INFECTION AND IMMUNITY, Sept. 1996, p. 3736–3743 Vol. 64, No. 9
0019-9567/96/$04.00+0

Two Pathogenicity Islands in Uropathogenic *Escherichia coli* J96: Cosmid Cloning and Sample Sequencing

DANA L. SWENSON,[1]† NICK O. BUKANOV,[2]‡ DOUGLAS E. BERG,[2] AND RODNEY A. WELCH[1]*

Department of Medical Microbiology and Immunology, University of Wisconsin—Madison, Madison, Wisconsin 53706,[1] and Departments of Molecular Microbiology and Genetics, Washington University School of Medicine, St. Louis, Missouri 63110[2]

Received 24 April 1996/Returned for modification 1 June 1996/Accepted 21 June 1996

Many of the virulence genes of pathogenic strains of *Escherichia coli* are carried in large multigene chromosomal segments called pathogenicity islands (PAIs) that are absent from normal fecal and laboratory K-12 strains of this bacterium. We are studying PAIs in order to better understand factors that govern virulence and to assess how such DNA segments are gained or lost during evolution. The isolation and sample sequencing of a set of 11 cosmid clones that cover all of one and much of a second large PAI in the uropathogenic *E. coli* J96 are described. These PAIs were mapped to the 64- and 94-min regions of the *E. coli* K-12 chromosome, which differ from the locations of three PAIs identified in other pathogenic *E. coli* strains. Analysis of the junction sequences with *E. coli* K-12-like DNAs showed that the insert at 94 min is within the 3′ end of a phenylalanine tRNA gene, *pheR*, and is flanked by a 135-bp imperfect direct repeat. Analysis of the one junction recovered from the insert at 64 min indicated that it lies near another tRNA gene, *pheV*. To identify possible genes unique to these PAIs, 100 independent subclones of the cosmids were made by *Pst*I digestion and ligation into a pBS+ plasmid and used in one-pass sample DNA sequencing from primer binding sites at the cloning site in the vector DNA. Database searches of the J96 PAI-specific sequences identified numerous instances in which the cloned DNAs shared significant sequence similarities to adhesins, toxins, and other virulence determinants of diverse pathogens. Several likely insertion sequence elements (IS*100*, IS*630*, and IS*911*) and conjugative R1 plasmid and P4 phage genes were also found. We propose that such mobile genetic elements may have facilitated the spread of virulence determinants within PAIs among bacteria.

Escherichia coli is a normal inhabitant of the intestines of most animals and humans. There are distinct *E. coli* strains that have the potential to cause a wide variety of extraintestinal diseases, including septicemia and meningitis in neonates and urinary tract infections in children and adults (46). These pathogenic isolates of *E. coli*, in contrast to common fecal strains, express traits that enable them to overcome host defenses, proliferate at extraintestinal sites, and cause tissue damage and disease (42). For the uropathogenic *E. coli*, several adhesins that are distinguished by host cell receptor specificities are important in the colonization of the urinary tract (21, 28, 32). These strains often express exotoxins, such as hemolysin and cytotoxic necrotizing factor type 1 (CNF-1) (10, 13, 59, 60). The genes for these colonization and virulence factors are linked, and in some cases duplicated, on the chromosomes of many uropathogenic *E. coli* isolates (13, 14, 24, 25, 35). The presence of genes for multiple virulence traits within regions of DNA unique to uropathogenic *E. coli* led Hacker et al. (22) to call these DNA segments pathogenicity islands (PAIs). Diarrheal *E. coli* strains also possess large chromosomal blocks of linked pathogenic determinants. For example, genes for the attaching and effacing phenotype of enteropathogenic and enterohemorrhagic *E. coli* strains are part of a discrete chromosomal segment called the locus of enterocyte effacement (38). Other diverse pathogenic bacteria that also possess large blocks of pathogenesis-associated genes include *Yersinia*, *Salmonella*, and *Dichelobacter* spp. and *Helicobacter pylori* (1, 16, 39, 44). Thus, the clustering of virulence genes appears to be a general phenomenon among bacteria (33).

Here we present evidence that at least two PAIs are present in *E. coli* J96 and that the PAIs are linked to tRNA loci but at sites different from those occupied by other known *E. coli* PAIs. Furthermore, using a sample sequencing approach, we identified potential J96 PAI genes for toxin and adhesin determinants that were previously unrecognized in uropathogenic *E. coli*. This analysis also led to the identification of DNA sequences for insertion elements and phage and plasmid genes within the *E. coli* J96 PAIs that may facilitate the dissemination of these virulence determinants in bacterial populations.

MATERIALS AND METHODS

Bacterial strains and plasmids. The uropathogenic strain *E. coli* J96 (O4:K6) (25, 35, 40) was used as the source of chromosomal DNA for construction of a cosmid library. *E. coli* J96-M1 is a spontaneous-deletion derivative of *E. coli* J96 (22) and was kindly supplied by Jorg Hacker. *E. coli* K-12 DH5α and DH12 (Gibco/BRL, Gaithersburg, Md.) were used as hosts for maintaining cosmid and plasmid clones. The Kohara phage gene library (30) was purchased from TaKaRa Biochemical Inc. (Berkeley, Calif.).

Lorist 6, a 5.2-kb moderate-copy-number cosmid vector with T7 and SP6 promoters close to the cloning site (20), was used for construction of the *E. coli* J96 cosmid library. pBS+, a high-copy-number plasmid (Stratagene, La Jolla, Calif.), was used for the cloning of *Pst*I fragments from cosmid DNAs. The P-fimbria-encoding plasmids pPAP5 (34) and pPAP16 (27) were kindly supplied by Scott Hultgren. The hemolysin-containing plasmids pSF4000 and pWAM589 were reported previously (59, 62). pDAR118 is an 8-kb *Sph*I subclone of cosmid 4 from the J96 library (57).

Bacteria were grown overnight at 37°C on Luria-Bertani agar or in Luria-Bertani broth. *E. coli*-harboring cosmid clones and pBS+ clones were grown with 25 μg of kanamycin per ml and 100 μg of ampicillin per ml, respectively. To assay for hemolysin production, cosmid-harboring *E. coli* cells were cultivated on sheep red blood agar supplemented with 10 mM $CaCl_2$ as described previously (60).

DNA isolation and recombination techniques. *E. coli* J96 and DH5α genomic

* Corresponding author. Phone: (608) 263-2700. Fax: (608) 262-8414. Electronic mail address: rawelch@facstaff.wisc.edu.

† Present address: Department of Molecular Microbiology, Washington University School of Medicine, St. Louis, MO 63310.

‡ Present address: Center for Advanced Biotechnology, Boston University, Boston, MA 02215.

DNAs were prepared as previously described (8). *E. coli* J96-M1 chromosomal DNA was prepared by extraction with guanidium thiocyanate (48). Cosmid and plasmid DNAs were extracted from *E. coli* host cells by alkaline lysis, and cosmid DNA used for subcloning was further purified by CsCl-ethidium bromide equilibrium density centrifugation (54).

Restriction endonucleases and other DNA-modifying enzymes were purchased from commercial sources and used according to the manufacturers' instructions. *Pst*I subclones from cosmid DNA were constructed and transformants were characterized and analyzed by standard protocols (54).

Cosmid library construction. A cosmid library of *E. coli* J96 DNA was constructed essentially as described by Bukanov and Berg (8). Briefly, DNA was digested with *Sau*3AI under conditions that generated fragments with an average size of 40 to 50 kb and electrophoresed through 1% agarose gels. Fragments of 35 to 50 kb were isolated and cloned into Lorist 6 vector that had been linearized with *Bam*HI and treated with bacterial alkaline phosphatase to block self-ligation. Cloned DNA was packaged in lambda phage particles in vitro by using a commercial kit (Amersham, Arlington Heights, Ill.), and cosmid-containing phage particles were used to transduce *E. coli* DH5α. Transductant colonies were transferred to 150 μl of Luria-Bertani broth supplemented with kanamycin in 96-well microtiter plates and grown overnight at 37°C with shaking.

Hybridization probe preparation. RNA probes specific for the ends of cloned cosmid DNAs (riboprobes) were generated by in vitro transcription as follows. Cosmid DNA (1 to 2 μg) was digested with *Eco*RV or *Hae*III (for probes from T7 and SP6 promoters, respectively) to limit transcript length. The resulting restriction fragments were treated with chloroform-isoamyl alcohol (24:1), ethanol precipitated, resuspended in 10 μl of distilled water, and used in in vitro transcription reactions with [α-^{32}P]dCTP and T7 or SP6 polymerases under conditions recommended by the supplier (Promega, Madison, Wis.). The labelled RNA was extracted with chloroform-isoamyl alcohol (24:1), ethanol precipitated, and then dissolved in 50 μl of distilled water for use in hybridization analysis.

DNA fragments for use as hybridization probes were purified from agarose by using GeneClean (Bio101, Vista, Calif.) and were labelled by using an enhanced chemiluminescence (ECL) direct labelling kit (Amersham) or a random-hexamer primer-labelling kit (Stratagene) and [α-^{32}P]dCTP (Amersham).

Colony and Southern blot hybridization. Overlapping cosmid clones were identified by colony and Southern blot hybridization as described previously (8). In brief, for colony blot hybridizations, RNA probes were hybridized at 50°C in 50% formamide followed by three washes at 65°C in 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% sodium dodecyl sulfate (SDS) and one wash in 0.1× SSC–0.1% SDS; DNA probes were hybridized at 42°C in 50% formamide followed by four washes at room temperature in 2× SSC–0.1% SDS and two washes in 1× SSC–0.1% SDS at 68°C. For Southern blot hybridizations, DNA probes were hybridized at 42°C in 50% formamide followed by one wash in 2× SSC–0.5% SDS at 20°C, one wash in 2× SSC–0.1% SDS at 20°C, and then two washes in 0.1× SSC–0.5% SDS at 68°C.

DNA fragments from cosmids were electrophoresed on 0.45% agarose gels, transferred to Hybond-N+ (Amersham) by capillary action, and alkali fixed to the membrane according to the manufacturer's instructions. Membrane prehybridizations, hybridizations, and washes were performed under high-stringency conditions in the presence of urea in accordance with the ECL protocol (Amersham). Hybridization results obtained with radioactively labelled probes were visualized by autoradiography, and ECL-labelled probes were visualized by luminography.

DNA sequence analysis and PCR. Sequencing of double-stranded DNAs from pBS+ subclones was carried out with the Sequenase 2.0 sequencing kit (U.S. Biochemical, Cleveland, Ohio). Double-stranded cosmid DNA was sequenced by using the Fmol sequencing system (Promega) and a Thermocycler 480 apparatus (Perkin-Elmer, Foster City, Calif.). Universal primers corresponding to the T7 and T3 promoters of pBS+ were used in all plasmid sequencing reactions. Oligonucleotides for DNA sequencing and PCR were synthesized as required with an Applied Biosystems DNA Synthesizer 481 (Perkin-Elmer) or were purchased from commercial sources. Database comparisons were performed with the Basic Local Alignment Search Tool (BLAST) program with similarity score (*S*-score) parameters as recommended by the designers (2). Alignments of junction sequences were performed with DNAStar software (DNAStar, Madison, Wis.).

RESULTS

Construction and ordering of the *E. coli* J96 cosmid library. Previous work had shown that *E. coli* J96 contains *hly*, *pap*, and *prs* gene clusters that are not present in *E. coli* K-12 (35, 61). To investigate the structure and function of such putative PAIs, a cosmid library was constructed by ligating size-selected (35- to 50-kb) *Sau*3AI partial-digestion fragments of J96 genomic DNA into the cosmid vector Lorist 6, and packaged cosmid clones were transduced into *E. coli* DH5α. Cosmid DNAs from 12 representative transductants were all 40 to 50 kb in length and exhibited different *Eco*RI digestion patterns, indicating that the library was random. Six hundred seventy-two kanamycin-resistant colonies were individually transferred to wells of microtiter plates for further analysis. Given a genome size of approximately 4,800 kb and cloned DNA fragments of approximately 40 kb, such a library should be approximately sixfold redundant.

To start identifying clones from the putative PAIs, the library was screened for hemolysin production on blood agar plates, and seven β-hemolytic clones were identified. Concurrently, nylon filters prepared with cosmid-containing bacteria were hybridized with pPAP5-derived probes to further identify PAI-containing cosmid clones. Fourteen clones exhibited *pap* or *pap*-like sequences, and three of these clones were also hemolytic. *Eco*RI, which generates fragments of 15 and 19.5 kb from *pap* and *prs*, respectively (43), was used to distinguish clones carrying genes for each type of pilus.

A walking method, involving hybridization of probes made from the ends of representative cloned DNAs to filters containing the full cosmid library, was used to order and identify overlapping PAI clones in the library. Clones that hybridized to *pap* probes and exhibited hemolytic activity were used first to make riboprobes; they were also hybridized to the cosmid library and to Southern blots containing *E. coli* DH5α and J96 chromosomal DNA to identify those that contained PAI-specific DNA and potential K-12 boundary sequences. Cosmid-derived riboprobes that hybridized with both *E. coli* K-12 and J96 DNAs likely contained either the junction between PAI-specific and K-12-type DNA or copies of large repeated sequences such as insertion sequence (IS) elements. The hybridization analysis identified two clusters of overlapping cosmids, and minisets used for subsequent analysis were assembled (Fig. 1). One miniset of four overlapping cosmid clones covers the *prs*-containing PAI (PAI V). A second miniset of seven overlapping cosmid clones covers much of the *pap*-containing PAI (PAI IV), including the right junction with K-12-type sequences (left and right designations are based on clockwise alignment of PAIs with the *E. coli* K-12 chromosome map). The PAIs are designated IV and V in accordance with previous publications (22, 39). The left junction for PAI IV is not represented in the cosmid library, and given the redundancy of the library, this gap may reflect selection against particular clones, as has been seen elsewhere (8, 18, 30). Attempts to close the gap by Southern hybridization of *E. coli* J96 DNA with a *pheV* probe and subsequent clone construction with DNA size selected on the basis of the corresponding hybridization signal were not successful. In addition, this gap could not be closed by either PCR strategies capable of amplifying >9-kb DNA species or direct DNA sequence analysis of J96 genomic-DNA preparations.

To determine the location of the PAI junctions relative to the K-12 chromosome map, RNA probes derived from cosmid clones carrying the K-12–J96 junction were hybridized to a filter containing the ordered Kohara phage library. The hybridization pattern indicated junction points for the two PAIs at 64 (*pheV*) and 94 (*pheR*) min on the K-12 chromosome. Intervening cosmids carrying only PAI-specific DNA did not hybridize to the Kohara phage filters. The size of PAI V, on the basis of the sizes of cloned DNAs and the predicted amount of DNA overlap (~20%), is estimated as 1.1×10^5 bp. We estimate that PAI IV is more than 1.7×10^5 bp, but because the size of the gap is unknown, the size of PAI IV cannot be more precisely determined at this time. Restriction endonuclease fragment maps of PAI-containing cosmids are shown in Fig. 2.

Construction and characterization of cosmid subclones. *Pst*I fragments were generated from the 11 individual cosmids

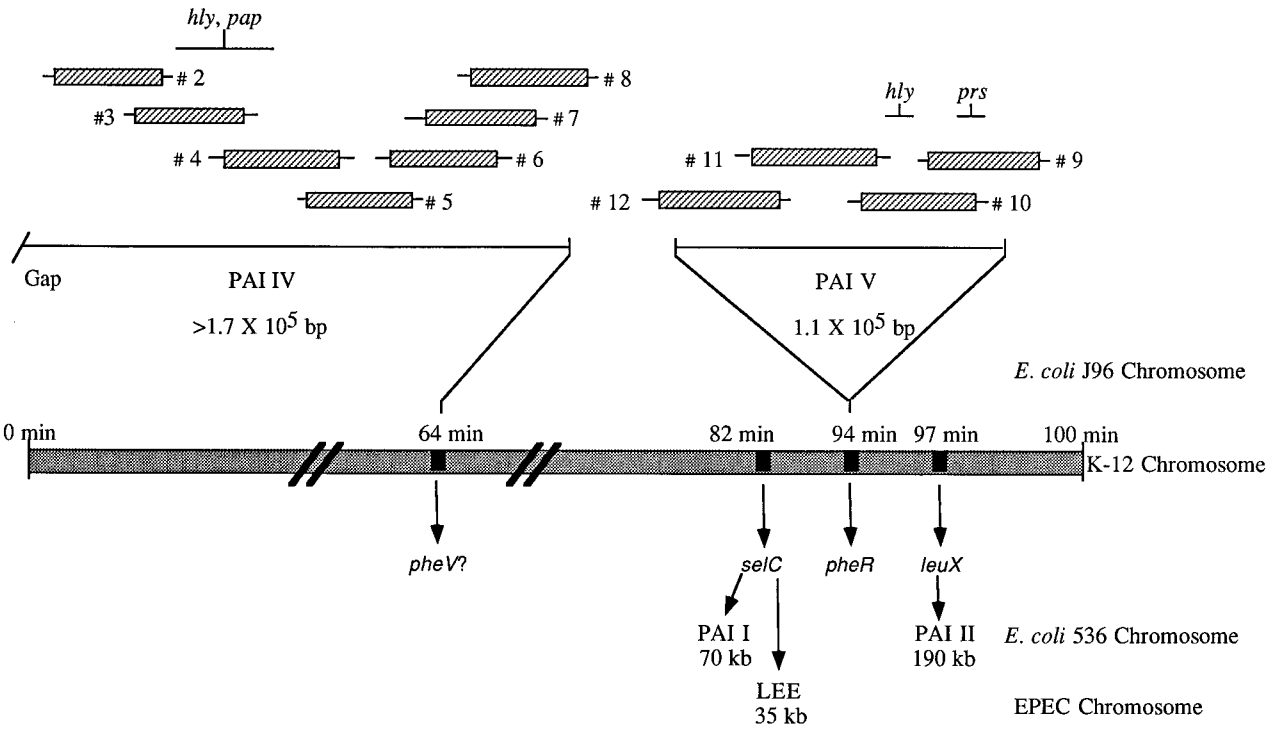

FIG. 1. Schematic diagram of cosmid clones derived from *E. coli* J96 PAI and map positions of known *E. coli* PAIs (not drawn to scale). The gray bar represents the *E. coli* K-12 chromosome (3), with minute demarcations for PAI junction points located above the bar. *E. coli* J96 overlapping cosmid clones are represented by hatched bars (overlap not drawn to scale), with the positions of *hly*, *pap*, and *prs* operons indicated above the bars. The PAIs and their estimated sizes are shown above and below the K-12 chromosome map. LEE, locus of enterocyte effacement; EPEC, enteropathogenic *E. coli*.

shown in Fig. 1 and ligated into *Pst*I-digested pBS+, and 100 independent clones carrying one or more *Pst*I fragments were isolated. The possible presence of multiple fragments in particular subclones did not complicate our studies because the determination of the precise gene order within a cosmid was not the object of this analysis. Approximately 100 to 200 nucleotides of sequence were determined from the ends of each clone by using vector T7 and T3 promoter sequences as universal primer binding sites, thereby generating two distinct sequences from each subclone. The sequences were compared with the database for analysis of similarity at both the protein and nucleotide levels by using the BLAST algorithm (2). Similarity results and *S*-score values for representative searches are summarized in Table 1.

To place selected *Pst*I fragments on the low-resolution restriction endonuclease maps obtained for the cosmid clones, probes were generated from several clones shown in Table 1 and Southern hybridizations were performed. The results are shown in Fig. 2. Probes derived from hemolysin-containing clones (589/B) hybridized to both PAIs as expected on the basis of earlier studies (26) and our initial cosmid mapping. The coding region for CNF-1 was found on PAI V (cosmid 10), linked to *hly* sequences, as previously described (14). An agglutinin-like sequence, identified recently in *E. coli* K99, was found on cosmid 9 (PAI V). Sequences similar to an activator component of two-component regulatory systems (clone 11-6) were found on the same fragment as the open reading frame (ORF) sequence matching the *Bacillus stearothermophilus* IS. Sequences similar to gene 32, which is proximal to the origin of transfer (*oriT*) of the *E. coli* R1 plasmid (clone 7-6) and bacteriophage P4-like sequences (clones 7-13 and 9-6), were hybridized against cosmids from both PAIs. The gene 32 probe hybridized to cosmids 6 and 7 (PAI IV), apparently on overlapping DNA fragments, and also to cosmid 12 (PAI V) (57). Sequences similar to those encoding a protein of unknown function in bacteriophage P4 (clone 7-13) hybridized to cosmids 6 and 7 (PAI IV) but not to cosmids representing the PAI V. Sequences similar to the *attP* and *int* regions of bacteriophage P4 (clone 9-6) were found in cosmid 9 (near the right PAI-specific junction of PAI V) but not in the respective PAI IV cosmids. Sequences similar to the insertion element IS*630* (clone 4-27) hybridized only to cosmid 4 (PAI IV). Sequences similar to the *Yersinia pestis*-like insertion element IS*100* were identified on two overlapping cosmids (11 and 12) from PAI V. A potential gene encoding an ORF (clone 10-2) similar to the *B. stearothermophilus* IS mentioned above was identified on cosmids 10 and 11 in PAI V.

Sequences at PAI–K-12 junctions. The precise junction points of PAI V with *E. coli* K-12 chromosomal-type sequences were determined by sequencing of appropriate clones carrying the two ends of PAI V. DNA sequences were determined from the ends of cosmid 9 and 12 inserts, and these sequences matched the K-12 *cad* gene region (94 min), in agreement with the Kohara phage mapping results. Oligonucleotide primers were synthesized to match the K-12 *cad* gene sequence and used to progressively walk toward and into the J96-specific DNA sequences from both junction regions. Both strands were sequenced by using appropriate primers (Fig. 3). This analysis showed that PAI V was near the *cad* genes and a phenylalanine tRNA gene, *pheR*, and revealed a 135-bp nearly perfect duplication at the left and right J96–K-12 junctions. An intact copy of *pheR* was present at the right junction, and a segment consisting of 25-bp of the 3′ end of *pheR* with a single internal base pair deletion was found at the left junction. Only the left

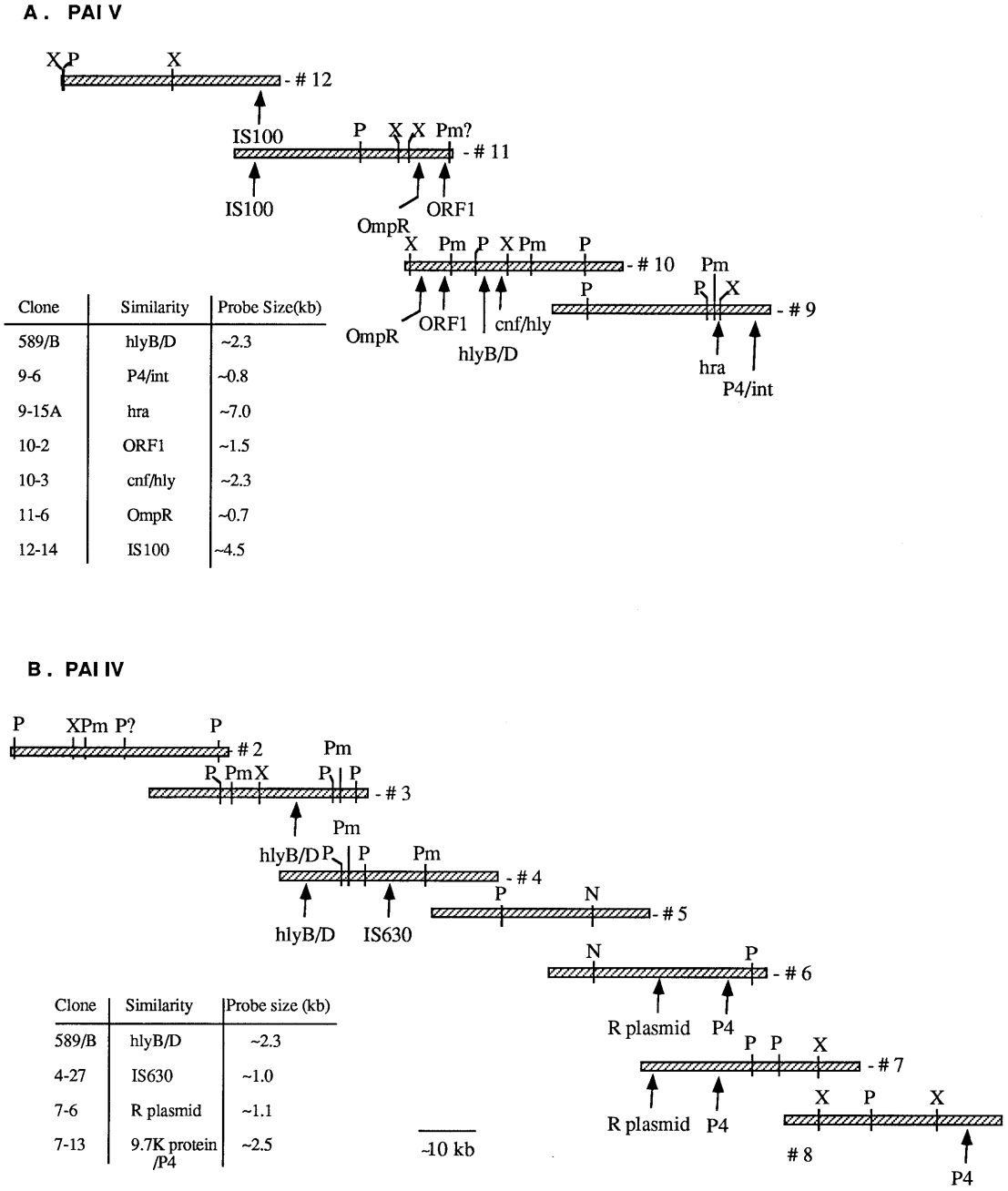

FIG. 2. Low-resolution restriction endonuclease map and summary of hybridization analysis for PAIs. The hatched bars represent overlapping cosmid clones derived from *E. coli* J96. (A) PAI V; (B) PAI IV. Letters indicate mapped restriction sites: N, *Not*I; P, *Pac*I; Pm, *Pme*I; and X, *Xho*I. The maps were aligned on the basis of common patterns of restriction endonuclease cleavage sites, and the exact extent of overlap is unknown. The tabular insets indicate the clones and database sequences with similarity from which probes were isolated. Arrows indicate fragments to which the probes hybridize; the precise locations and order of the sequences are unknown.

junction carries *cad* sequences, while the sequence between the *cad* and *pheR* genes is present at both junctions. Approximately 100 bp immediately adjacent to the duplicated regions within the PAI did not exhibit similarity to sequences in the database (57).

To analyze the DNA remaining after the loss of PAI V, the *pheR* region from the genome of J96-M1, a spontaneous-deletion derivative, was PCR amplified with primers matching a site in the K-12 *cad* gene and sequences downstream of *pheR*. This resulted in a 1.1-kb PCR product, as expected from the K-12 sequence. We sequenced approximately 300 bp of the PCR product through the *pheR* gene and found a single copy of the duplicated region carrying the single base pair deletion at the 3′ end of *pheR* that is characteristic of the left junction sequence (Fig. 3).

A similar strategy of DNA walking from the T7 promoter sequence on cosmid 8 identified the apparent right junction of PAI IV (57). The gap in the cosmid library prevented determination of the left junction. Comparison of PAI IV to the *E. coli* K-12 genome sequence indicated a match in the 65- to

TABLE 1. Summary of BLAST search

Clone[a]	Database sequence with similarity[b]	Accession no.[c]	Organism	Nucleotides entered	*S* score[d]	Probability
4-18	IS*911(P)*	U14003	*E. coli*	130	61	1.8×10^{-9}
4-27	IS*630* (N, P)	X05955	*Shigella sonnei*	119	327	1.2×10^{-35}
10-2	ORF1 (P)	D10543	*B. stearothermophilus*	110	87	4.1×10^{-5}
12-6	ORF1, IS*100* (P)	X78302	*Y. pestis*	233	603	6.9×10^{-72}
12-14	ORF2, IS*100* (P)	X78302	*Y. pestis*	76	380	4.3×10^{-24}
pSF4000	*rhs* (N, P)	L02370	*E. coli*	1,262	539	1.6×10^{-34}
9-15A	*hra* (N, P)	U07174	*E. coli*	305	668	1.3×10^{-89}
10-3	*cnf-1* (N, P)	X70670	*E. coli*	113	499	5.7×10^{-34}
10-3	*hly* (N)	M10133	*E. coli*	81	276	2.4×10^{-17}
589/B	*hlyB*, *hlyD*	M10133	*E. coli*	NA[e]	NA	NA
DAR118	SsaB (P)	M63481	*Streptococcus sanguis*	879	249	1.2×10^{-46}
9-8	*pmf* (N, P)	Z35428	*P. mirabilis*	341	220	2.1×10^{-8}
12-2	HlyB (P)	M22168	*Serratia marcescens*	322	136	7.6×10^{-11}
6-22	HlyA (P)	M22168	*S. marcescens*	376	88	2.1×10^{-5}
7-6	32K protein, R1 (N, P)	X15279	*E. coli*	285	300	3.0×10^{-15}
7-13	9.7K protein (P)	X51522	Bacteriophage P4	130	47	8.5×10^{-5}
9-6	*attP*, *int* (N, P)	X05947	Bacteriophage P4	123	196	1.6×10^{-7}
5-28	*att* (N)	Z11491	Bacteriophage P2	97	147	0.0017
11-6	OmpR (P)	X12374	*Salmonella typhimurium*	330	239	1.2×10^{-26}
11-6	PhoP (P)	X67676	*B. subtilis*	330	236	3.3×10^{-26}
8-7	NeuB (P)	U05248	*E. coli*	163	77	0.0013
8-6	*rfb* (N, P)	L14842	*Shigella flexneri*	87	200	4.4×10^{-9}
10-30	*nlpB* (N, P)	X57402	*E. coli*	145	711	1.1×10^{-51}
4-18	*pgtP* (N, P)	M21278	*S. typhimurium*	104	215	2.6×10^{-10}

[a] Clone name is derived as follows; first number indicates cosmid, second number indicates *Pst*I subclone. Identical clone numbers with brackets indicate search results derived from identical sequences with the same primer. 589/B is a *Pst*I B fragment derived from pWAM589.
[b] N, similarity at the nucleotide level; P, similarity at the protein level.
[c] Accession numbers of database sequences with similarity.
[d] BLAST S-score values are given only when N and/or P is shown.
[e] NA, not applicable.

68-min region (49) within 250 bp of a phenylalanine tRNA gene, *pheV*, which is in accord with the location inferred from the Kohara phage mapping. Precise localization of the junction will require further refinement of the *E. coli* K-12 sequence. The 150 bp of PAI-specific DNA adjacent to the PAI IV right junction were not similar to the sequences adjacent to the PAI V junction but instead indicated similarity to *kps* (capsular polysaccharide) gene sequences (56), which are present in certain pathogenic *E. coli* strains but not in strain K-12.

DISCUSSION

The present study contributes to the emerging concept that large sets of genes, some of which contribute to colonization and virulence, tend to be clustered in PAIs at relatively few loci in the chromosomes of pathogenic bacteria and that pathogenic and nonpathogenic strains of a given species tend to differ by the presence or absence of such PAIs (33). Using *E. coli* J96 as a prototype for uropathogenic *E. coli*, we have begun studying its PAIs in order to identify and better understand the mechanisms involved in extraintestinal *E. coli* diseases and how PAIs are acquired, lost, or changed during bacterial evolution. The *E. coli* J96 chromosome encodes several well-characterized factors involved in extraintestinal disease, including two types of P fimbriae (Pap and Prs), hemolysin, and the CNF-1 toxin (9, 36, 58). Previous studies indicated that the P-fimbrial and CNF-1 genes are linked to the *hly* operon in strain J96 (6, 14). Accordingly, sets of overlapping cosmid clones of two J96 *hly*-encoding PAIs and their J96–K-12-like junction regions were assembled. Two cosmid minisets were acquired, with one set containing four clones covering a complete PAI and a second set containing seven clones covering part, but not all, of a second PAI. Sample sequencing (31) of subclones from these cosmids revealed ORFs with significant amino acid sequence similarity to colonization and virulence gene products of other pathogens and the presence of insertion elements as well as phage and plasmid genes that are likely to be involved in the movement and recombination of the J96 PAIs.

Three different J96 PAI-specific regions matched most closely gene segments in the *Proteus mirabilis* MR/P adhesin (4), enterotoxigenic *E. coli* heat-resistant agglutinin (37), and the *Streptococcus parasanguis* adhesin family (17, 19, 55). These J96 PAI-specific DNA segments may express adhesins that were not previously known to occur in uropathogenic *E. coli*. This remarkable complexity of potential adhesins may help to explain the difficulty in demonstrating in vivo colonization or virulence phenotypes for *E. coli pap* knockout mutants in a murine model of ascending urinary tract infection (41). The extensive 879-bp match to the *S. parasanguis* adhesin family present in the PAI IV subclone is intriguing because it suggests that gram-positive and gram-negative bacteria may share common surface adherence determinants. The *Streptococcus sanguis* adhesin-like sequences are just 5′ of the *hly* operon present on the well-characterized pSF4000 hemolysin recombinant plasmid. Earlier DNA hybridization analysis using a probe corresponding to this pSF4000 region indicated that this portion of the PAI IV is rare among hemolytic, uropathogenic *E. coli* isolates (61). From this we conclude that the linkage relationships within PAI gene blocks can be dissimilar

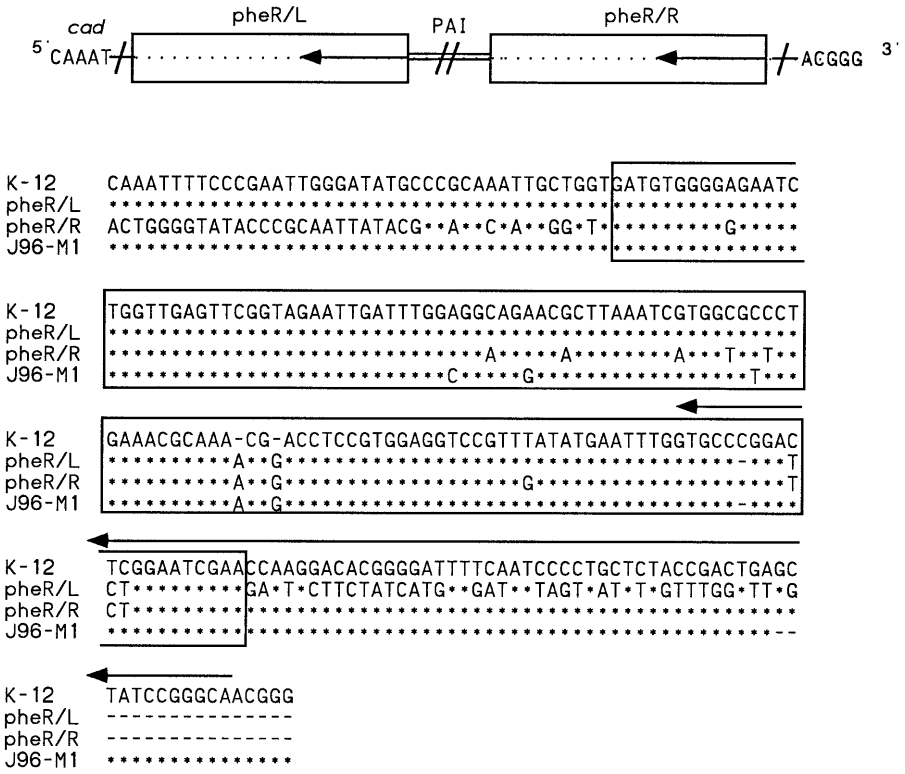

FIG. 3. Schematic diagram and nucleotide sequences of the junction sites of PAI V. In the schematic, the boxes represent the 135-bp direct repeat, with left and right junctions indicated (see below); 5′- and 3′-end nucleotides are shown for reference. The single diagonal lines represent DNA between these ends and the 135-bp direct repeat. The arrows represent *pheR* sequences, and the dotted line represents other repeated sequences. The double diagonal lines represent intervening PAI-specific DNA. The nucleotide sequences are as follows: *cad* and *pheR* region, derived from *E. coli* K-12; *pheR*/L and *pheR*/R, left and right junction sequences, respectively. J96-M1 is a deletion derivative of J96. Asterisks represent nucleotides identical to those in *E. coli* K-12, and dashes represent nucleotide deletions compared with *E. coli* K-12. *pheR* is indicated by horizontal arrows above the sequence in the direction of transcription. The 135-bp direct-repeat sequence is shown in boxes.

and that it is likely that uropathogenic *E. coli* PAIs undergo additional recombination processes that lead to additions and deletions of genes within the PAI.

The finding of sequences similar to those of the *Serratia marcescens* hemolysin structural and transport genes (50) is intriguing because neither of the parental cosmids carrying these sequences confers a hemolytic phenotype to *E. coli* K-12 (57). However, this particular hemolysin is part of a family of diverse cell surface and secreted bacterial proteins with members that act primarily as adhesins rather than cytolysins (5, 12, 47). Thus, at present the limited J96 sequence information available does not reveal the significance of these cosmids' nonhemolytic phenotype.

One of the important discoveries described here is that the chromosomal locations and junctions of the J96 PAIs are different from those described for the uropathogenic *E. coli* 536 PAIs and the enteropathogenic *E. coli* locus of enterocyte effacement. PAI V is flanked by a 135-bp imperfect direct repeat encompassing a region containing a tRNA gene, *pheR*. The deletion of PAI V from the chromosome leaves a single copy of the duplicated region at the deletion site, suggesting a homologous recombination process. One junction region of PAI IV was determined and is located near a tRNA gene, *pheV*. Although *E. coli* J96 PAI V and the PAIs of *E. coli* 536 (7) are located at tRNA loci, there are no obvious matches between possible PAI insertion target sequences. In *E. coli* 536, 16- and 18-bp direct repeats participate in excision of PAIs and a conserved 6-bp sequence is found at the junction regions (7, 29). The PAI V terminus-specific DNA sequences do not exhibit similarity to any sequences in the database; however, it is interesting that the PAI IV terminus-specific DNA matches closely the K1 polysialic capsular biosynthesis genes (*kps*). This is a curious discovery given that *E. coli* J96 expresses a K6-type capsule (23). It suggests that PAIs may be a potential source of genes used in recombination events that could lead to changes in expression of capsular antigens.

Although each of the uropathogenic *E. coli* PAIs studied to date share *pap*-like and *hly*-like determinants, differences in their chromosomal loci, sizes, and junction termini indicate diversity in PAI genetic biology and evolution. The site-specific insertion into tRNA genes of different genetic elements appears to be a general phenomenon among bacteria (11, 52). In consideration of the evolution of the *E. coli* PAIs, some have speculated that the involvement of tRNA genes in site-specific integration reflects shared sequence similarity of different genetic elements, such as temperate bacteriophage, to conserved sequences within the 3′ termini of tRNA genes (52). A virulence-associated region in *Dichelobacter nodosus* contains an integrase-like gene similar to that found in bacteriophages P4 and ϕR73 (11). Bacteriophage P4-like *int* and 9.7K protein gene sequences were identified in PAI V and PAI IV, respectively. It is interesting that our DNA hybridization results indicate that these two sequences are not present together in either of the PAI cosmid minisets. This suggests that complete P4-like genomes are probably not found in the individual J96 PAIs. The significance of this is unclear at present but again

suggests that the J96 PAIs have gone down separate evolutionary pathways. With regard to what type of genetic element may be responsible for horizontal transfer of PAIs, the identification of conjugative-plasmid-like sequences in the J96 PAIs supports the role of plasmids acting as possible vectors for horizontal transmission and evolution of PAIs.

The possible presence of at least three IS elements further suggests that the PAIs are the targets for diverse recombination processes. Interestingly, sequences similar to IS*100*, an insertion element identified in *Y. pestis* but not originally thought to occur in *E. coli* (51), were identified in PAI V. IS*100* has been implicated in plasmid-chromosome exchange and also in deletions in *Y. pestis* (15, 16). This IS-like element could provide a mechanism in *E. coli* J96 for PAI deletions or acquisition and integration of plasmid DNA as is found in *Y. pestis.*

The expression of PAI-associated genes appears to be critical for growth of uropathogenic *E. coli* in murine models of ascending urinary tract infection (45, 53). Our evidence for the existence of pathogen-specific colonization factors and virulence genes that were previously unknown in uropathogenic *E. coli* is an important advance in the study of the complex pathway that leads to urinary tract disease by this organism. The potential for transfer and acquisition of PAI-associated genes would be a distinct adaptive advantage for *E. coli* strains that grow at extraintestinal sites. Genetic analysis of these potential virulence genes and mobile DNA elements is under way in our laboratories.

ACKNOWLEDGMENTS

This work was supported by funds from the University of Wisconsin School of Medicine Research Committee and a Romnes Fellowship to R.A.W. and NIH grants AI138166 and DK48029 to D.E.B.

We thank Shahaireen Pellett and Tammy Robbins for assistance with DNA sequencing. We gratefully acknowledge Staffan Normark for his encouragement and support during the early phases of this work.

REFERENCES

1. **Akopayants, N. S., D. Kersulyte, S. Clifton, B. Roe, and D. E. Berg.** Unpublished data.
2. **Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
3. **Bachmann, B. J.** 1990. Linkage map of *Escherichia coli* K-12, edition 8. Microbiol. Rev. **54:**130–197.
4. **Bahrani, F. K., G. Massad, C. V. Lockatell, D. E. Johnson, R. G. Russell, J. W. Warren, and H. L. T. Mobley.** 1994. Construction of an MR/P fimbrial mutant of *Proteus mirabilis*: role in virulence in a mouse model of ascending urinary tract infection. Infect. Immun. **62:**3363–3371.
5. **Barenkamp, S. J., and E. Leininger.** 1992. Cloning, expression, and DNA sequence analysis of genes encoding nontypeable *Haemophilus influenzae* high-molecular-weight surface-exposed proteins related to filamentous hemagglutinin of *Bordetella pertussis*. Infect. Immun. **60:**1302–1313.
6. **Blum, G., V. Falbo, A. Caprioloi, and J. Hacker.** 1995. Gene clusters encoding the cytotoxic necrotizing factor type 1, Prs-fimbriae and alpha-hemolysin form the pathogenicity island II of uropathogenic *Escherichia coli* strain J96. FEMS Microbiol. Lett. **126:**189–196.
7. **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.
8. **Bukanov, N. O., and D. E. Berg.** 1994. Ordered cosmid library and high-resolution physical-genetic map of *Helicobacter pylori* strain NCTC11638. Mol. Microbiol. **11:**509–523.
9. **Caprioli, A., V. Falbo, L. G. Roda, F. M. Ruggeri, and C. Zona.** 1983. Partial purification and characterization of an *Escherichia coli* toxic factor that induces morphological cell alterations. Infect. Immun. **39:**1300–1306.
10. **Caprioli, A., V. Falbo, F. M. Ruggeri, L. Baldassarri, R. Bisicchia, G. Ippolito, E. Romoli, and G. Donelli.** 1987. Cytotoxic necrotizing factor production by hemolytic strains of *Escherichia coli* causing extraintestinal infections. J. Clin. Microbiol. **25:**146–149.
11. **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.
12. **Delisse-Gathoye, A.-M., C. Locht, F. Jacob, M. Raaschou-Nielsen, I. Heron, J.-L. Ruelle, M. De Wilde, and T. Cabezon.** 1990. Cloning, partial sequence, expression, and antigenic analysis of filamentous hemagglutinin gene of *Bordetella pertussis*. Infect. Immun. **58:**2895–2905.
13. **Falbo, V., M. Famiglietti, and A. Caprioli.** 1992. Gene block encoding production of cytotoxic necrotizing factor 1 and hemolysin in *Escherichia coli* isolates from extraintestinal infections. Infect. Immun. **60:**2182–2187.
14. **Falbo, V., T. Pace, L. Picci, E. Pizzi, and A. Caprioli.** 1993. Isolation and nucleotide sequence of the gene encoding cytotoxic necrotizing factor 1 of *Escherichia coli*. Infect. Immun. **61:**4909–4914.
15. **Fetherston, J. D., and R. D. Perry.** 1994. The pigmentation locus of *Yersinia pestis* KIM6+ is flanked by an insertion sequence and includes the structural genes for pesticin sensitivity and HMWP2. Mol. Microbiol. **13:**697–708.
16. **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.
17. **Fives-Taylor, P. M., F. L. Macrina, T. J. Pritchard, and S. S. Peene.** 1987. Expression of *Streptococcus sanguis* antigens in *Escherichia coli*: cloning of a structural gene for adhesion fimbriae. Infect. Immun. **55:**123–128.
18. **Fonstein, M., and R. Haselkorn.** 1993. Chromosomal structure of *Rhodobacter capsulatus* strain SB1003: cosmid encyclopedia and high-resolution physical and genetic map. Proc. Natl. Acad. Sci. USA **90:**2522–2526.
19. **Ganeshkumar, N., P. M. Hannam, P. E. Kolenbrander, and B. C. McBride.** 1991. Nucleotide sequence of a gene coding for a saliva-binding protein (SsaB) from *Streptococcus sanguis* 12 and possible role of the protein in coaggregation with actinomyces. Infect. Immun. **59:**1093–1099.
20. **Gibson, T. J., A. Rosenthal, and R. H. Waterston.** 1987. Lorist6, a cosmid vector with *Bam*HI, *Not*I, *Sca*I and *Hin*dIII cloning sites and altered neomycin phosphotransferase gene expression. Gene **53:**283–286.
21. **Hacker, J.** 1990. Genetic determinants coding for fimbriae and adhesins of extraintestinal *Escherichia coli*. Curr. Top. Microbiol. Immunol. **151:**1–27.
22. **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.
23. **Hagberg, L., R. Hull, S. Hull, S. Falkow, R. Freter, and C. S. Edén.** 1983. Contribution of adhesion to bacterial persistence in the mouse urinary tract. Infect. Immun. **40:**265–272.
24. **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.
25. **Hull, R., S. Bieler, S. Falkow, and S. Hull.** 1986. Chromosomal map position of genes encoding P adhesins in uropathogenic *Escherichia coli*. Infect. Immun. **51:**693–695.
26. **Hull, S., S. Bieler, and R. Hull.** 1988. Restriction fragment length polymorphism and multiple copies of DNA sequences homologous with probes for P-fimbriae and hemolysin genes among uropathogenic *Escherichia coli*. Can. J. Microbiol. **34:**307–311.
27. **Hultgren, S. J.** Personal communication.
28. **Hultgren, S. J., S. Abraham, M. Caparon, P. Falk, J. W. St. Geme III, and S. Normark.** 1993. Pilus and nonpilus bacterial adhesins: assembly and function in cell recognition. Cell **73:**887–901.
29. **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.
30. **Kohara, Y., K. Akiyama, and K. Isono.** 1987. The physical map of the whole *E. coli* chromosome: application of a new strategy for rapid analysis and sorting of a large genomic library. Cell **50:**495–508.
31. **Krishnan, B. R., I. Jamry, D. E. Berg, C. M. Berg, and D. D. Chaplin.** 1995. Construction of a genomic DNA 'feature map' by sequencing from nested deletions: application to the HLA class i region. Nucleic Acids Res. **23:**117–122.
32. **Krogfelt, K. A.** 1991. Bacterial adhesion: genetics, biogenesis, and role in pathogenesis of fimbrial adhesins of *Escherichia coli*. Rev. Infect. Dis. **13:** 721–735.
33. **Lee, C. A.** 1996. Pathogenicity islands and the evolution of bacterial pathogens. Infect. Agents Dis. **5:**1–7.
34. **Lindberg, F. P., B. Lund, and S. Normark.** 1984. Genes of pyelonephritogenic *E. coli* required for digalactoside-specific agglutination of human cells. EMBO J. **3:**1167–1173.
35. **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.
36. **Lund, B., B. I. Marklund, N. Stromberg, F. Lindberg, K. A. Karlsson, and S. Normark.** 1988. Uropathogenic *Escherichia coli* can express serologically identical pili of different receptor binding specificities. Mol. Microbiol. **2:**255–263.
37. **Lutwyche, P., R. Rupps, J. Cavanagh, R. A. J. Warren, and D. E. Brooks.** 1994. Cloning, sequencing, and viscometric adhesion analysis of heat-resistant agglutinin 1, an integral membrane hemagglutinin from *Escherichia coli*

O9:H10:K99. Infect. Immun. **62:**5020–5026.
38. **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.
39. **Mills, D. M., V. Bajaj, and C. A. Lee.** 1995. A 40kb 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.
40. **Minshew, B. H., J. Jorgensen, G. W. Counts, and S. Falkow.** 1978. Association of hemolysin production, hemagglutination of human erythrocytes, and virulence for chicken embryos of extraintestinal *Escherichia coli* isolates. Infect. Immun. **20:**50–54.
41. **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)βGal binding in virulence of a wild-type strain. Mol. Microbiol. **10:**143–155.
42. **Muhldorfer, I., and J. Hacker.** 1994. Genetic aspects of *Escherichia coli* virulence. Microb. Pathog. **16:**171–181.
43. **Normark, S.** Personal communication.
44. **Ochman, H., and E. A. Groisman.** 1995. The evolution of invasion by enteric bacteria. Can. J. Microbiol. **41:**555–561.
45. **O'Hanley, P., G. Lalonde, and G. Ji.** 1991. Alpha-hemolysin contributes to the pathogenicity of piliated digalactoside-binding *Escherichia coli* in the kidney: efficacy of an alpha-hemolysin vaccine in preventing renal injury in the BALB/c mouse model of pyelonephritis. Infect. Immun. **59:**1153–1161.
46. **Orskov, F., and I. Orskov.** 1992. *Escherichia coli* serotyping and disease in man and animals. Can. J. Microbiol. **38:**699–704.
47. **Palmer, K. L., and R. S. Munson.** 1995. Cloning and characterization of the genes encoding the haemolysin of *Haemophilus ducreyi*. Mol. Microbiol. **18:**821–830.
48. **Pitcher, D. G., N. A. Saunders, and R. J. Owen.** 1989. Rapid extraction of bacterial genomic DNA with guanidium thiocyanate. Lett. Appl. Microbiol. **8:**151–156.
49. **Plunkett, G., and F. Blattner.** Unpublished data.
50. **Poole, K., E. Schiebel, and V. Braun.** 1988. Molecular characterization of the hemolysin determinant of *Serratia marcescens*. J. Bacteriol. **170:**3177–3188.
51. **Portnoy, D. A., and S. Falkow.** 1981. Virulence-associated plasmids from *Yersinia enterocolitica* and *Yersinia pestis*. J. Bacteriol. **148:**877–883.
52. **Reiter, W.-D., P. Palm, and S. Yeats.** 1989. Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. Nucleic Acids Res. **17:**1907–1914.
53. **Ritter, A., G. Blum, L. Emody, M. Kerenyi, A. Bock, B. Heuhierl, 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.
54. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
55. **Sampson, J. S., S. P. O'Connor, A. R. Stinson, J. A. Tharpe, and H. Russell.** 1994. Cloning and nucleotide sequence analysis of *psaA*, the *Streptococcus pneumoniae* gene encoding a 37-kilodalton protein homologous to previously reported *Streptococcus* sp. adhesins. Infect. Immun. **62:**319–324.
56. **Silver, R. P., L. F. Wright, and Q. Zhao.** Unpublished data.
57. **Swenson, D. L., and R. A. Welch.** Unpublished data.
58. **Welch, R. A.** 1991. Pore-forming cytolysins of Gram-negative bacteria. Mol. Microbiol. **5:**521–528.
59. **Welch, R. A., E. P. Dellinger, B. Minshew, and S. Falkow.** 1981. Hemolysin contributes to virulence of extra-intestinal *Escherichia coli* infections. Nature (London) **294:**665–667.
60. **Welch, R. A., and S. Falkow.** 1984. Characterization of *Escherichia coli* hemolysins conferring quantitative differences in virulence. Infect. Immun. **43:**156–160.
61. **Welch, R. A., R. Hull, and S. Falkow.** 1983. Molecular cloning and physical characterization of a chromosomal hemolysin from *Escherichia coli*. Infect. Immun. **42:**178–186.
62. **Welch, R. A., and S. Pellett.** 1988. Transcriptional organization of the *Escherichia coli* hemolysin genes. J. Bacteriol. **170:**1622–1630.

Editor: A. O'Brien