JOURNAL OF BACTERIOLOGY, July 2008, p. 4470–4477 Vol. 190, No. 13
0021-9193/08/$08.00+0 doi:10.1128/JB.01937-07

A Genomic Islet Mediates Flagellar Phase Variation in *Escherichia coli* Strains Carrying the Flagellin-Specifying Locus *flk*▽†

Lu Feng,[1,3,5]* Bin Liu,[1,2,4] Yanqun Liu,[1,3] Yuli A. Ratiner,[6] Bo Hu,[1,3] Dan Li,[1,3] Xiaolin Zong,[1,3] Wei Xiong,[1,3] and Lei Wang[1,2,3,4,5]*

TEDA School of Biological Sciences and Biotechnology, Nankai University,[1] Tianjin Key Laboratory of Microbial Functional Genomics,[2] Tianjin Research Center for Functional Genomics and Biochip,[3] The Engineering and Research Center for Microbial Functional Genomics and Detection Technology, Ministry of Education,[4] and The Key Laboratory of Molecular Microbiology and Technology, Ministry of Education,[5] 23 Hongda Street, TEDA, Tianjin 300457, Peoples Republic of China, and Department of Microbiology, Mechnikov Research Institute for Vaccines and Sera, The Russian Academy of Medical Sciences, Moscow 105064, Russia[6]

Received 13 December 2007/Accepted 29 January 2008

The occurrence of unilateral flagellar phase variation was previously demonstrated in *Escherichia coli* strains carrying the non-*fliC* flagellin-specifying locus *flk*. In this study, we investigated the mechanism involved in this process. By using sequencing and sequence analysis, the *flk* region between the chromosomal genes *yhaC* and *rnpB* was characterized in all described *flk*-positive *E. coli* strains, including the H35 strain identified in this study (the other strains used are H3, H36, H47, and H53 strains), and this region was found to contain a putative integrase gene and flanking direct repeats in addition to the *flk* flagellin-specifying gene *flkA* and a *fliC* repressor gene, *flkB*, indicating that there is a typical genomic islet (GI), which was designated the *flk* GI. The horizontal transfer potential of the *flk* GI was indicated by detection of the excised extrachromosomal circular form of the *flk* GI. By generating *fliC*-expressing variants of H3 and H47 strains, unilateral flagellar phase variation in *flk*-positive strains was shown to be mediated by excision of the *flk* GI. The function of the proposed integrase gene was confirmed by deletion and a complementation test. The potential integration sites of the *flk* GI were identified. A general model for flagellar phase variation in *flk*-positive *E. coli* strains can be expressed as $fliC^{off} + flkA^{on} \rightarrow fliC^{on} + flkA^{none}$. This is the first time that a molecular mechanism for flagellar phase variation has been reported for *E. coli*.

Flagella are best known for conferring motility to bacteria, which allows the bacteria to swim toward attractants and away from repellents (5). Flagella also play a variety of other roles in many bacterial lifestyles, including bacterial pathogenesis and biofilm formation (27, 45). Pathogenic bacteria specifically produce flagella to promote colonization and invasion of mucosa (27). Flagellar antigen, also known as H antigen, is one of the major antigens in gram-negative bacteria. The serological variety of flagella is important for intrageneric differentiation of bacteria (12). Flagellin is the protein subunit of the flagellar filament and determines the specificity of the flagellar antigen.

Phase variation of antigenic expression, especially expression of surface structures, such as flagella, fimbria, capsular polysaccharide, and lipopolysaccharide, is a common strategy used by many bacteria for adaptation to particular environments (38). Phase variation of flagellar antigens was first described in *Salmonella enterica* serovar Typhimurium, in which two different flagellin genes, *fljB* and *fliC*, are alternatively expressed, giving rise to two different H phases (18, 50). Flagellar phase variation is thought to be related to bacterial survival in the presence of host defense systems and therefore is linked to bacterial virulence (19). The molecular mechanism for flagellar phase variation in *S. enterica* has been well studied (1, 20, 22, 40, 47, 51). Alternating expression of *fljB* and *fliC* is controlled by site-specific inversion of an approximately 996-bp DNA segment (H segment) containing a promoter for the cotranscription of *fljB* and *fljA* (encoding a posttranscriptional repressor of the unlinked gene *fliC*), and the *fljBA* promoter is turned on in one orientation but not in the other (51). The *hin* gene encoding a recombinase (Hin invertase) and a 26-bp inverted repeat sequence (*hix*) are responsible for the invertible recombination of the H segment (20, 40). Recently, flagellar phase variation caused by deletion of the *fljAB*-like operon in a z66 antigen-positive strain of *S. enterica* serovar Typhi was reported (17), but the exact mechanism involved is unclear.

A total of 53 different H type strains of *Escherichia coli* have been officially registered (26). In 44 of these strains the manifested flagellins are encoded at the *fliC* locus, while in all of the rest of them except the H35 type strain they are encoded at the *flk* (H3, H36, H47 and H53), *fll* (H44 and H55), *flm* (H54), or $fliC^{//}$ (H17) locus (4, 28–33, 43, 44, 46). H35 has a disrupted *fliC* gene (46), and its flagellin-specifying locus has not been determined. Although *E. coli* is generally considered monophasic (24), the occurrence of unilateral flagellar phase variation in *flk*-positive H3 and H47 strains has been reported (28, 32, 33). The *flk* region containing the flagellin-specifying gene *flkA* and the repressor gene *flkB* was located between chromosomal genes *rnpB* and *yhaC* in H3 and H53 strains, and the *flkB* product repressed the expression of *fliC* in

* Corresponding author. Mailing address: TEDA School of Biological Sciences and Biotechnology, Nankai University, 23 Hongda Street, TEDA, Tianjin 300457, Peoples Republic of China. Phone: 86-22-66229588. Fax: 86-22-66229596. E-mail for L. Feng: fenglu63@nankai.edu.cn. E-mail for L. Wang: wanglei@nankai.edu.cn.

† Supplemental material for this article may be found at http://jb.asm.org/.

▽ Published ahead of print on 25 April 2008.

E. coli and, in some cases, *S. enterica*; therefore, the *flkB* gene is functionally similar to *fljA* of *S. enterica* (43). The unilateral phase variation in *E. coli* H3 and H47 was expressed as $fliC^{off} + flkA^{on} \rightarrow fliC^{on} + flkA^{off}$ (33).

Genomic islets or islands (GIs) represent a large group of mobile elements in bacteria (39). GIs are known to encode many different functions and are related to bacterial virulence, antibiotic resistance, symbiosis, fitness, and adaptation (9, 39). The number of described GIs and GI-bearing hosts is constantly increasing (25). Some GIs have features of integrative and conjugative elements, and these elements are excised from the chromosome by site-specific recombination and are transferred to new hosts by conjugation (2, 7, 10). Recombination between two attachment sites, *attP* on the circular form of a GI and *attB* on a bacterial chromosome, that leads to the integration of the GI into the chromosome is mediated by an integrase (14, 15, 39). Integration of a GI generates two junctions in the chromosome, which are two direct repeats (DRs) at the left and right ends of the GI. These two DRs are the chimeras of *attB* and *attP*. Integrases also mediate the recombination between two flanking DRs of an integrated GI, resulting in excision of the GI from the chromosome via an extrachromosomal circular intermediate (10, 14, 16, 23, 39).

In this study, the mechanism involved in flagellar phase variation in *flk*-positive *E. coli* strains was investigated. H35 was identified as a new member of the *flk*-positive group. The regions between *yhaC* and *rnpB* in H35, H36, and H47 strains were sequenced, and the sequences were analyzed together with the corresponding sequences from H3 and H53 strains published previously. In all five *flk*-positive strains, the *flk* region was found to contain a putative integrase gene and flanking DRs at both ends in addition to the flagellin gene *flkA* and the repressor gene *flkB*, indicating that there is a typical GI, which was designated the *flk* GI. PCR was carried out to detect the excised extrachromosomal circular form of the *flk* GI. *fliC*-expressing variants of H3 and H47 were generated and examined to determine the presence of the *flk* GI, and the phase variation in the two parental strains was shown to be mediated by excision of the *flk* GI. The function of Orf486 as an integrase was confirmed by deletion and complementation tests. The frequency of flagellar phase transition was determined. The potential integration sites of the *flk* GI were identified. A general model for the mechanism mediating unilateral flagellar phase variation in *flk*-positive *E. coli* strains is presented below.

MATERIALS AND METHODS

Bacterial strains, media, and antisera. Strains used in this study are listed in Table 1. Flagellin-negative strain EJ34 was used as the host strain for construction of genomic libraries. Bacteria were grown in Luria-Bertani (LB) broth or agar A (BBI) supplemented with ampicillin (100 μg/ml) when necessary. A semisolid medium (motility agar plates) containing 0.2% agar in LB broth was used to enhance the bacterial motility and to screen for motile clones.

Antisera against *E. coli* H type strains were obtained from the Mechnikov Research Institute for Vaccines and Sera, Russian Academy of Medical Sciences, Moscow, Russia, and were used for H antigen identification by means of slide tests (34).

Construction of a genomic library. A genomic library was constructed by the shotgun cloning method (37). Genomic DNA partially digested with Sau3AI was ligated into pUC19, which was digested with BamHI and treated with calf intestinal alkaline phosphatase (Takara). The library was introduced into EJ34 by transformation, and the recombinant bacterial cells were screened for restoration of swimming motility on motility agar plates containing ampicillin.

Primers and PCR amplification. The primers used in this study are listed in Table S1 in the supplemental material. Each PCR was carried out using a 50-μl reaction mixture containing 2 mM $MgCl_2$, each deoxynucleoside triphosphate at a concentration of 0.2 mM, each primer at a concentration of 1 μM, and 2 U of *Taq* DNA polymerase (5 U of LA-*Taq* DNA polymerase for long-range PCR).

Sequence analysis. Sequencing was carried out by Tianjin Biochip Cooperation, China, using an ABI 3730 automated DNA sequencer. Sequence data were assembled using the Staden package (41). Artemis (36) was used for annotation. BLAST was used for searching databases, including the GenBank, COG (Clusters of Orthologous Groups), and Pfam protein motif databases. Sequence alignment and comparison were carried out using ClustalX (42). Phylogenetic analysis was conducted using MEGA, version 2.1 (21).

Selection of spontaneous variants with altered H antigen. Bacteria were grown on semisolid medium containing antiserum against the expressed H antigen, which provided strict conditions for selecting variants expressing an alternative flagellar antigen phase (30, 32). While the parental strain was immobilized, the variants remained motile.

Deletion of *orf486* from the *E. coli* H3 type strain. The *orf486* gene was replaced by a kanamycin resistance gene (*kan*) using the RED recombination system of phage lambda (8, 48). The *kan* gene was PCR amplified from plasmid pKD4 by using primers wl-11719 and wl-11720 binding to the 5′ and 3′ ends of the gene, and each primer contained 40 bp based on the H3 DNA, which flanks *orf486*. The PCR product was transformed into the H3 type strain (G1436) carrying pKD20, and kanamycin-resistant transformants were selected after induction of the RED genes by the protocol described by Datsenko and Wanner (8). PCR with primers specific for the DNA of the *kan* gene and the H3 *flk* region was carried out to confirm the replacement. To complement the *orf486*-deficient H3 mutant, the *orf486* gene was amplified from the H3 type strain using primers wl-10352 and wl-10353. The resulting PCR products were cloned into pUC18 to obtain plasmid pLW1330.

Determination of flagellar phase transition frequency in H3 strains. Bacteria were grown on LB agar for 24 h. A block of agar bearing a single colony was cut out and transferred to 15 ml of LB broth. After 6 h of incubation at 37°C, the culture was transferred into two centrifuge tubes. Fifty microliters of antiserum against the expressed H antigen (H3 antiserum in this study) was added to the first tube, and 50 μl of some other antiserum (H47 antiserum in this study) which did not contain antibodies to any surface antigens of the strain used (H3 in this study) was added to the second tube. The contents of each tube were gently mixed. After incubation at 37°C for 30 min, a small sample was taken from the second tube to make dilutions for plating to determine the cell number. The total number of bacterial cells was deduced.

The two tubes were centrifuged simultaneously (40 min, $40 \times g$). Sediment was produced in the first tube, and the second tube served as a control in which no sediment should have been produced. Immediately after centrifugation, 0.5 ml supernatant was taken from the first tube and used to prepare serial 10-fold dilutions in saline. For plating, 0.5 ml of each dilution was grown on semisolid medium containing antiserum against the expressed H antigen. The number of spreading colonies was determined, and the total number of flagellar phase variants was calculated.

Flagellar phase transition frequency was calculated by using the equation described previously (13, 49), as follows: flagellar phase transition frequency (per cell per generation) $= 1 - \sqrt[n]{1 - X}$, where n is the number of generations and X is (total number of flagellar phase variants)/(total number of bacterial cells). The number of generations (n) was determined by using the following equation: $n = (\log N - \log N_0)/\log 2$, where N is the total number of bacterial cells and N_0 is number of bacterial cells in the inoculum.

Nucleotide sequence accession numbers. The DNA sequences of *flk* regions between *rnpB* and *yhaC* of the *E. coli* H35, H36, and H47 type strains and *E. coli* strains H1171, H1173, H1174, and H1201 have been deposited in the GenBank database under accession numbers EF392692 to EF392698.

RESULTS

Characterization of the flagellin-specifying locus in H35, H36, and H47 strains. To localize the flagellin-specifying locus in the H35 strain and to examine the *flk* region in the H36 and H47 strains, a genomic library was constructed for each strain. From each library, a single recombinant plasmid (pLW1249, pLW1041, and pLW1248 carrying 8.5-, 4.1-, and 3.1-kb inserts, respectively) conferring motility to strain EJ34, which is a derivative of *E. coli* K-12 lacking flagellin, was isolated and con-

TABLE 1. *E. coli* strains and plasmids used in this study

Strain or plasmid[a]	Derivation, genotype, or phenotype	H antigen expressed	Flagellin genotype	Reference or source
Strains				
G1436	Bi7327-41, H3 type strain	H3	$fliC_{16}^{off}$ $flkA_3^{on}$	30
G1393	4370-53, H35 type strain	H35	$flkA_{35}^{on}$	This study
G1394	5017-53, H36 type strain	H36	$flkA_{36}^{on}$	28
G1397	1755-58, H47 type strain	H47	$fliC_{21}^{off}$ $flkA_{47}^{on}$	28
G1392	E480-68, H53 type strain	H53	$flkA_{53}^{on}$	28
H1171	Phase variant of G1436	H16	$fliC_{16}^{on}$ $flkA_3^{none}$	This study
H1172	Phase variant of G1436	H16	$fliC_{16}^{on}$ $flkA_3^{none}$	This study
H1173	Phase variant of G1436	H16	$fliC_{16}^{on}$ $flkA_3^{none}$	This study
H1174	Phase variant of G1436	H16	$fliC_{16}^{on}$ $flkA_3^{none}$	This study
H1175	Phase variant of G1436	H16	$fliC_{16}^{on}$ $flkA_3^{none}$	This study
H1176	Phase variant of G1436	H16	$fliC_{16}^{on}$ $flkA_3^{none}$	This study
H1201	Phase variant of G1397	H47	$fliC_{21}^{on}$ $flkA_{47}^{none}$	This study
H1202	Phase variant of G1397	H47	$fliC_{21}^{on}$ $flkA_{47}^{none}$	This study
G1438[b]	EJ34			11
G1370	DH5α			
G1067	H3 non-type strain	H3	$fliC_{16}^{off}$ $flkA_3^{on}$	
G1145	H47 non-type strain	H47	$fliC_{21}^{off}$ $flkA_{47}^{on}$	
H1692	*orf486*-deficient mutant of G1436	H3	$fliC_{16}^{off}$ $flkA_3^{on}$	This study
H1693	H1692 carrying pLW1330	H3	$fliC_{16}^{off}$ $flkA_3^{on}$	This study
H1694	G1436 carrying pLW1330	H3	$fliC_{16}^{off}$ $flkA_3^{on}$	This study
Plasmids				
pUC19	Cloning vector; Ap^r			Takara
pLW1249	pUC19 with a 8.5-kb insert containing the flagellin gene for H35			This study
pLW1041	pUC19 with a 4.1-kb insert containing the flagellin gene for H36			This study
pLW1248	pUC19 with a 3.1-kb insert containing the flagellin gene for H47			This study
pKD4	Contains a kanamycin resistance gene (*kan*); Km^r Ap^r			8
pKD20	RED recombinase expression plasmid; Ap^r			8
pUC18	Cloning vector; Ap^r			Takara
pLW1330	pUC18 containing *orf486*			This study

[a] *E. coli* H antigen type strains were obtained from the Institute of Medical and Veterinary Science, Adelaide, Australia, and the Mechnikov Research Institute for Vaccines and Sera, Russian Academy of Medical Sciences, Moscow, Russia. Other *E. coli* standard H type strains used have been described previously (12, 46).
[b] A derivative of *E. coli* K-12 which is nonflagellated and nonmotile because the expression of its single flagellin gene, *fliC*, is blocked by a mutation.

firmed to carry the respective flagellin gene by means of serotyping. Sequencing analysis revealed the presence of *flkA* and *flkB* in all three inserts. To test whether the *flk* regions in H35, H36, and H47 are located between *yhaC* and *rnpB* as they are in H3 and H53, PCRs were carried out to amplify the region between *yhaC* and *flkB* using primers wl-3622 and wl-5272, wl-5276, or wl-5274 and PCRs were carried out to amplify the region between *flkA* and *rnpB* using primers wl-3623 and wl-5273, wl-5277, or wl-5275. Sequencing analysis of the PCR products revealed the presence of the *flk* region between *rnpB* and *yhaC* in all three strains. The results indicated that H35 is a new member of the *flk*-positive group, and the *flk* region containing the *flkAB* operon is generally present between *yhaC* and *rnpB* in *flk*-positive *E. coli*.

Sequence analysis of the *yhaC-rnpB* region in H3, H35, H36, H47, and H53 strains. Sequences that were 13,317, 12,193, and 8,737 bp long were obtained for the region between *yhaC* and *rnpB* in the H35, H36, and H47 type strains, respectively, and these sizes are similar to the sizes obtained for the H3 (10,867 bp) and H53 (9,310 bp) type strains retrieved from the GenBank database (AB128916 and AB128917). *flkB*, *flkA*, *orf204*, *orf240*, and *orf486* in the same order were found in all five strains (Fig. 1). *orf486* encodes a putative integrase belonging to the tyrosine recombinase family (also known as the λ integrase family), *orf240* encodes a putative invertase belonging to the serine recombinase family, and *orf204* encodes a protein with an unknown function. An IS*1222* element was found between *flkA* and *flkB* in H36, suggesting that the cotranscription of these two genes may be disrupted, resulting in expression of only *flkA*. This explains the absence of repressor activity for *fliC* in H36 reported previously (28). The *flkB* gene in H35 was found to be disrupted by an IS*3* element, and therefore, the absence of repressor activity in H35 is also expected. Several insertion elements are present between *yhaC* and *flkB* in H3, H35, and H36 strains (Fig. 1).

In all five strains, three or four copies of an imperfect 23-bp DR, corresponding to the last 23 bp of the *rnpB* gene (5′-CG GCTTATCGGTCAGTTTCACCT-3′), were found at the right (DR1) and left (DR2, DR3, and DR4) ends of the *flk* region, and DR1 overlapped the 23 bp of *rnpB* (Fig. 1). One or two single-nucleotide polymorphic sites were found in some of the DRs (data not shown). In *E. coli* K-12, a 113-bp repeated DNA sequence beginning with the last 23 bp of the *rnpB* gene is present in the flanking region downstream of *rnpB* and is repeated several times (6, 35). Each 113-bp unit consists of the 23-bp sequence (DR in *flk*-positive strains) and the region for

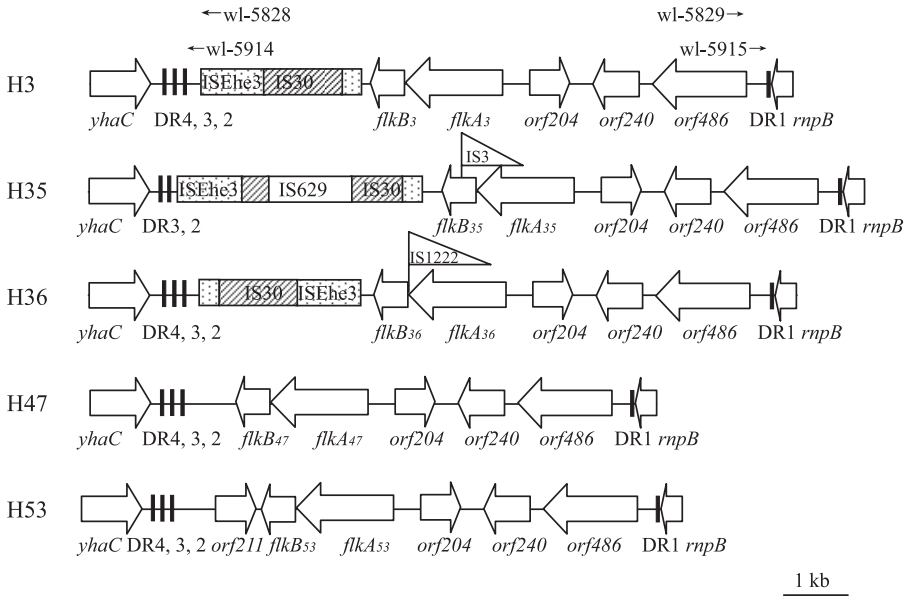

FIG. 1. Comparative map of the *flk* regions in *E. coli* H3, H53, H47, H36, and H35 strains. The orientation corresponds to that of the *E. coli* K-12 chromosome. Open arrows indicate the locations and orientations of open reading frames of putative genes. Insertion elements are indicated by rectangles or burgees. The imperfect 23-bp DRs are indicated by bars. *orf*204 in H53 strain was designated *orf210* in a previous study (43). The positions of the wl-5828/wl-5829 and wl-5914/wl-5915 primer pairs used for detection of the circular form of *flk* GI_{H3} are indicated by arrows.

intrinsic transcription termination. In *flk*-positive *E. coli* strains, the first 23-bp sequence (DR1) and the transcription termination region were found to be separated by the *flk* locus and the second 23-bp sequence (DR2) (Fig. 2).

In addition to the presence of the flanking DRs and a putative integrase gene (*orf486*), several other features of typical GIs were also found in the region, including the large size (6,951 to 11,651 bp) and lower G+C content of the region (42.2 to 46.2% versus 50.5% for the *E. coli* genome) and the absence of the region in closely related strains, such as *E. coli* K-12 (accession no. NC_000913) and O157:H7 (AE005174) (39). All these findings indicate that the *flk* regions between *yhaC* and *rnpB* in the five *flk*-positive strains are integrated GIs (referred to as *flk* GI_{H3}, *flk* GI_{H35}, *flk* GI_{H36}, *flk* GI_{H47}, and *flk* GI_{H53}) (Fig. 2). While DR1 and DR2 are the two flanking DRs of the *flk* GI, DR3 and DR4 appear to be intrinsic sequences in the host chromosome.

We also sequenced the *yhaC-rnpB* region in two nontype strains, *E. coli* O53:H3 strain G1067 and *E. coli* O156:H47 strain G1145, and each of the sequences obtained showed 99% DNA identity to the corresponding sequence in the corresponding type strain. This indicates that the *flk* GI is conserved in *E. coli* strains carrying the same H antigen.

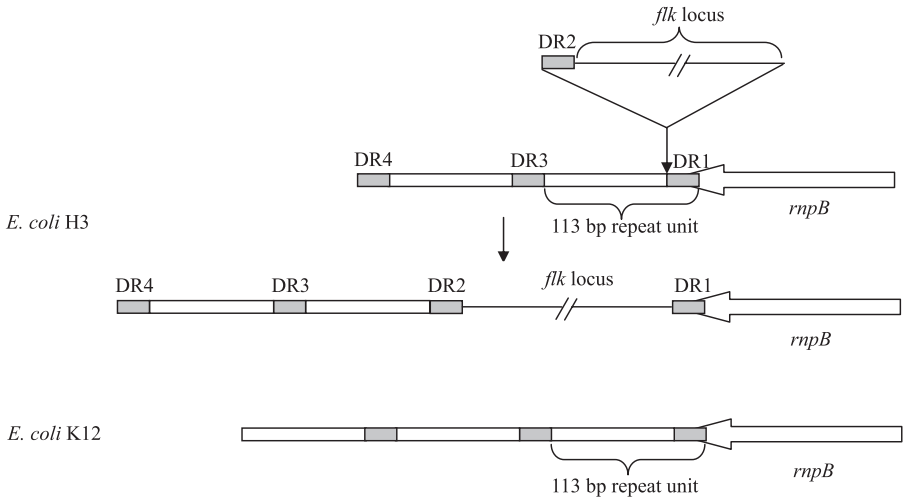

FIG. 2. Schematic representation of the 113-bp repeats downstream of the *rnpB* gene in *E. coli* strain K-12 and insertion of the *flk* locus in the same region in the H3 strain. The 113-bp repeats are indicated by open rectangles. The 23-bp sequences corresponding to the last 23 bp of the *rnpB* gene are indicated by shaded rectangles. The positions of DRs in H3 are indicated.

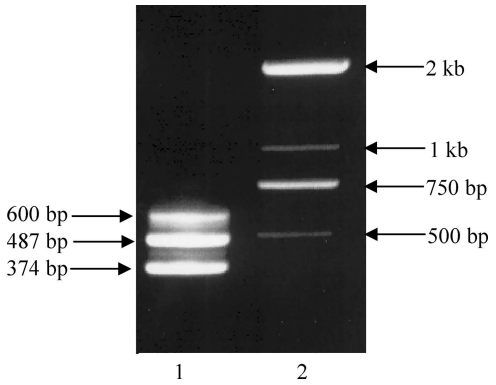

FIG. 3. Detection of the circular forms of *flk* GI_{H3} by two-step PCR. Lane 1, three PCR products obtained from the H3 type strain (374, 487, and 600 bp); lane 2, size markers. The wl-5828/wl-5829 and wl-5914/wl-5915 primer pairs were used.

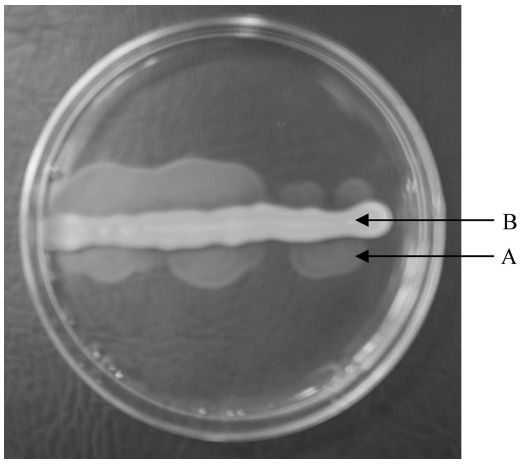

FIG. 5. Selection of spontaneous flagellar phase variants from *E. coli* H3 on motility agar containing antiserum against H3 antigen. Arrow A indicates spreading growth produced by the phase variants. Arrow B indicates immobilized growth of the H3 parental strain in the presence of H3 antiserum.

Detection of the excised extrachromosomal circular form of the *flk* GI. A two-step PCR was carried out with the H3 type strain to detect the extrachromosomal circular form of *flk* GI_{H3}. The first-round PCR was performed using primers wl-5828 and wl-5829 oriented toward the left and right ends of *flk* GI_{H3}, respectively. The second-round PCR was carried out using an aliquot of the products from the first round PCR as the template and primers wl-5914 and wl-5915 designed based on the sequence of the expected product from the first-round PCR. Three PCR products that were 600, 487, and 374 bp long were detected (Fig. 3), indicating that there was formation of three types of extrachromosomal circular forms. Sequence analysis of the three PCR products indicated that the recombination events occurred between DR1 and DR2, between DR1 and DR3, and between DR1 and DR4, respectively, and the three types of circular forms were designated *flk* GI_{H3}-1, *flk* GI_{H3}-2, and *flk* GI_{H3}-3 (Fig. 4). Using the same procedure but different primer pairs, three types of extrachromosomal circular forms of the *flk* GI were also detected in the H36, H47, and H53 type strains. However, only two types were detected in the H35 strain, consistent with the presence of one and two DRs at the two ends of *flk* GI_{H35} (data not shown).

Involvement of the *flk* GI in flagellar phase variation. By growing the H3 and H47 type strains on motility agar containing antiserum against H3 and H47, respectively, and screening for clones which exhibited spreading growth (Fig. 5), six and two spontaneous phase variants were obtained. All six H3 variants (strains H1171 to H1176) produced agglutination with H16 antiserum but not with H3 antiserum and were strongly immobilized when they were grown on semisolid motility medium containing H16 antiserum. Both H47 variants (strains H1201 and H1202) agglutinated with H21 antiserum but not with H47 antiserum and were strongly immobilized in the presence of H21 antiserum. The results are in agreement with the previous findings that the *fliC* genes in H3 and H47 strains encode H16 and H21 antigens, respectively (28, 29, 46). The strong immobilization of the variants by H16 or H21 antiserum

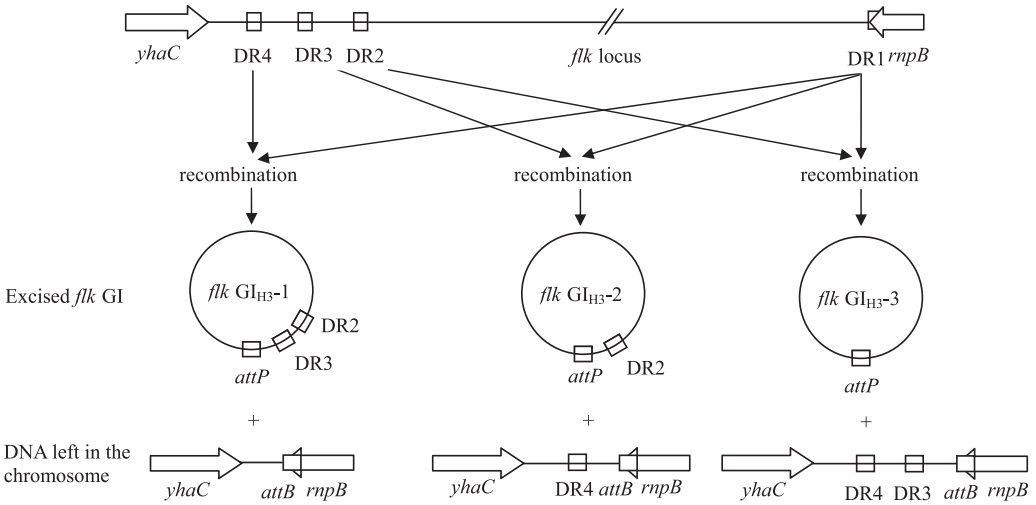

FIG. 4. Model for the excision of *flk* GI_{H3}. Recombination events for the generation of *flk* GI_{H3}-1, *flk* GI_{H3}-2, and *flk* GI_{H3}-3 are shown.

indicated that there was unilateral phase change, as reported for the two strains previously.

To find out whether the *flk* GI is still present in the variants, the *yhaC-rnpB* regions in all eight variant strains were sequenced using primers wl-3622 and wl-3623. In the H3 variants, the region is 500 bp long and contains two complete DRs (DR3 and DR4) in H1171, H1172, and H1175, is 387 bp long and contains one DR (DR4) in H1173 and H1176, and is 274 bp long and contains no DRs in H1174 (Fig. 4). The region is 507 bp long and contains two DRs (DR3 and DR4) in both H47 variant strains (H1201 and H1202). These results indicate that the phase variation of flagellar antigens in the H3 and H47 strains was caused by the excision of the *flk* GI and that recombination could occur between DR1 and DR2, between DR1 and DR3, and between DR1 and DR4.

No PCR products were obtained from any of the eight variant strains using the *flkA*-specific primer pairs (wl-5141/wl-5142 and wl-3820/wl-3821), indicating the absence of the *flkAB* operon and therefore the absence of the excised *flk* GI in the variants (data not shown). The excision and irrevocable loss of the *flk* GI provide an explanation for the unilateral phase variation observed in *flk*-positive strains.

Characterization of Orf486 as a functional integrase. To confirm the function of Orf486 as an integrase for the excision of the *flk* GI, an *orf486*-deficient mutant of the H3 type strain (H1692) and the corresponding transcomplementary strain (H1693) were generated. PCR was carried out to detect extrachromosomal circular forms of *flk* GI_{H3} in the two strains generated, and the expected PCR products indicative of the circular forms of the GI were detected only in transcomplementary strain H1693 and not in mutant strain H1692. By growing the two strains on motility agar containing antiserum against H3, spontaneous flagellar antigen phase variants with the H16 phenotype were detected for transcomplementary strain H1693 but not for the mutant. These results confirmed that *orf486* is a functional integrase gene required for excision of the *flk* GI from the chromosome.

The frequency of flagellar phase transition was also determined in H1692 (H3 type strain), H1693 (transcomplementary strain of *orf486* mutant), and H1694 (H3 type strain carrying pLW1330 containing *orf486*). The transition rates were determined to be 1.98×10^{-8} per bacterium per generation for H1692, 1.93×10^{-2} per bacterium per generation for H1693, and 2.83×10^{-2} per bacterium per generation for H1694. The increased transition frequencies in H1693 and H1694 could apparently be attributed to the overexpression of Orf486. The results further indicated that *orf486* is a functional integrase gene.

Identification of the potential integration sites. In all phase variant strains generated (H1171 to H1176 for H3 and H1201 and H1202 for H53), the last 23 bp of *rnpB* (corresponding to DR1 in the integrated *flk* GI) is retained (Fig. 4), indicating that the 3′-terminal 23 bp of the *rnpB* gene (5′-CGGCTTAT CGGTCAGTTTCACCT-3′) is the potential chromosome attachment site (*attB*) of the *flk* GI. The sequence of the potential *attB* site is conserved in all variants except H1174, which has a single-nucleotide polymorphic site at position 19. The potential *attB* site was also found in many *E. coli* strains of other types either by sequencing (H2, H8, H11, H16, H23, H27, and H55) or through genome searching (accession no. AE014075, BA000007, AE005174, NC_000913, AE005541, AP009048, CP000468, CP000247, CP000243, and AE014075), as well as in strains belonging to other species of the family *Enterobacteriaceae*, including *S. enterica*, *Shigella*, *Klebsiella pneumoniae*, and *Pantoea agglomerans* (data not shown). Therefore, this site is conserved not only in *E. coli* but also in other closely related species.

The potential *attP* site (5′-CGGCTTATCGGTCAGCTTC AACT-3′) in the extrachromosomal circular form of *flk* GI_{H3} was identified at the junction of the two ends of *flk* GI_{H3} (Fig. 4), which is identical to the sequence of DR1. Two single-nucleotide polymorphic sites between the *attP* and *attB* sites were found to be at positions 16 and 21. The *attP* site was formed by recombination between DR1 and one of the other DRs, but the exact mechanism involved is not clear. One or two single-nucleotide polymorphic sites were also detected in some of the DRs, including DR1, which had different residues at positions 16 and 21 than *flk* GI_{H3}. Presumably, the site polymorphism between the *attP* and *attB* genes might arise from site polymorphisms in DRs. To prove this, further studies are needed. Although not investigated in this study, polymorphism sites may also be present in the *attP* site of the *flk* GI in other *flk*-positive strains for the same reason.

In all five *flk*-positive strains, only one DR (DR1) was found to be at the right end of the *flk* GI, and the rest of the DRs were found to be at the left end, indicating that any *flk* GIs could integrate only into the *attB* site and not into other 23-bp DR sequences in the 3′ flanking region of *rnpB*.

Absence of the *flk* region in other H type strains. The region between *yhaC* and *rnpB* was amplified from 53 *E. coli* H type strains by PCR using primers wl-3622 and wl-3623. In *E. coli* K-12, which does not contain any non-*fliC* flagellin genes, this region contains a 615-bp intergenic DNA region (6). In comparison, the region is 8,737 to 13,317 bp long in the five *flk*-positive strains. Therefore, the presence of the *flk* region can be indicated by the size of the PCR product. The results showed that except for the five *flk*-positive strains described above, none of the strains gave PCR products larger than 1.5 kb (data not shown). PCRs were also performed using the primers targeting *flkAB* (wl-5136 and wl-5033) and *orf486* (wl-5830 and wl-5831), and the expected PCR products were detected for the five *flk*-positive strains but not for other strains (data not shown). Therefore, none of the type strains except those expressing H3, H35, H36, H47, and H53 antigens carries the *flk* region.

DISCUSSION

In this study, we examined the involvement of a GI, the *flk* GI, in flagellar phase variation in *E. coli* strains. A general model for the phase variation in the FlkA and FliC flagellins in *E. coli* is shown in Fig. 6. The principle is that when the *flk* GI is present in the chromosome, the *flkAB* operon is expressed to allow coproduction of the FlkA flagellin and the repressor protein FlkB, and the expression of *fliC* is repressed. When the *flk* GI is excised from the chromosome, *flkAB* is irreversibly deleted, and therefore the repression of the *fliC* is released and the FliC flagellin is produced. Therefore, the formula for the unilateral phase variation ($fliC^{off} + flkA^{on} \rightarrow fliC^{on} + flkA^{off}$) (33) should be revised to $fliC^{off} + flkA^{on} \rightarrow fliC^{on} + flkA^{none}$.

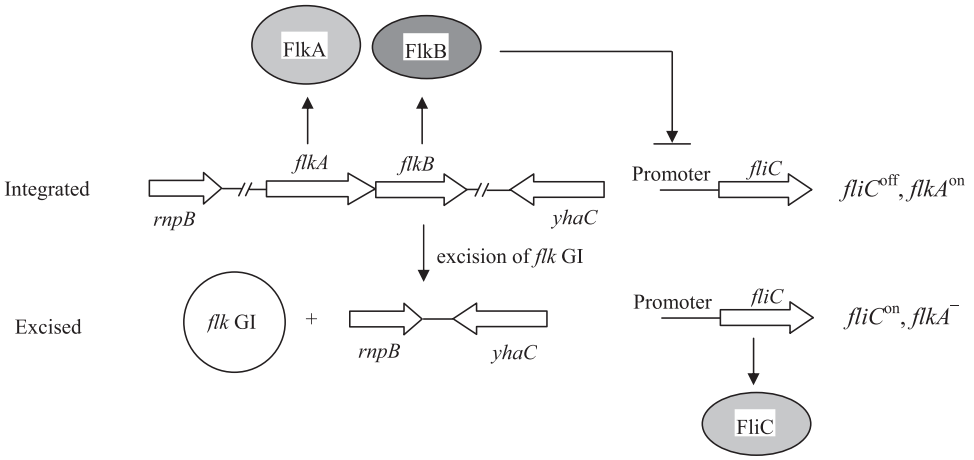

FIG. 6. Model for the phase variation between FliC and FlkA in *E. coli*. When the *flk* GI is present in the chromosome, the *flkAB* operon is expressed. FlkA flagellin is produced, and the *fliC* gene is repressed by FlkB. After the *flk* GI is excised from the chromosome, the *fliC* gene is expressed.

In the five *flk*-positive *E. coli* strains, the *fliC* genes in the H35, H36, and H53 strains are disrupted or defective (33, 43, 46). Therefore, any *flk*-negative variants of these strains would be nonmotile. Whether this can occur was not investigated here due to a lack of screening methods for nonmotile variants. However, it is more likely that such strains would become monophasic under physiological conditions, as the survival of the bacteria may be affected in the absence of a flagellin, and the flagellin encoded by *flkA* may offer a selective advantage to these strains in their particular environments. On the other hand, repression of *fliC* in these strains is not needed, and this is also reflected by the disruption of the repressor gene *flkB* by insertion elements in H35 and H36. The insertions must have happened after the bacteria gained the *flk* GI, so that the survival of bacteria would not be affected. Although the *flkB* gene is intact in H53, disruption of this gene is also expected sooner or later from an evolutionary point of view.

Although the principles for flagellar phase variation in *E. coli flk*-positive strains and *S. enterica* strains are rather similar and both organisms require cotranscription of a flagellin gene and a *fliC* repressor gene, different mechanisms are utilized. While the *flk* GI described here mediates only unilateral phase change owing to the irreversible loss of the *flk* GI, flagellar phase variation mediated by the site-specific inversion of the H segment in *S. enterica* is bilateral (1, 20, 22, 40, 47, 51). Although the flagellar phase variation in a z66-positive strain of *S. enterica* serovar Typhi is also caused by deletion of the *fljAB*-like operon (17), a mechanism different from that described here seems to be involved, as indicated by a recent report showing that the *fljB*-like gene in this strain is located on a novel liner plasmid (3). Still, mechanisms involved in flagellar phase variation in other *E. coli* strains carrying different non-*fliC* flagellin loci and showing flagellar phase variation, such as H17 carrying the *fliC*$^{//}$ locus (28, 32, 33), remain to be investigated.

In addition to the integrase gene (*orf486*) characterized, *orf204* encoding a putative invertase/resolvase is also present in the region. Site-specific recombination mediated by the invertase requires the presence of inverted repeats (14), such as the *hix* sequence in *S. enterica*, which are absent in the *flk* region. Therefore, the *orf204* gene is likely to be nonfunctional or not related to the *flk* region.

At this stage, we could not obtain any indication of the possible origin of the *flk* GI. The *flk* GI could have evolved from a single ancestor and been integrated into other hosts through recombination at the *attB* site. The presence of the 23-bp *attB* site in many bacterial species belonging to the family *Enterobacteriaceae* brings up the possibility of the presence of *flk*-like GIs in other bacterial species. The other possibility is that all five *flk* loci may have evolved from a common ancestor in which the *flk* GI had already been integrated at the *attB* site. The integration potential of the *flk* GI remains to be studied further. In future studies, it should also be worthwhile to investigate the factors or environmental conditions which can induce the flagellar phase variation and mechanisms for the maintenance of flagellar variants in the population.

ACKNOWLEDGMENTS

This work was supported by the National 863 program of China (grants 2006AA020703 and 2006AA06Z409), by the National Natural Science Foundation of China (NSFC) Key Programs (grants 30530010 and 20536040), by the Chinese National Science Fund for Distinguished Young Scholars (grant 30788001), by the NSFC General Program (grant 30670038), and by the Tianjin Municipal Fund for Science and Technology Innovation (grant 05FZZDSH00800).

REFERENCES

1. **Aldridge, P. D., C. Wu, J. Gnerer, J. E. Karlinsey, K. T. Hughes, and M. S. Sachs.** 2006. Regulatory protein that inhibits both synthesis and use of the target protein controls flagellar phase variation in *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **103:**11340–11345.
2. **Antonenka, U., C. Nolting, J. Heesemann, and A. Rakin.** 2005. Horizontal transfer of *Yersinia* high-pathogenicity island by the conjugative RP4 *attB* target-presenting shuttle plasmid. Mol. Microbiol. **57:**727–734.
3. **Baker, S., J. Hardy, K. E. Sanderson, M. Quail, I. Goodhead, R. A. Kingsley, J. Parkhill, B. Stocker, and G. Dougan.** 2007. A novel linear plasmid mediates flagellar variation in *Salmonella typhi*. PLoS Pathog. **3:**e59.
4. **Beutin, L., and E. Strauch.** 2007. Identification of sequence diversity in the *Escherichia coli fliC* genes encoding flagellar types H8 and H40 and its use in typing of Shiga toxin-producing *E. coli* O8, O22, O111, O174, and O179 strains. J. Clin. Microbiol. **45:**333–339.
5. **Blair, D. F.** 1995. How bacteria sense and swim. Annu. Rev. Microbiol. **49:**489–522.

6. **Blattner, F. R., G. I. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, et al.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1474.
7. **Burrus, V., G. Pavlovic, B. Decaris, and G. Guedon.** 2002. Conjugative transposons: the tip of the iceberg. Mol. Microbiol. **46:**601–610.
8. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA **97:**6640–6645.
9. **Dobrindt, U., B. Hochhut, U. Hentschel, and J. Hacker.** 2004. Genomic islands in pathogenic and environmental microorganisms. Nat. Rev. Microbiol. **2:**414–424.
10. **Doublet, B., D. Boyd, M. R. Mulvey, and A. Cloeckaert.** 2005. The *Salmonella* genomic island 1 is an integrative mobilizable element. Mol. Microbiol. **55:**1911–1924.
11. **Enomoto, M., and B. A. Stocker.** 1975. Integration, at hag or elsewhere, of H2 (phase-2 flagellin) genes transduced from *Salmonella* to *Escherichia coli*. Genetics **81:**595–614.
12. **Ewing, W. H.** 1986. Edwards and Ewing's identification of *Enterobacteriaceae*, 4th ed. Elsevier Science Publishing, Inc., New York, NY.
13. **Gally, D. L., J. A. Bogan, B. I. Eisenstein, and I. C. Blomfield.** 1993. Environmental regulation of the *fim* switch controlling type 1 fimbrial phase variation in *Escherichia coli* K-12: effects of temperature and media. J. Bacteriol. **175:**6186–6193.
14. **Grindley, N. D., K. L. Whiteson, and P. A. Rice.** 2006. Mechanisms of site-specific recombination. Annu. Rev. Biochem. **75:**567–605.
15. **Hallet, B., and D. J. Sherratt.** 1997. Transposition and site-specific recombination: adapting DNA cut-and-paste mechanisms to a variety of genetic rearrangements. FEMS Microbiol. Rev. **21:**157–178.
16. **Hochhut, B., and M. K. Waldor.** 1999. Site-specific integration of the conjugal *Vibrio cholerae* SXT element into *prfC*. Mol. Microbiol. **32:**99–110.
17. **Huang, X., V. Phung le, S. Dejsirilert, P. Tishyadhigama, Y. Li, H. Liu, K. Hirose, Y. Kawamura, and T. Ezaki.** 2004. Cloning and characterization of the gene encoding the z66 antigen of *Salmonella enterica* serovar Typhi. FEMS Microbiol. Lett. **234:**239–246.
18. **Iino, T.** 1969. Genetics and chemistry of bacterial flagella. Bacteriol. Rev. **33:**454–475.
19. **Ikeda, J. S., C. K. Schmitt, S. C. Darnell, P. R. Watson, J. Bispham, T. S. Wallis, D. L. Weinstein, E. S. Metcalf, P. Adams, C. D. O'Connor, and A. D. O'Brien.** 2001. Flagellar phase variation of *Salmonella enterica* serovar Typhimurium contributes to virulence in the murine typhoid infection model but does not influence *Salmonella*-induced enteropathogenesis. Infect. Immun. **69:**3021–3030.
20. **Johnson, R. C., and M. I. Simon.** 1985. Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. Cell **41:**781–791.
21. **Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei.** 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics **17:**1244–1245.
22. **Kutsukake, K., and T. Iino.** 1980. A trans-acting factor mediates inversion of a specific DNA segment in flagellar phase variation of *Salmonella*. Nature **284:**479–481.
23. **Lesic, B., S. Bach, J. M. Ghigo, U. Dobrindt, J. Hacker, and E. Carniel.** 2004. Excision of the high-pathogenicity island of *Yersinia pseudotuberculosis* requires the combined actions of its cognate integrase and Hef, a new recombination directionality factor. Mol. Microbiol. **52:**1337–1348.
24. **Macnab, R. M.** 1992. Genetics and biogenesis of bacterial flagella. Annu. Rev. Genet. **26:**131–158.
25. **Mantri, Y., and K. P. Williams.** 2004. Islander: a database of integrative islands in prokaryotic genomes, the associated integrases and their DNA site specificities. Nucleic Acids Res. **32:**D55–D58.
26. **Orskov, F., and I. Orskov.** 1984. Serotyping of *Escherichia coli*. Academic Press, London, United Kingdom.
27. **Ramos, H. C., M. Rumbo, and J. C. Sirard.** 2004. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends Microbiol. **12:**509–517.
28. **Ratiner, Y. A.** 1987. Different alleles of the flagellin gene *hagB* in *Escherichia coli* standard H test strains. FEMS Microbiol. Lett. **48:**97–104.
29. **Ratiner, Y. A.** 1967. Mutation of *E. coli* with regard to the H-antigen. I. Isolation of H-antigen mutants from test H-strains of *Escherichia* culture. Zh. Mikrobiol. Epidemiol. Immunobiol. **44:**23–29.
30. **Ratiner, Y. A.** 1982. Phase variation of the H antigen in *Escherichia coli* strain Bi7327-41, the standard strain for *Escherichia coli* flagellar antigen H3. FEMS Microbiol. Lett. **15:**33–36.
31. **Ratiner, Y. A.** 1983. Presence of two structural genes determining antigenically different phase-specific flagellins in *Escherichia coli* strains. FEMS Microbiol. Lett. **19:**37–41.
32. **Ratiner, Y. A.** 1985. Two genetic arrangement determining flagellar antigen specificities in two diphasic *Escherichia coli* strains. FEMS Microbiol. Lett. **29:**317–323.
33. **Ratiner, Y. A.** 1998. New flagellin-specifying genes in some *Escherichia coli* strains. J. Bacteriol. **180:**979–984.
34. **Ratiner, Y. A.** 1991. Serotyping of *Escherichia coli* flagellar antigens, p. 47–51. *In* G. Stein and R. Fünfstück (ed.), Harnwegsinfektion: Aktuale Gesichtspunkte zur pathogenese Diagnostik und Therapie. Wissenschaftliches Symposium, Frankfurt am Main, Germany.
35. **Reed, R. E., and S. Altman.** 1983. Repeated sequences and open reading frames in the 3′ flanking region of the gene for the RNA subunit of *Escherichia coli* ribonuclease P. Proc. Natl. Acad. USA **80:**5359–5363.
36. **Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell.** 2000. Artemis: sequence visualisation and annotation. Bioinformatics **16:**944–945.
37. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
38. **Saunders, J. R.** 1994. Population genetics of phase variable antigens, p. 247–268. *In* S. Baumberg, J. P. W. Young, E. M. H. Wellington, and J. R. Saunders (ed.), Population genetics of bacteria. Cambridge University Press, London, United Kingdom.
39. **Schmidt, H., and M. Hensel.** 2004. Pathogenicity islands in bacterial pathogenesis. Clin. Microbiol. Rev. **17:**14–56.
40. **Silverman, M., and M. Simon.** 1980. Phase variation: genetic analysis of switching mutants. Cell **19:**845–854.
41. **Staden, R.** 1996. The Staden sequence analysis package. Mol. Biotechnol. **5:**233–241.
42. **Thompson, J. D., D. G. Higgins, and T. J. Gibson.** 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:**4673–4680.
43. **Tominaga, A.** 2004. Characterization of six flagellin genes in the H3, H53 and H54 standard strains of *Escherichia coli*. Genes Genet. Syst. **79:**1–8.
44. **Tominaga, A., and K. Kutsukake.** 2007. Expressed and cryptic flagellin genes in the H44 and H55 type strains of *Escherichia coli*. Genes Genet. Syst. **82:**1–8.
45. **Van Houdt, R., and C. W. Michiels.** 2005. Role of bacterial cell surface structures in *Escherichia coli* biofilm formation. Res. Microbiol. **156:**626–633.
46. **Wang, L., D. Rothemund, H. Curd, and P. R. Reeves.** 2003. Species-wide variation in the *Escherichia coli* flagellin (H-antigen) gene. J. Bacteriol. **185:**2936–2943.
47. **Yamamoto, S., and K. Kutsukake.** 2006. FljA-mediated posttranscriptional control of phase 1 flagellin expression in flagellar phase variation of *Salmonella enterica* serovar Typhimurium. J. Bacteriol. **188:**958–967.
48. **Yu, D., H. M. Ellis, E. C. Lee, N. A. Jenkins, N. G. Copeland, and D. L. Court.** 2000. An efficient recombination system for chromosome engineering in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **97:**5978–5983.
49. **Ziebuhr, W., V. Krimmer, S. Rachid, I. Lossner, F. Gotz, and J. Hacker.** 1999. A novel mechanism of phase variation of virulence in *Staphylococcus epidermidis*: evidence for control of the polysaccharide intercellular adhesin synthesis by alternating insertion and excision of the insertion sequence element IS256. Mol. Microbiol. **32:**345–356.
50. **Zieg, J., M. Silverman, M. Hilmen, and M. Simon.** 1977. Recombinational switch for gene expression. Science **196:**170–172.
51. **Zieg, J., and M. Simon.** 1980. Analysis of the nucleotide sequence of an invertible controlling element. Proc. Natl. Acad. Sci. USA **77:**4196–4200.