INFECTION AND IMMUNITY, Feb. 2001, p. 937–948 Vol. 69, No. 2
0019-9567/01/$04.00+0 DOI: 10.1128/IAI.69.2.937–948.2001

afa-8 Gene Cluster Is Carried by a Pathogenicity Island Inserted into the $tRNA^{Phe}$ of Human and Bovine Pathogenic *Escherichia coli* Isolates

LILA LALIOUI AND CHANTAL LE BOUGUÉNEC*

Unité de Pathogénie Bactérienne des Muqueuses, Institut Pasteur, 75724 Paris Cedex 15, France

Received 8 August 2000/Returned for modification 21 September 2000/Accepted 14 November 2000

We recently described a new afimbrial adhesin, AfaE-VIII, produced by animal strains associated with diarrhea and septicemia and by human isolates associated with extraintestinal infections. Here, we report that the *afa-8* operon, encoding AfaE-VIII adhesin, from the human blood isolate *Escherichia coli* AL862 is carried by a 61-kb genomic region with characteristics typical of a pathogenicity island (PAI), including a size larger than 10 kb, the presence of an integrase-encoding gene, the insertion into a tRNA locus (*pheR*), and the presence of a small direct repeat at each extremity. Moreover, the G+C content of the *afa-8* operon (46.4%) is lower than that of the *E. coli* K-12/MG1655 chromosome (50.8%). Within this PAI, designated PAI I_{AL862}, we identified open reading frames able to code for products similar to proteins involved in sugar utilization. Four probes spanning these sequences hybridized with 74.3% of pathogenic *afa-8*-positive *E. coli* strains isolated from humans and animals, 25% of human pathogenic *afa-8*-negative *E. coli* strains, and only 8% of fecal strains ($P = 0.05$), indicating that these sequences are strongly associated with the *afa-8* operon and that this genetic association may define a PAI widely distributed among human and animal *afa-8*-positive strains. One of the distinctive features of this study is that *E. coli* AL862 also carries another *afa-8*-containing PAI (PAI II_{AL862}), which appeared to be similar in size and genetic organization to PAI I_{AL862} and was inserted into the *pheV* gene. We investigated the insertion sites of *afa-8*-containing PAI in human and bovine pathogenic *E. coli* strains and found that this PAI preferentially inserted into the *pheV* gene.

Pathogenic *Escherichia coli* strains have the potential to cause a wide variety of infectious diseases, including septicemia, newborn meningitis, and intestinal and urinary tract infections (UTIs). These strains carry virulence-associated genes, which may encode toxins, capsules, invasins, adhesins, and other virulence factors that enable them to overcome host defenses, to proliferate, and to cause tissue damage and disease. These determinants are usually clustered on the chromosome in pathogenicity islands (PAIs) (24). The PAIs of uropathogenic strains were the first to be described in *E. coli* species. At least four PAIs are present in the genome of uropathogenic *E. coli* (UPEC) strain 536. PAI I_{536} and PAI II_{536} encode the hemolysin and the P-related fimbrial adhesin, while PAI III_{536} encodes the S fimbrial adhesin (9, 24, 25, 33) PAI IV_{536} carries the *fyuA* (ferrin yersiniabactin uptake) and *irp1* through *irp5* (iron-repressible protein) genes originally found in the PAI (HPI) of various *Yersinia* species (24). Two PAIs were described in UPEC strain J96 and reported to encode the hemolysin and the P or P-related fimbrial adhesins. PAI II_{J96} also encodes the cytotoxic necrotizing factor 1 (CNF1) (8, 55). One PAI (PAI I_{CFT073}) has been identified in UPEC strain CFT073 and was reported to encode the hemolysin and the P fimbrial adhesin (23, 32). In diarrheagenic *E. coli* strains, several pathotypes have been reported to carry PAIs. Enteropathogenic *E. coli* (EPEC) strains carry the locus of enterocyte effacement (LEE) PAI (39, 40). Like EPEC, enterohemorrhagic *E. coli* (EHEC) strains are generally considered to contain the LEE (39, 44). Most of the enteroaggregative *E. coli* (EAEC) strains harbor the HPI of *Yersinia enterocolitica* (51). A potential PAI has been described in a prototypical enterotoxigenic *E. coli* (ETEC) strain, H10407, which contains a *tia* locus that mediates in vitro invasion into cultured intestinal epithelial cells (J. M. Fleckenstein, N. J. Snellings, E. A. Elsinghorst, and L. E. Lindler, Abstr. Meet. Microb. Pathogenesis Host Response, p. 37, 1997.)

Among the adhesins produced by pathogenic *E. coli* strains, afimbrial adhesins encoded by the *afa* family of gene clusters have been extensively studied (18, 19, 30, 34, 37, 42). We recently described a new *afa* operon (*afa-8*) encoding an afimbrial adhesin widespread among bovine *E. coli* isolates associated with diarrhea and/or septicemia (35) and human *E. coli* isolates associated with extraintestinal infections (20, 35; C. Le Bouguénec, L. Lalioui, L. du Merle, M. Jouve, P. Courcoux, S. Bouzari, R. Selvarangan, B. J. Nowicki, Y. Germani, A. Andremont, P. Gounon, and M. I. Garcia, submitted for publication). This gene cluster is chromosome or plasmid borne (20, 35), suggesting that it may be carried by a mobile element, facilitating its dissemination among pathogenic *E. coli* strains. Moreover, Garcia et al. (19) showed that the *afa-3* gene cluster, carried by human pathogenic *E. coli* strains, is flanked by insertion sequence elements and is able to translocate from a plasmid to the chromosome by an IS*1*-mediated recombination mechanism.

The aim of this study was to investigate the possible association of the *afa-8* operon with a PAI. Sequence analysis of the chromosomal regions downstream from the *afa-8* operon identified a potential P4 integrase gene and a phenylalanine-spe-

* Corresponding author. Mailing address: Pathogénie Bactérienne des Muqueuses, Institut Pasteur, 28 Rue du Dr Roux, 75724 Paris Cedex 15, France. Phone: 33 1 40613280. Fax: 33 1 40613640. E-mail: clb@pasteur.fr.

cific tRNA gene, consistent with the definition of PAIs. Partial characterization of the *afa-8*-containing PAI from *E. coli* AL862 indicated that this PAI is a 61-kb chromosomal region that carries the *afa-8* operon as the only known virulence determinant. Moreover, this *afa-8*-containing PAI has a distinctive feature: the ability to insert into the two $tRNA^{Phe}$ loci present on the chromosome of the same strain.

MATERIALS AND METHODS

Bacterial strains, cosmids, and culture conditions. Four partially characterized collections of human *E. coli* strains were used in this study. The first consisted of 44 isolates from urine specimens, 18 of which carried the *afa-8* operon (Le Bouguénec et al, submitted; C. Le Bouguénec, personal communication). The second collection consisted of 40 blood isolates from cancer patients (26), 14 of which carried the *afa-8* operon (Le Bougénec et al., submitted). The third collection consisted of 16 strains isolated from stool specimens from children with diarrhea (21). The fourth collection consisted of 35 *E. coli* strains isolated from the feces of healthy volunteers (26).

We also studied 39 strains isolated from calves (36 strains) and piglets (3 strains) with intestinal and extraintestinal disorders. All 39 strains have been reported to carry the *afa-8* operon (35; J. P. Girardeau, personal communication).

E. coli HB101 (10) was used as a host for maintaining cosmid clones. *E. coli* K-12/MG1655 (7) was used as a control for PCR assays.

The cosmid vector pHC79 (16) was used in cloning experiments.

E. coli strains were grown in Luria broth without glucose (10 g of tryptone, 5 g of yeast extract, and 5 g of NaCl per liter [pH 7.0]) or on Luria agar plates (containing 1.5% agar) at 37°C. *E. coli*-harboring cosmid clones were grown with 100 μg of carbenicillin per ml.

DNA analysis and genetic techniques. Total plasmid DNA was extracted by the Kado method (31). Recombinant cosmids were routinely isolated by alkaline lysis (38), and whole-cell DNA was prepared by cesium chloride gradient (34). Standard procedures were used for restriction endonuclease digestions and other common DNA manipulations (38). Pulsed-field gel electrophoresis of genomic DNA from *E. coli* AL862, using restriction enzyme *Not*I, was performed as previously described (11). Primers, sequences, and the predicted sizes of the PCR products are given in Table 1. The cycling conditions were initial denaturation at 95°C for 5 min followed by 30 cycles at 95°C for 30 s, 65°C for 30 s, and 72°C for 1 min. For amplification of the *iuC* gene from the aerobactin operon, annealing was performed at 55°C. For amplification of *afaE-8-pheR, afaE-8-yjdC, afaE-8-pheV*, and *afaE-8-yqgA* regions, each cycle consisted of 1 min at 94°C, 1 min at 65°C, and 2 min at 72°C.

Cosmid library. Genomic DNA was extracted from *E. coli* AL862 isolated from the blood of a cancer patient and partially digested with restriction endonuclease *Sau*3A. Restriction fragments (35 to 50 kb) were sized on a sucrose gradient (10 to 40%) and ligated to the *Bam*HI-digested and alkaline phosphatase-treated cosmid vector pHC79 DNA as previously described (38). Cosmids were packaged in vitro into phage lambda particles by using the λ DNA in vitro packaging module (Stratagene, Austin, Tex.) and used to infect *E. coli* HB101. Carbenicillin-resistant HB101 transductants were screened by colony hybridization.

The absence of rearrangements of the inserts from the recombinant cosmids during the molecular cloning processes was confirmed by Southern blot hybridization experiments.

Hybridization. Bacteria grown for 3 h on nitrocellulose filters were used for colony hybridization as described by Grunstein and Hogness (22). For Southern blot hybridization, total plasmid DNA and DNA restriction fragments were submitted to electrophoresis and transferred to nitrocellulose sheets (0.45-mm-diameter pore size; Schleicher and Schuell, Inc.) by the Southern blotting technique (53). Hybridization was performed under stringent conditions (65°C), with PCR products (Table 1) labeled with ^{32}P by using the Megaprime DNA labeling system (Amersham International) as probes and signals detected by autoradiography with Amersham Hyperfilm-MP.

DNA sequencing. Double-stranded DNA was sequenced by Big Dye Terminator chemistry (Perkin-Elmer Applied Biosystems, Foster, Calif.). For each cycle, the sequencing reaction mixture contained 16 μl of Big Dye Terminator mix, 13 pmol of primer, and 0.4 to 0.8 μg of DNA in a total volume of 40 μl. The cycling conditions were initial denaturation at 95°C for 5 min followed by 75 cycles at 95°C for 30 s, 55°C for 30 s, and 60°C for 4 min. Excess dye terminators were removed with a spin column (Millipore S.A., Molsheim, France), and reaction mixtures were dried in a vacuum system. Each sample was resuspended in 15 μl of template suppression reagent (TSR) and denatured by heating at 95°C for 2 min, and the entire volume was loaded on an ABI 310 automated DNA sequencing instrument (Perkin-Elmer). Sequence data were analyzed by ABI version 3.0.1b3 software. Sequences were screened for similarity to previously published sequences by using the computer programs BLASTN and BLASTX at the National Center for Biotechnology Information. Multiple alignments were performed with the CLUSTAL W program. We analyzed the partial sequence of the island for the presence of open reading frames (ORFs) of at least 45 codons.

Statistical analysis. Proportions were compared by using the chi-square test.

Nucleotide sequence accession number. The GenBank accession numbers for the sequences reported herein are AF072900, AF286670, and AF286671.

RESULTS

Analysis of the sequence of the *afa-8* operon. The recombinant cosmid pILL1211 was previously described as a cosmid carrying the *afa-8* gene cluster cloned from the chromosome of the bovine pathogenic *E. coli* strain 239KH89 (35). A 4.2-kb sequence from this cosmid was published (accession no. AF072900) and reported to carry the *afaC, afaD*, and *afaE* genes encoding the outer membrane protein anchor (AfaC), the invasin (AfaD), and the afimbrial adhesin (AfaE), respectively (Fig. 1) (35). In this study, we completed the genetic characterization of the *afa-8* operon by sequencing 2 kb upstream from the *afaC* gene. Computer analysis revealed three ORFs, ORF1, ORF2, and ORF3, which mapped to the same loci as and had similar sequences to the *afaF, afaA*, and *afaB* genes from the *afa-3* operon (19), respectively (Fig. 1). These results confirmed that the genetic organization of the *afa-8* operon was similar to that of the *afa-3* operon.

The first ORF, *afaF*, was transcribed in the opposite orientation to the other ORFs. It encoded a peptide of 62 amino acids (aa) with a molecular mass of 7.22 kDa, similar to those of the *E. coli* regulatory proteins DaaF (5) (67% identity and 74% similarity) and AfaF encoded by the *afa-3* operon (19) (66% identity and 72% similarity). Five hundred ninety-nine base pairs downstream from the *afaF* gene was the *afaA* ORF, encoding a peptide of 127 aa with a calculated molecular mass of 14.4 kDa. This peptide exhibited homologies with *E. coli* regulators such as DaaA (5) (58% identity and 65% similarity), AfaA encoded by the *afa-3* operon (19) (54% identity and 61% similarity), and PapB (3) (36% identity and 47% similarity). The *afaA* gene is followed by an 80-bp noncoding region and the *afaB* ORF, encoding a 255-aa protein similar to the periplasmic chaperone proteins involved in the biogenesis of bacterial adhesive structures. The predicted sequence of the AfaB product was similar to those of the AfaB protein involved in the production of the afimbrial adhesin AfaE-III (19) (66% identity and 71% similarity), the NfaE protein chaperone involved in the production of the nonfimbrial adhesin NFA-I (1) (63% identity and 69% similarity), and the AggD chaperone of EAEC (49) (59% identity and 66% similarity). A gram-negative pilus assembly chaperone motif (FPEDRESLQWLCVK GIPP) was found in the AfaB protein encoded by the *afa-8* operon (28, 29, 58). A putative signal sequence was identified in the AfaB protein, resulting in a predicted mature peptide of 232 aa with a deduced molecular mass of 24.9 kDa and a pI of 8.71.

Determination and analysis of the complete sequence of the *afa-8* operon (6,246 bp) showed that the G+C content of this operon (46.4%) was slightly different from that of the genome of *E. coli* MG1655 (50.8%) (7).

TABLE 1. PCR primers used in this study

Specificity	Nucleotide sequences of primers	Size of PCR product (bp)	Representative strain	Reference
afaE-8	5′-CTAACTTGCCATGCTGTGACAGTA-3′ 5′-TTATCCCCTGCGTAGTTGTGAATC-3′	302	239KH89	35
iuC	5′-AAACCTGGCTTACGCAACTGT-3′ 5′-ACCCGTCTGCAAATCATGGAT-3′	269	J96	6
sfaD/sfaE[a,b]	5′-CGGAGGAGTAATTACAAACCTGGCA-3′ 5′-CTCCGGAGAACTGGGTGCATCTTAC-3′	410	J96	36
cadC	5′-CCATTTTCAATCCAGTAAAGGG-3′ 5′-ATCAGCGCCAATACCGTGCTC-3′	692	MG1655	This study
cadC/pheR[a]	5′-AGCCGCGCTTTGGTACAGTAGC-3′ 5′-CCGAACTCAACCAGATTCTCCCC-3′	823	MG1655	This study
cadC/yjdC[a]	5′-AGCCGCGCTTTGGTACAGTAGC-3′ 5′-CAGATGGAACTGGTGCTGGAAGG-3′	1,168	MG1655	This study
afaE-8/int[a]	5′-GATTCACAACTACGCAGGGG-3′ 5′-CTGTCGCGTATTGACGGTTATAAAG-3′	1,102	239KH89	This study
afaE-8/pheR[a]	5′-GATTCACAACTACGCAGGGG-3′ 5′-CCGAACTCAACCAGATTCTCCCC-3′	1,850	239KH89	This study
afaE-8/yjdC[a]	5′-GATTCACAACTACGCAGGGG-3′ 5′-CAGATGGAACTGGTGCTGGAAGG-3′	2,255	AL862	This study
int/pheR[a]	5′-CTGAAGATGCCAGACTGTACGGC-3′ 5′-CCGAACTCAACCAGATTCTCCCC-3′	657	239KH89	This study
int/yjdC[a]	5′-CTGAAGATGCCAGACTGTACGGC-3′ 5′-CAGATGGAACTGGTGCTGGAAGG-3′	2,065	AL862	This study
pheR/yjdC[a]	5′-GGGGAGAATCTGGTTGAGTTCGG-3′ 5′-CAGATGGAACTGGTGCTGGAAGG-3′	495	AL862	This study
afaE-8/yqgA[a]	5′-GATTCACAACTACGCAGGGG-3′ 5′-CCGGATTACGCATCTGTGGCATT-3′	2,200[c]	AL862	This study
afaE-8/pheV[a]	5′GATTCACAACTACGCAGGGG-3′ 5′-ATTTGATTGACGAGACGAGGCGAA-3′	1,800[c]	AL862	This study
Probe A	5′-ATCAGATGCCTAAAGAAGGAGAAAC-3′ 5′-CAATACTCGGATAAGATGATTGC-3′	831	AL862	This study
Probe B	5′-TTTGATGAGCGATGTACTTTCCGAA-3′ 5′-GCAGATACAACGTGAACATACCGA-3′	991	AL862	This study
Probe C	5′-GGACGATAATGTGATCGTCTATAAG-3′ 5′-GTGGAAGATACTCATCTGCTACACG-3′	821	AL862	This study
Probe D	5′-CTGCTCGGCAATGTCTTTGGTGC-3′ 5′-CTGTGTACCAGATGCAAGGGCG-3′	1,119	AL862	This study
Probe E[d]	5′-CCAATCAGACAGTCTTATCCCATC-3′ 5′-GGGCGCAGGAAAGTCACCATCC-3′	474	AL862	This study

[a] Primer specific for a given region.
[b] Detect *sfa* and *foc* sequences.
[c] Sized approximately on agarose gel.
[d] Deduced from sequencing of 576 bp at the right end of the insert of cosmid 3 (Fig. 4). This sequence displays no significant similarities to any known sequence in the database.

Analysis of sequence of the region located downstream from the *afa-8* gene cluster. To determine whether the *afa-8* operon was associated with a PAI, we sequenced 2 kb downstream from this operon in pILL1211. The *afaE-8* gene was followed by a 300-bp noncoding region and a 1,263-bp ORF transcribed in the opposite orientation and encoding a putative protein of 421 aa. The deduced amino acid sequence of this putative protein displayed the highest percentage of identity and similarity with the *E. coli* MG1655 prophage P4 integrase (13) (67% identity and 74% similarity) and the bacteriophage P4 integrase (45) (54% identity and 66% similarity) (Fig. 2). This integrase-encoding gene (*int* gene) is intact and has two possible AUG start codons, but only the first is adjacent to a sequence resembling a ribosome-binding sequence (52; data

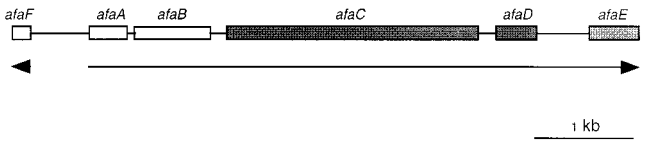

FIG. 1. Genetic organization of the *afa-8* operon. White boxes indicate ORFs sequenced in this study. Gray boxes indicate ORFs previously sequenced (35). Arrows show the direction of gene transcription.

not shown). Phage integration results from homologous recombination between the attachment site *attB* (20 bp) on the bacterial chromosome and an identical site (*attP*) on the phage chromosome (15). The site of bacteriophage P4 integration into the *E. coli* chromosome has been previously identified and was shown to reside within the *leuX* tRNA gene (45). Although not entirely identical to the *attP* gene, we identified, 197 bp downstream from the *int* gene, an *attB*-like site that displayed 14 identical nucleotides over a 20-bp sequence (Fig. 3). The *int* gene is followed by a noncoding region (219 bp) carrying a 136-bp sequence that is 95% identical to a region carrying the phenylalanine-specific tRNA-encoding gene (*pheR*) in *E. coli* MG1655 (Fig. 3). Only the 22 bp at the 3′ end of the 76 bp encoding the $tRNA^{Phe}$ were conserved (*pheR′*), with a single internal base pair deletion, and these residues carried the *attB*-like site (Fig. 3).

The G+C content of the *afa-8* operon and the association of this operon with a P4 integrase-encoding gene and the *pheR′* gene suggest that this gene cluster is carried by a PAI designated PAI $I_{239KH89}$.

In *E. coli* K-12, the *pheR* gene is preceded by the lysine decarboxylase regulatory gene *cadC* and is immediately followed by the hypothetical *yjdC* gene and maps to position 94 min in the chromosome (2, 7, 13). Surprisingly, sequence analysis of the region downstream from the *pheR′* gene in *E. coli* 239KH89 showed a noncoding region (900 bp) that was not similar to sequences in the database. Moreover, amplification and sequence analysis of the *cad-yjdC* region (Table 1) in strain 239KH89 showed the presence of an intact *pheR* gene, not disturbed by insertion (data not shown), suggesting that the chromosome of this strain carries at least two copies of the *pheR* gene, a complete copy between the *cad* and *yjdC* genes, and a truncated copy (*pheR′*) containing PAI $I_{239KH89}$.

To characterize the *afa-8*-containing PAI, we selected *E. coli* AL862, a human blood isolate, carrying the *afa-8* operon inserted into the *pheR* loci, near the *yjdC* gene, according to the PCR results (Table 1). This PAI was designated PAI I_{AL862}.

Determination of the right junction of PAI I_{AL862}. Three hundred recombinant clones from the cosmid library of *E. coli* AL862 were screened by colony hybridization with the *afaE-8* gene as a probe (Table 1) to identify the cosmids carrying the right junction of PAI I_{AL862} with *E. coli* K-12-type sequences. Two positive cosmids (cosmids 1 and 2) (Fig. 4A) were selected for further studies. Analysis of the sequences downstream from the *afa-8* operon on cosmid 1 revealed the presence of a 300-bp noncoding region followed by an integrase-encod-

```
MG1655    ..............................MHLLVHPNGSKYWRLQYRYEG..KQKMLALG
239KH89   ....MALTDAKIRAAKPTDKAYKLTDGAGMFLLVHPNGSRYWRLRYRILG..KEKTLALG
P4        MCPQMKLNARQVETAKPKDKTYKMADGGGLYLEVSAKGSKYWRMKYRRPSDKKEDRLAFG

MG1655    VYPEITLADARVRRDEARKLLANGVDPGDKKKNDKVEQSKARTFKEVAIEWHGTNKKWSE
239KH89   VYPEVSLSEARTKRDEARKLISEGIDPCEQKRVKKVVPDLQLSFEHIARRWHASNKQWAQ
P4        VWPTVTLAQARAKRDEAKKLLVQGIDPKVVQKEARAENSGAYTFEAIAREWHASNKRWSE

MG1655    DHAHRVLKSLEDNLFAALGERNIAELKTRDLLAPIKAVEMSGRLEVAARLQQRTTAIMRY
239KH89   SHSDKVLKSLETHVFPFIGNRDITTLNTPDLLIPVRAAEAKQIYEIASRLQQRISAVMRY
P4        DHRSRVLRYLELYIFPHIGSSDIRQLKTSHLLAPIKKVDASGKHDVAQRLQQRVTAIMRY

MG1655    AVQSGLIDYNPAQEMAGAVASCNRQHRPALELKRIPELLTKIDSYTGRPLTRWAIELTLL
239KH89   AVQSGIIRYNPALDMAGALTTVKRQHRPALNLSRLPELLSRIDGYKGQPVTRLAVMLNLL
P4        AVQNDYIDSNPASDMAGALSTTKARHYPALPSSRFPEFLARLAAYRGRVMTRIAVKLSLL

MG1655    IFIRSSELRFARWSEIDFEASIWTIPPEREPIPGVKHSHRGSKMRTTHLVPLSTQALAIL
239KH89   VFIRSSELRYARWSEIDIDNAMWTIPAEREPLPGVKFSHRGSKMRTPHLVPLSKQAVAIL
P4        TFVRSSELRFARWDEFDFDKSLWRIPAKREEIKGVRYSYRGMKMKEEHIVPLSRQAMILL

MG1655    KQIKQFYGAHDLIFIGDHDSHKPMSENTVNSALRVMGYDTKVEVCGHGFRTMACSSLVES
239KH89   TELQTWAGENGLIFTGAHDPRKPISENTVNKALRVMGYDTTREVCGHGFRAMACSALIES
P4        NQLKQISGDKELLFPGDHDATKVMSENTVNSALRAMGYDTKTEVCGHGFRTMARGALGES

MG1655    GLWSRDAVERQMSHMARNSVRAAYIHKAEHLEERRLMLQWWADFLDVNRERFISPFEYAK
239KH89   GLWSRDAVERQMSHQERNGVRAAYIHKAEHLEERRLMLQWWADFLDANRERFISPFEYAK
P4        GLWSDDAIERQLSHSERNNVRAAYIHTSEHLDERRLMMQWWADYLDMNRNKYISLMIIQN

MG1655    INNPLKQ............
239KH89   INNPLKQ............
P4        TKKYLNKNSYWLIFKMSVK
```

FIG. 2. Comparison of predicted protein sequences for integrases of prophage P4 of *E. coli* MG1655, *E. coli* 239KH89, and bacteriophage P4. Shaded residues are identical to those from the integrase of prophage P4 of *E. coli*. Gaps have been inserted to optimize the alignment.

```
E. coli MG1655      CAAAT............TGCTGGTGATGTGGGGAGAATCTGGTTGAGTTCGGTAGAATTG
E. coli 239KH89     ---T-ATACCCACACAA--AG-T-----------------------------------

E. coli MG1655      ATTTGGAGGCAGAACGCTTAAATCGTGGCGTCCTGAAACGAAAAACGGACCTCCGTGGAG
E. coli 239kh89     ---------A-----A--------A---T--T---------------------------

                                                <-------------------------------

E. coli MG1655      -------TATATGAATTTGGTGCCCGGACTCGGAATCGAACCAAGGACACGGGGATTT..
E. coli 239KH89     -------G----------------------------.---A-TG-TTATATAACCAA-TG

                    ---------------------------------------

E. coli MG1655      ..TCA.ATCCCCTGCTCTACCGACTGAGCTATCCGGGCAACGGGG
E. coli 239KH89     TTTTGTATCGA---T-T--T-C.T--TT-TT-TT---....-CCA
```

FIG. 3. Sequence alignment of the 219-bp region downstream from the *int* gene in *E. coli* 239KH89 with the region carrying the phenylalanine-specific tRNA gene (*pheR*) in *E. coli* MG1655. *pheR* is indicated by a horizontal arrow above the sequence in the direction of transcription. Underlined nucleotides represent the *attB*-like site. Dashes represent identical nucleotides. Gaps have been inserted to optimize the alignment.

ing gene identical (100% identity) to those described in *E. coli* 239KH89. Downstream from the *int* gene, we identified the 76-bp sequence encoding the $tRNA^{Phe}$ and carrying the *attB*-like site at its 3′ end (Fig. 5B), which was followed by the hypothetical *yjdC* gene.

Determination of the left junction of PAI I_{AL862}. To identify the left junction of PAI I_{AL862}, an internal fragment of the *cadC* gene was amplified (Table 1) and used as a probe to screen, by DNA hybridization, 200 recombinant cosmids from the cosmid library of *E. coli* AL862. Three positive cosmids (cosmids 3, 4, and 5) (Fig. 4A) were selected for further studies. An oligonucleotide primer, internal to the *cadC* gene, was used to sequence the PAI I_{AL862}-specific DNA sequences from the left junction. The *cadC* gene was found to be followed by a 136-bp nearly perfect duplication of the right junction (Fig. 4B and 5). A segment consisting of 22 bp of the 3′ end of the *pheR* gene with a single internal base pair deletion, carrying the *attB*-like site, was found at the left junction (Fig. 5B). A 204-bp noncoding region, immediately adjacent to the duplicated region within the PAI, showed no similarity to sequences in the database. This noncoding region is followed by two ORFs (ORF1 completely sequenced and ORF2 partially sequenced)

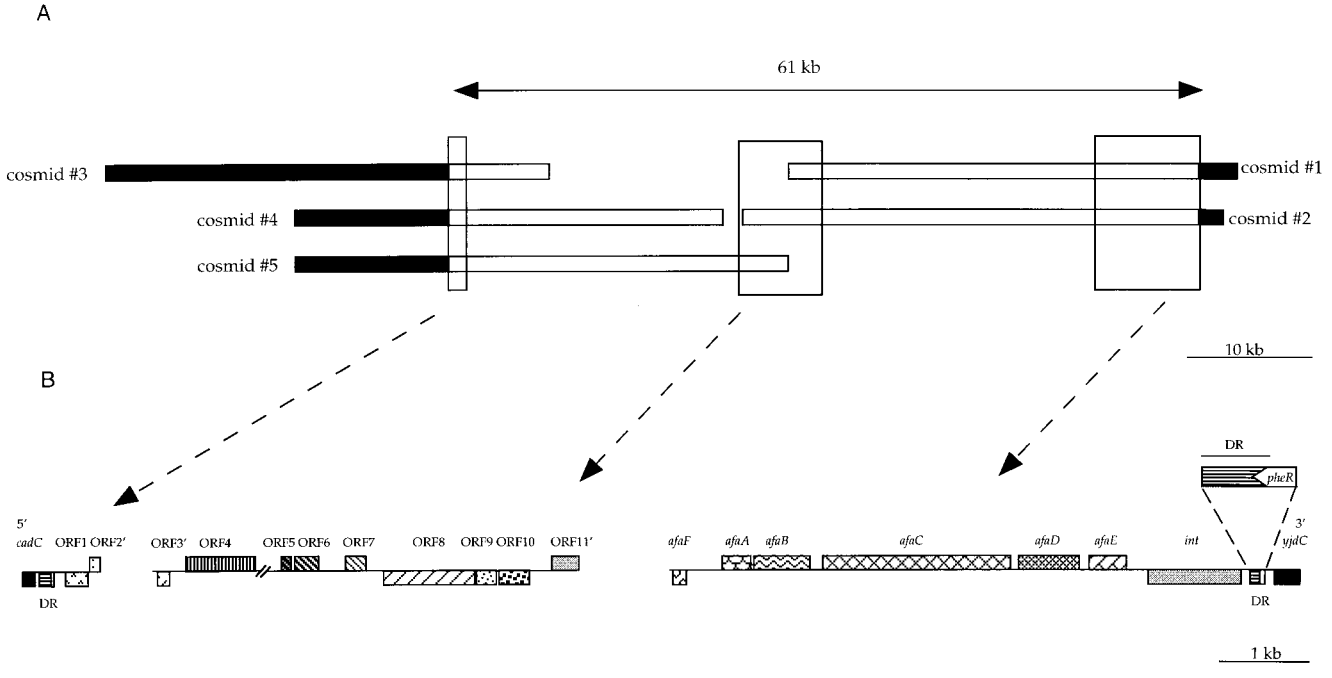

FIG. 4. Partial structure of PAI I_{AL862}. (A) Schematic diagram of the inserts of overlapping cosmids derived from PAI I_{AL862}. The solid bars represent the *E. coli* K-12 chromosome. PAI I_{AL862}-specific DNA is represented by white bars. The estimated size of PAI I_{AL862} is shown above. (B) ORFs deduced from the sequences at the left (5′) and right (3′) ends as well as the central region of PAI I_{AL862} are represented by boxes. Direction of transcription (left or right) is indicated by boxes (below and above the line, respectively). DR, direct repeat. Cosmids 1, 2, 3, 4, and 5 correspond to pILL1259, pILL1255, pILL1267, pILL1269, and pILL1270, respectively.

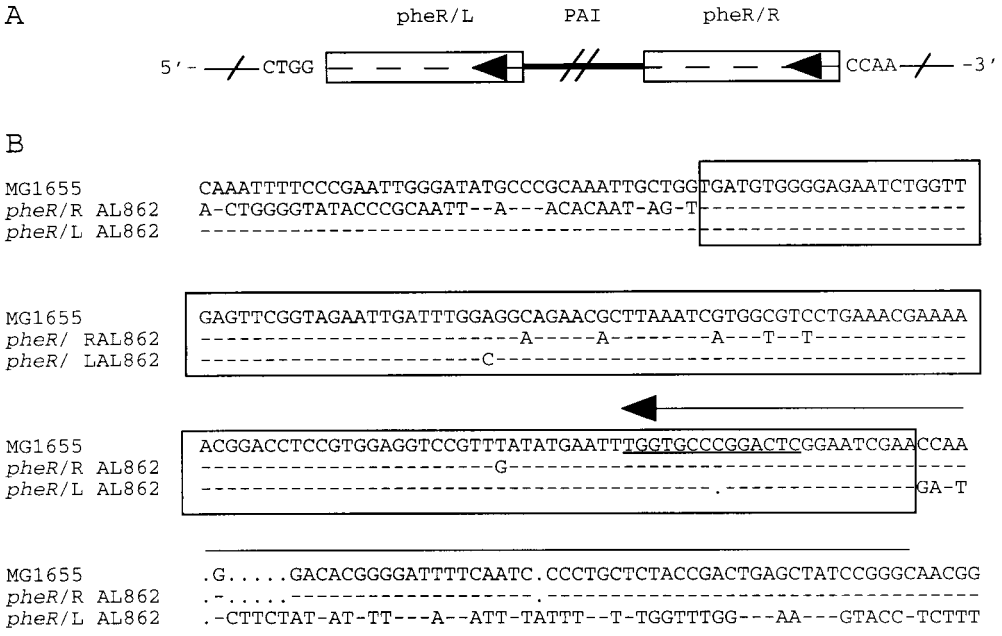

FIG. 5. Schematic diagram and nucleotide sequences of the junction sites of PAI I_{AL862}. (A) The boxes represent the 136-bp direct repeat with the left and right junctions indicated (see below); the 5′- and 3′-end nucleotides are shown for reference. The single diagonal lines represent the chromosomal DNA of *E. coli* K-12. The arrows represente *pheR* sequences. The double diagonal lines represent intervening PAI-specific DNA. (B) *pheR* regions were derived from *E. coli* MG1655. *pheR*/L and *pheR*/R, represent the left and right junction sequences of PAI I_{AL862}, respectively. *pheR* is indicated by a horizontal arrow above the sequence in the direction of transcription. Underlined nucleotides indicate the *attB*-like site. The 136-bp direct repeat sequence is shown in boxes. Dashes represent nucleotides identical to the sequence of *E. coli* MG1655. Gaps have been inserted to optimize the alignment.

(Fig. 4B). The sequence of the N terminus (48 aa over 88 aa) of the putative protein encoded by ORF1 was very similar (91% identity and 97% similarity) to the 61-aa product of ORF L12 previously described in a putative P4 family prophage of the LEE of the EHEC strain EDL933 (44) (Table 2). The product of ORF2′, which was partially sequenced (141 bp), was truncated by a stop codon and exhibited 80% identity (88% similarity) to the C terminus of a putative product of ORF L11 in the same prophage (Table 2).

Determination of the size of PAI I_{AL862}. A walking method, involving the hybridization of probes derived from the sequences of the ends of cosmids carrying the right and left junctions to filters containing cosmids from the library, was used to order and to identify overlapping clones harboring PAI sequences in the library. Positive hybridization was confirmed by PCR. One miniset of five overlapping cosmid clones covers the *afa-8*-containing PAI, including the right and left junctions and giving a total size of approximately 61 kb (Fig. 4A).

Virulence factors carried by PAI I_{AL862}. In addition to the *afa-8* operon, *E. coli* AL862 carries *sfa*/*foc* sequences encoding the fimbrial adhesin of the S family and the *iuC* gene from the aerobactin-encoding operon (Table 1). Hybridization and PCR assays showed that none of these determinants is carried by PAI I_{AL862}, suggesting that the afimbrial AfaE-VIII adhesin is the only known virulence factor encoded by this island.

Partial characterization of PAI I_{AL862} was initiated by sequencing the 3′ ends of the inserts of cosmids 1 and 2 and the right end of the insert of cosmid 5 (Fig. 4A). Analysis of the sequence of these regions led to the identification of seven complete ORFs (ORF4 to -10) and two partial ORFs (ORF3′ and -11′) (Fig. 4B and Table 2). Comparison of the products of these ORFs to the sequences from the database showed that PAI I_{AL862} contains a high density of genes that may be involved in sugar utilization: the ORF3′ to ORF4 products showed homologies with proteins involved in ribose metabolism, while ORF8 to ORF11′ products showed homologies with proteins of phosphotransferase systems. The ORF7 product was similar to the L13 IS*2*-like protein previously described in a putative P4 family prophage of the LEE PAI of the EHEC EDL933. The ORF5 and ORF6 products showed no similarity to any known protein in the database.

Although the putative proteins encoded by the central region of PAI I_{AL862} were similar to proteins mainly described in *E. coli* MG1655, the nucleotide sequences of this region showed no similarity to the sequence genome of this strain. Comparison of these sequences with unfinished genome sequences from the database revealed that the region carrying ORF3′ to ORF6 is highly similar (81 to 95% identity) to a chromosomal DNA sequence of *Salmonella enterica* serovar Typhimurium. Despite this high level of similarity, the G+C content of this region (39.7%) differs from that of the *Salmonella* genome (52%) (43). The region comprising ORF8 to ORF11′ was very similar (81% identity) to a chromosomal DNA sequence of *Yersinia pestis* and had a similar G+C content (46.6%) (4).

PAI I_{AL862} is duplicated on the chromosome of *E. coli* AL862. Hybridization assays with total DNA from *E. coli* AL862 digested with *Not*I, with the *afaE* gene from the *afa-8* operon used as a probe, revealed two hybridizing fragments greater than 300 kb in size, suggesting that there are two copies of this

TABLE 2. ORFs identified by partial sequencing of PAI I^{AL862}

ORF	Predicted size (aa)	Characteristic of homologous protein				
		Function	Species (strain)	Length (aa)	%Identity/ % similarity	Accession no. (reference)
ORF1	88	L12 putative protein	*Escherichia coli* (EDL933)	61	91/97	AAC31491 (44)
ORF2′	>47	L11 putative protein	*Escherichia coli* (EDL933)	76	80/88	AAC31490 (44)
ORF3′	>57	DeoR transcriptional repressor of deoxyribose operon	*Escherichia coli* (MG1655)	252	58/71	P06215 (41, 56, 57)
		DeoR transcriptional activator	*S. enterica* serovar Typhimurium	252	53/69	AAB80741
ORF4	314	RbsK putative ribokinase	*Schizosaccharomyces pombe*	318	36/46	O60116
		RbsK ribokinase	*Lactobacillus sakei*	302	34/42	AAD34338 (54)
		RbsK putative ribokinase	*Haemophilus influenzae*	306	33/43	P44331 (17)
		RbsK ribokinase	*Escherichia coli* (MG1655)	309	32/41	P05054 (27)
ORF5	48	No significant similarities				
ORF6	111	No significant similarities				
ORF7	94	L13 IS2-like protein	*Escherichia coli* (EDL933)	133	37/48	AAC31492 (44)
		Transposase	*Burkholderia glumae*	401	27/39	BAA24920
ORF8	418	Putative integral membrane protein	*Streptomyces coelicolor*	516	49/61	CAB52363 (47)
		SgaT putative transport protein	*Escherichia coli* (MG1655)	484	30/40	P39301 (13, 48)
ORF9	95	Hypothetical protein SCJ21.24c	*Streptomyces coelicolor*	>150	35/46	CAB52364 (47)
		B component (PtxB) of a putative phosphotransferase enzyme II	*Escherichia coli* (MG1655)	101	25/36	P39302 (13, 48)
ORF10	147	A component (PtyA) of a putative phosphotransferase enzyme II	*Escherichia coli* (MG1655)	147	40/52	P32058 (13, 48)
		A component (PtxA) of a putative phosphotransferase enzyme II	*Escherichia coli* (MG1655)	154	36/49	P39303 (13, 48)
		A component (SgcA) of a putative phosphotransferase enzyme II	*Escherichia coli* (MG1655)	143	36/47	P39363 (13, 48)
ORF11′	>124	Regulator of gluconate operon	*Escherichia coli* (MG1655)	313	38/47	AAC6463 (7)
		Gluconate repressor	*Pseudomonas aeruginosa*	315	36/47	AAD01801

operon on the chromosome of *E. coli* AL862 (data not shown). The absence of hybridization on total plasmid DNA of *E. coli* AL862 with the same probe confirmed the chromosomal location of the two copies of the *afa-8* operon. Moreover, analysis of *E. coli* AL862 cosmid library also showed other cosmids that hybridized with the regions carrying ORF4 to ORF11 from PAI I_{AL862} (Fig. 4B) and carried the *afa-8* operon, but tested negative for the *yjdC* gene. Sequence analysis of one of these cosmids (cosmid no. 6) (Fig. 6) confirmed the presence of an *int* gene, identical (100% identity over 440 bp) to that found in the right boundary of PAI I, followed by the phenylalanine-specific tRNA *pheV* gene and the hypothetical *yqgA* gene. In *E. coli* K-12, the *pheV* gene maps to position 67 min on the chromosome (7), between the *yqgA* and *yghD* genes (7, 46). We investigated whether the two copies of the *afa-8* operon were present on the chromosome of the same clone of *E. coli* AL862, by using PCR to test 12 single colonies for genetic associations of the *afaE-8* gene with *yjdC* and of the *afaE-8* gene with *yqgA* (Table 1). Both PCRs were positive for all 12 colonies, indicating that the chromosome of *E. coli* AL862 carries two *afa-8*-carrying PAIs: one inserted into the *pheR* gene (PAI I_{AL862}) and a second inserted into the *pheV* gene, designated PAI II_{AL862}.

The right junction of PAI II_{AL862} was determined by analysis of *E. coli* AL862 cosmid library. This revealed cosmids 7 and 8 (Fig. 6), which carried ORF1, ORF3′, and ORF4 previously described in the PAI I_{AL862}, but not the *cadC* gene. These data confirmed that the insertion sites of PAI I and PAI II are different and suggested that these two PAIs are similar. Moreover, reciprocal hybridizations between restriction fragments from the cosmids carrying PAI I (Fig. 4A) and the cosmids carrying PAI II (Fig. 6) indicated that the two sets of cosmids shared a region spanning PAI I. An oligonucleotide primer designated in ORF1 was used to identify the right boundary of PAI II. Sequence analysis of cosmid 7 (Fig. 6) revealed 448 bp, carrying the 3′ end of ORF1 followed by a 204-bp noncoding region identical (100% identity) to those found in the left junction of PAI I. Immediately downstream from the noncoding region, we identified a segment consisting of 22 bp of the 3′ end of the $tRNA^{Phe}$-encoding gene, which carries the *attB*-like site. Forty-three base pairs downstream from this short repeat, we found instead of the *yghD* gene another ORF, partially sequenced (138 bp), able to encode a putative protein similar to the N termini of integrases previously described in the LEE PAI of EHEC strain EDL933 (54% identity and 69% similarity) (44) and the HPI of *Yersinia pseudotuberculosis* (63% iden-

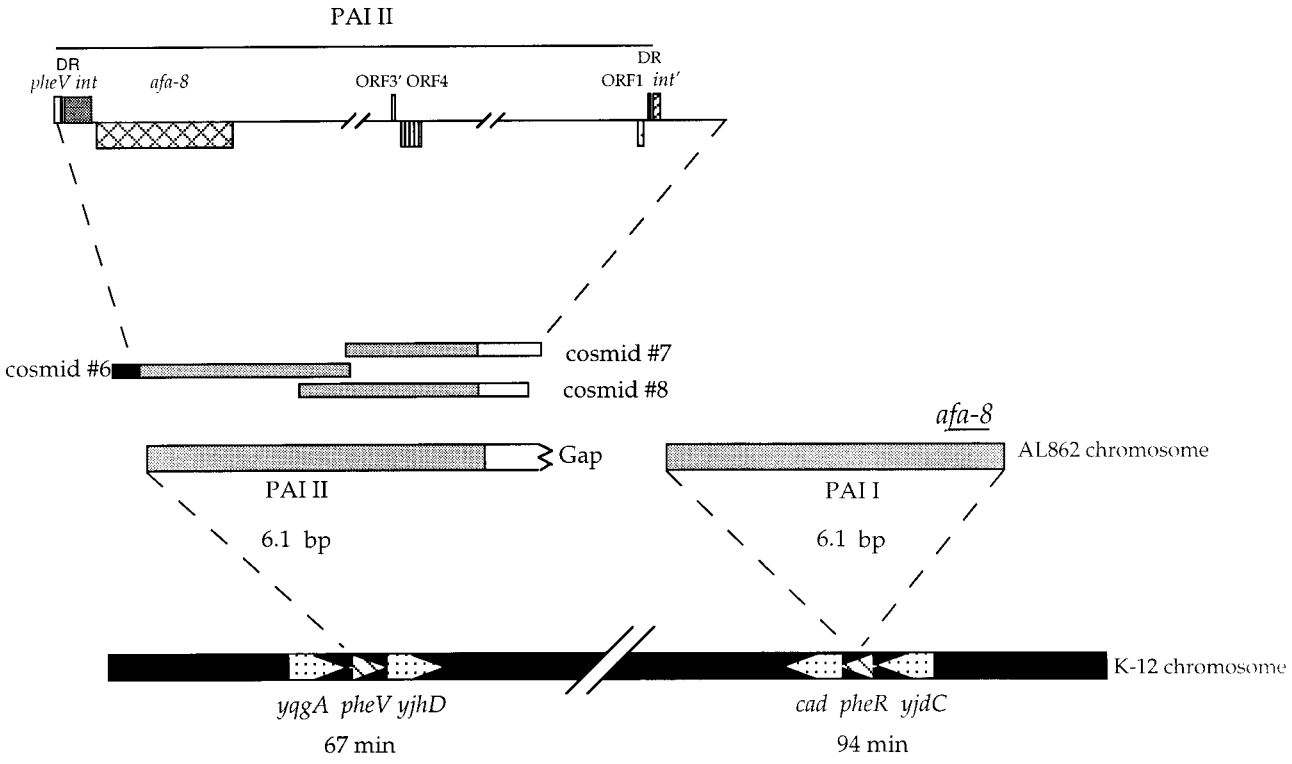

FIG. 6. Schematic diagram showing the positions of PAI I and PAI II from *E. coli* AL862. The gray bars represent sequences of PAI I and PAI II. The black bars represent the *E. coli* K-12 chromosome. The stippled arrows within the black bars indicate the $tRNA^{Phe}$ genes, and the boxes indicate the *yqgA, yjhD, cadC*, and *yjdC* genes, showing their orientation and their position on the *E. coli* K-12 chromosome. The PAIs and their estimated sizes are shown above the *E. coli* K-12 chromosome map. DR, direct repeat. Cosmids 7, 8, and 9 correspond to pILL1254, pILL1266, and pILL1271, respectively.

tity and 76% similarity) (11). It appears that the left junction of PAI II_{AL862} is not adjacent to the sequences described in *E. coli* K-12. Moreover, the sequence of the 3′ ends of cosmids 7 and 8 showed no similarity to sequences in the database.

Analysis of the distribution of *afa-8*-carrying PAI. To determine whether sequences within the PAI I_{AL862} were specific for pathogenic *E. coli* strains carrying the *afa-8* operon, we investigated the frequency of occurrence of the A, B, C, and D regions (Fig. 7 and Table 3). These regions were amplified (Table 1) and used as probes to screen by colony hybridization three collections of clinical isolates. These collections comprised 70 *afa-8*-positive strains isolated from animals (calves and piglets) with diarrhea or septicemia and from humans with UTI or septicemia; 68 *afa-8*-negative clinical isolates from humans with septicemia, UTI, or diarrhea; and 35 *afa-8*-negative strains isolated from healthy individuals. The distribution of the A, B, C, and D regions in strains from various origins and the correlation of these regions with the presence of the *afa-8* operon are shown in Fig. 7 and Table 3. The four probes were detected more frequently in pathogenic strains, regardless of whether or not these strains carried the *afa-8* operon, than in strains isolated from healthy individuals. The proportion of pathogenic strains testing positive with all of the probes was significantly higher for *afa-8*-positive strains (74.3%) than for *afa-8*-negative strains (25%) ($P = 0.05$). This difference was confirmed by hybridization experiments of ABCD-positive strains with probe E (Fig. 7), which reacted with 100% of the *afa-8*-positive strains and with only 65% of *afa-8*-negative strains. In contrast, the proportion of strains testing negative with all of the A, B, C, and D probes was significantly higher in

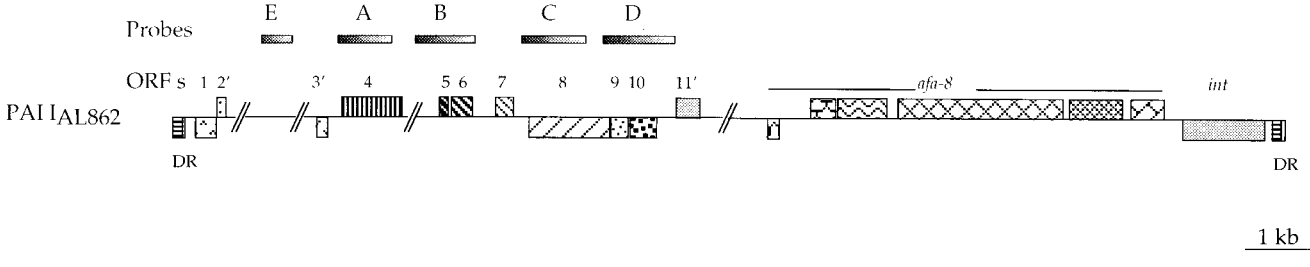

FIG. 7. Distribution of sequences from PAI I_{AL862} among pathogenic and nonpathogenic *E. coli* strains. Regions from PAI I_{AL862} used as probes are shown. Results of hybridization experiments are shown in Table 3.

TABLE 3. Results of hybridization studies

Strain type (*n*)	No. (%) of isolates that hybridized with probe(s)										
	ABCD	BCD	ABD	ABC	CD	AB	A	B	C	D	None
Pathogenic											
afa-8 positive (70)	52 (74.3)[a,b]	2 (2.9)	1 (1.4)	0 (0)	1 (1.4)	0 (0)	1 (1.4)	0 (0)	1 (1.4)	0 (0)	12 (17.1)[c]
afa-8 negative (68)	17 (25)[b]	0 (0)	0 (0)	2 (2.9)	0 (0)	3 (4.4)	1 (1.5)	1 (1.5)	0 (0)	1 (1.5)	43 (63.2)
Nonpathogenic (35)	3 (8.5)[b]	1 (3)	0 (0)	0 (0)	2 (5.5)	0 (0)	4 (11.5)	0 (0)	0 (0)	0 (0)	25 (71.5)

[a] Including strain 239KH89.
[b] Among the ABCD-positive strains, 100% of the *afa-8*-positive, 65% of the *afa-8*-negative, and 33% of the nonpathogenic strains also hybridized with the E probe.
[c] Strains carrying the *afa-8* gene cluster on a plasmid (this study; J. Gérardin, personal communication).

nonpathogenic strains and pathogenic *afa-8*-negative strains (71.5% and 63.2%, respectively) than in pathogenic *afa-8*-positive strains (17.1%). Interestingly, hybridization experiments with plasmid DNA isolated from the latter strains with the *afaE* probe indicated in all of these strains that the *afa-8* operon is plasmid borne. All of these data strongly suggested the presence of a genetic element similar to PAI I_{AL862} in most of the *afa-8*-positive pathogenic strains.

Variability in the chromosomal location of *afa-8*-carrying PAI. Sixteen strains (7 bovine diarrheagenic strains and 9 human blood isolates) that reacted with the A, B, C, and D probes and tested positive for the *afaE-8–int* genetic association were used to investigate the chromosomal location of the *afa-8*-containing PAI. Genetic association of the *afa-8* operon with either *pheR* or *pheV* or *yjdC* or *yqgA* was investigated with several set of primers, as indicated in Table 1. In one human strain, PCR results suggested that the *afa-8* operon inserted into the *pheR* gene next to the *yjdC* gene, indicating that, like PAI I_{AL862}, the *afa-8*-containing PAI mapped to position 94 min on the chromosome. In 13 strains (7 human and 6 bovine), PCR results suggested that the *afa-8* operon inserted into the *pheV* gene next to the *yqgA* gene, indicating that, like PAI II_{AL862}, the *afa-8*-containing PAIs mapped to position 67 min on the chromosome. In the two remaining strains (one human and one bovine), although the *afa-8* operon inserted into the *pheR* gene, it was not associated with the *yjdC* gene. These results suggested that as in *E. coli* 239KH89, *afa-8*-containing PAI was located at an unknown position on the chromosome. The chromosomal location of the *afa-8*-carrying PAI in strain 239KH89 was further investigated by sequencing of 4 kb downstream from the *pheR* gene, on cosmid pILL1211. This revealed a 947-bp noncoding region followed by nine ORFs. The products of ORF1, ORF2, ORF3, ORF4, ORF5, ORF6, and ORF7 displayed significant similarities to the peptides encoded by the L12 (78% identity), L11 (73% identity), L10 (73% identity), L9 (88% identity), L8 (90% identity), and L7 (95% identity) ORFs, respectively, previously described in the putative P4 family prophage carried by the LEE PAI of EHEC strain EDL933 (44). L7 and L8 were related to genes found only in the P4-like family of cryptic prophages from *E. coli* K-12, whereas L9 to L12 were completely unknown (44). The products of ORF2 and ORF4 were truncated by a stop codon and a frameshift, respectively. The products of ORF6 and ORF7 also matched two hypothetical proteins of *E. coli* MG1655, YeeW (63% identity) and YeeV (88% identity), respectively. The products of ORF8, ORF9, and ORF10 were similar to YeeU (89% identity), YeeT (94% identity), and YeeS (98% identity) of *E. coli* MG1655, respectively (7). The genes encoding the putative proteins YeeW to YeeS are contiguous on the chromosome of *E. coli* MG1655 and map to position 45 min.

All of these data taken together indicate the variability of the chromosomal location of *afa-8* containing PAIs and suggest that these PAIs are preferentially inserted into the *pheV* gene.

DISCUSSION

Genes encoding important virulence factors are often located on mobile genetic elements such as phages, plasmids, transposons, and PAIs. They may therefore be transferred from one cell to another, and this horizontal transfer represents a key genetic mechanism in the evolution of pathogens. *E. coli* represents an example of a pathogen that has developed by lateral gene transfer. Several PAIs carrying virulence genes introduced into the genome via lateral transfer have been described in pathogenic *E. coli* strains associated with intestinal or extraintestinal infections in humans. In previous studies, we have described a new *afa-8* operon, encoding an afimbrial adhesin (AfaE-VIII), carried by bovine *E. coli* strains associated with diarrhea or septicemia (35) and human *E. coli* isolates associated with extraintestinal infections (20, 35). This operon may be borne on a plasmid or on the chromosome (20, 35), suggesting that it is associated with mobility genes. In this report, we used *E. coli* AL862, a human blood isolate, as a prototype *afa-8*-carrying *E. coli* strain. We studied a putative PAI carrying this operon to identify and to improve our understanding of the mechanisms involved in dissemination of the *afa-8* operon among pathogenic isolates.

The results reported here indicate that the *afa-8* operon is carried by a genomic region that fits within the category of PAIs as defined by Hacker et al. (25). In strain AL862, the *afa-8* operon is (i) located within a 61-kb chromosomal region, (ii) in the vicinity of a mobility gene (*int* gene), or (iii) associated with the phenylalanine-specific tRNA gene (*pheR*). Moreover, the G+C content of the *afa-8* operon, which has been completely sequenced in the bovine pathogenic *E. coli* strain 239KH89, is slightly lower (46.4%) than that of the chromosome of *E. coli* MG1655 (50.8%) (7). This new PAI was designated PAI I_{AL862}. The presence of a putative integrase gene, highly similar to that of bacteriophage P4, and a *pheR* gene at the right extremity of this PAI, as well as a 14-bp sequence resembling the *attP* site of bacteriophage P4 at both the right and left extremities, strongly argues in favor of the hypothesis

that PAI I_{AL862} was acquired via horizontal transfer from a bacteriophage.

E. coli AL862 carries the *sfa*/*foc* sequences, encoding a fimbrial adhesin, and the *iuC* gene from the aerobactin-encoding operon, but none of these determinants is carried by PAI I_{AL862}, indicating that AfaE-VIII adhesin is the only known virulence factor encoded by this new island. However, the partial nucleotide sequence of PAI I_{AL862} revealed new ORFs with sequences similar to those of determinants encoding proteins involved in the utilization of various sugars. These regions showed heterogeneous G+C contents and were similar to sequences from different bacteria (*S. enterica* serovar Typhimurium and *Y. pestis*), suggesting a stepwise acquisition of these DNA fragments from heterogeneous sources, leading to the mosaic-like structure of this island. These sequences, as assayed by colony hybridization, are highly frequent in *afa-8*-positive strains, (81.5%), less frequent in human pathogenic *afa-8*-negative strains (25%), and generally absent from strains isolated from healthy individuals and considered to be nonpathogenic. We therefore suggest that these newly described genes are particularly found in pathogenic isolates and are preferentially associated with the *afa-8* operon. They probably define PAIs similar to PAI I_{AL862}. These new sequences may contribute to the survival of the strains in certain ecological niches and do not directly contribute to host damage and infection.

One of the interesting features revealed by analysis of the *E. coli* AL862 cosmid library is that the chromosome of this strain carries two *afa-8*-containing PAIs: PAI I, described above and located in the vicinity of the *pheR* gene; and PAI II, located in the vicinity of the *pheV* gene. Similarly, the PAI I and PAI II of UPEC strain J96 were inserted into the *pheV* and *pheR* genes, respectively. These two PAIs differ in size and in the virulence factors they encode (8, 9, 55). In contrast, the PAI I and PAI II of *E. coli* AL862 are similar in size and genetic organization. In addition, they have identical sequences at their extremities, suggesting at a first approximation that the *afa-8*-containing PAI is present in two copies on the chromosome of *E. coli* AL862. To our knowledge, this is the first report of such a phenomenon. Buchrieser et al. (11) previously reported insertion of the HPI into any of the three *asn* tRNA genes present on the bacterial chromosome of *Y. pseudotuberculosis*, but insertion occurred in different variants of the same serotype and not in the same variant. Further studies are necessary to confirm that PAI I and PAI II from *E. coli* AL862 are similar along their entire length. Unlike PAI I_{AL862}, both extremities of which were adjacent to the *E. coli* K-12 chromosome, the right junction of PAI II_{AL862} was adjacent to a putative integrase-encoding gene and to other unknown sequences. It is likely that PAI II_{AL862} is adjacent to an extrachromosomal segment that may define a putative PAI, designated PAI III_{AL862}. Further studies are required to determine whether this putative PAI carries the *aer* operon and the *sfa*/*foc* sequences. *E. coli* AL862 also carries the *fyuA, irp1*, and *irp2* genes (Girardeau, personal communication) found in the HPIs of various *Yersinia* species (11, 12, 14) and various pathotypes of *E. coli* (50, 51), as well as a disturbed *asnT* locus, suggesting that HPI has been acquired by this strain. All of these data suggest that the chromosome of *E. coli* AL862 carries at least four PAIs (PAIs I, II, and III and an HPI) and confirms the capacity of *E. coli* species to evolve by horizontal gene transfer.

Analysis of the chromosomal location of *afa-8*-containing PAI in pathogenic *E. coli* strains indicated that this PAI inserted preferentially into the *pheV* gene, rather than the *pheR* gene. It is well known that the *pheR* gene is followed by the *cadC* gene on the chromosome of the prototype *E. coli* strain MG1655 (7). However, analysis of the sequence of the insertion site of the *afa-8*-containing PAI in the bovine *E. coli* strain 239KH89 indicated that this PAI is adjacent to a truncated copy of the *pheR* gene (*pheR′*) and mapped to position 45 min on the chromosome. Sequence analysis suggested that this truncated copy of the *pheR* gene corresponded to the left junction of a remnant PAI carrying ORFs similar to those previously described in the putative prophage of the LEE PAI of *E. coli* EDL933, rather than to the right junction of the *afa-8*-containing PAI. The significance of the integration of this PAI into this truncated *pheR* gene is currently unclear, but suggests that, as for the PAI II of *E. coli* AL862, the adjacent regions of this PAI may have undergone considerable recombination over time. Whether such events occurred before or after the acquisition of *afa-8*-containing PAI is unclear. Differences in the chromosomal location of PAIs carrying the *afa-8* operon indicate diversity in PAI evolution. These PAIs probably have a common ancestor that inserted into the *pheR* and *pheV* loci of *E. coli*. The possibility that these PAIs also inserted into other tRNA-encoding genes should not be eliminated. The identification of an intact *int* gene, which presumably encodes a functional integrase protein, suggests that this integrase is involved in the mobility of these PAIs. In addition, direct repeat elements flanking PAIs may act as targets for specific recombinases, thereby playing an important role in the integration and/or excision of PAIs. Interestingly, PAI I_{AL862} is flanked by a 136-bp imperfect direct repeat carrying an *attB*-like site. We therefore suggest that PAI I_{AL862} has probably retained the capacity to excise from the chromosome.

In summary, we found that the *afa-8* operon of the human blood isolate AL862 is carried by a 61-kb PAI (PAI I_{AL862}) integrated into the *pheR* gene and possesses several characteristics suggestive of potential mobility. We also demonstrated that *E. coli* AL862 contains another *afa-8*-containing PAI, probably similar to PAI I, integrated into the *pheV* gene. Finally, we report that the *afa-8*-containing PAIs from human and bovine isolates are preferentially inserted into the *pheV* gene. Determination of the other genes carried by the *afa-8*-containing PAIs will be an interesting field for future research.

ACKNOWLEDGMENTS

We are grateful to A. Labigne, in whose unit this work was carried out, for continuing interest and helpful discussions. We also thank Y. Germani (Pasteur Institute, Bangui, Republic of Central Africa) for the gift of *E. coli* diarrheal isolates, J. P. Girardeau (INRA, Clermont Ferrand-Theix, France) and J. Mainil (Bacteriology, Faculty of Veterinary Medicine, University of Liege, Liege, Belgium) for the gift of animal *E. coli* isolates, and A. Andremont (Bichat Claude-Bernard Hospital, Paris, France) for the gift of human blood *E. coli* isolates. We thank J. Hacker for critical reading of the manuscript and E. Carniel, in whose laboratory the pulsed-field gel electrophoresis assays were performed, for helpful discussions. We also thank L. du Merle for technical assistance.

This work was supported by grant 1335 from the European Community program FAIR and a grant from the Programme de Recherche

Fondamentale en Microbiologie et Maladies Infectieuses et Parasitaires (PRFMMIP-MENRT). L. Lalioui received a fellowship from the Marcel Mérieux Fondation and the Fondation pour La Recherche Médicale.

REFERENCES

1. **Ahrens, R., M. Ott, A. Ritter, H. Hoschützky, T. Bühler, F. Lottspeich, G. J. Boulnois, K. Jann, and J. Hacker.** 1993. Genetic analysis of the gene cluster encoding nonfimbrial adhesin I from an *Escherichia coli* uropathogen. Infect. Immun. **61:**2505–2512.
2. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
3. **Baga, M., M. Goransson, S. Normark, and B. E. Uhlin.** 1985. Transcriptional activation of a *pap* pilus virulence operon from uropathogenic *Escherichia coli*. EMBO J. **4:**3887–3893.
4. **Bercovier, H., and H. H. Mollaret.** 1984. *Yersinia*, p. 498–506. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. Williams & Wilkins, Baltimore, Md.
5. **Bilge, S. S., J. M. Apostol, Jr., K. J. Fullner, and S. L. Moseley.** 1993. Transcriptional organization of the F1845 fimbrial adhesin determinant of *Escherichia coli*. Mol. Microbiol. **7:**993–1006.
6. **Bingen, E., B. Picard, N. Brahimi, S. Mathy, P. Desjardins, J. Elion, and E. Denamur.** 1998. Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. J. Infect. Dis. **177:**642–650.
7. **Blattner, F. R., G. R. Plunkett, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1474.
8. **Blum, G., V. Falbo, A. Caprioli, and J. Hacker.** 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. FEMS Microbiol. Lett. **126:**189–195.
9. **Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschäpe, and J. Hacker.** 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606–614.
10. **Boyer, H. W., and D. Roulland-Dussoix.** 1969. A complementation analysis of the restriction and modification of DNA in *Escherichia coli*. J. Mol. Microbiol. **41:**459–472.
11. **Buchrieser, C., R. Brosch, S. Bach, A. Guiyoule, and E. Carniel.** 1998. The high-pathogenicity island of *Yersinia pseudotuberculosis* can be inserted into any of the three chromosomal *asn* tRNA genes. Mol. Microbiol. **30:**965–978.
12. **Buchrieser, C., C. Rusniok, L. Frangeul, E. Couve, A. Billault, F. Kunst, E. Carniel, and P. Glaser.** 1998. The 102-kilobase *pgm* locus of *Yersinia pestis*: sequence analysis and comparison of selected regions among different *Yersinia pestis* and *Yersinia pseudotuberculosis* strains. Infect. Immun. **67:**4851–4861.
13. **Burland, V., G. R. Plunkett, H. J. Sofia, D. L. Daniels, and F. R. Blattner.** 1995. Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. Nucleic Acids Res. **23:**2105–2119.
14. **Carniel, E., I. Guilvout, and M. Prentice.** 1996. Characterization of a large chromosomal "high-pathogenicity island" in biotype 1B *Yersinia enterocolitica*. J. Bacteriol. **178:**6743–6751.
15. **Cheetham, B. F., and M. E. Katz.** 1995. A role for bacteriophages in the evolution and transfer of bacterial virulence determinants. Mol. Microbiol. **18:**201–208.
16. **Collins, J.** 1979. *Escherichia coli* plasmids packageable *in vitro* in lambda bacteriophage particles. Methods Enzymol. **68:**309–326.
17. **Fleischmann, R. D., M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick et al.** 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae*. Science **269:**496–512.
18. **Garcia, M. I., P. Gounon, P. Courcoux, A. Labigne, and C. Le Bouguénec.** 1996. The afimbrial adhesive sheath encoded by the *afa-3* gene cluster of pathogenic *Escherichia coli* is composed of two adhesins. Mol. Microbiol. **19:**683–693.
19. **Garcia, M.-I., A. Labigne, and C. Le Bouguenec.** 1994. Nucleotide sequence of the afimbrial-adhesin-encoding *afa-3* gene cluster and its translocation via flanking IS*1* insertion sequences. J. Bacteriol. **176:**7601–7613.
20. **Gérardin, J., L. Lalioui, E. Jacquemin, C. Le Bouguénec, and J. G. Mainil.** 2000. The *afa*-related gene cluster in necrotoxigenic and other *Escherichia coli* from animals belongs to the *afa-8* variant. Vet. Microbiol. **76:**155–184.
21. **Germani, Y., M. Morillon, E. Begaud, H. Dubourdieu, R. Costa, and J. Thevenon.** 1994. Two-year study of endemic enteric pathogens associated with acute diarrhea in New Caledonia. J. Clin. Microbiol. **32:**1532–1536.
22. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72:**3961–3965.
23. **Guyer, D. M., J.-S. Kao, and H. L. T. Mobley.** 1998. Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. Infect. Immun. **66:**4411–4415.
24. **Hacker, J., G. Blum-Oehler, B. Janke, G. Nagy, and W. Goebel.** 1999. Pathogenicity islands of extraintestinal *Escherichia coli*, p. 59–76. *In* J. Kaper and J. Hacker (ed.), Pathogenicity islands and other mobile virulence elements. ASM Press, Washington, D.C.
25. **Hacker, J., G. Blum-Oehler, I. Muhldorfer, and H. Tschape.** 1997. Pathogenicity islands of virulent bacteria: structure, function and impact on microbial evolution. Mol. Microbiol. **23:**1089–1097.
26. **Hilali, F., R. Ruimy, P. Saulnier, B. Barnabé, C. Le Bouguénec, M. Tibayrenc, and A. Andremont.** 2000. Prevalence of virulence genes and clonality in *Escherichia coli* strains that cause bacteremia in cancer patients. Infect. Immun. **68:**3983–3989.
27. **Hope, J. N., A. W. Bell, M. A. Hermodson, and J. M. Groarke.** 1986. Ribokinase from *Escherichia coli* K12. Nucleotide sequence and overexpression of the *rbsK* gene and purification of ribokinase. J. Biol. Chem. **15:**7663–7668.
28. **Hung, D. L., S. D. Knight, R. M. Woods, J. S. Pinkner, and S. J. Hultgren.** 1996. Molecular basis of two subfamilies of immunoglobulin-like chaperones. EMBO J. **15:**3792–3805.
29. **Jacob-Dubuisson, F., J. Pinkner, Z. Xu, R. Striker, A. Padmanhaban, and S. J. Hultgren.** 1994. PapD chaperone function in pilus biogenesis depends on oxidant and chaperone-like activities of DsbA. Proc. Natl. Acad. Sci. USA **91:**11552–11556.
30. **Jouve, M., M.-I. Garcia, P. Courcoux, A. Labigne, P. Gounon, and C. Le Bouguénec.** 1997. Adhesion to and invasion of HeLa cells by pathogenic *Escherichia coli* carrying the *afa*-3 gene cluster are mediated by the AfaE and AfaD proteins, respectively. Infect. Immun. **65:**4082–4089.
31. **Kado, C. I., and S.-T. Liu.** 1981. Rapid procedure for detection and isolation of large and small plasmids. J. Bacteriol. **145:**1365–1373.
32. **Kao, J.-S., D. M. Stucker, J. W. Warren, and H. L. T. Mobley.** 1997. Pathogenicity island sequences of pyelonephritogenic *Escherichia coli* CFT073 are associated with virulent uropathogenic strains. Infect. Immun. **65:**2812–2820.
33. **Knapp, S., J. Hacker, T. Jarchau, and W. Goebel.** 1986. Large unstable inserts in the chromosome affect virulence properties of uropathogenic *Escherichia coli* O6 strain 536. J. Bacteriol. **168:**22–30.
34. **Labigne-Roussel, A. F., D. Lark, G. Schoolnik, and S. Falkow.** 1984. Cloning and expression of an afimbrial adhesin (AFA-I) responsible for P blood group-independent, mannose-resistant hemagglutination from a pyelonephritic *Escherichia coli* strain. Infect. Immun. **46:**251–259.
35. **Lalioui, L., M. Jouve, P. Gounon, and C. Le Bouguénec.** 1999. Molecular cloning and characterization of the *afa-7* and *afa-8* gene clusters encoding afimbrial adhesins in *Escherichia coli* strains associated with diarrhea or septicemia in calves. Infect. Immun. **67:**5048–5059.
36. **Le Bouguénec, C., M. Archambaud, and A. Labigne.** 1992. Rapid and specific detection of the *pap, afa*, and *sfa* adhesin-encoding operons in uropathogenic *Escherichia coli* strains by polymerase chain reaction. J. Clin. Microbiol. **30:**1189–1193.
37. **Le Bouguénec, C., M. I. Garcia, V. Ouin, J.-M. Desperrier, P. Gounon, and A. Labigne.** 1993. Characterization of plasmid-borne *afa-3* gene clusters encoding afimbrial adhesins expressed by *Escherichia coli* strains associated with intestinal or urinary tract infections. Infect. Immun. **61:**5106–5114.
38. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **McDaniel, T. K., K. G. Jarvis, M. S. Donnenberg, and J. B. Kaper.** 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc. Natl. Acad. Sci. USA **92:**1664–1668.
40. **McDaniel, T. K., and J. B. Kaper.** 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol. Microbiol. **23:**399–407.
41. **Mortensen, L., G. Dandanell, and K. Hammer.** 1989. Purification and characterization of the *deoR* repressor of *Escherichia coli*. EMBO J. **8:**325–331.
42. **Nowicki, B., A. Labigne, S. Moseley, R. Hull, S. Hull, and J. Moulds.** 1990. The Dr hemagglutinin, afimbrial adhesins AFA-I and AFA-III, and F1845 fimbriae of uropathogenic and diarrhea-associated *Escherichia coli* belong to a family of hemagglutinins with Dr receptor recognition. Infect. Immun. **58:**279–281.
43. **Ochman, H., and J. G. Lawrence.** 1996. Phylogenetics and the amelioration of bacterial genomes, p. 2627–2637. *In* F. C. Neidhardt, R. Curtiss III, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, W. S. Reznikoff, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed., vol. 2. ASM Press, Washington, D.C.
44. **Perna, N. T., G. F. Mayhew, G. Pósfai, S. Elliott, M. S. Donnenberg, J. B. Kaper, and F. R. Blattner.** 1998. Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli*. O157:H7. Infect. Immun. **66:**3810–3815.
45. **Pierson, L. S. D., and M. L. Kahn.** 1987. Integration of satellite bacteriophage P4 in *Escherichia coli*. DNA sequences of the phage and host regions involved in site-specific recombination. J. Mol. Biol. **196:**487–496.
46. **Pittard, J., J. Praszkier, A. Certoma, G. Eggertsson, J. Gowrishankar,**

G. Narasaiah, and M. J. Whipp. 1990. Evidence that there are only two $tRNA^{Phe}$ genes in *Escherichia coli*. J. Bacteriol. **152:**6077–6083.
47. **Redenbach, M., H. M. Kieser, D. Denapaite, A. Eichner, J. Cullum, H. Kinashi, and D. A. Hopwood.** 1996. A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb *Streptomyces coelicolor* A3(2) chromosome. Mol. Microbiol. **21:**77–96.
48. **Reizer, J., A. Charbit, A. Reizer, and M. H. J. Saier.** 1996. Novel phosphotransferase system genes revealed by bacterial genome analysis: operons encoding homologues of sugar-specific permease domains of the phosphotransferase system and pentose catabolic enzymes. Genome Sci. Technol. **1:**53–75.
49. **Savarino, S. J., P. Fox, D. Yikang, and J. P. Nataro.** 1994. Identification and characterization of a gene cluster mediating enteroaggregative *Escherichia coli* aggregative adherence fimbria I biogenesis. J. Bacteriol. **156:**4949–4957.
50. **Schubert, S., A. Rakin, D. Fischer, J. Sorsa, and J. Heesemann.** 1999. Characterization of the integration site of *Yersinia* high-pathogenicity island in *Escherichia coli*. FEMS Microbiol. Lett. **159:**409–414.
51. **Schubert, S., A. Rakin, H. Karch, E. Carniel, and J. Heesemann.** 1998. Prevalence of the "high-pathogenicity island" of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. Infect. Immun. **66:** 480–485.
52. **Shine, J., and L. Dalgarno.** 1975. Determinant of cistron specificity in bacterial ribosomes. Nature **254:**34–38.
53. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:**503–515.
54. **Stentz, R., and M. Zagorec.** 1999. Ribose utilization in *Lactobacillus sakei*: analysis of the regulation of the *rbs* operon and putative involvement of a new transporter. J. Mol. Microbiol. Biotechnol. **1:**165–173.
55. **Swenson, D. L., N. O. Bukanov, D. E. Berg, and R. A. Welch.** 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. Infect. Immun. **64:**3736–3743.
56. **Valentin-Hansen, P., P. Hojrup, and S. Short.** 1985. The primary structure of the DeoR repressor from *Escherichia coli* K-12. Nucleic Acids Res. **13:**5927–5936.
57. **Valentin-Hansen, P., B. A. Svenningsen, A. Munch-Petersen, and K. Hammer-Jespersen.** 1978. Regulation of the *deo* operon in *Escherichia coli*: the double negative control of the *deo* operon by the *cytR* and *deoR* repressors in a DNA directed *in vitro* system. Mol. Gen. Genet. **159:**191–202.
58. **Van Rosmalen, M., and M. H. Saier, Jr.** 1993. Structural and evolutionary relationships between two families of bacterial extracytoplasmic chaperone proteins which function cooperatively in fimbrial assembly. Res. Microbiol. **144:**507–527.

Editor: J. T. Barbieri

ERRATA

afa-8 Gene Cluster Is Carried by a Pathogenicity Island Inserted into the tRNA^{Phe} of Human and Bovine Pathogenic *Escherichia coli* Isolates

Lila Lalioui and Chantal Le Bouguénec

Unité de Pathogénie Bactérienne des Muqueuses, Institut Pasteur, 75724 Paris Cedex 15, France

Volume 69, no. 2, p. 937–948, 2001. Page 939, Table 1, columns 2 and 3: Data for probes B and C should read as follows.

Probe B	5′-GGACGATAATGTGATCGTCTATAAG-3′	816
	5′-GTGGAAGATACTCATCTGCTACACG-3′	
Probe C	5′-TTTGATGAGCGATGTACTTTCCGAA-3′	992
	5′-GCAGATACAACGTGAACATACCGA-3′	

Novel *Aeromonas hydrophila* PPD134/91 Genes Involved in O-Antigen and Capsule Biosynthesis

Y. L. Zhang, E. Arakawa, and K. Y. Leung

Department of Biological Sciences, Faculty of Science, and Tropical Marine Science Institute, The National University of Singapore, Singapore 117543, and Department of Bacteriology, National Institute of Infectious Diseases, Tokyo, Japan 162-8640

Volume 70, no. 5, p. 2326–2335, 2002. Page 2329, Table 2, column 9, row 5 from the bottom: “72” should read “40a.” Page 2332, Table 4, column 9, row 5 from the bottom: “73” should read “25a.”