JOURNAL OF BACTERIOLOGY, Feb. 1990, p. 603–609
0021-9193/90/020603-07$02.00/0

Vol. 172, No. 2

Sequence Invariance of the Antigen-Coding Central Region of the Phase 1 Flagellar Filament Gene (*fliC*) among Strains of *Salmonella typhimurium*

NOEL H. SMITH* AND ROBERT K. SELANDER

Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802

Received 14 June 1989/Accepted 26 October 1989

Previous studies of the phase 1 flagellar filament protein (flagellin) in strains of five serovars of *Salmonella* indicated that the central region of the *fliC* gene encoding the antigenic part of the protein is hypervariable both between and within serovars. To explore the possible use of this variation as a source of information on the phylogenetic relationships of closely related strains, we used the polymerase chain reaction technique to sequence part of the central region of the phase 1 flagellar genes of seven strains of *Salmonella typhimurium* that were known to differ in chromosomal genotype, as indexed by multilocus enzyme electrophoresis. We found that the nucleotide sequences of the central region were identical in all seven strains and determined that both the previously published sequence of the *fliC* gene in *S. typhimurium* LT2 and a report of a marked difference in the amino acid sequence of the phase 1 flagellins of two isolates of this serovar are erroneous. Our finding that the *fliC* gene is not evolving by sequence drift at an unusually rapid rate is compatible with a model that invokes lateral transfer and recombination of the flagellin genes as a major evolutionary process generating new serovars (antigen combinations) of salmonellae.

Antigenic variation in the flagellar filament protein (flagellin) of bacteria of the genus *Salmonella* has been used as a basis for strain classification and identification for 50 years (16). The Kauffmann-White serotyping scheme now identifies more than 2,000 serovars of *Salmonella*, each representing a distinctive combination of O somatic (lipopolysaccharide) antigens and phase 1 and phase 2 flagellar antigens (20). A total of 50 phase 1 flagellar antigens occurring in 61 combinations are recognized (21), and this extensive serological diversity is reflected in variation in the tryptic peptide map and amino acid composition of the flagellins (25, 41).

Previous studies of the phase 1 flagellin gene (*fliC*) in strains of five *Salmonella* serovars demonstrated that the central region exhibits only 21 to 32% nucleotide sequence similarity between serovars, although there is virtual sequence identity at each end of the gene (11, 39, 40). In the serovar *Salmonella typhimurium*, the central region has been shown to code for the major epitope of the *i* antigen of the phase 1 flagellin (10), and the terminal regions of the coding sequence of the gene have been identified as being important for the polymerization and secretion of flagellin (7).

It has been suggested that antigenic diversity in the flagellins of *Salmonella* serovars is generated by rapid random drift of the amino acid sequence of the central region (39), and this hypothesis was supported by evidence that soluble tryptic peptides from the phase 1 flagellin proteins of two isolates of *S. typhimurium* laboratory strain LT2 differed in at least five amino acids (12). If the flagellins of cell lineages of the same strain can exhibit this level of divergence, we might expect to find that the flagellin genes contain the most highly variable coding sequences of the *Salmonella* genome. These hypervariable regions might then provide a powerful system for determining phylogenetic relationships among closely related strains and a rich source of markers for epidemiological research. The notion that the central region of the phase 1 flagellin gene is subject to unusually rapid evolution through sequence drift was made more plausible by a recent report that deletion of the central region of the flagellin gene in *Escherichia coli* does not impair flagellar activity (17). This finding implies that the central region of the flagellin has a simple spacing or sheathing function, in which case there might be relatively little functional constraint on its amino acid sequence.

We report here the results of a study of the antigen-coding central region of the phase 1 flagellin gene in *S. typhimurium* undertaken to assess the extent and rate of sequence drift and to determine the value of comparative sequencing in estimating evolutionary relationships among closely related strains. The polymerase chain reaction (PCR) (29) was used to sequence part of the central region in seven strains of *S. typhimurium* representing the total span of chromosomal genotypic diversity in this serovar, as previously indexed by multilocus enzyme electrophoresis (MLEE) (2). Contrary to expectations, the sequences in all seven strains proved to be identical. We determined both that the previously published sequence of the *fliC* gene of *S. typhimurium* LT2 (11) is partially incorrect and that the report of multiple amino acid differences between the phase 1 flagellins of two isolates of *S. typhimurium* LT2 (12) is also erroneous.

MATERIALS AND METHODS

Materials. Sequenase (Version 1.0) sequencing kits, T4 polynucleotide kinase (cloned), and T4 DNA ligase were purchased from U.S. Biochemical Corporation (Cleveland, Ohio); NuSieve agarose was from FMC BioProducts; Geneclean and Circleprep were from Bio 101, Inc.; Hybond-N, [γ-^{32}P]ATP (>5,000 Ci/mmol), and [α-^{35}S]dATP (>1,000 Ci/mmol) were from Amersham Corp. (Arlington Heights, Ill.); [α-^{32}P]dATP and [α-^{32}P]dCTP (3,000 Ci/mmol) were from ICN Biomedicals, Inc.; *Thermus aquaticus* (Taq) polymerase was from New England BioLabs, Inc. (Beverly, Mass.); pBluescript and rescue bacteriophage R408 were from Stratagene (La Jolla, Calif.); and DNA polymerase I

* Corresponding author.

TABLE 1. Properties of strains of *S. typhimurium* studied

Strain	Origin	Host	ET[a]	Allele at indicated enzyme locus[b]:						
				IDH	AP2	CAT	HEX	G6P	PGM	GLU
RKS4939	LT2 (laboratory strain)		Tm 1	3	10	4	3	3	3	3
RKS284	Mexico	Human	Tm 1	3	10	4	3	3	3	3
RKS821	Norway		Tm 3	2	7	4	3	3	2	3
RKS811	Finland		Tm 5	3	7	4	3	2.5	2	3
RKS209	United States	Avian	Tm 7	3	10	4	4	4	3	3
RKS829	Thailand		Tm 10	3	10	5	3	3	3	3
RKS151	United States	Equine	Tm 17	3	10	4	3	3	3	2

[a] Based on profile of alleles at 23 chromosomal enzyme loci assayed by MLEE (2).
[b] IDH, Isocitrate dehydrogenase; AP2, acid phosphatase-2; CAT, catalase; HEX, hexokinase; G6P, glucose-6-phosphate dehydrogenase; PGM, phophoglucomutase; and GLU, glutamate dehydrogenase. Sixteen other enzyme loci were each monomorphic (2).

was from Sigma Chemical Co. (St. Louis, Mo.). Restriction enzymes were purchased from Boehringer Mannheim Biochemicals (Indianapolis, Ind.), New England BioLabs, and Pharmacia LKB Biotechnology and used according to the instructions of the manufacturers, except for the addition of spermidine at 4 mM in all restriction digestions. All other chemicals were molecular biology grade or the highest grade available and were purchased from Sigma, except dextran sulfate, which was obtained from Pharmacia LKB Biotechnology, and deoxynucleotides, which were supplied by Boehringer Mannheim Biochemicals.

Bacterial strains and culture method. The strains of *S. typhimurium* studied are listed in Table 1, and the genetic relationships among them, estimated by an analysis of electrophoretically demonstrable allelic variation in 23 chromosomal genes encoding metabolic enzymes (2), are shown in Fig. 1.

Strains were grown in Luria-Bertani (LB) medium and stored as frozen stocks at −70°C in the presence of 7% dimethyl sulfoxide (1).

Nomenclature of flagellin genes. We used the new unified nomenclature for the flagellar genes of *S. typhimurium* and *E. coli* recently proposed by Iino et al. (9). The *fliC* gene of *Salmonella* serovars was previously known as H1 and the *fljB* gene as H2 (30). It has been suggested that these two genes arose via a duplication (37) and therefore have a similar sequence. To distinguish different antigenic types of flagellin, antigen designations of the Kauffmann-White serotyping scheme of classification (20) are indicated in parentheses. Thus, for example, *fliC(i)* refers to the H1 gene coding for the *i* antigen type of phase 1 flagellin.

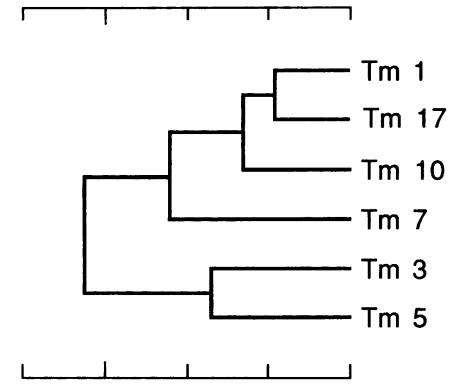

FIG. 1. Genetic relationships among six ETs of *S. typhimurium* estimated from allele profiles at 23 chromosomal metabolic enzyme loci obtained by MLEE analysis of 340 isolates from natural populations (2). Genetic distance between pairs of ETs is expressed as the proportion of enzyme loci at which dissimilar alleles occur (mismatches), and the dendrogram was generated from a matrix of pairwise genetic distances by the average-linkage method of clustering.

Chromosomal amplification by PCR. Chromosomal DNA was prepared from cultures grown overnight in LB medium by the method given by Silhavy et al. (34), with incorporation of the acetone wash recommended by Heath et al. (6). PCR amplification was done in a final volume of 100 μl containing 100 ng of chromosomal DNA and overlaid with 100 μl of mineral oil. The amplification buffer was 16.6 mM $(NH_4)_2SO_4$, 67 mM Tris hydrochloride (pH 8.8), 6.7 mM $MgCl_2$, 200 μg of bovine serum albumin per ml, 200 μM each dATP, dCTP, dGTP, dTTP, and 5% dimethyl sulfoxide. Primers were added to a final concentration of 1 μM. Samples were denatured at 95°C for 5 min before the addition of 2 U of Taq polymerase and β-mercaptoethanol to 10 mM. Amplification was done for 20 cycles on a DNA thermal cycler (The Perkin-Elmer Corp., Norwalk, Conn.). The optimal cycle was found to consist of a denaturation step at 90°C for 1 min, a primer annealing step at 55°C for 2 min, and an extension step at 70°C for 5 min. An autosequence extension of 15 s that permits the Taq polymerase to finish off each round of amplification was incorporated into the extension step, resulting in a cumulative increase of 15 s for the extension step of each cycle. After amplification, 10 μl of the sample was electrophoresed on a 1% NuSieve agarose gel; and by comparison with DNA of known concentration, we estimated that, on average, amplification yielded 1 μg of target sequence. Amplification samples were extracted with chloroform to assist the removal of mineral oil and were ethanol precipitated. As a control for contamination, we always included an amplification reaction containing all the elements mentioned above except chromosomal DNA.

Cloning and sequencing of PCR-generated products. The products of PCR amplification were sequentially digested with *Pst*I and *Cla*I and ligated to phagemid vectors (pBluescript) that had been digested with the same restriction enzymes. To facilitate sequencing in both directions, phagemids with inverted polylinker sequences (SK- and KS-type pBluescript) were used in separate ligations. The ligation products were transformed to *E. coli* JM109 which had been prepared by the method of Chung and Miller (4). Cells from the transformation experiment were plated on LB plates containing 100 μg of ampicillin per ml. Isopropyl-β-D-thiogalactopyranoside (5 mM) and 40 μg of 5-bromo-

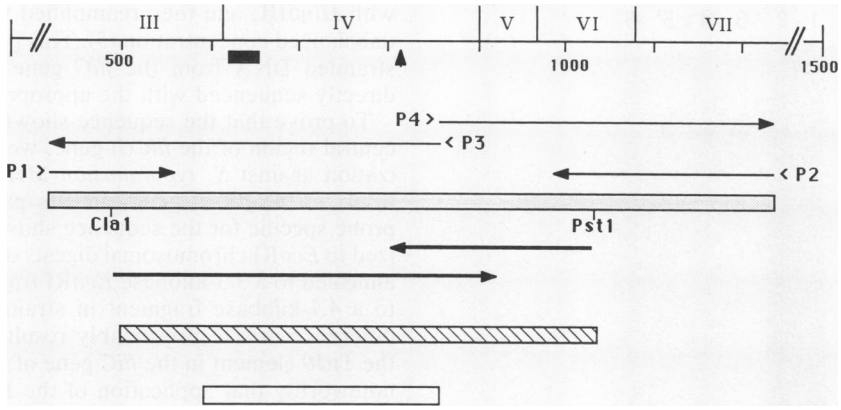

FIG. 2. Schematic representation of the central region of the *fliC(i)* gene of *S. typhimurium* and the sequencing strategy for strain RKS284, representing Tm 1. The upper part of the diagram indicates the length of the entire coding region of the *fliC(i)* gene sequenced by Joys (11) and identifies regions III to IV and the coordinates used by Wei and Joys (39). Region IV is the segment of the gene that contains the sequence of the major epitope of the *i* antigen (filled box) and shows less than 33% amino acid homology with other sequenced flagellins. The position of the recombination event hypothesized to have occurred in the course of the transduction of the *fliC(i)* gene by McDonough (25) is marked by an arrowhead. The boxes in the lower part of the diagram indicate regions sequenced in the present study for *S. typhimurium* ETs Tm 1 (stippled) and Tm 5 (hatched) and the direct sequencing of five other strains (open). Arrows above and below the stippled box show direct sequencing and cloned sequencing reactions for Tm 1 (six independent clones, three sequenced from each end). The positions and sequencing directions of primers 1, 2, 3, and 4 and the *Pst*I and *Cla*I sites used in cloning are shown.

4-chloro-3-indoyl-β-galactoside (X-Gal) per ml were included to detect cloned sequences that interfere with β-galactosidase expression from these vectors. White colonies were selected and analyzed for plasmid insert size by the rapid plasmid preparation technique of Kado and Liu (15). Single-stranded DNA was generated from these clones with Stratagene rescue phage R408 according to the procedure recommended by the manufacturer and sequenced with a Sequenase kit. The frequency of PCR-induced misincorporation errors (detected by sequencing a cloned sequence) was 0.1% (five mutations in 4,215 base pairs [bp] sequenced), which is roughly the same as a previously reported rate for bacterial PCR results (0.07%) (33). The sequencing strategy for *S. typhimurium* RKS284, representing Tm 1, is shown in Fig. 2.

Direct sequencing of PCR products. PCR amplification products were electrophoresed on 1% NuSieve agarose gels, and the amplified DNA was excised and extracted with Geneclean. Extracted DNA was digested with *Hin*dIII (to eliminate the product of *fljB* gene amplification), and approximately 50 ng of the digested product was subjected to another 20 cycles of amplification with unbalanced concentrations of primers (5). In sequencing with primer 1 and primer 4 (Fig. 2), the PCR primers were added to a final concentration of 10 nM primer 1 and 1 μM primer 2. To sequence with primer 2 and primer 3, the primers were added in the reverse concentration (Fig. 2). After amplification, the unwanted primers and nucleotides were removed by double ethanol precipitation and one-fourth of the sample was sequenced with the Sequenase kit and primer 1, 2, 3, or 4, which had been end labeled by standard procedures (1).

Southern analysis. Chromosomal DNA was prepared as described above, and 0.5 μg was digested with *Eco*RI, electrophoresed on a 0.8% agarose gel, capillary transferred to nylon membranes (Hybond-N), hybridized, and probed by standard methods (23, 35).

Primers and probes. Oligonucleotide primers for PCR and direct sequencing were synthesized by betacyanoethyl-1-phosphoramidite chemistry and solid-phase synthesis with a Biosearch 7500 DNA synthesizer. Primer 1 is 5′-AAAG TCCTGGCGCAGGACAACAC-3′, primer 2 is 5′-ATAGC GGAGTTGAAACGGTTCTG-3′, primer 3 is 5′-GCTGC TGTCAATTGCGGCTTT-3′, and primer 4 is 5′-GCAAATG CTGATTTGACAG-3′.

Probes for Southern analysis were obtained by restriction enzyme digestion of plasmid DNA purified with a Circleprep kit. The probes were separated by electrophoresis in NuSieve agarose gels, lightly stained with ethidium bromide, and excised while visualized under UV light. Probes were extracted from the agarose with Geneclean and nick translated by standard protocols (1). The *fliC(i)* probe used in the Southern analysis was a 157-bp *Ssp*I to *Rsa*I DNA fragment. The *fljB(1,2)* probe was a 550-bp *Acc*I-*Hin*dIII fragment, and the *hin* probe was a 745-bp *Cla*I-*Hin*dIII fragment obtained from plasmid pES201 (3), which was kindly supplied by K. Hughes and M. I. Simon.

RESULTS

PCR amplification of central region of *fliC* gene. The PCR technique was used to generate 900-bp sequences from the central, most variable region of the *fliC* (phase 1) flagellin gene (Fig. 1). Primers were chosen from the conserved N-terminal (primer 1) and C-terminal (primer 2) regions, based on the published sequences of five flagellin genes (11, 39, 40). The published sequence available for the first 60 bp of the *fljB* gene (37) suggested that this locus is so divergent from *fliC* that our primers would not anneal to it. However, this assumption proved to be incorrect (see below).

Chromosomal DNA was prepared for PCR amplification from strain RKS284, which represents Tm 1, the most commonly occurring multilocus enzyme genotype (electrophoretic type [ET]) in natural populations of *S. typhimurium*, and is indistinguishable from the laboratory strain LT2 by MLEE analysis of 24 enzymes (2; unpublished data). After 20 cycles of PCR amplification, a DNA fragment of the expected size was observed by agarose gel electrophoresis. However, restriction enzyme digestion of this fragment with *Pst*I revealed that two different sequences had been amplified in apparently equimolar amounts. The major band

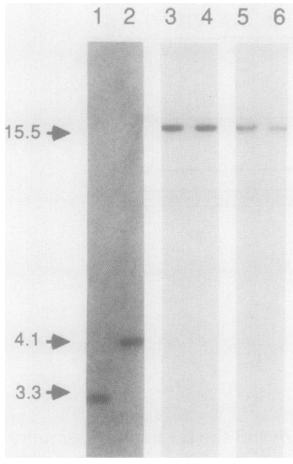

FIG. 3. Southern hybridization of *Eco*RI-digested chromosomal DNA from *S. typhimurium* LT2 (lanes 1, 3, and 5) and MH111 (*fliC*::Tn*10*; lanes 2, 4, and 6). Lanes 1 and 2 were hybridized with a *fliC*(*i*)-specific probe (250 to 407 bp; Fig. 4); lanes 3 and 4 were hybridized with a *fljB*(*1,2*)-specific probe; and lanes 5 and 6 were hybridized with a *hin*-specific probe. The size of each band in kilobases estimated by comparison with molecular weight markers of known size is indicated.

obtained from one of these sequences by *Pst*I digestion was 860 bp, while the major band of the other sequence was 651 bp. Further digestion with *Cla*I reduced the size of both of these bands by 67 bp to 793 bp (large *Cla*I-*Pst*I fragment) and 584 bp (small *Cla*I-*Pst*I fragment). The restriction enzyme pattern for this region of the published sequence of the *fliC*(*i*) gene is compatible with the small *Cla*I-*Pst*I fragment (11). We demonstrated that the large *Cla*I-*Pst*I fragment was derived from the *fljB*(*1,2*) gene (see below).

The two sequences amplified by PCR were isolated by cloning *Cla*I-*Pst*I restriction digestions of each fragment into phagemid vectors (pBluescript), and transformants from this ligation were then screened by agarose gel electrophoresis and sorted into large and small clones equivalent to the large and small bands observed after *Cla*I-*Pst*I digestion of the PCR amplification products.

Sequencing PCR products. Clones of the large *Cla*I-*Pst*I fragment were sequenced and found to have a sequence generally similar to previously published flagellin sequences (11, 39, 40) but quite distinct from the sequence reported for the *fliC*(*i*) gene (11). This sequence was identified as that of the *fljB*(*1,2*) (phase 2 flagellin) gene by Southern hybridization of a probe specific to the large *Cla*I-*Pst*I fragment to the same 15.5-kilobase *Eco*RI chromosomal fragment to which a probe from the *hin* gene also annealed (Fig. 3). The *hin* gene is immediately adjacent to the *fljB* gene of *S. typhimurium*, which is located more than 16 min from the *fliC* locus (30). [The sequence of the *fljB*(*1,2*) gene will be reported elsewhere.]

The sequence of the small *Cla*I-*Pst*I fragment is given in Fig. 4 (bp 41 to 625). Three independent clones were sequenced from each end; and from the overlap region of this sequencing, new primers were designed (primers 3 and 4) and used to sequence single-stranded DNA generated from PCR-amplified DNA by the unbalanced primer method (5). To obtain the sequence from each PCR primer to the *Pst*I and *Cla*I sites used for cloning, a *Hin*dIII site identified in the *fljB* gene PCR product was used to exclude the *fljB* gene PCR product from the unbalanced primer reaction. The products of a PCR amplification were gel purified, digested with *Hin*dIII, and then reamplified with primers 1 and 2 in unbalanced concentration (5). This procedure yielded single-stranded DNA from the *fliC* gene only, which was then directly sequenced with the appropriate PCR primer.

To prove that the sequence shown in Fig. 4 is indeed the central region of the *fliC*(*i*) gene, we did a Southern hybridization against *S. typhimurium* MH111, which has a Tn*10* insert in the *fliC*(*i*) gene, and its parent strain LT2 (8). A probe specific for the sequence shown in Fig. 4 was hybridized to *Eco*RI chromosomal digests of each strain. The probe annealed to a 3.3-kilobase *Eco*RI fragment in strain LT2 but to a 4.1-kilobase fragment in strain MH111 (Fig. 3). This difference in size presumably results from the presence of the Tn*10* element in the *fliC* gene of strain MH111. It is also noteworthy that application of the PCR technique to chromosomal DNA from strain MH111 resulted in amplification of the *fljB* gene only, presumably because the 9.3-kilobase Tn*10* insert in the *fliC* gene placed the PCR primers too far apart for effective amplification.

Sequencing the *fliC*(*i*) genes of diverse strains. The sequence of the first 348 bp of the *fliC*(*i*) gene of strain RKS284 shown in Fig. 4 is identical to the previously published sequence for strain LT2 (11), but the promoter-distal sequence is distinctly different. We initially attributed this difference to our having sequenced the *fliC* gene from a strain of *S. typhimurium* other than LT2. To further examine interstrain sequence variation, we repeated the amplification, restriction digestion, and cloning procedure with chromosomal DNA prepared from a strain of Tm 5, which is the multilocus enzyme genotype of the *S. typhimurium* most distantly related to that of LT2 (Tm 1, Fig. 1). Three independent clones of the small *Cla*I-*Pst*I fragment of the *fliC*(*i*) gene of a strain of Tm 5 were sequenced for their entire length, two on one strand and one on the other strand, with the aid primers 3 and 4. Surprisingly, the sequence of the 584-bp *Cla*I-*Pst*I fragment of the *fliC*(*i*) gene of Tm 5 was found to be identical to that of Tm 1 (Fig. 4).

That the sequences of the central region of the *fliC*(*i*) gene in these two genetically divergent strains of *S. typhimurium* are identical suggested that this region is not evolving rapidly and that the previously published sequence of the *fliC*(*i*) gene is incorrect. To test these hypotheses, we directly sequenced a region of the *fliC*(*i*) gene in LT2 and four other strains of *S. typhimurium* that represent the total span of genotypic diversity detected among 340 isolates of this serovar by MLEE (Fig. 1). Primers 1 and 2 were used to amplify the central region of the *fliC* gene, and single-stranded DNA was generated from *Hin*dIII-digested amplification products by the unbalanced primer method. The single-stranded DNA derived from the *fliC* sequence was directly sequenced with primer 3. The region sequenced is the most variable part of *fliC*(*i*) among the allelic forms of the gene occurring in different serovars (39). We obtained 260 bp of sequence commencing 60 bp from primer 3 and extending toward primer 1 (124 to 384 bp, Fig. 4) from strains of ETs Tm 3, Tm 8, Tm 10, Tm 17, and LT2 (representing Tm 1). No variation in the sequence of the 260-bp region was observed among the five strains of *S. typhimurium*.

DISCUSSION

Using a combination of cloning and direct sequencing of PCR-generated DNA, we sequenced 852 bp from the central region of the *fliC*(*i*) gene of *S. typhimurium* RKS284, representing ET Tm 1. For the first 348 bp (except for a G-to-C change at 211), this sequence is the same as that previously

```
                                        27                  ClaI            54
ACC ATC CAG GTT GGT GCC AAC GAC GGT GAA ACT ATC GAT ATC GAT CTG AAG CAG
Thr Ile Gln Val Gly Ala Asn Asp Gly Glu Thr Ile Asp Ile Asp Leu Lys Gln
                                        81                              108
ATC AAC TCT CAG ACC CTG GGT CTG GAT ACG CTG AAT GTG CAA CAA AAA TAT AAG
Ile Asn Ser Gln Thr Leu Gly Leu Asp Thr Leu Asn Val Gln Gln Lys Tyr Lys
                                        135                             162
GTC AGC GAT ACG GCT GCA ACT GTT ACA GGA TAT GCC GAT ACT ACG ATT GCT TTA
Val Ser Asp Thr Ala Ala Thr Val Thr Gly Tyr Ala Asp Thr Thr Ile Ala Leu
                                        189                             216
GAC AAT AGT ACT TTT AAA GCC TCG GCT ACT GGT CTT GGT GGT ACT GAC CAG AAA
Asp Asn Ser Thr Phe Lys Ala Ser Ala Thr Gly Leu Gly Gly Thr Asp Gln Lys
                                        243                             270
ATT GAT GGC GAT TTA AAA TTT GAT GAT ACG ACT GGA AAA TAT TAC GCC AAA GTT
Ile Asp Gly Asp Leu Lys Phe Asp Asp Thr Thr Gly Lys Tyr Tyr Ala Lys Val
                                        297                             324
ACC GTT ACG GGG GGA ACT GGT AAA GAT GGC TAT TAT GAA GTT TCC GTT GAT AAG
Thr Val Thr Gly Gly Thr Gly Lys Asp Gly Tyr Tyr Glu Val Ser Val Asp Lys
                                        351                             378
ACG AAC GGT GAG GTG ACT CTT GCT GGC GGT GCG ACT TCC CCG CTT ACA GGT GGA
Thr Asn Gly Glu Val Thr Leu Ala Gly Gly Ala Thr Ser Pro Leu Thr Gly Gly
                                        405                 Primer 4    432
CTA CCT GCG ACA GCA ACT GAG GAT GTG AAA AAT GTA CAA GTT GCA AAT GCT GAT
Leu Pro Ala Thr Ala Thr Glu Asp Val Lys Asn Val Gln Val Ala Asn Ala Asp
                Primer 3        459                             486
TTG ACA GAG GCT AAA GCC GCA TTG ACA GCA GCA GGT GTT ACC GGC ACA GCA TCT
Leu Thr Glu Ala Lys Ala Ala Leu Thr Ala Ala Gly Val Thr Gly Thr Ala Ser
                                        513                             540
GTT GTT AAG ATG TCT TAT ACT GAT AAT AAC GGT AAA ACT ATT GAT GGT GGT TTA
Val Val Lys MET Ser Tyr Thr Asp Asn Asn Gly Lys Thr Ile Asp Gly Gly Leu
                                        567                             594
GCA GTT AAG GTA GGC GAT GAT TAC TAT TCT GCA ACT CAA AAT AAA GAT GGT TCC
Ala Val Lys Val Gly Asp Asp Tyr Tyr Ser Ala Thr Gln Asn Lys Asp Gly Ser
                                        621 PstI                        648
ATA AGT ATT AAT ACT ACG AAA TAC ACT GCA GAT GAC GGT ACA TCC AAA ACT GCA
Ile Ser Ile Asn Thr Thr Lys Tyr Thr Ala Asp Asp Gly Thr Ser Lys Thr Ala
                                        675                             702
CTA AAC AAA CTG GGT GGC GCA GAC GGC AAA ACC GAA GTT GTT TCT ATT GGT GGT
Leu Asn Lys Leu Gly Gly Ala Asp Gly Lys Thr Glu Val Val Ser Ile Gly Gly
                                        729                             756
AAA ACT TAC GCT GCA AGT AAA GCC GAA GGT CAC AAC TTT AAA GCA CAG CCT GAT
Lys Thr Tyr Ala Ala Ser Lys Ala Glu Gly His Asn Phe Lys Ala Gln Pro Asp
                                        783                             810
CTG GCG GAA GCG GCT GCT ACA ACC ACC GAA AAC CCG CTG CAG AAA ATT GAT GCT
Leu Ala Glu Ala Ala Ala Thr Thr Thr Glu Asn Pro Leu Gln Lys Ile Asp Ala
                                        837
GCT TTG GCA CAG GTT GAC ACG TTA CGT TCT GAC CTG GGT GCG
Ala Leu Ala Gln Val Asp Thr Leu Arg Ser Asp Leu Gly Ala
```

FIG. 4. Sequence of the central region of the *fliC(i)* gene of *S. typhimurium* RKS284. The sequence commences and ends 4 bp from the PCR primers used to generate it. The positions of primer 4 and the complement of primer 3 are underlined, and the *Cla*I and *Pst*I restriction sites used in cloning are marked. The region sequenced in all *S. typhimurium* strains is shown in bold face. Base 1 corresponds to base 445 of the previously published sequence of this region (11). The sequence shown corresponds to the previously identified regions III, IV, V, VI, and VII of flagellin genes (39).

reported for the *fliC(i)* gene of *S. typhimurium* LT2 (also representing Tm 1) and includes the sequence coding for the pentapeptide that is the major antigenic epitope of the *i* antigen (10). But then the two sequences abruptly diverge for the remaining stretch of 93 bases (349 to 442 bp) of the central region; and the promoter-distal 411 bp of the two sequences show 29 bp differences, a degree of divergence comparable to that reported for this region of the *fliC* gene between strains of different serovars of *Salmonella* (39). Comparative sequencing of 584 bp from the same region of a genetically distant strain of *S. typhimurium* (Tm 5) and 260 bp for each of five other genotypically distinct strains of *S. typhimurium* (including LT2) revealed that the sequence of the central region of the *fliC(i)* gene is invariant in all strains. Our findings indicate that the earlier reported sequence of the *fliC(i)* gene of *S. typhimurium* is partially incorrect, and we believe that this error can be traced to the procedure used in cloning the gene.

The initial investigations of the structure of flagellin proteins were based on an analysis of amino acid composition. To facilitate these studies, McDonough (25) transduced the *fliC(i)* gene from *S. typhimurium* LT2 to a nonflagellated strain of *S. paratyphi B*. It was this transductant strain (designated SL877) from which a *fliC* gene, supposedly that of *S. typhimurium* LT2, subsequently was cloned and sequenced. But if recombination had occurred between the resident chromosomal *S. paratyphi B fliC(b)* locus and the transduced *fliC(i)* gene of *S. typhimurium*, a chimeric *fliC* locus containing the sequences for the major determinant of the *i* antigen (10) would be present on the chromosome of strain SL877. Intragenic recombination of this type, mediated by transduction, has been demonstrated for other flagellin genes (41), including mutant *fliC* genes in strains of *S. typhimurium* (14).

We suggest that it was a chimeric *fliC(i)/fliC(b)* locus in strain SL877 that was then cloned, sequenced, and mistakenly reported as the *fliC(i)* gene of *S. typhimurium* (11). Moreover, an earlier study reporting multiple amino acid differences between phase 1 flagellins of two isolates of *S. typhimurium* LT2 involved a comparison of strain SL877 (the putative hybrid) and a second *S. paratyphi B* strain (SL167) that had also been transduced with the *fliC(i)* gene of *S. typhimurium* LT2 (12, 25). On the assumption that strain SL167 is a clean transductant, we conclude that the amino acid differences recorded in the comparison of these two strains (12) reflect differences between the promoter-distal *S. paratyphi B* region of the hybrid protein of strain SL877 and the actual *S. typhimurium* phase 1 flagellin of strain SL167.

It is noteworthy that the coding sequences of all the novel tryptic peptides reported for the flagellin of strain SL167 occur in the promoter-distal region of the sequence shown in Fig. 4 (678 to 745 bp).

The methods later used by Wei and Joys (39, 40) to sequence the *fliC(a)*, *fliC(c)*, *fliC(d)*, and *fliC(r)* genes of the serovars *S. paratyphi A*, *S. cholerae-suis*, *S. muenchen*, and *S. rubislaw*, respectively, avoided the possibility of recombination by cloning directly from a strain of each serovar.

Our findings emphasize a need for caution in interpreting experimental results based on the transduction of flagellin genes to new genetic backgrounds (13, 19, 25). However, we believe that our correction of the sequence of the *fliC(i)* gene does not alter the basic conclusions reached by other workers who used the earlier reported sequence for comparative purposes (18, 24, 38, 39), because these conclusions were based on the sequences at the ends of the gene, which apparently are highly conserved in all *Salmonella* serovars (39).

It has been suggested that the extensive antigenic diversity displayed by the flagellins of *Salmonella* serovars is generated by random mutation of a weakly constrained coding sequence (39). However, the data presented here suggest that the sequence of the antigen-coding central region of the phase 1 flagellin gene is, in fact, subject to a considerable degree of selective constraint. As in all *Salmonella* flagellin genes (25), this region contains no cysteine or tryptophan residues and only a few histidine residues, whereas threonine, alanine, and glycine constitute almost 40% of the total amino acids. Further evidence of constraint on the sequence of the antigen-coding region of the flagellin genes is provided by the observation that certain in-frame inserts in the antigen-coding region of the *fliC(d)* gene interfered with flagellar function (28). And the constraint hypothesis is also supported by the codon adaptation index (31, 32) of the central region, as it measures codon bias. The sequence shown in Fig. 4 has a moderate codon adaptation index of 0.41. An association between a high codon adaptation index and a low rate of sequence evolution has been demonstrated for genes that are highly expressed, presumably because of selective constraints on substitutions at synonymous sites (31). A moderate codon adaptation index for the *fliC* gene is not unexpected, because this locus is highly expressed, involving expenditure of approximately 2% of the biosynthetic energy of the cell (22).

Our demonstration that the sequence of the phase 1 flagellin gene is invariant among genotypically distinct strains of *S. typhimurium* recovered from diverse hosts and in several geographic regions is compatible with the concept that *S. typhimurium* is a complex of cell lineages that has evolved relatively recently from a common ancestor (2). Evidence of a similar clonal structure has been reported from comparative sequencing of parts of the *trp* operon in groups of strains of *E. coli* that are as closely related (as indexed by MLEE) as the *S. typhimurium* strains described here (26, 27, 36). Hence, we conclude that the antigen-coding region of the *fliC(i)* gene in *S. typhimurium* is evolving by sequence drift at a rate no faster than parts of the *trp* genes of *E. coli*.

MLEE analysis of strains of eight serovars of *Salmonella* by Beltran et al. (2) demonstrated that certain phase 1 antigens are expressed in common by strains that are distantly related in overall chromosomal genotype. Conversely, in other cases, strains expressing different phase 1 antigens were nearly identical in multilocus enzyme genotype. These findings were interpreted as evidence that recombination following horizontal transfer of flagellar antigen genes is a significant process in the evolution of *Salmonella* serovars. If in the salmonellae as a whole, as we have shown here for diverse strains of *S. typhimurium*, sequence drift occurs at a relatively slow rate, it is unlikely that this process alone has generated the antigenic differences between strains that are closely related in overall chromosomal genotype (e.g., those of *S. typhimurium* and *S. heidelberg*) (2). It is, therefore, probable that lateral transfer and recombination of flagellin sequences is a major evolutionary mechanism generating new serovars in *Salmonella*. We will address this possibility in future studies.

ACKNOWLEDGMENTS

We thank Pilar Beltran for information on the multilocus enzyme genotypes of strains and for assistance in many other ways. We are indebted to Christine M. Doyle for technical assistance and to Terence M. Joys for much helpful discussion.

This research was supported by Public Health Service grant AI-22144 from the National Institute of Allergy and Infectious Diseases (to R.K.S.). N.H.S. was supported by a postdoctoral fellowship in molecular studies of evolution (88-3-8 ME) from the Alfred P. Sloan Foundation.

LITERATURE CITED

1. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl (ed.).** 1987. Current protocols in molecular biology. John Wiley & Sons, Inc., New York.
2. **Beltran, P., J. M. Musser, R. Helmuth, J. J. Farmer III, W. M. Frerichs, I. K. Wachsmuth, K. Ferris, A. C. McWhorter, J. G. Wells, A. Cravioto, and R. K. Selander.** 1988. Toward a population genetic analysis of *Salmonella*: genetic diversity and relationships among strains of serotypes *S. cholerae-suis*, *S. derby*, *S. dublin*, *S. enteritidis*, *S. heidelberg*, *S. infantis*, *S. newport* and *S. typhimurium*. Proc. Natl. Acad. Sci. USA **85**:7753–7757.
3. **Bruist, M. F., and M. I. Simon.** 1984. Phase variation and the Hin protein: in vivo activity measurements, overproduction, and purification. J. Bacteriol. **159**:71–79.
4. **Chung, C. T., and R. H. Miller.** 1988. A rapid and convenient method for the preparation and storage of competent bacterial cells. Nucleic Acids Res. **16**:3580.
5. **Gyllensten, U. B., and H. A. Erlich.** 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA DQA locus. Proc. Natl. Acad. Sci. USA **85**:7652–7656.
6. **Heath, L. S., G. L. Sloan, and H. E. Heath.** 1986. A simple and generally applicable method for releasing DNA from bacterial cells. Appl. Environ. Microbiol. **51**:1138–1140.
7. **Homma, M., H. Fujita, S. Yamaguchi, and T. Iino.** 1987. Regions of *Salmonella typhimurium* flagellin essential for its polymerization and excretion. J. Bacteriol. **169**:291–296.
8. **Homma, M., and T. Iino.** 1985. Locations of hook-associated proteins in flagellar structures of *Salmonella typhimurium*. J. Bacteriol. **162**:183–189.
9. **Iino, T., Y. Komeda, K. Kutsukake, R. M. Macnab, P. Matsumura, J. S. Parkinson, M. I. Simon, and S. Yamaguchi.** 1988. New unified nomenclature for the flagellar genes of *Escherichia coli* and *Salmonella typhimurium*. Microbiol. Rev. **52**:533–535.
10. **Joys, T. M.** 1976. Identification of an antibody binding site in the phase-1 flagellar protein of *Salmonella typhimurium*. Microbios **15**:221–228.
11. **Joys, T. M.** 1985. The covalent structure of the phase-1 flagellar filament protein of *Salmonella typhimurium* and its comparison with other flagellins. J. Biol. Chem. **260**:15758–15761.
12. **Joys, T. M., J. F. Martin, H. L. Wilson, and V. Rankis.** 1974. Differences in the primary structure of the phase-1 flagellins of two strains of *Salmonella typhimurium*. Biochim. Biophys. Acta **351**:301–305.
13. **Joys, T. M., and V. Rankis.** 1972. The primary structure of the phase-1 flagellar protein of *Salmonella typhimurium*. J. Biol.

Chem. **247**:5180–5193.

14. **Joys, T. M., and B. A. D. Stocker.** 1963. Mutation and recombination of flagellar antigen *i* of *Salmonella typhimurium*. Nature (London) **197**:413–414.
15. **Kado, C. I., and S.-T. Liu.** 1971. Rapid procedure for detection and isolation of large and small plasmids. J. Bacteriol. **145**: 1365–1373.
16. **Kauffmann, F.** 1964. Das Kauffmann-White-schema, p. 21–66. *In* E. van Oye (ed.), The world problem of salmonellosis. Dr. W. Junk Publishers, The Hague.
17. **Kuwajima, G.** 1988. Construction of a minimum-size functional flagellin of *Escherichia coli*. J. Bacteriol. **170**:3305–3309.
18. **Kuwajima, G., J.-I. Asaka, T. Fujiwara, T. Fujiwara, K. Node, and E. Kondoh.** 1986. Nucleotide sequence of the *hag* gene encoding flagellin of *Escherichia coli*. J. Bacteriol. **168**:1479–1483.
19. **Langman, R. E.** 1972. The occurrence of antigenic determinants common to flagella of different *Salmonella* strains. Eur. J. Immunol. **2**:582–586.
20. **Le Minor, L.** 1984. *Salmonella* Lignieres 1900, 389, p. 427–458. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore.
21. **Le Minor, L., and M. Y. Popoff.** 1987. Antigenic formulas of the *Salmonella* serovars. WHO Collaborating Centre for Reference and Research on *Salmonella*. Institut Pasteur, Paris.
22. **Macnab, R. M.** 1987. Motility and chemotaxis, p. 732–759. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 1. American Society for Microbiology, Washington, D.C.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **Martin, J. H., and D. C. Savage.** 1988. Cloning, nucleotide sequence, and taxonomic implications of the flagellin gene of *Roseburia cecicola*. J. Bacteriol. **170**:2612–2617.
25. **McDonough, M. W.** 1965. Amino acid composition of antigenically distinct *Salmonella* flagellar proteins. J. Mol. Biol. **12**: 342–355.
26. **Milkman, R. D., and I. P. Crawford.** 1983. Clustered third-base substitutions among wild strains of *E. coli*. Science **221**:378–380.
27. **Milkman, R., and A. Stoltzfus.** 1988. Molecular evolution of the *Escherichia coli* chromosome. II. Clonal segments. Genetics **120**:359–366.
28. **Newton, S. M. C., C. O. Jacob, and B. A. D. Stocker.** 1989. Immune response to cholera toxin epitope inserted in *Salmonella* flagellin. Science **244**:70–72.
29. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Erlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239**:487–491.
30. **Sanderson, K. E., and J. A. Hurley.** 1987. Linkage map of *Salmonella typhimurium*, p. 877–918. *In* F. C. Neidhardt, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology, vol. 2. American Society for Microbiology, Washington, D.C.
31. **Sharp, P. M., and W.-H. Li.** 1987. The rate of synonymous substitution in enterobacterial genes is inversely related to codon-bias. Mol. Biol. Evol. **4**:222–230.
32. **Sharp, P. M., and W.-H. Li.** 1987. The codon adaptation index—a measure of directional codon usage bias, and its potential applications. Nucleic Acids Res. **15**:1281–1295.
33. **Shyamala, V., and G. F.-L. Ames.** 1989. Amplification of bacterial genomic DNA by the polymerase chain reaction and direct sequencing after asymmetric amplification: application to the study of periplasmic permeases. J. Bacteriol. **171**:1602–1608.
34. **Silhavy, T. J., M. L. Berman, and L. W. Enquist.** 1984. Experiments with gene fusions. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
36. **Stoltzfus, A., J. F. Leslie, and R. Milkman.** 1988. Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between *trp* and *tonB*. Genetics **120**:345–358.
37. **Szekely, E., and M. Simon.** 1983. DNA sequence adjacent to flagellar genes and evolution of flagellar-phase variation. J. Bacteriol. **155**:74–81.
38. **Trachtenberg, S., and D. J. DeRosier.** 1988. Three-dimensional reconstruction of the flagellar filament of *Caulobacter crescentus*. J. Mol. Biol. **202**:787–808.
39. **Wei, L.-N., and T. M. Joys.** 1985. Covalent structure of three phase-1 flagellar filament proteins of *Salmonella*. J. Mol. Biol. **186**:791–803.
40. **Wei, L.-N., and T. M. Joys.** 1986. The nucleotide sequence of the H-1^r gene of *Salmonella rubislaw*. Nucleic Acids Res. **14**:8227.
41. **Yamaguchi, S., and T. Iino.** 1969. Genetic determination of the antigenic specificity of flagellar protein in *Salmonella*. J. Gen. Microbiol. **55**:59–74.