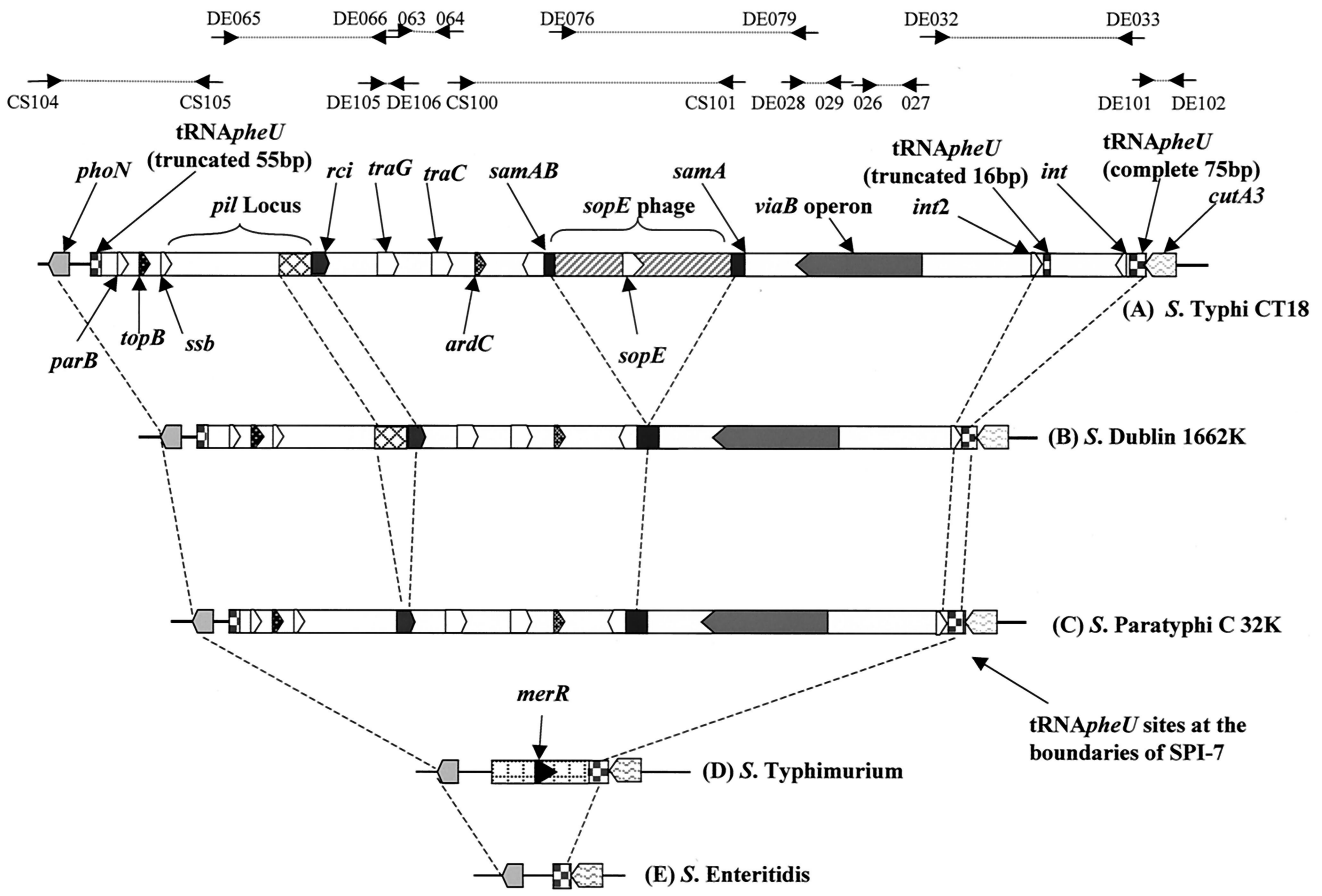

FIG. 2. SPI-7 of *S. enterica* serovar Typhi and the related SPI-7 regions of other *S. enterica* serovars, showing the sites of the PCR primer pairs used in relation to the *S. enterica* serovar Typhi strain CT18 sequence and the resulting products obtained, as indicated by dashed lines. (A) SPI-7 region from *S. enterica* serovar Typhi with the areas of interest labeled, including the gene clusters encoding the *viaB* exopolysaccharide, the SopE bacteriophage, and the type IVB pilus locus. (B) SPI-7 region in *S. enterica* serovar Dublin strain 1622K, which is notably lacking the region between *int1* and *int2* and also the SopE bacteriophage. (C) SPI-7 in *S. enterica* serovar Paratyphi C isolate 32K. (D) tRNA^{pheU} region in *S. enterica* serovar Typhimurium strain LT2 harboring a gene resembling *merR* at the *phoN* (upstream) end of this tRNA. (E) Intact tRNA^{pheU} as found, for example, in *S. enterica* serovar Enteritidis, with no additional DNA inserted between the *cutA3* and *phoN* genes.

X. axonopodis pv. citri and *P. aeruginosa* SG17M with *S. enterica* serovar Typhi SPI-7, while Table 3 presents the synteny analysis data. For *X. fastidosa*, the synteny observed with SPI-7 matched that described by Larbig et al. for the *P. aeruginosa* SG17M and C islands (28).

Distribution of DNA insertions in tRNA^{pheU} in other *S. enterica* serovars that do not express Vi exopolysaccharide. In order to ascertain the frequency of insertions occurring at the tRNA^{pheU} site in a range of Vi-negative *S. enterica* strains, DNA prepared from 30 different serovars were analyzed by using oligonucleotide primers designed to generate PCR products across the site. The isolates selected were from SARB (Table 1). Initially, PCR primers were designed by using DNA sequences in the *phoN* gene (primer SB001) and the *cut3A* gene (primer SB002) at the 5′ and 3′ ends of SPI-7, respectively. Two serovars, *S. enterica* serovar Typhimurium and *S. enterica* serovar Saintpaul, both yielded 3.7-kb PCR products. This result supports multilocus enzyme electrophoresis data which showed that *S. enterica* serovar Typhimurium and *S. enterica* serovar Saintpaul are very closely related (44). *S. enterica* serovar Typhimurium strain LT2 harbors a single complete tRNA^{pheU} with the addition of a gene related to *merR* adjacent to and downstream of the *phoN* gene. The PCR products described here were the predicted size when they were compared to the previously published *S. enterica* serovar Typhimurium strain LT2 sequence (31).

The following four Vi-negative *S. enterica* isolates yielded no detectable PCR product: SARB 1 (*S. enterica* serovar Agona), SARB 9 (*S. enterica* serovar Derby), SARB 59 (*S. enterica* serovar Senftenberg), and SARB 61 (*S. enterica* serovar Stanleyville). To detect any possible insertion at this site, PCR primers based on DNA sequences internal to SPI-7 were used to try to amplify each region of SPI-7 separately. The results were negative, as were the results obtained with Southern probes for specific loci within SPI-7. Further analysis of these four isolates with PCR primers specific for a sequence farther away from the tRNA eventually suggested that the initial negative PCR was due to sequence variation at the tRNA^{pheU}-*phoN* junction and thus to failure of the PCR primers to bind. This possibility is being investigated further in order to clarify

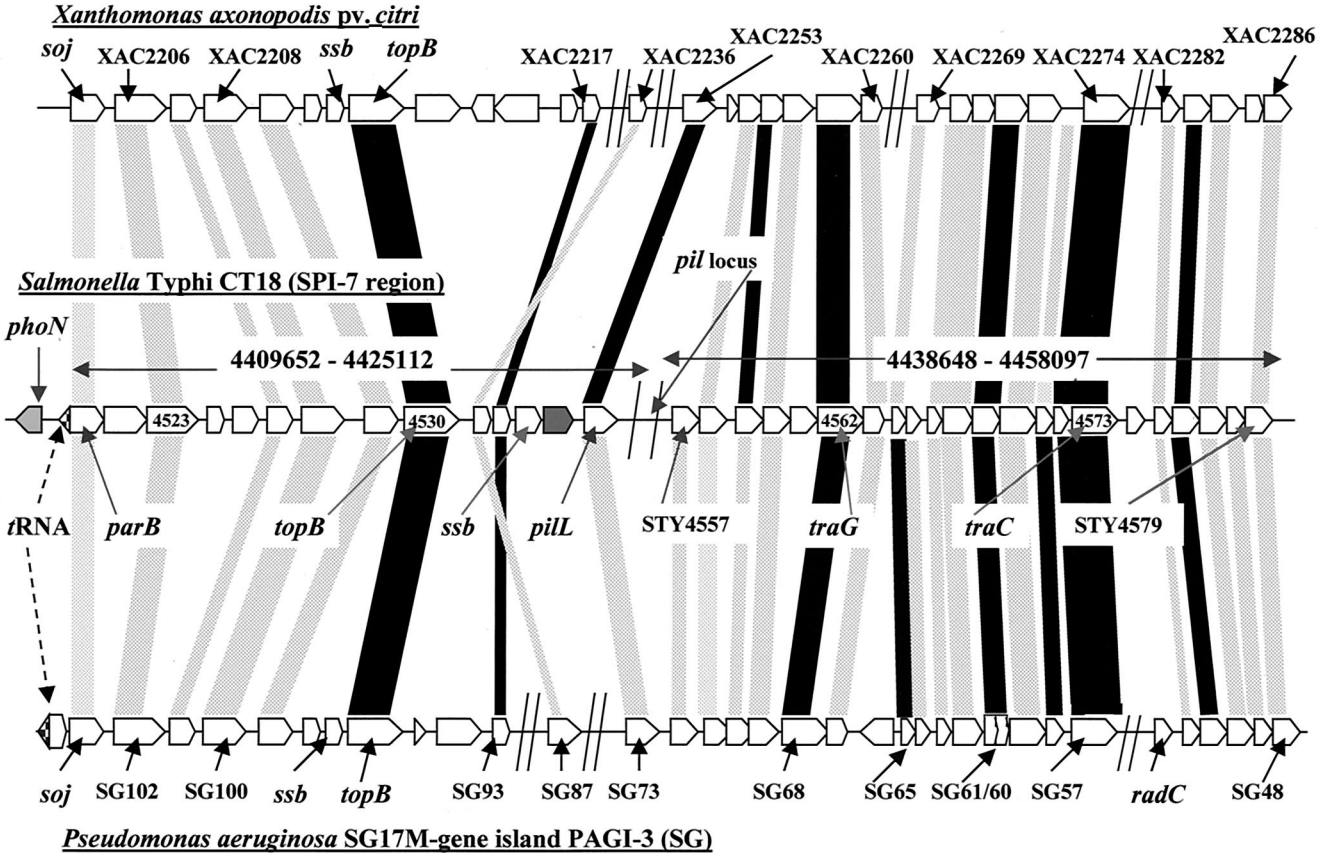

FIG. 3. Synteny between sections of SPI-7 and related regions in the genomes of other bacteria. The diagram was constructed by using The Sanger Institute programs Artemis and ACT in conjunction with TBLASTX protein-versus-protein analysis. The synteny among *S. enterica* serovar Typhi strain CT18 SPI-7, *X. axonopodis* pv. citri, and *P. aeruginosa* SG17M gene island PAGI-3 is illustrated. The accession numbers are AL62783, AE008923, and AF440524, respectively. The grey shading indicates CDSs exhibiting 25 to 44% amino acid identity, and the black boxes indicate CDSs with 45 to 70% amino acid identity. The gene designations for *traC*, *traG*, and *parB* are putative designations with respect to SPI-7 in *S. enterica* serovar Typhi. The cargo genes of the *P. aeruginosa* SG17M island are not shown, but the gene designations are SG01 to SG47. The term cargo gene is used in the context suggested by Larbig et al. (28) and refers to the variable region on the gene islands that include a range of metabolic enzyme gene clusters in *P. aeruginosa* and *viaB* in *S. enterica* serovar Typhi. The cargo regions occupy similar locations in both islands with respect to the putative Dtr regions and the integrase gene adjacent to one of the tRNA duplications and a *parB*-like gene at the other end.

these observations. Twenty-four serovars yielded PCR products that were approximately 2 kb long, indicating that there was no insertion at the $tRNA^{pheU}$ locus. Interestingly, one of these 24 serovars, represented by SARB 64, had previously been reported to be Vi positive (8). In our hands, however, serological analysis confirmed that the SARB 64 isolate described here was indeed Vi negative. Consequently, it is possible that the entire SPI-7 locus was precisely deleted during lab storage.

DISCUSSION

In this study we confirmed that the *viaB* operon in Vi-positive *S. enterica* serovars, including *S. enterica* serovar Typhi, *S. enterica* serovar Paratyphi C, and *S. enterica* serovar Dublin, is encoded on a common but mosaic pathogenicity island termed SPI-7. The *viaB* operon is of particular interest for several reasons. The ability to express Vi polysaccharide has an unusual and restricted distribution in *S. enterica*. In addition, purified Vi polysaccharide is currently used as a component antigen in human typhoid vaccines, and consequently the basis of acquisition and stability of the *viaB* operon is of immense practical interest. The *viaB* operon of *S. enterica* serovar Typhi strain CT18 is characterized by the presence of an intact IS*1* element (Fig. 2E). Analysis of this IS*1* element (see annotation data in accession number AL62783) revealed that it lies between the normally overlapping *tviD* and *tviE* genes (STY4659 and STY4656, respectively), leading to disruption of the Shine-Dalgarno ribosome binding site for *tviE* that is encoded within the last 13 bases of *tviD* (21). An 8-bp duplication is present on either side of the IS*1* inverted repeats, as expected. An IS*1* element has also been found to insert within a hot spot of the *viaB* region of *C. freundii* isolates and is responsible for the Vi-negative negative phenotype of such isolates (35), and it is possible that the Vi-negative phenotype which we observed for *S. enterica* serovar Typhi strain CT18 was also due to the presence of this IS*1* element. In *C. freundii* the Vi phenotype is reversible, hinting that the IS*1* element can be excised under some circumstances, and this may also be the case for *Salmonella* strains like *S. enterica* serovar Typhi strain CT18.

Sequencing of the *viaB*-associated DNA from Vi-positive *S. enterica* serovar Paratyphi C and *S. enterica* serovar Dublin strains showed that these *S. enterica* serovars harbor an SPI-7 related element with at least 99% similarity at the DNA and protein levels but lack the SopE phage. The SopE phage may have been acquired by *S. enterica* serovar Typhi after the basic SPI-7 element was acquired. There were only a small number of other deletions or insertions that were evident from this comparison. At this time we do not know if these deletions or insertions are unique to the Vi-positive *S. enterica* serovar Paratyphi C or *S. enterica* serovar Dublin strains which we sequenced. *S. enterica* serovar Typhi strain CT18 possesses a unique 15-kb region upstream of the *viaB* operon which is thought to represent a further integration event and may explain the presence of two integrase genes in *S. enterica* serovar Typhi SPI-7. The integrase gene nearest the intact $tRNA^{pheU}$ boundary of *S. enterica* serovar Typhi SPI-7 shows homology to the P4 phage family of integrase genes. The intact integrase gene (STY4666) that is approximately 15 kb into SPI-7 is present in all three serovars and is most similar to an integrase gene from *A. actinomycetemcomitans*. From these data we surmised that the original SPI-7 was most similar to that found in *S. enterica* serovar Dublin and that there was a common source of SPI-7 for all three serovars. The fact that the DNA sequences of the common regions of SPI-7 are so highly conserved suggests that SPI-7 acquisition by the three serovars was a relatively recent event. The very marked differences among the G+C contents of the regions of SPI-7 described here support the concept that the regions are mosaics, which were formed by serial acquisition of DNA. The G+C content of the *viaB* locus is 44% (Fig. 1E), and the G+C content of the region between the *viaB* locus and the intact $tRNA^{pheU}$ is 47% (Fig. 1F). These regions in particular may have been acquired in the pathogenicity island later.

Our detailed analysis of the overall gene complement of SPI-7 provides additional evidence that SPI-7 was originally obtained by horizontal transfer, perhaps in the form of a conjugative transposon. The type IVB pilus (50) could have originally constituted the mating pair formation system for a conjugative transposon or plasmid, but many expected genes required for DNA transfer were not identified previously. We now believe that these genes may be located between the previously uncharacterized region bordered by *samB* and *rci* of SPI-7. The regions on each side of the *pil* locus contain both a putative DNA transfer system and genes potentially directly associated with single-stranded DNA during the conjugative process, such as *ssb*,*ardC*, and *topB* homologues. Putative *traC* and *traG* genes have been identified in the region that lies between *samB* and *rci*, as well as the hypothetical genes mentioned above. It is of interest that in the transfer region of the *Bacteriodes* conjugative transposon, CTnDOT (7), putative functions could be assigned to only two genes, *traC* and *traG*. This finding may be related to particular important domains, such as ATPase activity and the site of the gene products within the conjugative machinery of the host bacteria (e.g., as coupling proteins linking Dtr and Mpf functions in the conjugative apparatus).

The presence of other genes, such as *samAB* and *parB*, supports the putative origin of SPI-7 as a conjugative transposon that evolved from a conjugative plasmid by loss of plasmid replication genes (i.e., *repA*) and other genes for plasmid maintenance. The presence of intact *samAB* genes is of interest since these genes encode DNA repair enzymes, which are frequently found on plasmids, such as pLT, the virulence plasmid of *S. enterica* serovar Typhimurium strain LT2 (31). Putative DNA repair genes have also been found in the conjugative integrating element R391 (5) and in the SG17M island (*radC*) (28) of *P. aeruginosa*.

We also obtained evidence that many of the SPI-7 gene products encoded on either side of the *pil* locus exhibit significant levels of sequence similarity to predicted proteins from a number of soil bacteria and phytobacteria, including *B. fungorum, R. metallidurans, P. fluorescens*, *P. putida* plasmid pWWO, and *P. aeruginosa*, as well as the plant pathogen *X. axonopodis* pv. citri and its close relative *X. fastidosa*. The apparent similarity extends past the gene sequences as analysis revealed that many of the genes showing homology to SPI-7 were in clusters, which also shared a high level of conservation of gene order and orientation with SPI-7 genes. The similarity of many SPI-7 genes to genes from soil bacteria and phytobacteria raises the possibility that SPI-7 and *viaB* may have originated from such sources. These bacteria produce a wide range of exopolysaccharides, some of which are encoded on plasmids (18). Enteric bacteria could encounter some of the bacteria containing genes that exhibit gene synteny with SPI-7 in the gut via contaminated food (food containing soil residue, for example). It is known, for example, that the soil bacterium *B. fungorum* is capable of surviving and growing in an acid environment (10) and thus may be able to survive in the acidic contents of the stomach. Whatever the source is, the potential transmissibility of these islands is emphasized by their insertions in tRNA sites such as $tRNA^{pheU}$ for SPI-7 and $tRNA^{glyU}$ for *P. aeruginosa* SG17M and in *viaB* for *C. freundii.* In these specific cases, an integrase is located near the tRNA site. This is also the case for the symbiosis island of *M. loti* that is located at a $tRNA^{phe}$ site (17, 46). It is known that phage integrases use tRNA genes as chromosomal attachment sites, and it is likely that these phage-derived integrases may perform the same function.

In summary, the synteny of SPI-7 with bacteria from soil and plants indicates that the SPI-7 regions may have a specific function related to DNA transfer, mobilization, and possible stabilization of the island and hence to the conservation seen. A common distant ancestor can also be postulated for these regions based upon the extensive synteny observed. This example of horizontal transfer in conjunction with recent studies with the ubiquitously adapted organism *P. aeruginosa* (28) underlies the diverse interactions possible between bacteria from environmental and enteric sources. The ongoing sequencing of saprophytic bacteria and phytobacteria will undoubtedly reveal more examples of such genetic exchange. Experiments are under way to examine the pilus locus of SPI-7 to assess if it can still act as a mating pore formation structure for any putative conjugative transposon. Further evidence that SPI-7 may be a functional element is provided by the observation that in *S. enterica* serovar Typhi strain SARB 64 SPI-7 is absent even though this isolate was originally described as a Vi-positive strain (8). The SARB 64 isolate which we have is Vi negative but is also definitely an *S. enterica* serovar Typhi isolate. We are currently investigating the genome of this organism in more detail using DNA microarrays.

The fact that Vi production is associated with a relatively unstable and potentially mobilizable element may have some consequences for the utility of Vi as an antigenic component of human typhoid vaccines. Vi production appears to have been a relatively recent acquisition by *S. enterica*. Interestingly, *S. enterica* serovar Typhi harbors mutations in three genes, *wcaA*, *wcaD*, and *wcaK*, associated with colonic acid biosynthesis, and Vi production may have replaced this function in *S. enterica* serovar Typhi. However, if sufficient immunological pressure is exerted by Vi vaccination, it is possible that Vi could be lost by *S. enterica* serovar Typhi or that Vi could be replaced by an alternative but immunologically distinct capsular locus which could replace the SPI-7-associated *viaB* locus. Although this may seem unlikely, it seems prudent to monitor the Vi status of primary *S. enterica* serovar Typhi isolates and the stability of the *viaB* locus during Vi vaccination trials.

ACKNOWLEDGMENTS

This work was supported by The Wellcome Trust and the BBSRC.

We thank Ken Sanderson, Salmonella Genetic Stock Centre, Calgary, Canada, and M. Y. Popoff, Pasteur Institute, Paris, France, for generous advice and for supplying the strains used in this study. Ian Dodd of the University of Adelaide gave us invaluable advice concerning all aspects of bacteriophage biology. We also thank the DOE Joint Genome Institute for sequence data for the saprophytic bacteria mentioned in this paper (http://www.jgi.doe.gov/JGI_microbial/html/index.html)

REFERENCES

1. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Arricau, N., D. Hermant, H. Waxin, C. Ecobichon, P. S. Duffey, and M. Y. Popoff.** 1998. The RcsB-RcsC regulatory system of *Salmonella typhi* differentially modulates the expression of invasion proteins, flagellin and Vi antigen in response to osmolarity. Mol. Microbiol. **29:**835–850.
3. **Ausubel, F. B. R., R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1987. Current protocols in molecular biology, p. 112–116. Wiley, New York, N.Y.
4. **Belogurov, A. A., E. P. Delver, O. V. Agafonova, N. G. Belogurova, L. Y. Lee, and C. I. Kado.** 2000. Antirestriction protein Ard (type C) encoded by IncW plasmid pSa has a high similarity to the "protein transport" domain of TraC1 primase of promiscuous plasmid RP4. J. Mol. Biol. **296:**969–977.
5. **Boltner, D., C. MacMahon, J. T. Pembroke, P. Strike, and A. M. Osborn.** 2002. R391: a conjugative integrating mosaic comprised of phage, plasmid and transposon elements. J. Bacteriol. **184:**5158–5169.
6. **Bonfield, J. K., K. F. Smith, and R. A. Staden.** 1995. A new DNA sequence assembly program. Nucleic Acids Res. **23:**4992–4999.
7. **Bonheyo, G., D. Graham, N. B. Shoemaker, and A. A. Salyers.** 2001. Transfer region of a bacteroides conjugative transposon, CTnDOT. Plasmid **45:**41–51.
8. **Boyd, F. E., F.-S. Wang, P, Beltran, S. A. Plock, K. Nelson, and R. K. Selander.** 1993. *Salmonella* reference collection B (SARB): strains of 37 serovars of subspecies I. J. Gen. Microbiol. **139:**1125–1132.
9. **Buchrieser, C., P. Glaser, C. Rusniok, H. Nedjari, H. D'Hauteville, F. Kunst, P. Sansonetti, and C. Parsot.** 2000. The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. Mol. Microbiol. **38:**760–771.
10. **Curtis, P., C. H. Nakatsu, and A. Konopka.** 2002. Aciduric proteobacteria isolated from pH 2.9 soil. Arch. Microbiol. **178:**65–70.
11. **Daniels, E. M., R. Schneerson, W. M. Egan, S. C. Szu, and J. B. Robbins.** 1989. Characterization of the *Salmonella paratyphi* C Vi polysaccharide. Infect. Immun. **57:**3159–3164.
12. **da Silva, A. C., J. A. Ferro, F. C. Reinach, C. S. Farah, L. R. Furlan, R. B. Quaggio, C. B. Monteiro-Vitorello, M. A. Van Sluys, N. F. Almeida, L. M. Alves, A. M. do Amaral, M. C. Bertolini, L. E. Camargo, G. Camarotte, F. Cannavan, J. Cardozo, F. Chambergo, L. P. Ciapina, R. M. Cicarelli, L. L. Coutinho, J. R. Cursino-Santos, H. El-Dorry, J. B. Faria, A. J. Ferreira, R. C. Ferreira, M. I. Ferro, E. F. Formighieri, M. C. Franco, C. C. Greggio, A. Gruber, A. M. Katsuyama, L. T. Kishi, R. P. Leite, E. G. Lemos, M. V. Lemos, E. C. Locali, M. A. Machado, A. M. Madeira, N. M. Martinez-Rossi, E. C. Martins, J. Meidanis, C. F. Menck, C. Y. Miyaki, D. H. Moon, L. M. Moreira, M. T. Novo, V. K. Okura, M. C. Oliveira, V. R. Oliveira, H. A. Pereira, A. Rossi, J. A. Sena, C. Silva, R. F. de Souza, L. A. Spinola, M. A. Takita, R. E. Tamura, E. C. Teixeira, R. I. Tezza, M. Trindade dos Santos, D. Truffi, S. M. Tsai, F. F. White, J. C. Setubal, and J. P. Kitajima.** 2002. Comparison of the genomes of two *Xanthomonas* pathogens with differing host specificities. Nature **417:**459–463.
13. **Dodd, I. B., and J. B. Egan.** 1996. The *Escherichia coli* retrons Ec67 and Ec86 replace DNA between the *cos* site and a transcription terminator of a 186-related prophage. Virology **219:**115–124.
14. **Dodd, I. B., B. Kalionis, and J. B. Egan.** 1990. Control of gene expression in the temperate coliphage 186. VIII. Control of lysis and lysogeny by a transcriptional switch involving face-to-face promoters. J. Mol. Biol. **214:**27–37.
15. **Dow, J. M., and M. J. Daniels.** 2000. *Xylella* genomics and bacterial pathogenicity to plants. Yeast **17:**263–271.
16. **Fernandez-Prada, C. M., D. L. Hoover, B. D. Tall, A. B. Hartman, J. Kopelowitz, and M. M. Venkatesan.** 2000. *Shigella flexneri* IpaH(7.8) facilitates escape of virulent bacteria from the endocytic vacuoles of mouse and human macrophages. Infect. Immun. **68:**3608–3619.
17. **Finan, T. M.** 2002. Evolving insights: symbiosis islands and horizontal gene transfer. J. Bacteriol. **184:**2855–2856.
18. **Finan, T. M., S. Weidner, K. Wong, J. Buhrmester, P. Chain, F. J. Vorholter, I. Hernandez-Lucas, A. Becker, A. Cowie, J. Gouzy, B. Golding, and A. Puhler.** 2001. The complete sequence of the 1,683-kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. Proc. Natl. Acad. Sci. USA **98:**9889–9894.
19. **Groisman, E. A., M. H. Saier, Jr., and H. Ochman.** 1992. Horizontal transfer of a phosphatase gene as evidence for mosaic structure of the *Salmonella* genome. EMBO J. **11:**1309–1316.
20. **Hansen-Wester, I., and M. Hensel.** 2002. Genome-based identification of chromosomal regions specific for *Salmonella* spp. Infect. Immun. **70:**2351–2360.
21. **Hashimoto, Y., N. Li, H. Yokoyama, and T. Ezaki.** 1993. Complete nucleotide sequence and molecular characterization of ViaB region encoding Vi antigen in *Salmonella typhi*. J. Bacteriol. **175:**4456–4465.
22. **Hochhut, B., K. Jahreis, J. W. Lengeler, and K. Schmid.** 1997. CTnscr94, a conjugative transposon found in enterobacteria. J. Bacteriol. **179:**2097–2102.
23. **Hornick, R. B., S. E. Greisman, T. E. Woodward, H. L. DuPont, A. T. Dawkins, and M. J. Snyder.** 1970. Typhoid fever: pathogenesis and immunologic control. N. Engl. J. Med. **283:**686–691.
24. **Houng, H. S., K. F. Noon, J. T. Ou, and L. S. Baron.** 1992. Expression of Vi antigen in *Escherichia coli* K-12: characterization of ViaB from *Citrobacter freundii* and identity of ViaA with RcsB. J. Bacteriol. **174:**5910–5915.
25. **Hsu, M. Y., M. Inouye, and S. Inouye.** 1990. Retron for the 67-base multicopy single-stranded DNA from *Escherichia coli*: a potential transposable element encoding both reverse transcriptase and Dam methylase functions. Proc. Natl. Acad. Sci. USA **87:**9454–9458.
26. **Hull, R. A., R. E. Gill, P. Hsu, B. H. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33:**933–938.
27. **Lamont, I., A. M. Brumby, and J. B. Egan.** 1989. UV induction of coliphage 186: prophage induction as an SOS function. Proc. Natl. Acad. Sci. USA **86:**5492–5496.
28. **Larbig, K. D., A. Christmann, A. Johann, J. Klockgether, T. Hartsch, R. Merkl, L. Wiehlmann, H. J. Fritz, and B. Tummler.** 2002. Gene islands integrated into $tRNA^{Gly}$ genes confer genome diversity on a *Pseudomonas aeruginosa* clone. J. Bacteriol. **184:**6665–6680.
29. **Liu, S. L., and K. E. Sanderson.** 1995. Genomic cleavage map of *Salmonella typhi* Ty2. J. Bacteriol. **177:**5099–5107.
30. **Mandel, A. D., L. S. Baron, and C. E. Buckler.** 1959. Role of Vi in *Salmonella paratyphi* C infections. Proc. Soc. Exp. Biol. Med. **100:**653–656.
31. **McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du, S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson.** 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. Nature **413:**852–856.
32. **Mirold, S., W. Rabsch, M. Rohde, S. Stender, H. Tschape, H. Russmann, E. Igwe, and W. D. Hardt.** 1999. Isolation of a temperate bacteriophage encoding the type III effector protein SopE from an epidemic *Salmonella typhimurium* strain. Proc. Natl. Acad. Sci. USA **96:**9845–9850.
33. **Nakayama, K., S. Kanaya, M. Ohnishi, Y. Terawaki, and T. Hayashi.** 1999. The complete nucleotide sequence of phi CTX, a cytotoxin-converting phage of *Pseudomonas aeruginosa*: implications for phage evolution and horizontal gene transfer via bacteriophages. Mol. Microbiol. **31:**399–419.
34. **Neufing, P. J., K. E. Shearwin, J. Camerotto, and J. B. Egan.** 1996. The CII protein of bacteriophage 186 establishes lysogeny by activating a promoter upstream of the lysogenic promoter. Mol. Microbiol. **21:**751–761.
35. **Ou, J. T., C. J. Huang, H. S. Houng, and L. S. Baron.** 1992. Role of IS1 in the conversion of virulence (Vi) antigen expression in Enterobacteriaceae. Mol. Gen. Genet. **234:**228–232.
36. **Pancetti, A., and J. E. Galan.** 2001. Characterization of the *mutS*-proximal region of the *Salmonella typhimurium* SPI-1 identifies a group of pathogenicity island-associated genes. FEMS Microbiol Lett. **197:**203–208.
37. **Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain,**

C. Churcher, K. L. Mungall, S. D. Bentley, M. T. Holden, M. Sebaihia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connerton, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. Nature **413:**848–852.

38. **Parry, C. M., T. T. Hien, G. Dougan, N. J. White, and J. J. Farrar.** 2002. Typhoid fever. N. Engl. J. Med. **347:**1770–1782.
39. **Pickard, D., J. Li, M. Roberts, D. Maskell, D. Hone, M. Levine, G. Dougan, and S. Chatfield.** 1994. Characterization of defined *ompR* mutants of *Salmonella typhi*: *ompR* is involved in the regulation of Vi polysaccharide expression. Infect. Immun. **62:**3984–3993.
40. **Porwollik, S., R. M. Wong, and M. McClelland.** 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. Proc. Natl. Acad. Sci. USA **99:**8956–8961.
41. **Robbins, J. D., and J. B. Robbins.** 1984. Reexamination of the protective role of the capsular polysaccharide (Vi antigen) of *Salmonella typhi*. J. Infect. Dis. **150:**436–449.
42. **Romling, U., J. Wingender, H. Muller, and B. Tummler.** 1994. A major *Pseudomonas aeruginosa* clone common to patients and aquatic habitats. Appl. Environ. Microbiol. **60:**1734–1738.
43. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945.
44. **Selander, R. K., P. Beltran, N. H. Smith, R. Helmuth, F. A. Rubin, D. J. Kopecko, K. Ferris, B. D. Tall, A. Cravioto, and J. M. Musser.** 1990. Evolutionary genetic relationships of clones of *Salmonella* serovars that cause human typhoid and other enteric fevers. Infect. Immun. **58:**2262–2275.
45. **Simpson, A. J., F. C. Reinach, P. Arruda, F. A. Abreu, M. Acencio, R. Alvarenga, L. M. Alves, J. E. Araya, G. S. Baia, C. S. Baptista, M. H. Barros, E. D. Bonaccorsi, S. Bordin, J. M. Bove, M. R. Briones, M. R. Bueno, A. A. Camargo, L. E. Camargo, D. M. Carraro, H. Carrer, N. B. Colauto, C. Colombo, F. F. Costa, M. C. Costa, C. M. Costa-Neto, L. L. Coutinho, M. Cristofani, E. Dias-Neto, C. Docena, H. El-Dorry, A. P. Facincani, A. J. Ferreira, V. C. Ferreira, J. A. Ferro, J. S. Fraga, S. C. Franca, M. C. Franco, M. Frohme, L. R. Furlan, M. Garnier, G. H. Goldman, M. H. Goldman, S. L. Gomes, A. Gruber, P. L. Ho, J. D. Hoheisel, M. L. Junqueira, E. L. Kemper, J. P. Kitajima, J. E. Krieger, E. E. Kuramae, F. Laigret, M. R. Lambais, L. C. Leite, E. G. Lemos, M. V. Lemos, S. A. Lopes, C. R. Lopes, J. A. Machado, M. A. Machado, A. M. Madeira, H. M. Madeira, C. L. Marino, M. V. Marques, E. A. Martins, E. M. Martins, A. Y. Matsukuma, C. F. Menck, E. C. Miracca, C. Y. Miyaki, C. B. Monteriro-Vitorello, D. H. Moon, M. A. Nagai, A. L. Nascimento, L. E. Netto, A. Nhani, Jr., F. G. Nobrega, L. R. Nunes, M. A. Oliveira, M. C. de Oliveira, R. C. de Oliveira, D. A. Palmieri, A. Paris, B. R. Peixoto, G. A. Pereira, H. A. Pereira, Jr., J. B. Pesquero, R. B. Quaggio, P. G. Roberto, V. Rodrigues, A. J. D. M. Rosa, V. E. de Rosa, Jr., R. G. de Sa, R. V. Santelli, H. E. Sawasaki, A. C. da Silva, A. M. da Silva, F. R. da Silva, W. A. da Silva, Jr., J. F. da Silveira, et al.** 2000. The genome sequence of the plant pathogen *Xylella fastidiosa*. The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing and Analysis. Nature **406:**151–157.M. R. A. J.
46. **Sullivan, J. T., J. R. Trzebiatowski, R. W. Cruickshank, J. Gouzy, S. D. Brown, R. M. Elliot, D. J. Fleetwood, N. G. McCallum, U. Rossbach, G. S. Stuart, J. E. Weaver, R. J. Webby, F. J. De Bruijn, and C. W. Ronson.** 2002. Comparative sequence analysis of the symbiosis island of *Mesorhizobium loti* strain R7A. J. Bacteriol. **184:**3086–3095.
47. **Venkatesan, M. M., M. B. Goldberg, D. J. Rose, E. J. Grotbeck, V. Burland, and F. R. Blattner.** 2001. Complete DNA sequence and analysis of the large virulence plasmid of *Shigella flexneri*. Infect. Immun. **69:**3271–3285.
48. **Virlogeux, I., H. Waxin, C. Ecobichon, J. O. Lee, and M. Y. Popoff.** 1996. Characterization of the *rcsA* and *rcsB* genes from *Salmonella typhi*: *rcsB* through *tviA* is involved in regulation of Vi antigen synthesis. J. Bacteriol. **178:**1691–1698.
49. **Zhang, X. L., C. Morris, and J. Hackett.** 1997. Molecular cloning, nucleotide sequence, and function of a site-specific recombinase encoded in the major 'pathogenicity island' of *Salmonella typhi*. Gene **202:**139–146.
50. **Zhang, X. L., I. S. Tsui, C. M. Yip, A. W. Fung, D. K. Wong, X. Dai, Y. Yang, J. Hackett, and C. Morris.** 2000. *Salmonella enterica* serovar typhi uses type IVB pili to enter human intestinal epithelial cells. Infect. Immun. **68:**3067–3073.