INFECTION AND IMMUNITY, July 1997, p. 2812–2820 Vol. 65, No. 7
0019-9567/97/$04.00+0

Pathogenicity Island Sequences of Pyelonephritogenic *Escherichia coli* CFT073 Are Associated with Virulent Uropathogenic Strains

JYH-SHYANG KAO, DEBRA M. STUCKER, JOHN W. WARREN, AND HARRY L. T. MOBLEY*

Division of Infectious Diseases, University of Maryland School of Medicine, Baltimore, Maryland 21201

Received 16 December 1996/Returned for modification 20 February 1997/Accepted 9 April 1997

Urinary tract infection is the most frequently diagnosed kidney and urologic disease, and *Escherichia coli* is by far the most common etiologic agent. Defined blocks of DNA termed pathogenicity islands have been found in uropathogenic strains to carry genes not generally found in fecal strains. We have identified one of these regions of DNA within the chromosome of the highly virulent *E. coli* CFT073, isolated from the blood and urine of a woman with acute pyelonephritis. This strain, which is cytotoxic for cultured renal cells and causes acute pyelonephritis in transurethrally infected CBA mice, contains two distinct copies of the *pap* operon and is hemolytic. One *pap* operon was localized on a cosmid clone which was used to identify three overlapping cosmid clones. By using restriction mapping, DNA hybridization, sequencing, and PCR amplification, a region of approximately 50 kb was found to be present in this uropathogenic strain and to have no corresponding sequences in *E. coli* K-12. This gene block also carries hemolysin genes *hlyCABD*. The pathogenicity island begins 7 bp downstream of *dadX* (catabolic alanine racemase; 26.55 min) and ends at a position in the K-12 genome 75 bp downstream of the *metV* tRNA gene (62.74 min); this suggests that a chromosomal rearrangement has occurred relative to the K-12 linkage map. The junctions of the pathogenicity island were verified by PCR amplification directly from the genomic DNA of strain CFT073. DNA sequencing within the boundaries of the junctions revealed genes not previously identified in *E. coli* or in some cases bearing no known homologs. When used as probes for DNA hybridization, these sequences were found significantly more often in strains associated with the clinical syndromes of cystitis (82%) and acute pyelonephritis (79%) than in fecal strains (19%; $P < 0.001$).

The urinary tract is among the most common sites of bacterial infection, and *Escherichia coli* is by far the most common infecting agent at this site (21). Individuals at high risk for symptomatic urinary tract infection (UTI) include neonates, preschool girls, sexually active women, and elderly women and men. In 1991, the last year for which pertinent information is available for the United States, UTIs were the cause of 9.6 million physician visits (33) and were noted on 1.5 million hospital discharge summaries (32).

The most serious UTI is acute pyelonephritis. This infection in one or both kidneys usually results from the ascent of organisms from the bladder via the ureter and is distinguished from other UTIs clinically, pathologically, and by characteristics of the causative organisms. Certain *E. coli* phenotypes have been found more frequently in the urine of patients with acute pyelonephritis than in the urine of patients with cystitis or asymptomatic bacteriuria or in the feces of normal individuals (reviewed by Johnson [18]). *E. coli* strains from a small group of O serotypes have characteristics epidemiologically associated with acute pyelonephritis in the normal urinary tract. These include P fimbriae, hemolysin, aerobactin, serum resistance, and encapsulation. These observations have been interpreted as evidence for the existence of a set of virulence factors that allow specific clones of *E. coli* to cause pyelonephritis.

Recently, examples of clustered virulence factors, virulence cassettes, or pathogenicity islands (PAI) in a number of pathogenic bacterial species have been described. Enteropathogenic and enterohemorrhagic *E. coli* and *Citrobacter freundii* (26), meningitis-associated *E. coli* (4, 5), *Helicobacter pylori* (1, 22), *Salmonella typhimurium* (27), and *Yersinia pestis* (9) all can possess clustered virulence genes that are not found in less virulent strains of these species (23, 31). Indeed, Hacker and colleagues (14, 20) were the first to describe and name such a PAI in a virulent uropathogen. Other uropathogenic strains of *E. coli* have also been identified as having PAIs, which are large stretches of DNA containing genes not found in the genome of fecal *E. coli* strains (38) or clustered virulence determinants (6, 15, 25). Our laboratory has identified a distinct PAI in the highly virulent pyelonephritogenic *E. coli* CFT073 and has determined the precise insertion points of these sequences within the genome. Some of the genes found within these new sequences have not been previously identified in *E. coli*, in some cases have no recognized homologs, and are associated with virulent strains.

MATERIALS AND METHODS

Bacterial strains. *E. coli* CFT073 was isolated from the blood and urine of a woman admitted to the University of Maryland Medical System for the treatment of acute pyelonephritis. This *hly*$^+$ *pap*$^+$ *sfa*$^+$ *pil*$^+$ strain is highly virulent in the CBA mouse model of ascending UTI (29) and is cytotoxic for cultured human renal proximal tubular epithelial cells (28). It is phenotypically positive for the production of P fimbriae, hemolysin, and type 1 fimbriae. *E. coli* DH5α (2) was used as the recipient for gene bank and recombinant clones.

Four collections of *E. coli* strains were established from humans with appropriate clinical syndromes. The first consists of 61 isolates from the urine or blood of patients (41 women and 20 men) who were admitted to the University of Maryland Medical Systems with acute pyelonephritis (bacteriuria of $\geq 10^5$ CFU/ml, pyuria, fever, and no other source of infection) (28). The second collection consists of 38 isolates from the urine of women with cystitis. These isolates were kindly provided by A. Stapleton (University Washington) and B. Foxman (University of Michigan) (10). The third collection consists of 61 isolates from the urine of 26 patients with long-term urinary catheters in place. Each was isolated during the first week of a new epidemiologically defined episode of *E. coli* bacteriuria (39). The fourth collection consists of 28 control strains of *E. coli* from the feces of healthy women (20 to 50 years old) who had not had a

* Corresponding author. Mailing address: Division of Infectious Diseases, University of Maryland School of Medicine, 10 S. Pine St., Baltimore, MD 21201. Phone: (410) 706-7560. Fax: (410) 706-8700. E-mail: hmobley@umabnet.ab.umd.edu.

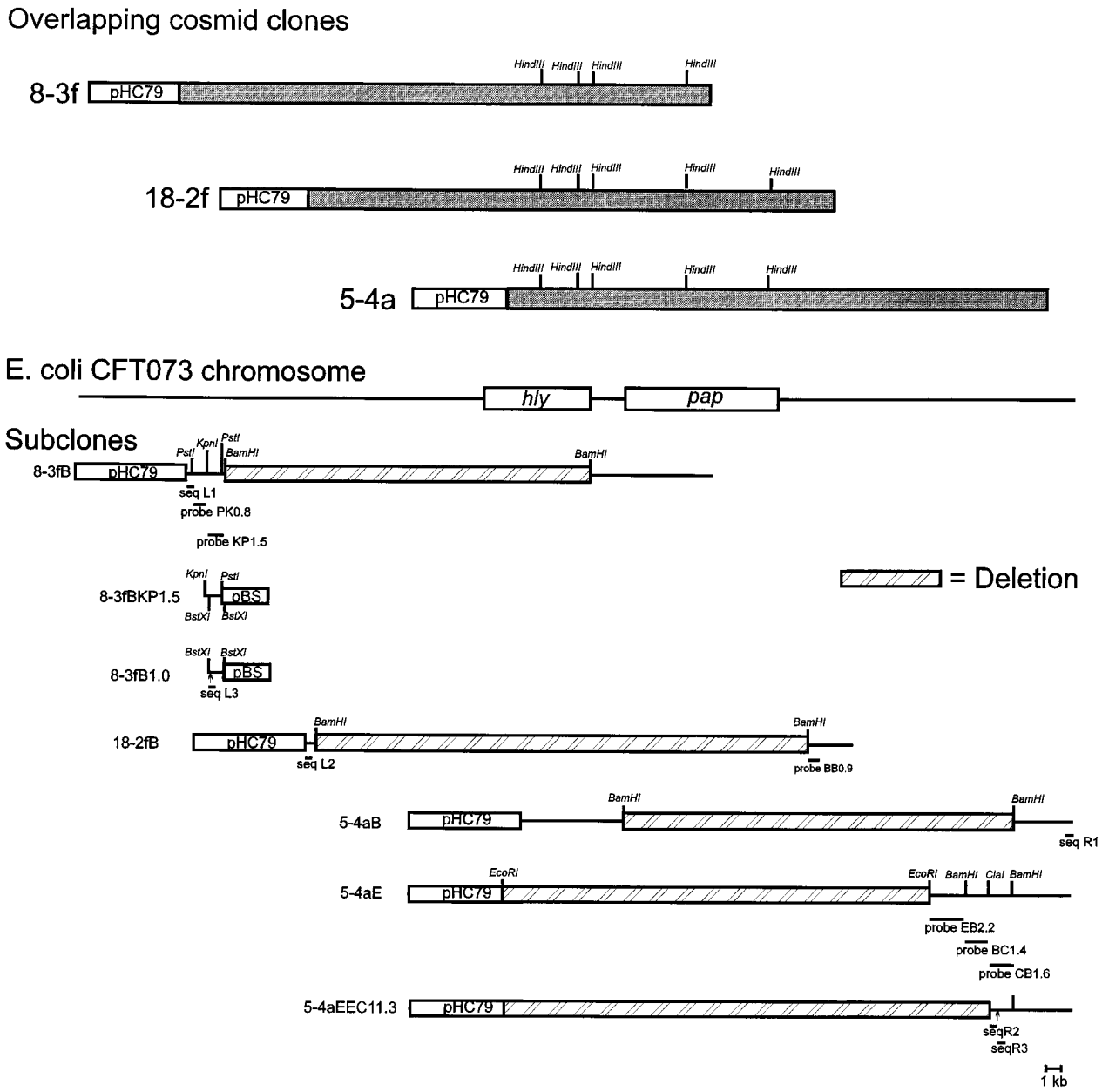

FIG. 1. Isolation of overlapping cosmid clones covering the 50-kb PAI and constructs used for sequencing and Southern hybridization. The three overlapping cosmid clones were identified by probing and restriction mapping of *pap hly* clones. *Hin*dIII sites common to these cosmid clones are shown. Subclones of the cosmid clones were constructed, and restriction fragments (shown below the subclones) isolated from these constructs were used as probes to sequentially approach the junctions between the K-12 DNA sequence and the uropathogen sequence. Nucleotide sequencing of specific fragments (seq) was used to locate the precise junctions.

symptomatic UTI or known bacteriuria within the previous 6 months and who had not experienced diarrhea or received antibiotics within the preceding 1 month.

Cosmid library. A cosmid library was constructed with partially *Sau*3A-digested genomic DNA isolated from *E. coli* CFT073. The DNA was ligated into *Bam*HI-digested pHC79. The ligation mixture was packaged in vitro with the Gigapack lambda packaging kit (Stratagene) and used to infect *E. coli* DH5α. Transformants were selected on Luria agar containing ampicillin (200 μg/ml). Isolated colonies were organized in the wells of 96-well microtiter plates and stored at −70°C in Luria broth containing 15% (vol/vol) glycerol. The library has been used previously for the isolation of *pap* and *hly* sequences (29).

Nucleotide sequencing and analysis. Double-stranded DNA was used as a template for sequencing by the dideoxy chain termination method (36). The reactions were run with reagents from the Prism Ready Reaction Dye Deoxy Termination Kit (Applied Biosystems) in conjunction with *Taq* polymerase. A model 373A DNA sequencer (Applied Biosystems) was used, and the sequence was determined in each direction. DNAsis software (Hitachi, version 2.1) was used for analysis of the DNA sequence for base composition, identification of open reading frames (ORFs) and restriction sites, and other basic analyses. Apparent homologies between ORFs both outside and inside the PAI were sought in GenBank by using the Wisconsin package (version 8.1; Genetics Computer Group, Inc.).

PCR. PCR amplification was done by a standard method (2). The left junction was amplified with primers 697 (5′-TGTCGGCGTTCGTTGTC-3′) and 698 (5′-CTGGTGCTGGGGCTATT-3′), using 25 cycles of 94°C for 2 min, 53°C for 30 s, 72°C for 40 s, 94°C for 30 s, and 72°C for 3 min. The right junction was amplified with primers 682 (5′-CCAGCAGGAAGTTGCGGA-3′) and 684 (5′-AACACCATCTTTGCGCTTT-3′), using 25 cycles of 94°C for 2 min, 52°C for 30 s, 72°C for 100 s, 94°C for 30 s, and 72°C for 3 min.

Southern hybridization. DNA restriction fragments isolated from cosmid clones were used as gene probes to determine whether sequences were present in genomic DNA preparations of *E. coli* CFT073 and DH5α. The fragments were

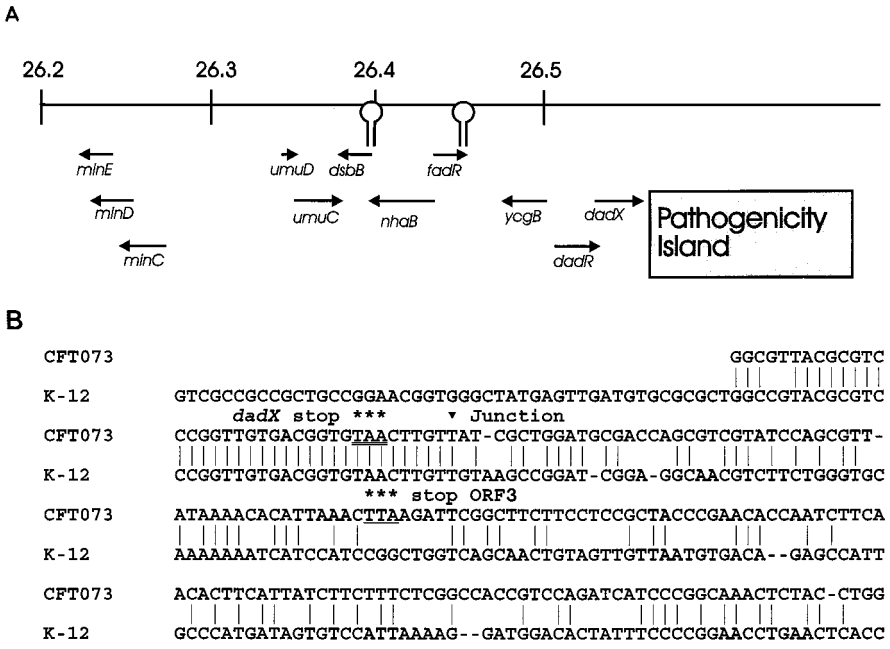

FIG. 2. Localization of the left junction of the PAI in *E. coli* CFT073. (A) The junction, located immediately downstream (7 bp) of *dadX*, is shown in relation to conserved genes within the K-12 genome. (B) Alignment of the nucleotide sequence of the left PAI junction with the corresponding K-12 genome sequence (22). Identical nucleotides are joined by vertical lines. The stop codons for *dadX* and ORF3 are doubly underlined and singly underlined, respectively. The left junction of the PAI is marked with an arrowhead.

labeled with the enhanced chemiluminescence system (Amersham). Southern blots were prepared by standard methods (35) and developed with the enhanced chemiluminescence system (Amersham) as specified by the manufacturer.

Dot blot hybridization. *E. coli* strains were cultured in 200 μl of Luria broth in 96-well microtiter plates. Dot blots were prepared with these cultures as described by Kafatos et al. (19).

Nucleotide sequence accession numbers. The sequences of the ORFs within the boundaries of the left and right junctions of the PAI of *E. coli* CFT073 (see Fig. 7) have been assigned GenBank accession numbers AF003741 and AF003742, respectively.

RESULTS

PAI identification in *E. coli* CFT073. *E. coli* CFT073 is a highly virulent pyelonephritogenic strain isolated from the blood and urine of a woman with the clinical symptoms that characterize acute pyelonephritis (28). There are two complete and functional copies of the *pap* operon encoding two distinguishable P fimbrial types (29). Based on studies with *E. coli* 536 (7, 13, 16, 30, 34) and J96 (38), we hypothesized that each of the *pap* gene clusters was associated with a distinct PAI. To identify and characterize the first PAI, a cosmid gene bank was screened for *pap*-positive clones. Of 1,000 cosmid clones screened, 39 reacted with *pap*-specific probes. Of these clones, 10 had common restriction sites; 3 of these were subjected to extensive restriction mapping and were found to overlap and cover a region of approximately 50 kb (Fig. 1).

Three cosmid clones, 8-3f, 18-2f, and 5-4a, shared significant regions of overlap in which they have identical restriction sites. Cosmids 8-3f and 18-2f overlap by 27 kb; cosmids 18-2f and 5-4a overlap by 23 kb; all three cosmids overlap by 14 kb and include the same 6.6-kb *Hin*dIII fragment found within only one of the two *pap* gene clusters present in this strain. In addition, two of the cosmids (8-3f and 18-2f) carry the entire hemolysin gene cluster (*hlyCABD*) and produce an active hemolysin; the third cosmid (5-4a) carries a portion of the *hly* gene cluster of which there is only a single copy on the CFT073 chromosome.

Left junction. To localize the left junction (where "left" is defined as the sequences associated with lower numbers on the linkage map of *E. coli* K-12), deletion mutations of cosmid clones 8-3f were constructed (Fig. 1). To determine whether the sequence at the extreme left junction of clone 8-3fB was inside the PAI or was included in K-12 DNA, the nucleotide sequence was determined (seqL1) with primers based on cosmid vector pHC79. A sequence of 330 bp was determined. A search of GenBank revealed a perfect match with *E. coli* K-12 sequences corresponding to *dadA* (D-amino acid dehydrogenase) and *dadX* (alanine racemase) (24); these sequences were associated with the left-hand flanking region, which is present between 26 and 27 min on the linkage map. This region of cosmid subclone 8-3fB was thus identified as being outside the PAI. To identify sequences not present on the K-12 map (i.e., inside the PAI), two restriction fragments (probes PK0.8 and KP1.5 [Fig. 1]) were isolated, labeled, and hybridized with chromosomal DNA isolated from *E. coli* DH5α and CFT073 which had been singly digested with each of five different restriction enzymes (data not shown). Probe PK0.8 hybridized to fragments from both strains but more strongly to CFT073 DNA. Probe KP1.5, a fragment immediately to the right of probe PK0.8, also hybridized to DNA from both strains. However, the hybridization signal to DH5α was less intense than the strong signal observed for CFT073, suggesting that the junction between K-12-like DNA and the PAI of CFT073 may be found within this fragment. Nucleotide sequencing of subclone p8-3fB1.0 revealed that 7 bp downstream from *dadX*, sequence identity between K-12 DNA and CFT073 DNA ceases and sequence unique to CFT073 begins (Fig. 2). This region represents the left junction of the CFT073 PAI and maps to 26.75 min on the K-12 linkage map.

Right junction. A similar strategy to that described above was applied to identify the right junction (where "right" is defined as the sequences associated with higher numbers on

the linkage map of E. coli K-12) of the PAI. To locate sequences that were outside the boundary of the PAI, sequencing of a subclone of cosmid 5-4a, namely, 5-4aB (Fig. 1), identified the *recB* ORF present in the K-12 genome (60.6 min on the K-12 linkage map). Southern hybridization with probes derived from p18-2fB and p5-4aE was used to localize a restriction fragment containing the junction (Fig. 3). Probes BB0.9, EB2.2, and BC1.4 all hybridized only to CFT073 and not DH5α DNA, indicating that these sequences were inside the PAI. However, probe CB1.6, a 1.6-kb *Cla*I-*Bam*HI fragment, hybridized to both CFT073 and DH5α DNA, indicating that the junction resided within this fragment.

The precise location of the right junction was confirmed by two sequencing reactions (seqR2 and seqR3 [Fig. 4]). Sequence analysis of the right junction indicated that the junction occurs within the stop codon of a hypothetical ORF f447 in the K-12 genome (Fig. 4A). The beginning of the CFT073 sequence, however, retains a stop codon which has changed from TGA in K-12 to TAA in CFT073. This position would be 75 bp downstream of the *metV* tRNA gene in the K-12 genome.

metV and ORF f447 map at 62.74 min; *metV* is the third of three tandem formylmethionine tRNA genes, *metZ*, *metW*, and *metV*, on the *E. coli* chromosome. That the left junction of the PAI maps to 26.75 min and the right junction maps to 62.74 min demonstrates that a chromosomal rearrangement has occurred in strain CFT073 relative to the K-12 linkage map.

PCR amplification of the junctions. To verify that the location and sequence of the left and right junctions, determined from sequencing of cosmid clones, correspond to those in the chromosome of CFT073 and were not artifacts of construction of the cosmid gene bank, we used PCR to amplify the left and right junction regions directly from the CFT073 chromosome (Fig. 5). For the left junction, primers 697 and 698 were used to amplify a 1.4-kb fragment from CFT073 genomic DNA. This corresponded to a fragment of identical size amplified from cosmid 8-3f, which overlaps the left junction. As expected, such a fragment could not be amplified from DH5α or from cosmid 5-4a, which overlaps the right junction. For the right junction, primers 682 and 684 were used to amplify a 3.1-kb fragment from CFT073 genomic DNA. A fragment of identical size was amplified from cosmid 5-4a, which overlaps the right junction. No such fragment was amplified from DH5α or from cosmid 8-3f, which overlaps the left junction. For both left and right junctions, the nucleotide sequence derived from the cosmid clones exactly matched the nucleotide sequence of the PCR products amplified from chromosomal DNA of wild-type strain CFT073 (data not shown). These experiments confirm that the junctions of the PAI characterized with cosmid clones 8-3f and 5-4a are also present in wild-type *E. coli* CFT073.

Direct repeats at the junction. Other PAIs of *E. coli* uropathogens have been associated with direct repeats of sequences present just inside and just outside the PAI (7, 20, 38). Such a sequence, CCAGCGTCG, has been located 14 bp downstream of the left junction and 12 bp downstream of the right junction (Fig. 6). These sequences in strain CFT073 share little or no homology with the repeated sequences reported by Blum et al. (7) or Swenson et al. (38).

Novel genes inside the PAI. To determine whether newly described genes were present within the boundaries of the PAI, these regions were subjected to nucleotide sequencing. Inside the left junction, 1.7 kb was determined and revealed an ORF of 1,609 bp, predicting a polypeptide of 57.8 kDa (Fig. 7). Inside the right junction, 3.6 kb was determined and revealed ORFs of 882 and 1,578 bp, predicting polypeptides of 34.1 and 57.6 kDa, respectively (Fig. 7). Each of the three ORFs was sequenced to completion. None of the three predicted polypeptides have been described for any *E. coli* strain, suggesting that they represent novel genes in this species. However, homology searches reveal a similarity of the 34.1- and 57.6-kDa polypeptides to an antiterminator in the *sac* gene cluster of *Bacillus subtilis* and a sugar-specific component of a phosphotransferase transport system used to import sugars, respectively.

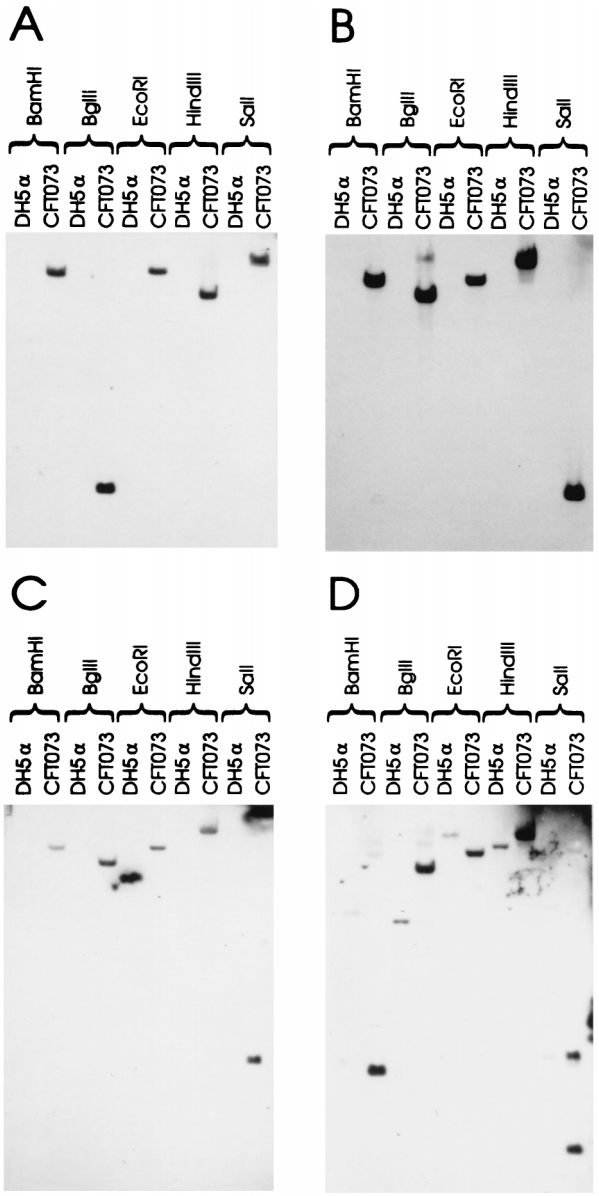

FIG. 3. Southern blots of *E. coli* chromosomal DNA hybridized with probes derived from cosmids 18-2fB and 5-4aE (right flanking region). Southern blots were prepared as described in the legend to Fig. 2 and hybridized with each of four probes. (A) Probe BB0.9 (see Fig. 1 for the locations of all the probes); (B) probe EB2.2; (C) probe BC1.4. These three probes hybridized only to DNA from strain CFT073 but not to strain DH5α, suggesting that the junction was to the right of this point. (D) Probe CB1.6. This probe hybridized to DNA from CFT073 and DH5α, indicating that the junction was within this 1.6-kb fragment.

PAI sequences are associated with virulent strains. To determine whether sequences found within the junctions of the PAI were common to other uropathogenic strains of *E. coli*, a probe was isolated and used to hybridize dot blots of genomic

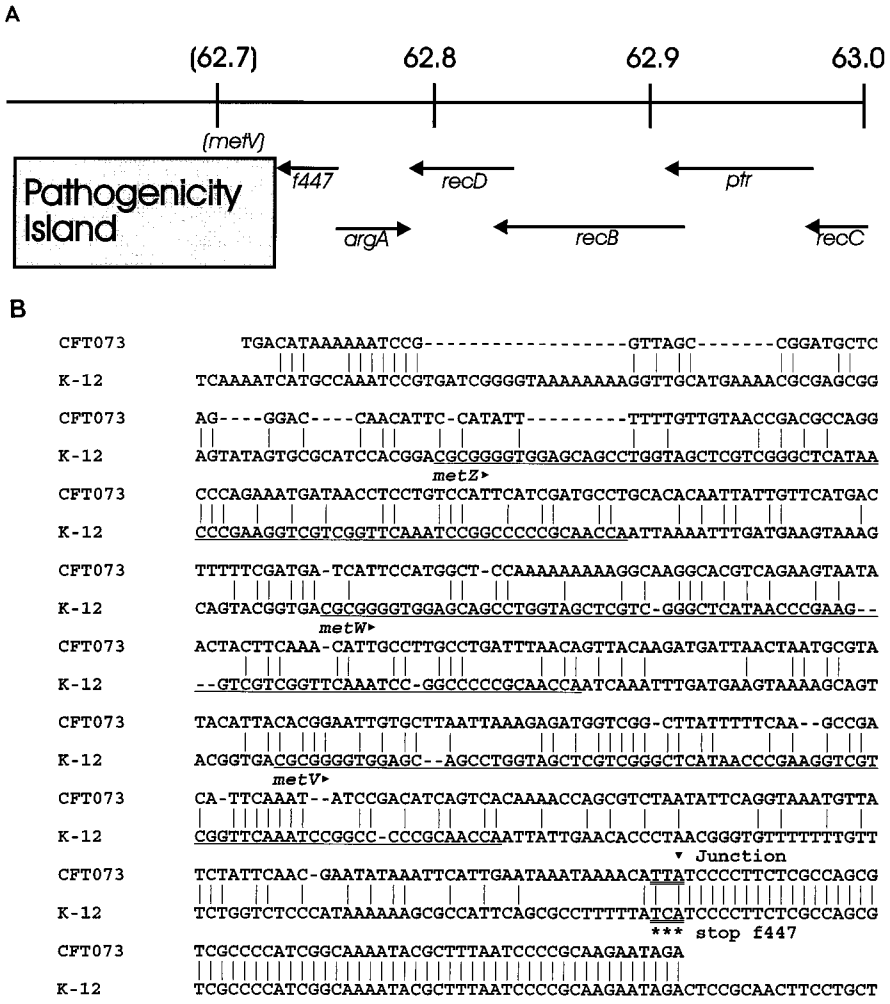

FIG. 4. Localization of the right junction by nucleotide sequencing of 5-4aB (seq R1). This sequence identified *recB*. (A) From sequencing and restriction mapping, the right junction can be localized upstream of *argA* (*N*-acetylglutamate synthetase). (B) Alignment of the nucleotide sequence of the PAI junction with the corresponding K-12 genome sequence. Identical nucleotides are joined by vertical lines. Sequences encoding the formylmethionine-tRNA genes, *metZ*, *metW*, and *metV*, are underlined. The right junction of the PAI is marked with an arrowhead. The stop codon for ORF f447 is doubly underlined.

DNA isolated from clinical isolates (Table 1). The probe was a 2.1-kb PCR fragment amplified from a region of the pathogenicity island that is 0.8 kb to the left of the right junction (the location is shown in Fig. 7). A high proportion of isolates from patients with cystitis or acute pyelonephritis reacted with the probe (approximately 80%). These proportions were significantly higher than for strains from patients with catheter-associated bacteriuria (46%) or from fecal strains (19%) ($P < 0.001$), indicating that PAI sequences from strain CFT073 are found in other virulent uropathogenic strains.

Arrangement of PAI sequences in other virulent strains. In all descriptions of PAIs thus described for uropathogenic *E. coli*, the organization of sequences was analyzed only for the strain under study. To determine whether our findings could be generalized to other strains, a Southern blot of *Hin*dIII-digested genomic DNA from 12 strains was hybridized with an 11-kb *Hin*dIII fragment derived from within the CFT073 PAI (located between the *pap* operon and the right junction) (Fig. 8). Five *E. coli* strains (including CFT073) isolated from the urine or blood of patients with acute pyelonephritis were analyzed along with five *E. coli* strains isolated from the urine of women with cystitis. These were compared to *E. coli* SH1 (a *lac*-negative derivative of the classic uropathogenic strain J96) and *E. coli* DH5α (representative of K-12 genomic sequences).

The Southern blot (Fig. 8) revealed that homologous PAI-related sequences were present in all the isolates from patients with pyelonephritis, including SH1 (J96), and in four of five strains from women with cystitis but were absent in DH5α. While none of the strains shared identical patterns, three of the strains from patients with pyelonephritis (CFT9, CFT26, and CFT47) had similar patterns with respect to two prominent bands. The pattern for strain CFT073 was distinctly different from that for all other strains. Strain SH1 was not similar to other strains from patients with pyelonephritis but was similar to strain 11 from a patient with cystitis. We conclude that while sequences homologous to those found in the CFT073 PAI are present in other virulent strains, their arrangement in the chromosome differs among strains.

DISCUSSION

A PAI of 50 kb has been isolated and characterized from the highly virulent pyelonephritogenic *E. coli* CFT073. The PAI is

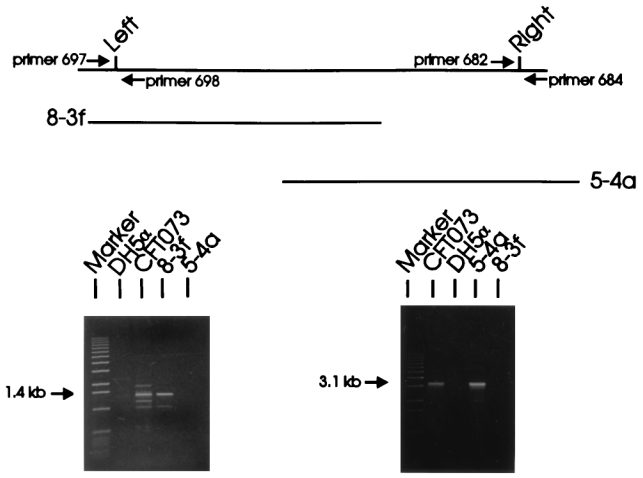

FIG. 5. PCR amplification of the left and right junctions of the PAI. Pairs of oligonucleotide primers (697 and 698; 682 and 684) were used to amplify fragments containing the junctions of the PAI. Cosmid 8-3f spans the left junction; cosmid 5-4a spans the right junction. Templates for the reactions (indicated above each gel) included genomic DNA from *E. coli* DH5α and CFT073 and purified cosmids 8-3f and 5-4a. Amplified DNA was electrophoresed on agarose gels along with standard DNA markers. The sizes of the predicted products are marked beside the gels for the left junction (1.4 kb) and right junction (3.1 kb).

marked with both *pap* sequences encoding P fimbriae and the hemolysin gene cluster. While PAIs have been described for two other strains, 536 (14, 20) and J96 (38), the sequences described here have features that are distinct from those previously reported. First, at 50 kb, it is the smallest of the five PAIs yet described for uropathogenic *E. coli*, which previously ranged from 70 to 190 kb (6, 20, 38). Second, the site of insertion is not specifically within a tRNA gene. It is, however, within 75 bp of the *metV* gene. In the other strains, amino acid tRNA genes are interrupted, but insertion of a PAI at this site has not been reported. In addition, the specific genes found within the boundaries of the junctions of the PAI share no homology with genes previously reported to reside in an *E. coli* uropathogen PAI or on the K-12 genome (38; search of the GenBank database). Finally, insertion of the PAI appears to be associated with a chromosomal rearrangement in strain CFT073.

While the discovery and description of PAIs in two uropathogenic strains has allowed previous investigators (14, 20, 38) to postulate that these sequences contain genes that contribute to virulence, it has not been possible to make generalizations regarding the frequency at which these sequences occur in different strains. However, based on the hybridization of a strain collection with a unique probe derived from just inside the right junction of the PAI (Table 1; Fig. 7), we have demonstrated that the presence of these PAI sequences is a feature of *E. coli* strains associated with more severe clinical syndromes. Hybridization studies revealed that approximately 80% of strains isolated from patients with cystitis or acute pyelonephritis were positive for this PAI probe. This alone is an interesting observation, since it has not been determined whether strains that cause acute pyelonephritis represent clones that are distinct from those that cause cystitis. These data provide evidence for similarity between the two types of strains, but of course they represent a limited snapshot of the genome. This high frequency was significantly different from that of fecal strains, since only 19% of the strains reacted, suggesting that fecal *E. coli* strains in general do not possess these sequences. Strains from elderly nursing home patients with catheter-associated bacteriuria (often asymptomatic) were intermediate in their reactivity, with 46% of the strains found to be positive for hybridization. This is the first report to survey a large collection of clinical isolates of uropathogenic *E. coli* for the presence of PAI sequences.

In all other reports, insertion of the PAI is associated with interruption of a specific sequence. That is, sequences to the left of the PAI are continued immediately following the right junction of the PAI. The PAI of CFT073, on the other hand, shows a site of insertion at 26.55 min at the left junction and 62.74 min at the right junction. This observation strongly suggests that a rearrangement has occurred within the chromosome of CFT073. The left junction of the PAI was found 7 bp downstream of the termination codon for *dadX* (alanine racemase), which localizes to 26.55 min on the *E. coli* chromosome (3). The right junction was found within the stop codon of ORF f447, a hypothetical gene in the K-12 genome which is upstream of *argA* (which encodes *N*-acetylglutamate synthase) (8). The right junction was also 75 bp downstream of the *metV* tRNA gene. Strangely, *argA* maps to 62.75 min on the *E. coli* chromosome. It is possible, however, that the presence of two copies of the *pap* operon served as a crossover point at some point in time for a chromosomal rearrangement. Studies by pulsed-field gel electrophoresis may be necessary to resolve this question. We can rule out, however, the possibility that this is a cloning artifact. The three cosmid clones that cover the 50-kb PAI share significant overlapping sequences (Fig. 1). Indeed, all three cosmids overlap by 14 kb and possess iden-

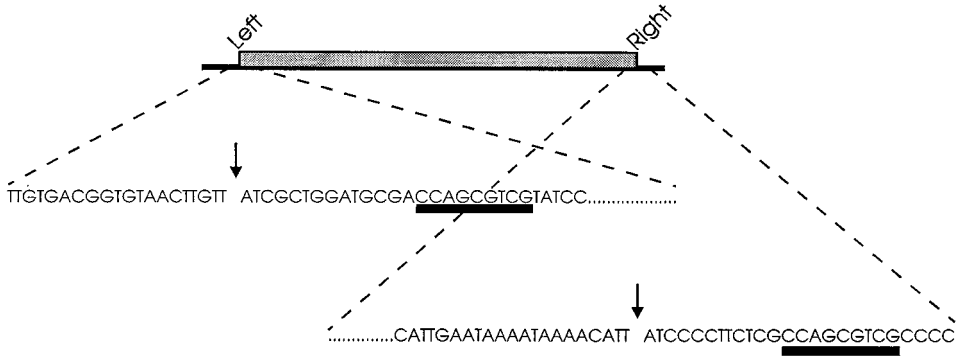

FIG. 6. Direct repeats adjacent to the junction of the pathogenicity island. Nucleotide sequences of the junctions of the PAI (arrows) reveal direct repeats of 9 bp (underlined). One repeat is located 14 bp downstream of the left junction (i.e., just inside the PAI), and the other repeat is located 12 bp downstream of the right junction (i.e., just outside the PAI).

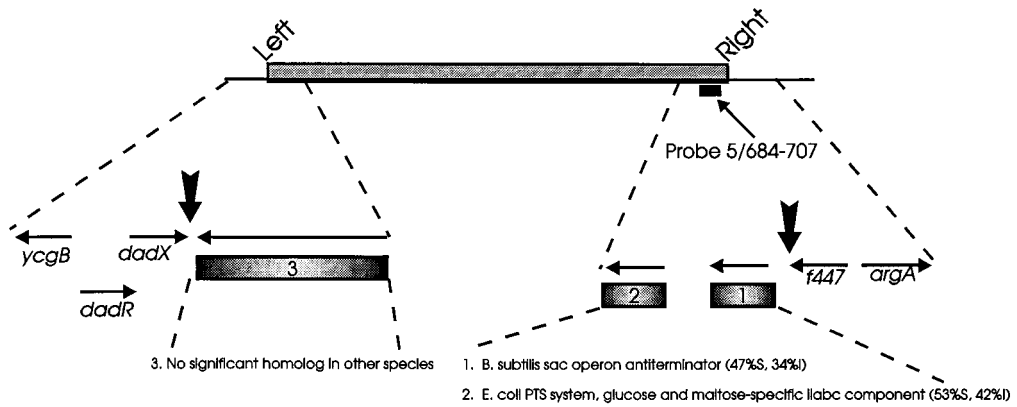

FIG. 7. ORFs within the junctions of the PAI. The nucleotide sequence within the boundaries of the junctions (thick vertical arrows) of the PAI of *E. coli* CFT073 predicts one ORF inside the left junction and two inside the right junction (stippled boxes). Genes previously determined for the K-12 genome are indicated by horizontal arrows above their gene designation. Homologs (best matches) of putative genes predicted by the PAI sequences are listed below the genetic map of these newly described ORFs. Numbers preceding homolog names correspond to the positions of the predicted ORFs. Percent similarity (%S) and percent identity (%I) with respect to the predicted amino acid sequences are shown in parentheses. The accession numbers for homologs are as follows: 1, P26212; 2, P19642. Probe 5/684-707, a 2.1-kb PCR product amplified from a region of 0.8 kb inside the right junction, was used as a probe to survey an *E. coli* strain collection (Table 1). PTS, phosphotransferase.

tical restrictions sites within the region of overlap. In addition, all three clones carry a 6.6-kb *Hin*dIII fragment associated with *pap* gene sequences, which differentiate it from the other copy of the *pap* gene cluster marked by a 3.8-kb *Hin*dIII fragment (29).

In other strains, the sites of insertion are always associated with a tRNA gene. Previously, PAIs have been reported to be inserted into tRNA genes *selC* (70-kb PAI I in strain 536 [12]), *leuX* (190-kb PAI II in strain 536 [12]), *pheV* (170-kb PAI IV in strain J96 [38]), and *pheR* (110-kb PAI V in strain J96 [38]). In this instance, the PAI was inserted near but not within the *metV* gene, another tRNA gene. This represents a different site from that previously reported in other strains.

The stability of PAIs differs in uropathogenic strains. The basis for this stability (or instability) appears to be due to the presence of direct repeats found near the junctions of the PAIs. Long direct repeats, such as those found in strain 536, are associated with instability and spontaneous deletion of the PAI from strain 536. Such deletion mutants can be isolated at high frequencies (10^{-3}) (14). The phenotypes of P fimbriae and hemolysin production, however, are stable within strain CFT073; this may be due to the presence of only short direct repeats (Fig. 6). The direct repeats of CFT073 may be of sufficient size to have allowed initial integration of the PAI at this site but are not sufficiently large to allow frequent pairing and subsequent deletion of these sequences from the chromosome of strain CFT073.

It is likely that strain CFT073 contains a second PAI, which is associated with the other copy of the P fimbrial genes present in this strain (29). The two other well-characterized uropathogenic strains, 536 and J96, for which PAIs have been identified both have two separate PAIs of different sizes (12, 38).

For years, the focus of studies of uropathogenic *E. coli* strains has been P fimbriae and hemolysin. These proteins represent prominent phenotypic markers of uropathogenic strains, particularly the strains associated with upper UTI. Because of their ease of assay (agglutination or hemolysis of erythrocytes), research has focused intensely on these gene products as major players in the model for the pathogenesis of upper UTI. While these genes may represent highly visible markers for the presence of PAIs in uropathogenic strains, they may themselves not be sufficient or even necessary for the virulence of these organisms. We reported previously that construction of isogenic mutants of strain CFT073 in which both copies of the *pap* gene cluster were mutated resulted in no measurable decrease in virulence in the CBA mouse model of ascending UTI (29). This study included an analysis of more than 100 mice challenged with the parent or *pap*-negative mutant at four different inoculum concentrations ranging from 10^5 to 10^9 CFU. A similar lack of effect was observed when the virulence of wild-type CFT073 and a Tn*phoA* mutant with a mutation within *hlyD* (hemolysin-negative phenotype) were compared. No measurable loss of virulence due to the hemolysin mutation was observed in the mouse model (17).

The lack of effect of such mutations, coupled with the presence of PAIs in pyelonephritogenic strains that encompass hundreds of kilobases of DNA, suggests that the genes truly associated with pathogenicity may reside in these uncharacterized sequences present in the PAI of strain CFT073 and other pyelonephritogenic isolates. Clearly, novel genes are encoded within the PAI of CFT073. These genes appear to be distinct from homologs identified by Swenson et al., who described new PAI-associated genes by sample sequencing (38). This provides

TABLE 1. Hybridization of the probe derived from within the PAI of *E. coli* CFT073 with genomic DNA from clinical isolates of *E. coli*

Source of isolate	No. of strains positive for hybridization with probe derived from within PAI/total no.[a]	% of hybridization-positive strains	*P*[b]
Acute pyelonephritis	53/67	79.1	
Cystitis	31/38	81.6	NS[c]
Catheter-associated bacteriuria[d]	23/40	46.9	<0.001
Fecal sample	5/27	18.5	<0.001

[a] A 2.1-kb probe (probe 5/684-707) which was PCR amplified from a region within the pathogenicity island that is 0.8 kb inside the right junction was used for dot blot hybridization with genomic DNA from strains isolated from the indicated source; the location of the probe is shown in Fig. 7.
[b] Compared to acute pyelonephritis isolates, using a chi-square test.
[c] NS, not significant ($P > 0.3$).
[d] First-week isolate.

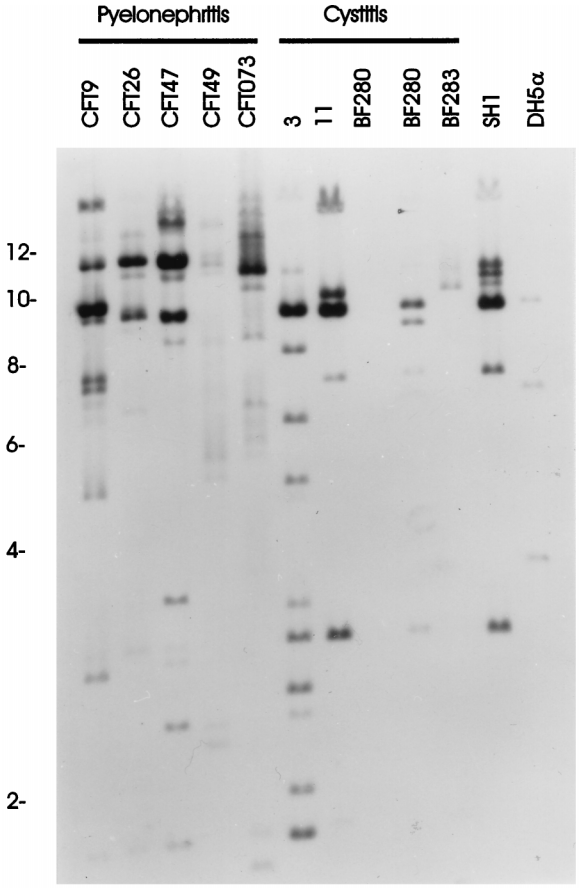

FIG. 8. Southern blot of *Hin*dIII-digested chromosomal DNA from *E. coli* strains hybridized with a probe from the CFT073 pathogenicity island. Chromosomal DNA was isolated from the *E. coli* strains, digested with *Hin*dIII, electrophoresed on an agarose gel, and used to prepare a Southern blot. The blot was hybridized with an 11-kb *Hin*dIII fragment isolated from within the CFT073 PAI (see the text). Strain SH1 is a *lac*-negative derivative of uropathogenic strain J96. Strain DH5α was used as representative of K-12 genomic sequence.

evidence that the PAIs of one uropathogenic strain may differ from those of another. As we and others continue our studies of the PAIs of uropathogenic *E. coli*, it is likely that we will be able to place strains into one of several categories, each representing defined mechanisms of pathogenesis in the development of UTI and acute pyelonephritis. This is reminiscent of diarrheogenic strains of *E. coli* that were once grouped together but have now been found to have distinct mechanisms of pathogenesis. For example, strains classified as enterotoxigenic, enteropathogenic, and enterohemorrhagic appear to have distinct sets of genes associated with their ability to cause diarrhea (11). It is likely that we will also be able to subdivide uropathogenic strains of *E. coli* into distinct classes of pathogens once the PAIs of individual strains are further characterized.

ACKNOWLEDGMENT

This work was supported by Public Health Service grant DK47920 from the National Institutes of Health.

REFERENCES

1. **Akopyants, N. S., D. Kersulyte, and D. E. Berg.** 1995. CagII, a new multigene locus associated with virulence in *Helicobacter pylori*. Gut **37**(Suppl. 1):A1.
2. **Ausubel, F. M., R. Brent, R. E. Kingston, D. D. Moore, J. G. Seidman, J. A. Smith, and K. Struhl.** 1995. Current protocols in molecular biology, vol. 2, p. 10.8.1–10.8.17. John Wiley & Sons, Inc., New York, N.Y.
3. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
4. **Bloch, C. A., and C. K. Rode.** 1996. Pathogenicity island evaluation in *Escherichia coli* K1 by crossing with laboratory strain K-12. Infect. Immun. **64:** 3218–3223.
5. **Bloch, C. A., C. K. Rode, V. Obreque, and K. Y. Russell.** 1994. Comparative genome mapping with mobile physical map landmarks. J. Bacteriol. **176:** 7121–7125.
6. **Blum, G., V. Falbo, A. Caprioli, and J. Hacker.** 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and α-hemolysin form the pathogenicity island II of the uropathogenic *Escherichia coli* strain J96. FEMS Microbiol. Lett. **126:**189–196.
7. **Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschape, and J. Hacker.** 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. Infect. Immun. **62:**606–614.
8. **Brown, K., P. W. Finch, I. D. Hickson, and P. T. Emmerson.** 1987. Complete nucleotide sequence of the *Escherichia coli argA* gene. Nucleic Acids Res. **15:**10586.
9. **Fetherston, J. D., P. Schuetze, and R. D. Perry.** 1992. Loss of the pigmentation phenotype in *Yersinia pestis* is due to the spontaneous deletion of 102 kb of chromosomal DNA which is flanked by a repetitive element. Mol. Microbiol. **6:**2693–2704.
10. **Foxman, B., L. Zhang, K. Palin, P. Tallman, and C. F. Marrs.** 1995. Bacterial virulence characteristics of Escherichia coli isolates from first-time urinary tract infection. J. Infect. Dis. **171:**1514–1521.
11. **Guerrant, R. L.** 1990. Principles and syndromes of enteric infection, p. 837–851. *In* G. L. Mandell, R. G. Douglas, Jr., and J. E. Bennett (ed.), Principles and practice of infectious diseases, 3rd ed. Churchill Livingstone, Inc., New York, N.Y.
12. **Hacker, J.** 1990. Genetic determinants coding for fimbriae and adhesins of extraintestinal *Escherichia coli*. Curr. Top. Microbiol. Immunol. **151:**1–27.
13. **Hacker, J., L. Bender, M. Ott, J. Wingender, B. Lund, R. Marre, and W. Goebel.** 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur *in vitro* and *in vivo* in various extraintestinal *Escherichia coli* isolates. Microb. Pathog. **8:**213–225.
14. **Hacker, J., S. Knapp, and W. Goebel.** 1983. Spontaneous deletions and flanking regions of the chromosomally inherited hemolysin determinant of an *Escherichia coli* O6 strain. J. Bacteriol. **154:**1145–1152.
15. **High, N. J., B. A. Hales, K. Jann, and G. J. Boulnois.** 1988. A block of urovirulence genes encoding multiple fimbriae and hemolysin in *Escherichia coli* O4:K12:H$^-$. Infect. Immun. **56:**513–517.
16. **Hughes, C., J. Hacker, H. Duvel, and W. Goebel.** 1987. Chromosomal deletions and rearrangements cause coordinate loss of haemolysis, fimbriation and serum resistance in a uropathogenic strain of *Escherichia coli*. Microb. Pathog. **2:**227–230.
17. **Johnson, D. E., H. L. T. Mobley, and J. W. Warren.** Unpublished data.
18. **Johnson, J. R.** 1991. Virulence factors in *Escherichia coli* urinary tract infection. Clin. Microbiol. Rev. **4:**80–128.
19. **Kafatos, F. C., C. W. Jones, and A. Efstratiadis.** 1979. Determination of nucleic acid sequence homologies and relative concentrations by a dot blot hybridization procedure. Nucleic Acids Res. **7:**1541–1552.
20. **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.
21. **Kunin, C. M.** 1987. Detection, prevention and management of urinary tract infections, 4th ed. Lea & Febiger, Philadelphia, Pa.
22. **Lange, N., S. Censino, Z. Xiang, M. Borodovsky, R. Rappuoli, and A. Covacci.** 1995. The cagI chromosomal region of *Helicobacter pylori* is type I-specific and encodes exporter molecules and invasion genes, p. 63. *In* 8th International Workshop on Campylobacters, Helicobacters and Related Organisms.
23. **Lee, C. A.** 1996. Pathogenicity islands and the evolution of bacterial pathogens. Infect. Agents Dis. **5:**1–7.
24. **Lobocka, M., J. Hennig, J. Wild, and T. Klopotowski.** 1994. Organization and expression of the *Escherichia coli* K-12 *dad* operon encoding the smaller subunit of D-amino acid dehydrogenase and the catabolic alanine racemase. J. Bacteriol. **176:**1500–1510.
25. **Low, D., V. David, D. Lark, G. Schoolnik, and S. Falkow.** 1984. Gene clusters governing the production of hemolysin and mannose-resistant hemagglutination are closely linked in *Escherichia coli* serotype O4 and O6 isolates from urinary tract infections. Infect. Immun. **43:**353–358.
26. **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.
27. **Mills, D. M., V. Bajaj, and C. A. Lee.** 1995. A 40 kb chromosomal fragment encoding *Salmonella typhimurium* invasion genes is absent from the corresponding region of the *Escherichia coli* K-12 chromosome. Mol. Microbiol. **15:**749–759.
28. **Mobley, H. L. T., D. M. Green, A. L. Trifillis, D. E. Johnson, G. R. Chippendale, C. V. Lockatell, B. D. Jones, and J. W. Warren.** 1990. Pyelonephri-

togenic *Escherichia coli* and killing of cultured human renal proximal tubular epithelial cells: role of hemolysin in some strains. Infect. Immun. **58:**1281–1289.
29. **Mobley, H. L. T., K. G. Jarvis, J. P. Elwood, D. I. Whittle, C. V. Lockatell, R. G. Russell, D. E. Johnson, M. S. Donnenberg, and J. W. Warren.** 1993. Isogenic P-fimbrial deletion mutants of pyelonephritogenic Escherichia coli: the role of αGal(1-4) Gal binding in virulence of a wild-type strain. Mol. Microbiol. **10:**143–155.
30. **Morschhauser, J., V. Vetter, L. Emody, and J. Hacker.** 1994. Adhesin regulatory genes within large, unstable DNA regions of pathogenic *Escherichia coli*: cross-talk between different adhesin gene clusters. Mol. Microbiol. **11:** 555–566.
31. **Muhldorfer, I., and J. Hacker.** 1994. Genetic aspects of *Escherichia coli* virulence. Microb. Pathog. **16:**171–181.
32. **National Center for Health Statistics.** 1994. Detailed diagnoses and procedures. National Hospital Discharge Survey 1991. National Center for Health Statistics, Centers for Disease Control, Atlanta, Ga.
33. **National Center for Health Statistics.** Unpublished data from 1991 National Ambulatory Medical Care Survey.
34. **Ritter, A., G. Blum, L. Emody, M. Kerenyi, A. Bock, B. Neuhierl, W. Rabsch, F. Scheutz, and J. Hacker.** 1995. tRNA genes and pathogenicity islands: influence on virulence and metabolic properties of uropathogenic *Escherichia coli*. Mol. Microbiol. **17:**109–121.
35. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
36. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
37. **Sedgwick, S. G., M. Robson, and F. Malik.** 1988. Polymorphism in the *umuDC* region of *Escherichia* species. J. Bacteriol. **170:**1610–1616.
38. **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.
39. **Warren, J. W., J. H. Tenney, H. M. Hoopes, H. L. Muncie, and W. C. Anthony.** 1982. A prospective microbiologic study of bacteriuria in patients with chronic indwelling urethral catheters. J. Infect. Dis. **146:**719–723.

Editor: A. O'Brien