INFECTION AND IMMUNITY, Nov. 1997, p. 4606–4614 Vol. 65, No. 11
0019-9567/97/$04.00+0
Copyright © 1997, American Society for Microbiology

Use of a Novel Approach, Termed Island Probing, Identifies the *Shigella flexneri she* Pathogenicity Island Which Encodes a Homolog of the Immunoglobulin A Protease-Like Family of Proteins

K. RAJAKUMAR,[1]* C. SASAKAWA,[2] AND B. ADLER[1]

Department of Microbiology, Monash University, Clayton, Victoria 3168, Australia,[1] and Department of Bacteriology, Institute of Medical Science, University of Tokyo, Minato-ku, Tokyo 108, Japan[2]

Received 25 April 1997/Returned for modification 26 June 1997/Accepted 25 August 1997

The *she* gene of *Shigella flexneri* 2a, which also harbors the internal enterotoxin genes *set1A* and *set1B* (F. R. Noriega, GenBank accession no. U35656, 1995) encodes a homolog of the virulence-related immunoglobulin A (IgA) protease-like family of secreted proteins, Tsh, EspC, SepA, and Hap, from an avian pathogenic *Escherichia coli*, an enteropathogenic *E. coli*, *S. flexneri* 5, and *Haemophilus influenzae*, respectively. To investigate the possibility that this locus was carried on a larger deletable element, the *S. flexneri* 2a YSH6000T *she* gene was insertionally disrupted by allelic exchange using a Tn*10*-derived *tetAR*(B) cassette. Then, to detect loss of the *she* locus, the tetracycline-resistant derivative was plated onto fusaric acid medium to select for tetracycline-sensitive revertants, which were observed to arise at a frequency of 10^{-5} to 10^{-6}. PCR and pulsed-field gel electrophoresis analysis confirmed loss of the *she*::*tetAR*(B) locus in six independent tetracycline-sensitive isolates. Sample sequencing over a 25-kb region flanking *she* identified four insertion sequence-like elements, the group II intron-like sequence Sf.IntA, and the 3′ end of a second IgA protease-like homolog, *sigA*, lying 3.6 kb downstream and in an orientation inverted with respect to *she*. The deletion was mapped to chromosomal *Not*I fragment A and determined to have a size of 51 kb. Hybridization with flanking probes confirmed that at least 17.7 kb of the 51-kb deletable element was unique to the seven *she*$^{+}$ strains investigated, supporting the conclusion that *she* lay within a large pathogenicity island. The method described in this study, termed island probing, provides a useful tool to further the study of pathogenicity islands in general. Importantly, this approach could also be of value in constructing safer live attenuated bacterial vaccines.

The term pathogenicity island (PAI) was first coined to describe large, unstable, virulence-associated, chromosomal segments in uropathogenic *Escherichia coli* (10). Since then, the definition has been broadened to refer to regions of the chromosome foreign to the host bacterium and which encode one or more virulence determinants (28, 40). PAIs ranging in size from a 1.6-kb single gene island encoding *sifA* of *Salmonella typhimurium* (69) to large 35- to 190-kb multigene elements have now been identified in many species, including uropathogenic *E. coli* (10, 11, 72), enteropathogenic *E. coli* (EPEC) (38, 39), *S. typhimurium* (44, 64), *Yersinia pestis* (22), *Yersinia enterocolitica* (12), *Helicobacter pylori* (13), *Vibrio cholerae* (5), *Listeria monocytogenes* (27), and *Dichelobacter nodosus* (6, 9, 14).

In *Shigella* species, the plasmid-encoded *mxi-spa* cluster of virulence genes, which encodes a type III secretion system, has been referred to as a PAI, as it is homologous to three other type III gene clusters that are borne on true chromosomal PAIs (40, 62, 64). Recently a second putative PAI mapping to chromosomal *Not*I fragment D of *Shigella flexneri* 2a YSH6000 was identified (58). Loss of this 99-kb element, which probably arose following the integration of an NR1-like resistance plasmid (59), resulted in the deletant YSH6000T, which exhibited reduced contact hemolytic activity (58), a recognized virulence-associated phenotype in *Shigella* (61).

As watery diarrhea often precedes the bloody, mucoid dysentery symptomatic of invasive shigellosis, it had been hypothesized that *Shigella* produces an enterotoxin (20). Recently, two enterotoxins were identified. The toxins ShET1 (20) and ShET2 (47) result in fluid accumulation in isolated rabbit ileal loops and are expressed optimally under iron-limiting conditions. ShET2 is encoded on the large 140-MDa virulence plasmid of *S. flexneri* 2a. It is homologous to the enteroinvasive *E. coli* (EIEC) enterotoxin EIET (47). However, ShET1 is produced by the tandem chromosomal genes *set1B* and *set1A*, which encode the 7- and 20-kDa subunits of the 55-kDa holotoxin. Convalescent sera obtained following *S. flexneri* 2a infection bound to immobilized ShET1 (20). Indeed, the finding that *set1A* and *set1B* are conserved among *S. flexneri* 2a strains, the most prevalent endemic strain in the developing world, rare in other *S. flexneri* serotypes, and probably absent in EIEC isolates led Noriega et al. (49) to suggest that the toxin may play a role in the communicability of shigellosis.

Interestingly, *set1A* and *set1B* are contained entirely within a third, oppositely orientated open reading frame (ORF), designated *she* (48). *she* encodes a 1,372-amino-acid predicted protein, ShMu, with putative hemagglutinin and mucinase activities (48). ShMu exhibits high levels of similarity to Tsh (56), EspC (70), SepA (3), and Hap (71) from an avian-pathogenic *E. coli*, an EPEC strain, *S. flexneri* 5, and *Haemophilus influenzae*, respectively (48, 57). Although these latter proteins lack immunoglobulin A (IgA) protease activity, they have been

* Corresponding author. Mailing address: Department of Microbiology, Monash University, Wellington Road, Clayton, Melbourne, Victoria 3168, Australia. Phone: 61 3 9905 4823. Fax: 61 3 9905 4811. E-mail: Kumar.Rajakumar@med.monash.edu.au.

TABLE 1. Bacterial strains and plasmids used

Baterial strain or plasmid	Relevant characteristics[a]	Reference or source
Bacterial strains		
S. flexneri		
YSH6000	Wild-type *S. flexneri* 2a Japanese isolate, Ap^r Cm^r Sm^r Tc^r	63
YSH6000T	Spontaneous variant of YSH6000 exhibiting a 99-kb chromosomal deletion, Ap^s Cm^s Sm^s Tc^s	58
SBA1336	Derivative of YSH6000T harboring a Tc^r cassette within *she*	This study
SBA1341	Tc^s derivative of SBA1336 exhibiting a 51-kb chromosomal deletion	This study
SBA1342	Tc^s derivative of SBA1336 exhibiting a 51-kb chromosomal deletion	This study
SBA1343	Tc^s derivative of SBA1336 exhibiting a 51-kb chromosomal deletion	This study
SBA1317	Australian isolate serotype 2a	D. Lightfoot
SBA1318	Australian isolate serotype 2a	D. Lightfoot
212-83	South American isolate serotype Y	66
219-83	South American isolate serotype Y	66
SBA1173	Serotype 1b	D. Lightfoot
SBA1299	Japanese isolate serotype 1a	C. Sasakawa
SBA1300	Japanese isolate serotype 2b	C. Sasakawa
SBA1316	Australian isolate serotype 2b	D. Lightfoot
E. coli		
DH5α	F^- φ80 Δ*lacZ* M15 *recA1*	Bethesda Research Laboratories
SY327λ*pir*	F^- *araD* Δ(*lac pro*) *argE recA56 nalA* [λ*pir*]	C. Sasakawa
SM10λ*pir*	*thi thr leu tonA lacY supE recA*::RP4-2-Tc::Mu [λ*pir*] Km^r	42
SBA635	SM10λ*pir* harboring pSBA389	This study
Plasmids		
pWSK129	pSC101-based low-copy-number vector, Km^r Δ*lacZ*	77
pWSK29	pSC101-based low-copy-number vector, Ap^r Δ*lacZ*	77
pUC4-KIXX	pUC4K derivative containing Km^r and Bm^r genes from Tn*5*	Pharmacia
pGP704	*ori* R6K, *mob* RP4, derivative of pJM703.1, Ap^r	42
R100-1	Derivative of NR1 displaying derepressed conjugal transfer activity, Tc^r	78
pSBA381	2.6-kb PCR fragment containing Δ*she* from YSH6000T cloned into T-tailed *Eco*RV site in pWSK129, Km^r	This study
pSBA383	2.2-kb PCR fragment containing *tetAR*(B) from R100-1 cloned into *Hin*dIII sites of pUC4-KIXX, Ap^r Tc^{rb}	This study
pSBA385	2.2-kb *Sma*I *tetAR*(B) cassette from pSBA383 cloned into *Asp*700 site of pSBA381, Km^r Tc^r	This study
pSBA389	4.0-kb *Bam*HI/*Sal*I fragment of pSBA385 cloned into *Bgl*II/*Sal*I sites of pGP704, Ap^r Tc^r	This study
pSBA414	14.4-kb *Bam*HI fragment of SBA1336 bearing the 3′ truncated *she* gene with an inserted *tetAR*(B) cassette cloned into *Bam*HI site of pWSK29, Ap^r Tc^r	This study
pSBA415	19.9-kb *Sal*I fragment of SBA1336 bearing *she*::*tetAR*(B) cloned into *Sal*I site of pWSK29, Ap^r Tc^r	This study
pSBA449	9.4-kb *Bam*HI/*Hin*dIII fragment of R100-1 bearing *tetAR*(B) genes of Tn*10* cloned into *Bam*HI/*Hin*dIII sites of pWSK29, Ap^r Tc^r	This study

[a] Ap, ampicillin; Bm, bleomycin; Cm, chloramphenicol; Km, kanamycin; Sm, streptomycin; Tc, tetracycline.
[b] pSBA383, a high-copy-number plasmid, conferred only low-level tetracycline resistance (<1 μg/ml).

assigned to an IgA protease-like family based on their similarity to the IgA proteases of *Neisseria gonorrheae* and *H. influenzae* and because of apparent similarities in their modes of secretion (3, 56, 70, 71). SepA probably plays a role in tissue invasion, as a defined *S. flexneri sepA* mutant caused less mucosal atrophy and tissue inflammation in a ligated rabbit ileal loop model (3). Tsh conferred upon *E. coli* K-12 the ability to agglutinate chicken erythrocytes, suggesting a possible role for this protein in adherence of avian-pathogenic *E. coli* to its target tissue (56). Similarly, Hap conferred upon a laboratory strain of *H. influenzae* the ability to adhere to and invade Chang epithelial cells (71). EspC is also likely to play a part in virulence, as its secretion is coregulated with that of other secreted EPEC virulence factors (70).

The limited distribution of *she* among *Shigella* strains (49), the unusual arrangement of overlapping genes (18), and the presence within *she* of toxin genes which themselves are frequently associated with accessory elements (11, 14, 40) suggested to us that *she* itself may be encoded within a larger PAI. In this study, a novel approach was used to explore this hypothesis.

MATERIALS AND METHODS

Bacterial strains, plasmids, and growth conditions. Bacterial strains and plasmids used in this study are listed in Table 1. Strains were routinely grown at 37°C in either 2YT or LB medium (2) with the addition of ampicillin (100 μg/ml), kanamycin (50 μg/ml), or tetracycline (10 μg/ml) when necessary.

Preparation and manipulation of DNA. Genomic DNA was isolated by the small-scale preparation method of Ausubel et al. (2). Plasmid DNA was isolated by a modification of the alkaline lysis method (45). Standard restriction mapping and cloning procedures (2) using DNA-modifying enzymes supplied by Boehringer Mannheim were employed. Transformation of *E. coli* DH5α was achieved following electroporation (68) with a Bio-Rad Gene Pulser. Nucleotide sequencing was performed with a PRISM Ready Reaction DyeDeoxy Terminator cycle sequencing kit (Perkin-Elmer Corporation), and the reactions were analyzed on an Applied Biosystems model 373A DNA sequencing system.

Conjugation and selection of tetracycline-sensitive derivatives of SBA1336. Volumes of 0.7 ml of overnight cultures of *S. flexneri* 2a YSH6000T and *E. coli* SBA635 [harboring the suicide plasmid construct bearing the 5′ and 3′ truncated *she* gene with a *tetAR*(B) insertion cassette] were mixed; the bacteria were pelleted, washed with antibiotic-free medium, and plated onto Trypticase soy agar. Following the overnight plate matings, bacteria were harvested and plated onto *Shigella*-specific minimal medium (M9 medium [2] supplemented with glucose 10 mM and nicotinic acid 10 μg/ml) containing tetracycline (10 μg/ml) and incubated at 37°C for 48 to 72 h. Tetracycline-resistant transconjugants were patched onto ampicillin and tetracycline plates to identify the ampicillin-sensitive, tetracycline-resistant transconjugant SBA1336. Tetracycline-sensitive deriv-

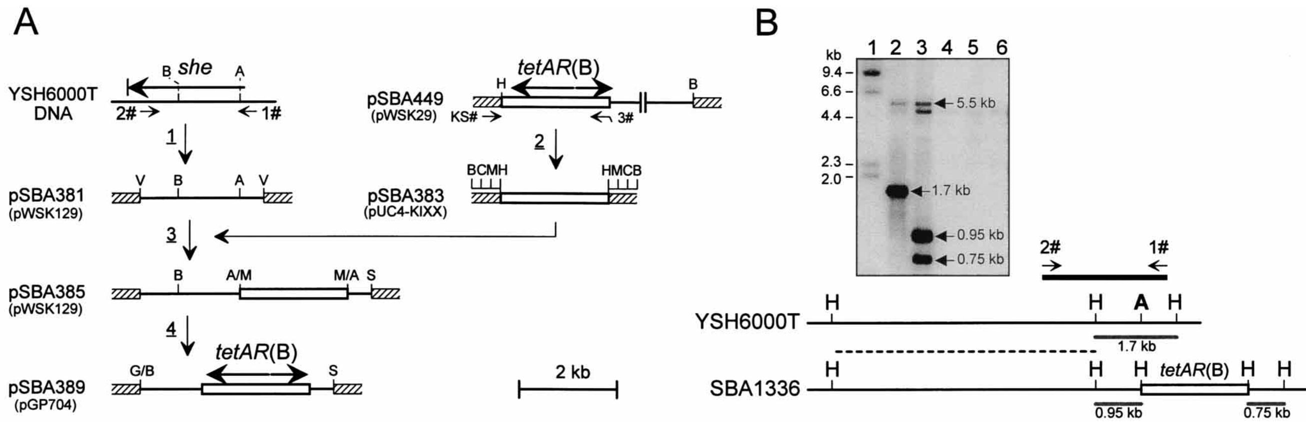

FIG. 1. Construction of the *tetAR*(B) cassette and the insertional *she* mutant SBA1336. The 5′ two-thirds of *she* and the *tetAR*(B) genes, where indicated, are represented as thick black arrows. Vector DNA is shown as open hatched boxes, and the *tetAR*(B) cassette is shown as a black lined box. A, B, C, G, H, M, and S indicate cleavage sites for the enzymes *Asp*700, *Bam*HI, *Sac*I, *Bgl*II, *Hin*dIII, *Sma*I, and *Sal*I, respectively, while V represents the T-tailed *Eco*RV site. Only relevant restriction sites are indicated. (A) 1, the 2.6-kb fragment encoding most of the 5′ two-thirds of *she* was amplified by PCR from YSH6000T DNA, using oligonucleotides 1 and 2, and cloned into the T-tailed *Eco*RV site of pWSK129, generating pSBA381. 2, the 2.2-kb fragment encoding *tetAR*(B) was amplified by PCR from pSBA449, using the pBluescript oligonucleotide KS and oligonucleotide 3, containing an incorporated *Hin*dIII site, digested with *Hin*dIII, and cloned into the *Hin*dIII sites of pUC4-KIXX, replacing the kanamycin and bleomycin genes in this vector. The resulting plasmid, pSBA383, contained the *tetAR*(B) cassette flanked by convenient *Hin*dIII, *Sma*I, *Sac*I, and *Bam*HI sites. 3, a *Sma*I-flanked *tetAR*(B) cassette was cloned into the *Asp*700 site of pSBA381, generating plasmid pSBA385. 4, The 4.0-kb *Bam*HI/*Sal*I fragment of pSBA385 was cloned into the *Bgl*II/*Sal*I sites of the suicide vector pGP704, and the resulting construct, pSBA389, was conjugally transferred into YSH6000T to select for the insertional mutant SBA1336. (B) Southern hybridization analysis of *Hin*dIII-digested genomic DNA with a *she* probe (black rectangle), generated by PCR amplification from YSH6000T DNA by using the oligonucleotide primers 1 and 2. Lane 1, digoxigenin-labeled λ *Hin*dIII size markers; lane 2, YSH6000T; lane 3, SBA1336; lane 4, SBA1341; lane 5, SBA1342; lane 6, SBA1343. Arrows indicate the 5.5- and 1.7-kb and the 5.5-, 0.95-, and 0.75-kb hybridizing bands in YSH6000T and SBA1336, respectively. Additional bands are the result of nonspecific hybridization with the cryptic plasmids in these strains. The sizes of λ *Hin*dIII markers are shown on the left. The line drawings below represent the restriction maps of the *she* loci in YSH6000T and SBA1336. The dashed line indicates the 5.5-kb hybridizing fragments common to both YSH6000T and SBA1336, while the grey lines represent the remaining hybridizing fragments. The *Asp*700 site shown in bold on the YSH6000T map indicates the insertion point in SBA1336 of the *tetAR*(B) cartridge which contains *Hin*dIII sites at both ends.

atives of SBA1336 were selected by plating dilutions of SBA1336 grown in LB broth alone onto LB agar supplemented with fusaric acid (12 μg/ml), chlortetracycline (50 μg/ml), and 0.1 mM $ZnCl_2$ (35).

Southern hybridization. Following electrophoresis, DNA was transferred onto positively charged nylon membranes (Boehringer Mannheim) by using a vacuum blotting apparatus (TE80 Transvac; Hoefer). Overnight hybridization and subsequent washing steps were performed under conditions of high stringency at a temperature of 65°C, using the protocol supplied with the Boehringer Mannheim digoxigenin labeling and detection kit. Probes were labeled by PCR amplification with digoxigenin as specified by Boehringer Mannheim. The 2.6-kb *she* probe was generated by using primers 1 and 2, which flank most of the proximal two-thirds of the gene. The 0.4-kb left-flanking and 0.4-kb right-flanking probes were generated by using primer pairs 4-SK and 5-T7, respectively. The templates used in the amplification reactions were YSH6000T genomic DNA for the *she* probe and pSBA415 for the latter two probes (Fig. 1 and 2A). Chemiluminescence detection with Lumigen PPD (Boehringer Mannheim) was used.

PFGE. High-molecular-weight DNA was prepared as described by Smith and Cantor (67) except that the detergent Brij 58 was omitted from the lysis solution. Low-melting-temperature agarose plugs (30 μl) containing approximately 2 μg of DNA were preequilibrated with digestion buffer before incubation for 16 h with 15 U of *Not*I (New England Biolabs), *Bln*I (Amersham Life Sciences), or *Sfi*I (New England Biolabs) in a 250-μl reaction mixture as recommended by the enzyme manufacturer. After digestion, the agarose plugs were incubated in ES (0.5 M EDTA [pH 9.5], 1% lauryl sarcosine) at 50°C for 2 h and equilibrated in TE (10 mM Tris [pH 8.0], 1 mM EDTA) at room temperature before pulsed-field gel electrophoresis (PFGE) in a Bio-Rad CHEF apparatus in agarose–1% gels and a running buffer consisting of 0.5× TBE (44.5 mM Tris, 44.5 mM boric acid, 1 mM EDTA). Gels were electrophoresed at 150 V for 40 h with a ramped pulse time of 15 to 25 s at 14°C.

PCR amplification. PCR amplifications were performed with the following oligonucleotide primers (GenBank accession numbers for sequences used to design the oligonucleotide primers are shown after the sequences): primers 1 (5′-CTCTTAAATATTGCCCCGTCACC-3′; U35656, 202 to 224 bp) and 2 (5′-TTCCAGTGCTGCGTTATCGCC-3′; U35656, 2795 to 2775 bp), which bind 2.6 kb apart on the *she* gene; primer 3 (5′-ATTT**AAGCTT**TTGAATCTAAAGG GTGG-3′; J01830, 817 to 844 bp), which anneals to a site adjacent to the *tetR* gene and contains an incorporated *Hin*dIII site (shown in bold); the outwardly directed primers 4 (5′-GCTGTGCCCATTTCCAGGC-3′; U97486, 370 to 352 bp) and 5 (5′-GCAGCGTGGTCAGGGTGTG-3′; U97490, 160 to 178 bp), which anneal to sites 0.4 kb from the left and right flanks, respectively, of the 19.9-kb *Sal*I fragment in pSBA415; and primers 6 (5′-TTCAAACTGCTTCCC TTCGCC-3′; U97489, 798 to 816 bp) and 7 (5′-ACGCCAGAGAAATAACTG GC-3′; U97491, 357 to 338 bp), which anneal 4.3 kb apart to sites within Sf.IntA and IS2, respectively (Fig. 1 and 2A). The pBluescript primers KS and SK as well as T7 were also used. PCR amplifications were performed with *Taq* polymerase (Boehringer Mannheim) in a reaction volume of 25 to 50 μl. Amplification reactions comprised 30 cycles of 94, 60, and 70°C for 44-s denaturation, 40-s annealing, and 1- to 4-min extension steps, respectively.

Computer analysis. Sequencing chromatograms were analyzed by using the Sequencher program (GeneCodes Corporation, Ann Arbor, Mich.). Nucleotide sequence and amino acid similarity searches with sequences in the databases were performed by using the BlastN, BlastP, or BlastX program (1) available on the GenomeNet World Wide Web server. G+C content was determined by using the Gene Runner program (Hastings Software, Inc.). Multiple protein sequence alignments were performed with Pileup, while pairwise comparisons were performed with Bestfit (Genetics Computer Group [Madison, Wis.] package).

Nucleotide sequence accession numbers. The nucleotide sequence data reported in this study have been lodged under GenBank accession no. U97486, U97487, U97488, U97489, U97490, U97491, U97492, and U97493 (Fig. 2A).

RESULTS

Construction of a *she* mutant. The oligonucleotide primers 1 and 2 were designed to amplify a 2.6-kb fragment spanning most of the 5′ two-thirds of *she* (48) (Fig. 1A). PCR amplification using as the template genomic DNA from *S. flexneri* 2a YSH6000T resulted in a 2.6-kb fragment, confirming the presence of the *she* gene in this strain and revealing that it was not borne on the previously identified 99-kb deletable element absent in YSH6000T (58, 59). The 2.6-kb PCR fragment was cloned into the T-tailed *Eco*RV site of pWSK129 to construct plasmid pSBA381 (Fig. 1A). A 9.4-kb *Hin*dIII/*Bam*HI fragment of R100-1 containing the Tn*10*-derived tetracycline resistance genes *tetR*(B) and *tetA*(B) (78) was cloned into the corresponding sites in pWSK29 to generate pSBA449. This was done to facilitate the construction of a *tetAR*(B) cassette to be used for both positive and negative selection (35). A 2.2-kb PCR fragment containing the *tetAR*(B) genes was then generated from pSBA449 by using the pBluescript primer KS, annealing to a site adjacent to the vector-borne *Hin*dIII site, and a second primer designed to incorporate into the PCR product

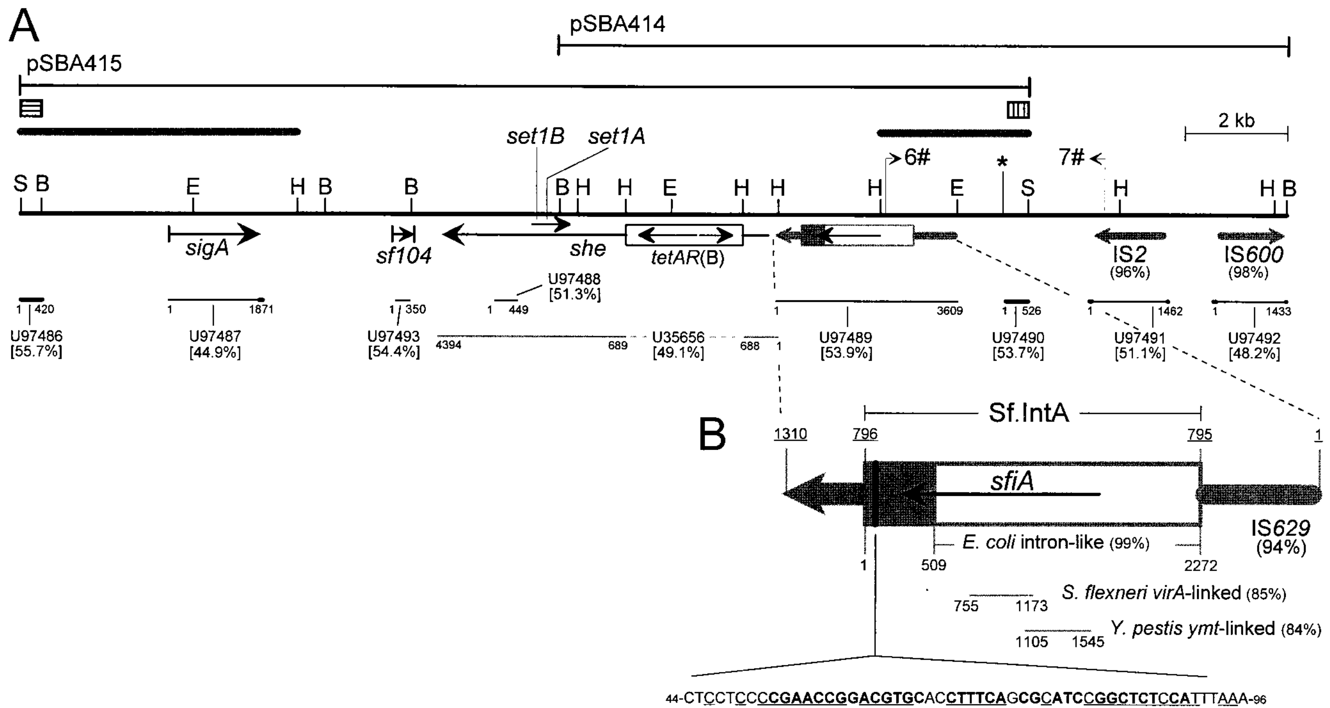

FIG. 2. Genetic organization of the *S. flexneri* 2a strain SBA1336 *she*-bearing region. (A) Genetic organization and restriction map of the *she*-bearing region showing the relative placement of pSBA414 and pSBA415. Genes are represented as black arrows, IS elements are represented as thick grey arrows, the Tn*10*-derived *tetAR*(B) cassette is shown as a box outlined in black, and the group II intron-like sequence Sf.IntA is shown as a grey partly shaded box. Vertical lines on arrows representing genes indicate that the ORF remains open at that end. The minimum percentage identities of the IS*2* and IS*600* elements to GenBank sequences V00610 and X05952, respectively, are shown in parentheses. *sf104* encodes a homolog of the *A. tumefaciens orf104* (GenBank accession no. X00493) hypothetical gene product. Other examples of this family of hypothetical proteins include those encoded within the following GenBank sequences: AE000077, AE000079, and AE000093 (*Rhizobium* sp. strain NGR234); M10204 and M19352 (*A. tumefaciens*); Y00551 (*Y. pseudotuberculosis*); and X60106 (*E. coli*). The hatched boxes represent probes corresponding to the left (horizontal hatch) and right (vertical hatch) ends of pSBA415, while the thick grey lines beneath these boxes indicate the 5.4- and 2.9-kb *Hin*dIII/*Sal*I fragments of SBA1336 that hybridized with these probes. The black lines below the restriction map indicate regions of the locus sequenced in this study (single-stranded sequence data only), with corresponding GenBank accession numbers, nucleotide positions, and percentage G+C content (in brackets) of each stretch of sequence as indicated. The thickened regions of these black lines represent stretches of sequence exhibiting no similarity to sequences in the databases. The lowermost thin grey line represents the original *she* sequence which was derived from a *S. flexneri* 2a strain M4243A (48). B, E, H, and S indicate cleavage sites for *Bam*HI, *Eco*RI, *Hin*dIII, and *Sal*I, respectively. The thin arrows above the map indicate the binding sites of oligonucleotide primers 6 and 7. The asterisk marks the location of the additional *Hin*dIII site in SBA1317 and SBA1318. (B) An expanded view of the IS*629* element harboring the 2,272-bp group II intron-like sequence Sf.IntA. *sfiA* encodes a putative protein exhibiting high levels of similarity to reverse transcriptase-like proteins encoded within the introns of fungi (GenBank accession no. X55026, U41288, and X57546), plants (GenBank accession no. M68929), and bacteria (GenBank accession no. U77945, U50902, X98606, and X71404). Nucleotides 1 to 508 of the intron-like sequence (shaded grey) exhibit only low-level similarity to sequences in the databases with the exception of a 53-bp stretch shown below which exhibits high similarity to a number of fungal group II introns (GenBank accession no. Z69899, X55026, and V00694) and to two mycobacterial sequences (GenBank accession no. U66101 and U47864). Underlined nucleotides are conserved among the introns, and nucleotides shown in bold are common to the two mycobacterial sequences. Nucleotides 509 to 2272 exhibit similarity to a chromosomally borne *E. coli* intron-like sequence (GenBank accession no. AE000133), and parts of this sequence also exhibit similarity to sequences linked to the bacterial plasmid-borne virulence determinants, *virA* (GenBank accession no. D26468) and *ymt* (GenBank accession no. X92727). Numbers below the schematic representation of this element correspond to nucleotide positions of the *S. flexneri* intron-like sequence, while the underlined numbers shown above refer to the nucleotide positions on the original IS*629* element (GenBank accession no. X51586). The minimum percentage identities of the *Shigella* sequences to corresponding GenBank sequences are shown in parentheses.

a second flanking *Hin*dIII site, primer 3 (Fig. 1A). The 2.2-kb PCR product was digested with *Hin*dIII and cloned into the *Hin*dIII sites of pUC4-KIXX, replacing the kanamycin and bleomycin resistance genes in this vector and generating plasmid pSBA383. Next, the 2.2-kb *Sma*I-flanked *tetAR*(B)-bearing fragment of pSBA383 was cloned into the *Asp*700 site of pSBA381, disrupting the *she* gene and resulting in the plasmid construct pSBA385. The 4.0-kb *Bam*HI/*Sal*I fragment of pSBA385 was then ligated into the compatible *Bgl*II/*Sal*I sites of pGP704, and the resulting plasmid, pSBA389 (Fig. 1A), was conjugally transferred into YSH6000T. Tetracycline-resistant transconjugants were patched onto medium containing ampicillin or tetracycline to identify the tetracycline-resistant, ampicillin-sensitive transconjugant SBA1336. PCR analysis using primers 1 and 2 and Southern hybridization with a *she* probe confirmed that allelic exchange had occurred in SBA1336, with the presence of the *tetAR*(B) cassette inserted into the *Asp*700 site in SBA1336 giving rise to the altered hybridization profile shown in Fig. 1B.

***she* lies within a deletable region of the genome.** To detect loss of the *she* locus, SBA1336 was grown to a density of about 10^8 ml^{-1} in LB broth, diluted, and plated onto LB medium supplemented with fusaric acid and incubated for 16 h at 37°C to select for tetracycline-sensitive derivatives. Three tetracycline-sensitive isolates, SBA1341, SBA1342, and SBA1343, were chosen at random and analyzed by PCR and Southern hybridization as described previously, confirming that the *she* locus had been deleted in all three cases (Fig. 1B). Because of the possibility that these three isolates were siblings, a further five tetracycline-sensitive isolates derived from independent cultures were investigated by PCR analysis and found to have lost the *she* locus in each case.

The rate of reversion to tetracycline sensitivity was calculated from data gained in duplicate experiments involving ex-

ponential growth of SBA1336 in nonselective LB medium for between six and nine generations. Appropriate dilutions of the culture were plated onto LB or fusaric acid medium to calculate the number of generations of growth and the proportion of tetracycline-sensitive revertants. Based on these data, the rate of reversion of SBA1336 to tetracycline sensitivity when assayed by plating onto fusaric acid medium was calculated to be approximately 3×10^{-6} to 1×10^{-5} per cell per generation.

Cloning and sample sequencing of a 25-kb region flanking the *she* locus. The *tetAR*(B) tag in SBA1336 was used to facilitate the rapid cloning of large regions flanking *she*. A 19.9-kb *Sal*I fragment and an overlapping 14.4-kb *Bam*HI fragment from SBA1336 were cloned into pWSK29, resulting in plasmids pSBA415 and pSBA414, respectively. Restriction mapping and alignment of the two fragments revealed that the clones spanned a 25-kb region flanking the *she* gene (Fig. 2A). Subclones were generated to allow for a rapid sample sequencing approach for analysis of the flanking regions.

The sequence data obtained identified two complete insertion sequence (IS) elements, IS*600* (37) and IS*2* (26), and a disrupted insertion sequence, IS*629* (36). In addition, a 350-bp stretch of sequence 0.6 kb downstream of *she* contained an ORF designated *sf104* (*S. flexneri* homolog of *Agrobacterium tumefaciens orf104*) (32), open at both ends, which encoded a putative protein exhibiting up to 68% similarity to hypothetical proteins encoded on several IS-like elements from *Rhizobium* (23) and *Agrobacterium* spp. (Fig. 2A). Possible IS-like elements linked to the *E. coli csvR* and *Yersinia pseudotuberculosis yopH* plasmid-borne virulence genes also encode products exhibiting similarity to that of *sf104* (32). Database searches also revealed that IS*600* and IS*629* (or the closely related IS*3411*) (30) were often found associated with other virulence determinants. IS*600*-like sequences map close to the Shiga toxin gene of *Shigella dysenteriae* 1 (33) and are also contained within the previously described 99-kb *Shigella* PAI (57, 59). IS*629*-like sequences are linked to the *S. flexneri* virulence plasmid-borne *virA* (74), *sopA* (19), *sepA* (19), and *ipaH* (75) and to the urease (16), Shiga-like toxin I (55), and CS1 pilus (24, 31) loci of strains of *E. coli*.

A 2,272-bp stretch of sequence lying within IS*629* exhibited 99% identity over the 3′ 1,764 bp to an *E. coli* chromosomal group II intron-like sequence (Fig. 2B); 419- and 441-bp stretches of this latter sequence also exhibited significant similarity to sequences linked to the *S. flexneri virA* (74) and the *Y. pestis ymt* (15) genes, both of which are plasmid-borne virulence determinants. In addition, both the *Shigella* and the *E. coli* intron-like sequences contained an ORF (truncated in the case of *E. coli*) that encoded a putative protein which exhibited significant amino acid sequence similarity to a large family of reverse transcriptase-like proteins, many of which are encoded on fungal, plant, and bacterial group II introns (21, 41) (Fig. 2B). The remaining 508 bp of sequence lying within IS*629* exhibited only low-level similarity to sequences in the databases with the exception of a 53-bp stretch of sequence toward the 5′ end which exhibited 77 to 81% identity to sequences found within a number of fungal group II introns. In addition, an internal 39 bp of this sequence showed 84% identity to two mycobacterial sequences (Fig. 2B).

Located 3.6 kb downstream of *she* is a 1.7-kb stretch of sequence exhibiting significant similarity to regions toward the 3′ end of *she*, *tsh*, *espC*, and *sepA*. The ORF contained within this sequence was still open at the 5′ end. It was oriented in an opposite direction to *she* and encoded a putative protein exhibiting between 38 and 72% similarity with the C-terminal ends of ShMu, Tsh, EspC, SepA, and Hap (Figs. 2A and 3), hence its designation as *sigA* (*Shigella* IgA-like protease homolog).

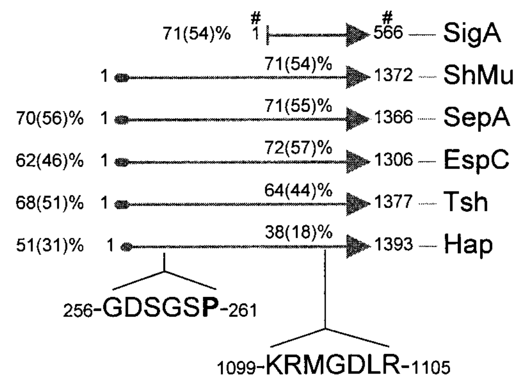

FIG. 3. Schematic representation of ShMu and the carboxy-terminal end of a putative SigA protein and a comparison of these proteins with the other members of the IgA protease-like family encoded by the *sepA*, *espC*, *tsh*, and *hap* genes (GenBank accession no. Z48219, U69128, L27423, and U11024). Amino acid similarity (identity) to ShMu is shown to the left, while similarity (identity) to SigA is shown above the corresponding line drawing. The amino acid sequences GDSGSP and KRMGDLR represent the catalytic site of IgA proteases and the putative carboxy-terminal cleavage motif of the helper domain, respectively. The first of these motifs is common to ShMu, SepA, EspC, Tsh, and Hap, with the exception of the proline residue (P) which is substituted with a glycine in EspC, while the second motif is conserved in all but the Hap protein. The thickened grey lines at the ends corresponding to the amino termini of the proteins represent the identified *sec*-dependent signal peptides (3, 48, 56, 57, 70, 71, 76). Amino acid positions are as indicated. The hatches (#) indicate that amino acid positions shown corresponding to SigA refer only to the predicted carboxy terminus of this putative protein. The numbers shown with the two motifs correspond to positions in the ShMu protein.

Finally, there were two identified long stretches of sequence of 420 and 526 bp along the 25-kb flanking region which exhibited no significant similarity at either a nucleotide or amino acid level to sequences in the databases. The G+C contents of these novel spans of sequence were 55.7 and 53.7%, respectively (Fig. 2A).

Size of the *she*-bearing deletable element. Southern hybridization of *Hin*dIII/*Sal*I-digested genomic DNA from YSH6000T and SBA1336 with the 0.4-kb probes corresponding to the left and right flanks of pSBA415 was performed. With both strains, 2.9- and 5.4-kb hybridizing fragments were observed (Fig. 2A and 4). However, the probes failed to hybridize with any *Hin*dIII/*Sal*I fragments in the three *she* deletants, SBA1341, SBA1342, and SBA1343, indicating loss of a fragment of at least 19.9 kb from the YSH6000T derivative, SBA1336. Hybridization of PFGE-resolved *Not*I-, *Bln*I-, or *Sfi*I-digested YSH6000T DNA with the *she* probe revealed that *she* mapped to the 1,080-kb chromosomal *Not*I fragment A (52, 53), a 218-kb *Bln*I fragment, and a 326-kb *Sfi*I fragment in this strain (data not shown). As expected, the probe did not hybridize with the negative control comprising *Not*I-digested SBA1341 DNA. A comparison of the PFGE profiles of *Bln*I- or *Sfi*I-digested YSH6000T, SBA1341, SBA1342, and SBA1343 DNA revealed that one of the two 218-kb *Bln*I fragments (218-kb doublet in Fig. 5, lane C, compared to single 218-kb *Bln*I fragments in lanes D through F) and the 326-kb *Sfi*I fragment of YSH6000T were reduced in size to 167 and 275 kb, respectively, in all three deletants, indicating that an equivalent deletion event involving the loss of 51 kb of chromosome had occurred in each case (Fig. 5). Importantly, PFGE of *Sfi*I-digested DNA from the additional five independent tetracycline-sensitive revertants confirmed that the same 51-kb deletion had also occurred in these strains.

The flanking DNA is unique to *she*$^+$ strains. In total, three distinct *S. flexneri* 2a, two *S. flexneri* 2b, one *S. flexneri* 1a, one

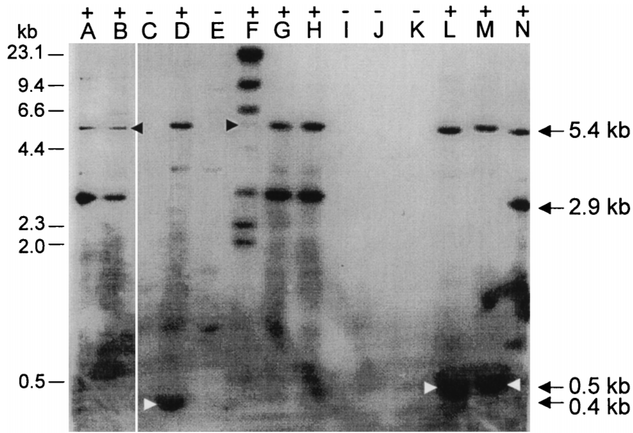

FIG. 4. Southern hybridization analysis of *Shigella* strains, using left and right *she*-flanking probes. *Hin*dIII/*Sal*I-digested genomic DNA from *S. flexneri* strains (serotype in parentheses below) was probed simultaneously with the 0.4-kb probes corresponding to the left and right ends of pSBA415. Lane A, 212-83 (Y); lane B, SBA1300 (2b); lane C, SBA1299 (1a); lane D, SBA1316 (2b); lane E, SBA1173 (1b); lane F, digoxigenin-labeled λ *Hin*dIII size markers [and 25% loading of YSH6000 (2a) *Hin*dIII/*Sal*I-digested DNA as a Southern hybridization sensitivity control]; lane G, YSH6000T (2a) positive control; lane H, SBA1336 (2a); lane I, SBA1341 (2a); lane J, SBA1342 (2a); lane K, SBA1343 (2a); lane L, SBA1317 (2a); lane M, SBA1318 (2a); lane N, 219-83 (Y). The arrows on the right indicate the sizes and positions of the 5.4- and 2.9-kb (lanes A, B, F, G, H, and N), 0.5-kb (lanes L and M), and 0.4-kb (lane D) hybridizing bands, while the arrowheads within the panel highlight the less distinct bands. The plus and minus symbols at the top indicate the presence or absence of *she* [or *she*::*tetAR*(B)] as assessed by PCR analysis using the *she*-specific oligonucleotides 1 and 2. The sizes of λ *Hin*dIII markers are shown on the left.

S. flexneri 1b, and two *S. flexneri* Y strains were investigated for the presence of the *she* gene by PCR analysis using primers 1 and 2. Genomic DNA from all 2a strains and both 2b and Y strains tested gave rise to amplification of a 2.6-kb fragment, confirming the presence of *she* in these strains. However, no amplification product was detected with the sole serotype 1a and 1b strains investigated. Hybridization with the left and right flanking probes resulted in detection of a 5.4-kb and a 2.9-kb (YSH6000T, 212-83, 219-83, and SBA1300), 0.5-kb (SBA1317 and SBA1318), or 0.4-kb (SBA1316) *Hin*dIII/*Sal*I fragment in all seven she^+ strains characterized. No hybridization was detected with the two *she* strains, demonstrating that at least 17.7 kb {19.9 kb [minimum size of deletion in SBA1336] − 2.2 kb [size of *tetAR*(B) cassette]} of the 51 kb *she*-bearing deletable element was unique to she^+ strains (Fig. 4). PCR amplification of the region hybridizing with the right flanking probe, using primers 6 and 7 (Fig. 2A), from YSH6000T, SBA1316, SBA1317, or SBA1318 resulted as expected in a 4.3-kb fragment with YSH6000T but in larger 5.6-kb fragments with SBA1317 and SBA1318. No amplification product was detected with SBA1316. Mapping of the *Sal*I and *Hin*dIII sites on the PCR amplification products showed identical relative placement of the *Sal*I site and oligonucleotide 7 priming site in all three fragments and revealed the presence of an additional *Hin*dIII site mapping 0.5 kb to the left of the *Sal*I site in SBA1317 and SBA1318, accounting for the distinct pattern of hybridization observed with these she^+ strains (Fig. 2A and 4).

DISCUSSION

In this study, *she* and the overlapping enterotoxin genes *set1A* and *set1B* were shown to be borne on a much larger 51-kb deletable chromosomal element which we have designated the *she* PAI. However, we cannot completely exclude the possibility that the *she* PAI had inserted into an inherently unstable region of the *Shigella* chromosome, and it could therefore be less than 51 kb in size. In addition, based on the findings of this study and the previously reported epidemiological survey of Noriega et al. (49), this locus was found to be prevalent in *S. flexneri* 2a strains and at least occasionally represented in a minimum of four other serotypes, Y, 1a, 2b, and 3b. As observed with use of fusaric acid selective medium, the *she* PAI was relatively unstable, deleting at a rate of 10^{-5} to 10^{-6}, consistent with the reported rates of loss of about 10^{-4} to 10^{-5} of similar elements in uropathogenic *E. coli* and *Y. pestis* (54). Like many other PAIs (28, 40), the DNA flanking *she* contained several IS elements, IS*600*, IS*2*, and a disrupted copy of IS*629*. The stretch of sequence encoding *sf104* was probably contained within a fourth IS-like element related to mobile elements in the phylogenetically distant plant-associated bacteria of the genera *Rhizobium* (23) and *Agrobacterium* (32). The 2,272-bp sequence that appears to have inserted precisely into IS*629* is likely, based on its features, to be an example of a bacterial group II intron, hence its designation as Sf.IntA and that of the gene contained within as *sfiA* (*S. flexneri* intron A). Other bacterial group II introns also exhibit a propensity to target accessory elements, such as plasmids (43, 65), IS elements (32), and conjugative transposons (46). Indeed, an analysis of the regions flanking the related *E. coli* intron suggests that it may have been truncated at one end following an IS*5*-mediated deletion event (25, 57). Interestingly, sequences related to IS*600*, IS*629* (or the closely related IS*3411*), *sf104*, and Sf.IntA were found to be closely linked to virulence determinants in *S. flexneri*, *E. coli*, *Y. pestis*, and *Y. pseudotuberculosis*, many of which were plasmid encoded.

The 51-kb PAI identified was mapped to the 1,080-kb chromosomal *Not*I fragment A (52, 53). However, this PAI was unlikely to encompass *thyA* (53) or *ompR-envZ* (4), the other virulence-associated loci mapping to this region of the chromosome (52), as the *she* deletants were not thymine auxotrophs (57) and the two-component regulator genes *ompR-envZ* are likely to be common to all *Shigella* and *E. coli* strains, unlike the reported distribution of the *she* locus (49).

The unusual genetic organization of this region, comprising the two convergent homologous genes *she* and *sigA*, was prob-

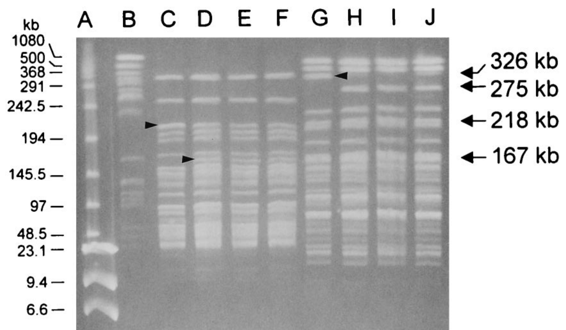

FIG. 5. PFGE profile of *Not*I-, *Bln*I-, or *Sfi*I-digested DNA of *Shigella* strains. Lane A, λ concatemers and λ *Hin*dIII size markers; lane B, YSH6000T (*Not*I); lane C, YSH6000T (*Bln*I); lane D, SBA1341 (*Bln*I); lane E, SBA1342 (*Bln*I); lane F, SBA1343 (*Bln*I); lane G, YSH6000T (*Sfi*I); lane H, SBA1341 (*Sfi*I); lane I, SBA1342 (*Sfi*I); lane J, SBA1343 (*Sfi*I). The arrows on the right indicate the positions and sizes of the 218-kb *Bln*I (lane C [doublet]) and the 326-kb *Sfi*I (lane G) fragments of YSH6000T and the corresponding 167-kb *Bln*I and 275-kb *Sfi*I fragments in the three deletants SBA1341, SBA1342, and SBA1343. Arrowheads highlight the locations of these bands. The positions of *Not*I fragment A (1,080 kb), *Not*I fragment B (500 kb), *Not*I fragment C (368 kb) (52), and the standard DNA size markers are shown on the left.

ably common to all seven *she*$^{+}$ strains investigated, as a 5.4-kb *Hin*dIII/*Sal*I fragment hybridized with the flanking probes in all cases (Fig. 2A and 4). The larger PCR amplification products obtained with SBA1317 or SBA1318 DNA by using primers 6 and 7 suggest the presence of an additional 1.3-kb element mapping to this region in these strains. Indeed, this is not surprising, as PAIs appear to be hot spots for the capture of other mobile elements (13, 28, 40). In addition, the lack of an amplification product from SBA1316 by using the same primer pair and its unique hybridization profile with the flanking probes suggest yet another variation in this *S. flexneri* 2b strain. Consistent with the hypothesis that this region comprised foreign DNA, at least 0.9 kb of the sequence obtained in this study exhibited no significant similarity to sequences in the databases. Furthermore, the two stretches of novel sequence identified exhibited G+C contents significantly different from the 49 to 53% G+C content of the *Shigella* chromosome (51) (Fig. 2A).

Based on Southern hybridization data, *S. flexneri* 2a was reported to harbor a *sepA*-like sequence on its virulence plasmid (3). Hence, the simultaneous presence of *she* is an example of the phenomenon of duplication of virulence-associated systems in bacteria (64). In addition, the identification of the ORF *sigA* suggests a possible third functional homolog in this strain. The presence of multiple IgA protease-like homologs in *S. flexneri* 2a may account for the observed effect on virulence of a *sepA* mutation alone, which resulted in attenuation in a ligated rabbit ileal loop model but not in in vitro virulence assays (3). Further data on *she* (48) and *sigA*, which we are currently characterizing, will clarify the contribution to virulence of each of these determinants in *Shigella*.

There is considerable evidence to suggest that the genes encoding SepA and EspC also map to PAIs. SepA is encoded on the large virulence plasmid of *S. flexneri* 5. As this plasmid is known to integrate reversibly into the chromosomal *metB* gene (80), *sepA* may be chromosomally resident on occasion. In addition, the relative placement of *sepA* and the *mxi-spa* cluster of genes differs in these two serotypes, supporting the proposition that the plasmid is composed of a mosaic of DNA fragments (3) comprising distinct PAIs which have integrated into the original plasmid replicon. Indeed, the recently identified *sopA* which encodes an outer membrane serine protease exhibiting high similarity to the phage-encoded OmpT lies about 12 kb downstream of *sepA*. Like the *she* PAI, the *sopA-sepA* locus harbors an IS*629*-like sequence, further supporting the suggestion that it may comprise part of a plasmid-borne PAI. The identification of *espC*-like sequences in *Citrobacter freundii* and *Hafnia alvei* strains exhibiting an attaching-and-effacing (AE) phenotype characteristic of EPEC strains, but not in nonpathogenic strains of these bacteria (70), supports the notion that *espC* is borne on an AE pathogen-specific region of DNA. Indeed, all other AE genes are borne on the 35.4-kb LEE (locus of enterocyte effacement) PAI (38, 39). However, *espC* appears to map elsewhere on the chromosome (70). In addition, the association in *H. pylori* of the PAI-borne *cagA* gene and the production of VacA, a vacuolating cytotoxin which is exported similarly to the IgA protease-like proteins, raises the possibility that *vacA* is also PAI encoded (17). However, as *vacA* maps some distance from *cagA* (17), it would have to lie within a second PAI.

Molecular epidemiological studies are leading to the identification of new PAIs (28). The application of techniques such as signature tagging (64), subtractive hybridization (34), representational difference analysis (13), differential display of mRNA (29), comparative genome mapping (8), conjugation-mediated directed tagged interval DNA replacement (7), and cosmid cloning combined with sample sequencing (72) offers the promise of accelerating the pace of chromosomal virulence gene discovery. The approach taken in this study, which we have termed island probing, adds to this armory by facilitating the detection of spontaneous chromosomal deletion, a feature common to many recently acquired members of the horizontal gene pool (28, 40, 73). Island probing offers the prospect of quickly addressing the question of the existence of other IgA protease-like PAIs. Equally, the method can be applied to investigate other virulence-associated genes which are likely to have been acquired via horizontal gene transfer. The use of a *tetAR*(B) cassette offers, in addition to the conventional uses of mutant construction and gene tagging, the added advantage of a simple one-step selection for tetracycline-sensitive revertants. Following random *tetAR*(B) mutagenesis, this technique can also be applied to survey the *Shigella*, *E. coli*, or *Salmonella* genomes for other spontaneously deletable regions. Further investigation of these tagged loci may lead to the identification of new PAIs. Use of an alternative negative selectable marker such as *sacB* (60, 79) in tandem with a suitable resistance gene would broaden the potential applicability of this method.

A similar approach could also be of value in the construction of live attenuated bacterial vaccines. Indeed, Noriega et al. (50) have suggested the potential benefit of mutagenizing the ShET1 genes in a Δ*guaB-A* Δ*virG S. flexneri* 2a candidate vaccine strain, CVD 1205. Deletion of the entire *she*-bearing PAI may prove to be even more efficacious in minimizing the potential reactogenicity of this vaccine. Importantly, the curing of redundant PAIs from vaccine strains will reduce the risk of undesirable horizontal gene transfer events (40). Indeed, recently the complete LEE PAI was cloned into an avirulent *E. coli* strain, resulting in a single-step acquisition of the AE phenotype and highlighting the risk of emergence of new pathogens by way of inheritance of PAIs (38).

ACKNOWLEDGMENTS

We are grateful to D. Lightfoot for providing and serotyping some of the *Shigella* strains used in this study.

This work was supported by a research grant from the National Health and Medical Research Council, Canberra, Australia.

ADDENDUM

While this manuscript was under review, a new homolog of the IgA protease-like family of proteins, EspP, was identified in enterohemorrhagic *E. coli* O157:H7. EspP was shown to cleave pepsin A and human coagulation factor V; the authors suggested that this latter activity may contribute to the colonic mucosal hemorrhage often accompanying *E. coli* O157 infection (11a). EspP exhibits 60% similarity to ShMu and 75% similarity to the predicted 566-amino-acid carboxy terminus of SigA (57).

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. **Ausubel, F. M., R. Brent, R. E. Kingston, D. M. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1991. Current protocols in molecular biology. Greene Publishing Associates and Wiley-Interscience, New York, N.Y.
3. **Benjelloun-Touimi, Z., P. J. Sansonetti, and C. Parsot.** 1995. SepA, the major extracellular protein of *Shigella flexneri*: autonomous secretion and involvement in tissue invasion. Mol. Microbiol. **17:**123–135.
4. **Bernardini, M. L., A. Fontaine, and P. J. Sansonetti.** 1990. The two-component regulatory system OmpR-EnvZ controls the virulence of *Shigella flexneri*. J. Bacteriol. **172:**6274–6281.
5. **Bik, E. M., A. E. Bunschoten, R. D. Gouw, and F. R. Mooi.** 1995. Genesis of the novel epidemic *Vibrio cholerae* O139 strain: evidence for horizontal transfer of genes involved in polysaccharide synthesis. EMBO J. **14:**209–216.
6. **Billington, S. J., J. L. Johnston, and J. I. Rood.** 1996. Virulence regions and

virulence factors of the ovine footrot pathogen, *Dichelobacter nodosus*. FEMS Microbiol. Lett. **145:**147–156.
7. **Bloch, C. A., and C. K. Rode.** 1996. Pathogenicity island evaluation in *Escherichia coli* K1 by crossing with laboratory strain K-12. Infect. Immun. **64:** 3218–3223.
8. **Bloch, C. A., C. K. Rode, V. Obreque, and K. Y. Russell.** 1994. Comparative genome mapping with mobile physical map landmarks. J. Bacteriol. **176:** 7121–7125.
9. **Bloomfield, G. A., G. Whittle, M. B. McDonagh, M. E. Katz, and B. F. Cheetham.** 1997. Analysis of sequences flanking the *vap* region of *Dichelobacter nodosus*: evidence for multiple integration events, a killer system, and a new genetic element. Microbiology **143:**553–562.
10. **Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschape, and J. Hacker.** 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606–614.
11. **Blum, G., V. Falbo, A. Caprioli, and J. Hacker.** 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. FEMS Microbiol. Lett. **126:**189–195.
11a. **Brunder, W., H. Schmidt, and H. Karch.** 1997. EspP, a novel extracellular protease of enterohaemorrhagic *Escherichia coli* O157:H7 cleaves human coagulation factor V. Mol. Microbiol. **24:**767–778.
12. **Carniel, E., I. Guilvout, and M. Prentice.** 1996. Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. J. Bacteriol. **178:**6743–6751.
13. **Censini, S., C. Lange, Z. Xiang, J. E. Crabtree, P. Ghiara, M. Borodovsky, R. Rappuoli, and A. Covacci.** 1996. *cag*, a pathogenicity island of *Helicobacter pylori*, encodes type I-specific and disease-associated virulence factors. Proc. Natl. Acad. Sci. USA **93:**14648–14653.
14. **Cheetham, B. F., and M. E. Katz.** 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbiol. **18:**201–208.
15. **Cherepanov, P. A., T. G. Mikhailova, G. A. Karimova, N. M. Zakharova, I. U. V. Ershov, and K. I. Volkovoi.** 1991. Cloning and detailed mapping of the *fra-ymt* region of the *Yersinia pestis* pFra plasmid. Mol. Gen. Mikrobiol. Virusol. **12:**19–26.
16. **Collins, C. M., and D. M. Gutman.** 1992. Insertional inactivation of an *Escherichia coli* urease gene by IS*3411*. J. Bacteriol. **174:**883–888.
17. **Cover, T. L.** 1996. The vacuolating cytotoxin of *Helicobacter pylori*. Mol. Microbiol. **20:**241–246.
18. **Dar, M. E., and A. S. Bhagwat.** 1993. Mechanism of expression of DNA repair gene *vsr*, an *Escherichia coli* gene that overlaps the DNA cytosine methylase gene, *dcm*. Mol. Microbiol. **9:**823–833.
19. **Egile, C., H. d'Hauteville, C. Parsot, and P. J. Sansonetti.** 1997. SopA, the outer membrane protease responsible for polar localisation of IcsA in *Shigella flexneri*. Mol. Microbiol. **23:**1063–1073.
20. **Fasano, A., F. R. Noriega, D. R. Maneval, Jr., S. Chanasongcram, R. Russell, S. Guandalini, and M. M. Levine.** 1995. *Shigella* enterotoxin 1: an enterotoxin of *Shigella flexneri* 2a active in rabbit small intestine in vivo and in vitro. J. Clin. Invest. **95:**2853–2861.
21. **Ferat, J.-L., and F. Michel.** 1993. Group II self-splicing introns in bacteria. Nature **364:**358–361.
22. **Fetherston, J. D., P. Schuetze, and R. D. Perry.** 1992. Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. Mol. Microbiol. **6:**2693–2704.
23. **Freiberg, C., R. Fellay, A. Bairoch, W. J. Broughton, A. Rosenthal, and X. Perret.** 1997. Molecular basis of symbiosis between Rhizobium and legumes. Nature **387:**394–401.
24. **Froehlich, B. J., A. Karakashian, L. R. Melsen, J. C. Wakefield, and J. R. Scott.** 1994. CooC and CooD are required for assembly of CS1 pili. Mol. Microbiol. **12:**387–401.
25. **Galas, D. J., and M. Chandler.** 1989. Bacterial insertion sequences, p. 109–162. *In* D. E. Berg, and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
26. **Ghosal, D., H. Sommer, and H. Saedler.** 1979. Nucleotide sequence of the transposable DNA-element IS2. Nucleic Acids Res. **6:**1111–1122.
27. **Gouin, E., J. Mengaud, and P. Cossart.** 1994. The virulence gene cluster of *Listeria monocytogenes* is also present in *Listeria ivanovii*, an animal pathogen, and *Listeria seeligeri*, a nonpathogenic species. Infect. Immun. **62:**3550–3553.
28. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. **23:**1089–1097.
29. **Heard, P. L., C. S. Lewis, and G. Chaudhuri.** 1996. *Leishmania mexicana amazonensis*: differential display analysis and cloning of mRNAs from attenuated and infective forms. J. Eukaryot. Microbiol. **43:**409–415.
30. **Ishiguro, N., and G. Sato.** 1988. Nucleotide sequence of insertion sequence IS*3411*, which flanks the citrate utilization determinant of transposon Tn*3411*. J. Bacteriol. **170:**1902–1906.
31. **Jordi, B. J. A. M., A. H. M. van Vliet, G. A. Willshaw, B. A. M. van der Zeijst, and W. Gaastra.** 1991. Analysis of the first two genes of the CS1 fibrial operon in human enterotoxigenic *Escherichia coli* of serotype O139:H28. FEMS Microbiol. Lett. **80:**265–270.
32. **Knoop, V., and A. Brennicke.** 1994. Evidence for a group II intron in *Escherichia coli* inserted into a highly conserved reading frame associated with mobile DNA sequences. Nucleic Acids Res. **22:**1167–1171.
33. **Kozlov, Y. U. V., A. A. Kabishev, E. V. Lukyanov, and A. A. Bayev.** 1988. The primary structure of the operons coding for *Shigella dysenteriae* toxin and temperature phage H30 Shiga-like toxin. Gene **67:**213–221.
34. **Mahairas, G. G., P. J. Sabo, M. J. Hickey, D. C. Singh, and C. K. Stover.** 1996. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. J. Bacteriol. **178:**1274–1282.
35. **Maloy, S. R., and W. D. Nunn.** 1981. Selection for loss of tetracycline resistance by *Escherichia coli*. J. Bacteriol. **145:**1110–1112.
36. **Matsutani, S., and E. Ohtsubo.** 1990. Complete sequence of IS*629*. Nucleic Acids Res. **18:**1899.
37. **Matsutani, S., H. Ohtsubo, Y. Maeda, and E. Ohtsubo.** 1987. Isolation and characterization of IS elements repeated in the bacterial chromosome. J. Mol. Biol. **196:**445–455.
38. **McDaniel, T. K., and J. B. Kaper.** 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol. Microbiol. **23:**399–407.
39. **McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg, and J. B. Kaper.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. USA **92:**1664–1668.
40. **Mecsas, J., and E. J. Strauss.** 1996. Molecular mechanisms of bacterial virulence: type III secretion and pathogenicity islands. Emerg. Infect. Dis. **2:**271–288.
41. **Michel, F., and J.-L. Ferat.** 1995. Structure and activities of group II introns. Annu. Rev. Biochem. **64:**435–461.
42. **Miller, V. L., and J. J. Mekalanos.** 1988. A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *toxR*. J. Bacteriol. **170:**2575–2583.
43. **Mills, D. A., L. L. Mckay, and G. M. Dunny.** 1996. Splicing of a group II intron involved in the conjugative transfer of pRS01 in lactococci. J. Bacteriol. **178:**3531–3538.
44. **Mills, D. M., V. Bajaj, and C. A. Lee.** 1995. A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. Mol. Microbiol. **15:**749–759.
45. **Morelle, G.** 1989. A plasmid extraction procedure on a miniprep scale. Focus **11:**7–8.
46. **Mullany, P., M. Pallen, M. Wilks, J. R. Stephen, and S. Tabaqchali.** 1996. A group II intron in a conjugative transposon from the Gram-positive bacterium, *Clostridium difficile*. Gene **174:**145–150.
47. **Nataro, J. P., J. Seriwatana, A. Fasano, D. R. Maneval, L. D. Guers, F. Noriega, F. Dubovsky, M. M. Levine, and J. G. Morris, Jr.** 1995. Identification and cloning of a novel plasmid-encoded enterotoxin of enteroinvasive *Escherichia coli* and *Shigella* strains. Infect. Immun. **63:**4721–4728.
48. **Noriega, F. R.** 1995. ShMu, a protein of *S. flexneri* 2a with hemagglutinin and mucinase activities is encoded by an open reading frame (*she*) that forms an antisense gene pair with the operon encoding *Shigella* enterotoxin 1 (ShET1). GenBank accession no. U35656.
49. **Noriega, F. R., F. M. Liao, S. B. Formal, A. Fasano, and M. M. Levine.** 1995. Prevalence of *Shigella* enterotoxin 1 among *Shigella* clinical isolates of diverse serotypes. J. Infect. Dis. **172:**1408–1410.
50. **Noriega, F. R., G. Losonsky, C. Lauderbaugh, F. M. Liao, J. Y. Wang, and M. M. Levine.** 1996. Engineered Δ*guaB-A* Δ*virG Shigella flexneri* 2a strain CVD 1205: construction, safety, immunogenicity, and potential efficacy as a mucosal vaccine. Infect. Immun. **64:**3055–3061.
51. **Normore, W. M.** 1973. Guanine-plus-cytosine (GC) composition of DNA of bacteria, fungi, algae and protozoa, p. 585–740. *In* A. I. Laskin and H. A. Lechevalier (ed.), CRC handbook of microbiology. II. Microbial composition. CRC Press, Cleveland, Ohio.
52. **Okada, N., C. Sasakawa, T. Tobe, K. A. Talukder, K. Komatsu, and M. Yoshikawa.** 1991. Construction of a physical map of the chromosome of *Shigella flexneri* 2a and the direct assignment of nine virulence-associated loci identified by Tn*5* insertions. Mol. Microbiol. **5:**2171–2180.
53. **Okada, N., C. Sasakawa, T. Tobe, M. Yamada, S. Nagai, K. A. Talukder, K. Komatsu, S. Kanegasaki, and M. Yoshikawa.** 1991. Virulence associated chromosomal loci of *Shigella flexneri* identified by random Tn*5* insertion mutagenesis. Mol. Microbiol. **5:**187–195.
54. **Ott, M.** 1993. Dynamics of the bacterial genome: deletions and integrations as mechanisms of bacterial virulence modulation. Zentralbl. Bakteriol. **278:** 457–468.
55. **Paton, A. W., J. C. Paton, P. N. Goldwater, M. W. Heuzenroeder, and P. A. Manning.** 1993. Sequence of a variant Shiga-like toxin type-I operon of *Escherichia coli* O111:H^-. Gene **129:**87–92.
56. **Provence, D. L., and R. Curtiss III.** 1994. Isolation and characterization of a gene involved in hemagglutination by an avian pathogenic *Escherichia coli* strain. Infect. Immun. **62:**1369–1380.

57. **Rajakumar, K., and B. Adler.** Unpublished data.
58. **Rajakumar, K., C. Sasakawa, and B. Adler.** 1996. A spontaneous 99-kb chromosomal deletion results in multi-antibiotic susceptibility and an attenuation of contact hemolysis in *Shigella flexneri* 2a. J. Med. Microbiol. **45:**64–75.
59. **Rajakumar, K., D. Bulach, J. Davies, L. Ambrose, C. Sasakawa, and B. Adler.** 1997. Identification of a chromosomal *Shigella flexneri* multi-antibiotic resistance locus which shares sequence and organizational similarity with the resistance region of the plasmid NR1. Plasmid **37:**159–168.
60. **Ried, J. L., and A. Collmer.** 1987. An *nptI-sacB-sacR* cartridge for constructing directed, unmarked mutations in Gram-negative bacteria by marker exchange-eviction mutagenesis. Gene **57:**239–246.
61. **Sansonetti, P. J., A. Ryter, P. Clerc, A. T. Maurelli, and J. Mounier.** 1986. Multiplication of *Shigella flexneri* within HeLa cells: lysis of the phagocytic vacuole and plasmid-mediated contact hemolysis. Infect. Immun. **51:**461–469.
62. **Sasakawa, C.** 1997. Early stages of *Shigella* interaction with host cells. J. Infect. Chemother. **3:**63–72.
63. **Sasakawa, C., K. Kamata, T. Sakai, S. Y. Murayama, S. Makino, and M. Yoshikawa.** 1986. Molecular alteration of the 140-megadalton plasmid associated with loss of virulence and Congo red binding activity in *Shigella flexneri*. Infect. Immun. **51:**470–475.
64. **Shea, J. E., M. Hensel, C. Gleeson, and D. W. Holden.** 1996. Identification of a virulence locus encoding a second type III secretion system in *Salmonella typhimurium*. Proc. Natl. Acad. Sci. USA **93:**2593–2597.
65. **Shearman, C., J.-J. Godon, and M. Gasson.** 1996. Splicing of a group II intron in a functional transfer gene of *Lactococcus lactis*. Mol. Microbiol. **21:**45–53.
66. **Silva, R. M., S. Saadi, and W. K. Maas.** 1988. A basic replicon of virulence-associated plasmids of *Shigella* spp. and enteroinvasive *Escherichia coli* is homologous with a basic replicon in plasmids of IncF groups. Infect. Immun. **56:**836–842.
67. **Smith, C. L., and C. R. Cantor.** 1987. Purification, specific fragmentation, and separation of large DNA molecules. Methods Enzymol. **155:**449–467.
68. **Smith, M., J. Jessee, T. Landers, and J. Jordan.** 1990. High efficiency bacterial electroporation: 1×10^{10} *E. coli* transformants/μg. Focus **12:**38–40.
69. **Stein, M. A., K. Y. Leung, M. Zwick, F. G. Portillo, and B. B. Finlay.** 1996. Identification of a *Salmonella* virulence gene required for formation of filamentous structures containing lysosomal membrane glycoproteins within epithelial cells. Mol. Microbiol. **20:**151–164.
70. **Stein, M., B. Kenny, M. A. Stein, and B. B. Finlay.** 1996. Characterization of EspC, a 110-kilodalton protein secreted by enteropathogenic *Escherichia coli* which is homologous to members of the immunoglobulin A protease-like family of secreted proteins. J. Bacteriol. **178:**6546–6554.
71. **St. Geme, J. W., III, M. L. de la Morena, and S. Falkow.** 1994. A *Haemophilus influenzae* IgA protease-like protein promotes intimate interaction with human epithelial cells. Mol. Microbiol. **14:**217–233.
72. **Swenson, D. L., N. O. Bukanov, D. E. Berg, and R. A. Welch.** 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. Infect. Immun. **64:**3736–3743.
73. **Thomas, C. M.** 1996. Meeting report of the second workshop of the European Science Foundation (ESF) network on the molecular biology and ecology of plasmid-mediated gene transfer, Potsdam, Germany. Microbiol. **142:**3317–3318.
74. **Uchiya, K., T. Tobe, K. Komatsu, T. Suzuki, M. Watarai, I. Fukuda, M. Yoshikawa, and C. Sasakawa.** 1995. Identification of a novel virulence gene, *virA*, on the large plasmid of *Shigella*, involved in invasion and intercellular spreading. Mol. Microbiol. **17:**241–250.
75. **Venkatesan, M. M., J. M. Buysse, and A. B. Hartman.** 1991. Sequence variation in two *ipaH* genes of *Shigella flexneri* 5 and homology to the LRG-like family of proteins. Mol. Microbiol. **5:**2435–2445.
76. **von Heijne, G.** 1984. How signal sequences maintain cleavage specificity. J. Mol. Biol. **173:**243–251.
77. **Wang, R. F., and S. R. Kushner.** 1991. Construction of versatile low-copy-number vectors for cloning, sequencing and gene expression in *Escherichia coli*. Gene **100:**195–199.
78. **Womble, D. D., and R. H. Rownd.** 1988. Genetic and physical map of plasmid NR1: comparison with other IncFII antibiotic resistance plasmids. Microbiol. Rev. **52:**433–451.
79. **Wu, S. S., and D. Kaiser.** 1996. Markerless deletions of *pil* genes in *Myxococcus xanthus* generated by counterselection with the *Bacillus subtilis sacB* gene. J. Bacteriol. **178:**5817–5821.
80. **Zagaglia, C., M. Casalino, B. Colonna, C. Conti, A. Calconi, and M. Nicoletti.** 1991. Virulence plasmids of enteroinvasive *Escherichia coli* and *Shigella flexneri* integrate into a specific site on the host chromosome: integration greatly reduces expression of plasmid-carried virulence genes. Infect. Immun. **59:**792–799.

Editor: J. T. Barbieri