INFECTION AND IMMUNITY, Feb. 1989, p. 314–321
0019-9567/89/020314-08$02.00/0

Vol. 57, No. 2

Structure and Copy Number of Gene Clusters Related to the *pap* P-Adhesin Operon of Uropathogenic *Escherichia coli*

MICHEL ARTHUR,[1] CRAIG CAMPANELLI,[1] ROBERT D. ARBEIT,[2] CHEUNG KIM,[1] SUZANNE STEINBACH,[2] CANDICE E. JOHNSON,[3] ROBERT H. RUBIN,[4] AND RICHARD GOLDSTEIN[1]*

Section of Molecular Genetics and Epidemiology, Maxwell Finland Laboratory for Infectious Diseases, Boston University Schools of Medicine and Public Health, Boston, Massachusetts 02118[1]; Departments of Medicine and Pediatrics, Boston University School of Medicine, Boston, Massachusetts 02118[2]; Department of Pediatrics, Cleveland Metropolitan General Hospital and Case Western Reserve University School of Medicine, Cleveland Ohio 44109[3]; and Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114[4]

Received 13 July 1988/Accepted 17 October 1988

The structurally related *pap* and *prs* operons of the uropathogenic *Escherichia coli* isolate J96 encode a P and an F adhesin that mediate bacterial attachment to the human P blood group antigen and the Forssman antigen, respectively. Using probes prepared from different segments of the *pap* operon, Southern blot hybridizations were performed to characterize *pap*-related sequences of 30 *E. coli* clinical isolates expressing different adhesin phenotypes. Gene clusters encoding P and F adhesins displayed no restriction site polymorphism in sequences homologous to the *papH*, *papC*, and *papD* genes that encode proteins essential to the transport and polymerization of the subunits of the P-pilus adhesin. In contrast, *pap*-related genetic elements associated with a null phenotype either lacked homology to the *papH*, *papC*, and *papD* genes or displayed a restriction site polymorphism in this region. Sequences within and surrounding the J96 *papG* and *prsG* adhesin genes that determine the binding specificities to the P and F antigens, respectively, were not conserved. However, gene clusters encoding different binding specificities could not be distinguished based on such restriction site polymorphisms. The majority of clinical isolates had more than one copy of *pap*-related sequences that involved gene clusters similar to the J96 *pap* operon, as well as genetic elements that were related only to a part of this operon. The implications of this unexpected copy number polymorphism with respect to possible recombination events involving *pap*-related sequences are discussed.

Escherichia coli can express a wide variety of adhesins with different antigenic and binding specificities. Adhesins with P-adhesin-binding specificity mediate bacterial attachment to the P blood group antigen, a globoside that is present on human erythrocytes and uroepithelial cells (9, 12, 14). The ability of particular clones of *E. coli* to express a P adhesin, and thereby to adhere to the uroepithelium, is considered critical to the pathogenesis of ascending urinary tract infections (22).

The pap (pili associated with pyelonephritis) operon encoding a P fimbrial adhesin has been cloned from chromosomal DNA of the pyelonephritic *E. coli* isolate J96 (8, 14) (Fig. 1). The P pili encoded by the *pap* operon are heteropolymers consisting of about 1,000 major pilin subunits (PapA [4]) and a few copies of minor pilin subunits located either at the base (papH [3]) or at the tip of the pilus (PapE, PapF, and PapG [13, 14]). Among the latter, PapG appears to be the actual adhesin which interacts with the receptor (14). The PapC and PapD proteins are required for the polymerization and transport of the pilus subunit, respectively (19). PapB and PapI proteins are implicated in the transcriptional activation of the operon (2).

Strain J96 has an additional genetic element that is structurally related to the *pap* operon, but it is in a different chromosomal location (7, 14). This element, designated *prs* (pap-related sequence), is closely related to the *pap* operon but encodes a distinct adhesin, which is referred to as the F adhesin because it mediates preferential binding to the Forssman antigen, which is present on sheep erythrocytes and the human renal pelvis (14, 16). Extensive restriction mapping identified structural differences between the *pap* and *prs* gene clusters only in the *papG* region. Transcomplementation analysis demonstrated that a single protein, the PapG or the PrsG adhesin, determines the binding specificity to the human P blood group or the Forssman antigen, respectively. This was the only functional difference that was detected between the two gene clusters.

In the accompanying epidemiologic survey (1), colony hybridizations were performed to characterize the distribution of nucleotide sequences related to the *pap* operon among 137 *E. coli* clinical isolates. Nucleotide sequences related to the *pap* operon were found more frequently among isolates from patients with pyelonephritis than among isolates from patients with cystitis or isolates from fecal specimens. In addition, the adhesin phenotypes associated with *pap*-related sequences differed among the three groups of isolates. Most pyelonephritis *pap* probe-positive isolates coexpressed a P and an F adhesin, while expression of only an F adhesin was the most frequent phenotype among *pap* probe-positive fecal isolates. These results suggest that *pap*-related adhesins with distinct binding specificities, either alone or in combination, may affect differently the ability of *E. coli* to cause ascending urinary tract infections. Since such differences could not be resolved by colony hybridization, in this report we describe the characterization of the structure of *pap*-related sequences by Southern blot hybridization. Based on this analysis we established correlations between the adhesin phenotype and the variability in defined regions of individual *pap*-related gene clusters, as well as the copy number of *pap*-related sequences.

* Corresponding author.

MATERIALS AND METHODS

Bacterial strains. *E. coli* J96 and 30 of the 137 clinical isolates described in the accompanying report (1) were further characterized by Southern blot hybridization (21). These 30 isolates were characterized both for the expression of P and F adhesins and for the presence of nucleotide sequences related to the *pap* operon (Table 1).

DNA purification. *E. coli* DNA was obtained by bacterial lysis in the presence of sodium lauryl sarcosine (8) and was purified by cesium chloride-ethidium bromide density gradient centrifugation (17).

Southern blot analysis. DNA restriction fragments were separated by electrophoresis in 0.8% agarose horizontal slab gels (25 by 16 by 0.7 cm). Transfer to Zeta-probe membranes was done as recommended by the manufacturer (Bio-Rad Laboratories, Richmond, Calif.). Construction of the probes (Fig. 1) is described in the accompanying report (1). Hybridizations were performed at 42°C in 50% formamide (17). The restriction fragments used to generate the probes are indicated in Fig. 1. Between successive hybridizations, the probes that bound to the membranes were removed by alkaline treatment, as recommended by the manufacturer. Three different types of analyses were carried out. (i) For strain J96 plus the 30 clinical isolates, *Pst*I restriction fragments were resolved by gel electrophoresis to determine comigration patterns of different strains. Hybridization was performed sequentially with the *pap-11* and *papEFG* probes. The results obtained for strain J96 and 16 of the 30 clinical isolates are displayed in Fig. 2.

(ii) For strain J96 plus the 30 clinical isolates, three restriction digests (*Pst*I, *Eco*RI, and *Pst*I-*Eco*RI) were electrophoresed in adjacent lanes, to determine both the number of *Eco*RI fragments carrying *pap*-related gene clusters and their respective conserved *Pst*I sites. Hybridization was performed with the *pap-11*, *papL*, *papHC*, *papCD*, and *papEFG* probes. The results obtained for 3 of the 30 isolates are displayed in Fig. 3.

(iii) For strain J96 plus 7 of the 30 clinical isolates, *Pst*I and *Pst*I-*Bgl*II restriction fragments were resolved by gel electrophoresis, to detect the presence or absence of a unique *Bgl*II site found in the *papG* gene of strain J96. Hybridization was performed with the *papEFG* probe (Fig. 4).

RESULTS

As described in detail in the accompanying report (1), four hybridization classes were found among 137 *E. coli* clinical isolates based on colony hybridization by using different segments of the *pap* operon as probes. Among class I isolates, positive hybridization was detected with all four *pap* probes (*papL*, *papHC*, *papCD*, and *papEFG*), which together encompass almost the entire *pap* operon (Fig. 1). Among the bacteria of class I, four phenotypes were detected based on agglutination assays for Pap-related adhesins: (i) expression of only a P adhesin, (ii) expression of only an F adhesin, (iii) expression of both adhesins, and (iv) expression of neither adhesin. In the accompanying report (1) we provide a detailed discussion of the adhesin specificities that could account for these phenotypes, as defined by agglutination profiles of clinical isolates, by using erythrocytes and latex beads coated with purified receptors. No expression of Pap-related adhesins was detected in any of the clinical isolates belonging to the three other hybridization classes defined by a positive colony hybridization signal with the *papL* and *papEFG* probes (class II), the *papL* probe only (class III), or none of the probes (class IV). Here we describe the use of Southern blot hybridization to characterize further a subset of 30 of these clinical isolates (Table 1). This subset included representatives of all four colony hybridization classes and of all four Pap-related adhesin phenotypes.

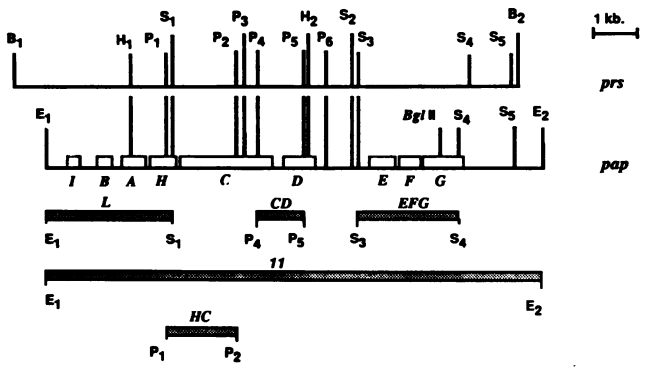

FIG. 1. Probes used to detect nucleotide sequences related to different parts of the *pap* operon. Detailed maps of the *pap* (P adhesin) and *prs* (F adhesin) operons have been presented previously (14, 16). Restriction endonuclease cutting sites are denoted by vertical lines, genes are denoted by open horizontal boxes, and the DNA fragments used to generate probes are denoted by horizontal shaded bars, with the restriction endonuclease cutting site noted at each end. The *pap* probes are labeled according to the positions of the *pap* genes, with the exceptions of the *papL* probe (left end) and the *pap-11* probe (11-kb *Eco*RI fragment spanning the entire *pap* operon). Restriction endonuclease abbreviations: B, *Bam*HI; E, *Eco*RI; H, *Hin*dIII; P, *Pst*I; S, *Sma*I. The subscript numbers following restriction enzyme designations refer to consecutive cutting sites.

No discrepancies were detected between the results derived from colony versus Southern blot hybridizations. This was the case for hybridizations with the *pap* probes described in this report, in addition to the *hlyA*, *pil*, and *afaI* probes described in the accompanying epidemiologic survey (1) (unpublished data). This analysis indicates that colony hybridization is a sensitive and specific technique for the detection of nucleotide sequences related to virulence operons among *E. coli* clinical isolates.

Restriction site polymorphism in *Pst*I sites internal to the *pap* operon. Because there are six *Pst*I restriction endonuclease sites at homologous positions within the *pap* and *prs* operons of *E. coli* J96 (Fig. 1), we chose this enzyme to evaluate the restriction polymorphism of DNA sequences within *pap*-related gene clusters. Digestion of the chromosomal DNA of *E. coli* J96 with *Pst*I should yield five

TABLE 1. Properties of the 30 clinical isolates examined in this study

Hybridization class	Positive hybridization with the following probes:	No. expressing the following adhesins:			
		P only	F only	P and F	Neither P nor F
I[a]	*papL*, *papHC*, *papCD*, *papEFG*	3	5	12	2
II	*papL*, *papEFG*	0	0	0	3
III	*papL*	0	0	0	3
IV	None	0	0	0	2

[a] Strain J96 belongs to hybridization class I and expresses the P and F adhesins.

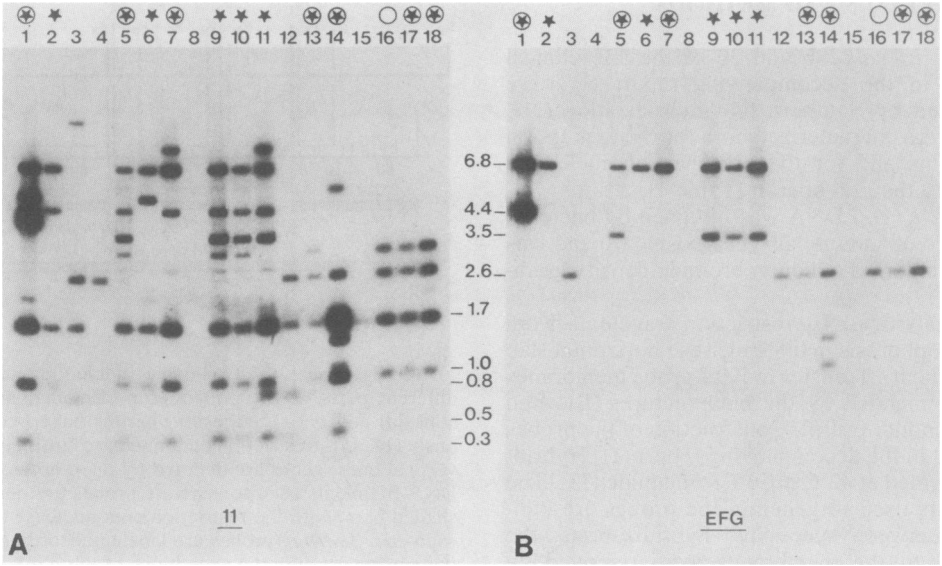

FIG. 2. Restriction polymorphism in *pap*-related genetic elements. The DNAs of 18 strains were digested with *Pst*I, electrophoresed in a 0.8% agarose gel, and transferred to a membrane. (A) Hybridization with the *pap-11* probe; (B) hybridization of the same membrane with the *papEFG* probe. The adhesin phenotype of each strain is symbolized above each lane: circle, P adhesin; star, F adhesin, circled star, P and F adhesins. Lane 1, *E. coli* J96. Strains expressing neither adhesin included two class I isolates detected by the four internal *pap* probes (lanes 3 and 12) and one class III isolate detected only by the *papL* probes (lane 4). Lane 15, Insufficient amounts of DNA based on ethidium bromide staining of DNA. Between the two autoradiograms the sizes of relevant restriction fragments are given in kilobases.

fragments internal to the *pap* operon plus five fragments internal to the *prs* operon. Homologous fragments from the two operons would be expected to comigrate. Among these fragments, the 1.7-kilobase (kb) (delineated by the *Pst*I sites 1 and 2) and 1.0-kb (*Pst*I sites 3 and 4; Fig. 1) fragments were well resolved by the gel system used in this study (see J96 restriction profile, Fig. 2, lane 1). Because the 1.7- and 1.0-kb fragments were not contiguous, their presence indicates the conservation of four *Pst*I sites. Among the remaining three predicted *Pst*I fragments of 0.5, 0.3, and 0.1 kb, the first two approached the lower limits of resolution and sensitivity of the procedure, while the last one was not detectable.

For strain J96 and each of the 20 class I clinical isolates that phenotypically expressed either or both a P and an F adhesin, the *pap-11* probe which spanned the entire *pap* operon detected both 1.7- and 1.0-kb *Pst*I fragments (Fig. 2, lanes 1, 2, 5, 6, 7, 9, 10, 11, 13, 14, 16, 17, and 18). To confirm that these DNA fragments of the 20 clinical isolates carried related sequences, we purified and labeled the 1.7-kb (*papHC* probe) and 1.0-kb (*papCD* probe) *Pst*I fragments of the J96 *pap* operon. In all 20 isolates the *papHC* and *papCD* probes detected a 1.7- and a 1.0-kb *Pst*I band, respectively (Fig. 3, lanes 3, 6, and 9). In two of the isolates, the *papCD* probe also resolved a 3.2-kb fragment (Fig. 3, lane 6). Our results also indicate that, in accord with the restriction maps of the J96 *pap* and *prs* operons (Fig. 1), none of the conserved 1.0- and 1.7-kb fragments found in these 20 isolates were sensitive to *Eco*RI (Fig. 3, lanes 2, 5, and 8).

The two remaining class I isolates were detected by all four *pap* probes but were phenotypically negative for P- and F-adhesin expression (Table 1). The Southern blots of *Pst*I restriction digests for these two isolates lacked the 1.0-kb fragments discussed above, but they displayed a new 0.8-kb fragment which hybridized with the *pap-11* probe (Fig. 2, lanes 3 and 12). The *papCD* probe hybridized only to this new fragment, while the *papHC* probe detected the conserved 1.7-kb band (data not shown).

Thus, the four *Pst*I sites delineating the 1.0- and 1.7-kb *Pst*I fragments of the J96 *pap* and *prs* operons were conserved in all 20 class I isolates that expressed Pap-related adhesins. In contrast, a *Pst*I restriction site polymorphism was identified in both of the class I isolates that did not express Pap-related adhesins.

Copy number polymorphism of *pap*-related gene clusters. The structurally related *pap* and *prs* operons of *E. coli* J96 are carried by two distinct *Eco*RI fragments of 15 and 19 kb, respectively (7). As expected, these two *Eco*RI fragments were detected by the *papL*, *papHC*, *papCD*, and *papEFG* probes used in this study (data not shown). In the *Pst*I digest and the *Eco*RI-*Pst*I double digests, the *papHC* and *papCD* probes detected only the 1.7- and 1.0-kb *Pst*I bands described above because of the conservation of *Pst*I restriction sites in the J96 *pap* and *prs* operons. If, as in J96, *pap*-related gene clusters of clinical isolates do not contain an internal *Eco*RI site and different clusters are located on different-sized *Eco*RI fragments, then the number of *Eco*RI restriction fragments detected by all of the four *pap* probes should reveal the number of *pap*-related gene clusters. Because the *pilC* probe described in the accompanying paper (1) was always found to hybridize with a single band in the Southern blot restriction profiles, it proved possible to rule out partial digestion as an explanation for the detection of multiple bands by the *pap* probes (unpublished data). Based on this process of deduction, four different copy number conditions were determined to exist. (i) There are multiple *pap*-related gene clusters carried by distinct *Eco*RI fragments. Of the 12 class I clinical isolates expressing both a P and an F adhesin (Table 1), 8 had restriction profiles similar to that described in detail for strain J96 (Fig. 3, lanes 1 to 3). In each strain two *Eco*RI fragments ranging from 12 to 32 kb were detected by all four *pap* probes. In the *Pst*I and *Eco*RI-*Pst*I restriction

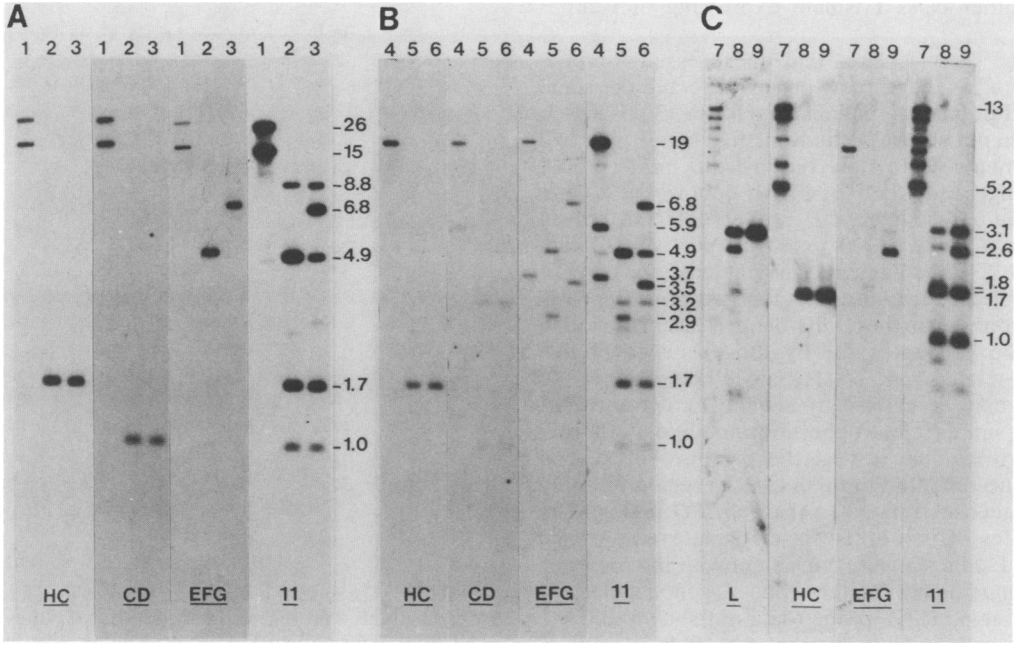

FIG. 3. Estimation of the number of *pap*-related operons in three different clinical isolates. (A and C) Southern blots of two isolates expressing both P and F adhesins. (B) Southern blot of an isolate expressing only an F adhesin. The chromosomal DNAs of three strains were digested with *Eco*RI (lanes 1, 4, and 7), *Pst*I-*Eco*RI (lanes 2, 5, and 8), or *Pst*I (lanes 3, 6, and 9). Successive hybridizations were performed with the probes noted at the bottom of each lane of the autoradiograms.

digests, the *papHC* and *papCD* probes specifically detected only the 1.7- and 1.0-kb fragments, respectively, indicating the conservation of internal *Pst*I sites in the two gene clusters. The most likely explanation for these restriction profiles is the presence of two gene clusters related to the entire *pap* operon. By the same criteria, one additional class I isolate coexpressing P and F adhesins had three *pap*-related gene clusters.

(ii) There is a single *pap*-related gene cluster carried by a single *Eco*RI fragment. In eight class I clinical isolates the four *pap* probes detected only a single *Eco*RI fragment. The *papL*, *papHC*, *papCD*, and *papEFG* probes detected single but distinct fragments in the *Pst*I digest and in the *Eco*RI-*Pst*I double digests (data not shown). A likely explanation for such results is that these isolates have a single gene cluster related to the entire *pap* operon. These eight class I isolates included three of the five isolates expressing only an F adhesin, the three isolates expressing only a P adhesin, and the two adhesin-negative isolates (Table 1).

The restriction profile of 17 class I isolates described above in (i) and (ii) can be accounted for by the presence of complete *pap*-related gene clusters carried by distinct *Eco*RI fragments. For these 17 isolates coexpression of P- and F-adhesin binding specificities was associated with the presence of two or three gene clusters. In contrast a single gene cluster was found to be associated with expression of either binding specificities or a null phenotype.

(iii) There are multiple copies of *pap*-related sequences that may involve genetic elements related to only a part of the *pap* operon. The remaining 5 of the 22 class I isolates (Table 1) had more complex restriction profiles with the following features. First, the *pap-11* probe detected multiple *Eco*RI restriction fragments, but these fragments were not all detected by the complete set of probes, i.e., *papL*, *papHC*, *papCD*, and *papEFG*. Such findings suggest either the existence of an *Eco*RI site(s) within (or between) *pap*-related sequences or the presence of genetic elements related to only a part of the *pap* operon. Second, the copy number of *pap*-related gene clusters was greater than the number of *pap*-related adhesins that were expressed. To illustrate these features, we will describe in detail the restriction profiles of one of the two isolates expressing only an F adhesin and one of the three isolates expressing both P and F adhesins.

One isolate expressing only an F adhesin had a 19-kb *Eco*RI restriction fragment detected by all the *pap* probes (Fig. 3, lane 4), indicating the presence of a gene cluster related to the entire *pap* operon. Nucleotide sequences homologous to the *papEFG* probe were carried on a 6.8-kb *Pst*I fragment and a 4.9-kb *Pst*I-*Eco*RI fragment (Fig. 3, lanes 5 and 6). The J96 *prs* operon is carried by an *Eco*RI fragment with a similar size (7), and as determined below the *prsEFG* genes are also carried by a 6.8-kb *Pst*I and a 4.9-kb *Pst*I-*Eco*RI fragment. These observations, in combination with resolution in this isolate of the conserved 1.7- and 1.0-kb *Pst*I fragments, suggest that the F adhesin of this clinical isolate is encoded by a gene cluster similar to the *E. coli* J96 *prs* operon. In addition, for this isolate the *papEFG* probe detected a second *pap*-related genetic element carried by restriction fragments of 3.7 kb (*Eco*RI), 3.5 kb (*Pst*I), and 2.9 kb (*Pst*I-*Eco*RI). Likewise, the *papCD* probe detected additional fragments of 3.7 kb in the *Eco*RI digest and 3.2 kb both in the *Pst*I digest and the double digests. The fact that the *papCD* and *papEFG* probes detected distinct restriction fragments in each of the restriction profiles, in addition to the gene cluster carried by the 19-kb *Eco*RI fragment, indicates the presence of two copies of *pap*-related sequences. Because the *papHC* probe detected single bands in the three restriction digests, these sequences are likely to include a genetic element that is related to only a part of the *pap* operon. This complex restriction profile was likewise de-

tected in one other class I isolate expressing only an F adhesin.

For one of the three isolates that coexpressed P and F adhesins, the *papCD* and *papHC* probes detected the same set of five *Eco*RI fragments ranging from 5.2 to 13 kb (Fig. 3, lanes 7 to 9; data not shown). The same five bands plus two additional fragments were detected by the *papL* probe. When the *Pst*I digest and the *Pst*I-*Eco*RI double digest were analyzed, the *papHC* and *papCD* probes detected only a single 1.7-kb or a single 1.0-kb band, respectively. Thus, each of the large *Eco*RI fragments carry at least one copy of the conserved *Pst*I fragments. In contrast, the *papEFG* probe detected only a single *Eco*RI band. These restriction profiles could be accounted for by the existence of five copies of a *pap*-related gene cluster, since homology to the *papEFG* probe may be carried by several *Eco*RI and *Pst*I fragments with similar electrophoretic mobilities. Alternatively, these profiles may reveal the presence of several copies of sequences related to the *papHCD* region but only one copy of sequences related to the *papEFG* region. The restriction profiles of two other class I isolates expressing both a P and an F adhesin were consistent with the presence of two *pap*-related gene clusters plus a genetic element related only to the *papEFG* probe (data not shown).

(iv) There are genetic elements that are related to only a part of the *pap* operon. The three class II isolates did not express either a P or an F adhesin and were detected only by the *papL* and *papEFG* probes (Table 1). Since the *papEFG* probe is internal to the *pap* operon, these results demonstrate the occurrence of a genetic element that is related to only a part of the *pap* operon. For each of these three isolates, a homology with the *papEFG* and *papL* probes was found in single bands in the *Pst*I, *Eco*RI, and *Pst*I-*Eco*RI digests. There was no restriction fragment which was detected by both probes, suggesting that sequences homologous to the two probes were not clustered (data not shown).

The three class III isolates did not express either adhesin and were detected only by the *papL* probe (Table 1). In each of these three isolates the *papL* probe detected a single band in the *Pst*I, *Eco*RI, and *Pst*I-*Eco*RI digests. The *papL* probe contained sequences encoding part of the PapA major pilin subunit, the regulatory region of the *pap* operon, and chromosomal sequences of *E. coli* J96 surrounding the operon (Fig. 1). The positive signal observed with *papL* may therefore reflect a partial homology to any of these regions.

Correlation between the restriction polymorphism in sequences detected by the *papEFG* probe and the expression of a P versus an F adhesin. In the clinical isolate *E. coli* J96, the *papG* (P adhesin) gene but not the *prsG* (F adhesin) gene contains a *Bgl*II site (Fig. 1). In agreement with the published maps of the *pap* and *prs* operons, only one of the two *Pst*I fragments detected by the *papEFG* probe in J96 DNA was cleaved by *Bgl*II (Fig. 4, lanes 4a and 4b). The *papG* gene was therefore carried by a 4.4-kb *Pst*I fragment that gave a 2.6-kb *Pst*I-*Bgl*II fragment in the double digest. This junction fragment which contained J96 sequences external to the *pap* operon did not contain an *Eco*RI site (group A in Table 2). In contrast, the *prsG* gene was carried by a 6.8-kb *Pst*I junction fragment which did not contain a *Bgl*II site and gave a 4.9-kb *Eco*RI-*Pst*I junction fragment.

To determine whether the *Bgl*II site present in the *papG* gene was conserved in gene clusters encoding a P adhesin, we compared the *Pst*I and the *Pst*I-*Bgl*II restriction profiles of seven clinical isolates and J96 (Fig. 4). To determine whether sequences surrounding the *pap* and *prs* operons were conserved in gene clusters associated with different

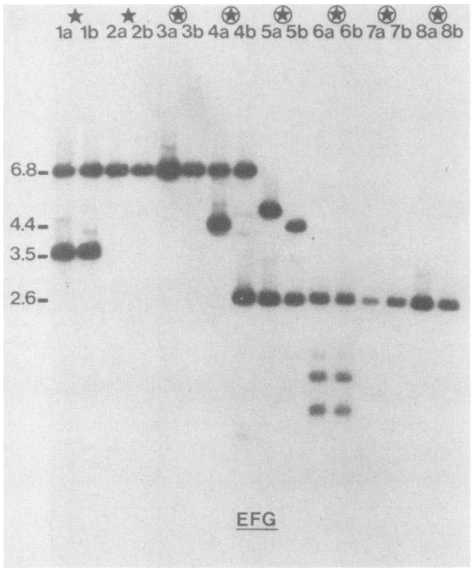

FIG. 4. Absence of a *Bgl*II site in adhesin-related genes. DNAs of eight strains were digested either with *Pst*I (lanes 1a to 8a) or with *Pst*I-*Bgl*II (lanes 1b to 8b). Hybridization was with the *papEFG* probe. Lanes 4a and 4b, *E. coli* J96. The adhesion phenotype of each isolate is symbolized above each set of lanes: star, F adhesin; circled star, P and F adhesins.

adhesin phenotypes, we compared the *Pst*I and *Eco*RI-*Pst*I restriction profiles of J96 and the 22 class I isolates (Table 2).

In one isolate expressing both P and F adhesins, the *papEFG* probe detected a single 6.8-kb band in the *Pst*I and *Pst*I-*Bgl*II digests (Fig. 4, lanes 3a and 3b) and a single 4.9-kb band in the *Eco*RI-*Pst*I double digest (group B in Table 2). In this strain, as in strain J96, a P and an F adhesin may be encoded by distinct gene clusters since two *Eco*RI fragments were detected by all the *pap* probes (see above; Fig. 3, lanes 1 to 3). In contrast to strain J96, the two gene clusters gave fragments with the same electrophoretic mobilities when the *Pst*I, *Pst*I-*Eco*RI, and *Pst*I-*Bgl*II digests were characterized with the *papEFG* probe. In each digest, the bands comigrated with homologous fragments of the *prs* operon (Fig. 4, lanes 3a, 3b, 4a, and 4b, and groups A and B in Table 2).

TABLE 2. Restriction fragments detected by the *papEFG* probe in strain J96 and in the 22 class I and the 3 class II clinical isolates

Group (class[a])	No. of isolates	No. of isolates expressing the following adhesins:				Size (kb) of fragment detected by the *papEFG* probe	
		P	F	P and F	Neither P nor F	*Pst*I	*Pst*I-*Eco*RI
A (I)	1 (J96)	0	0	1	0	4.4 (*pap*) 6.8 (*prs*)	4.4 (*pap*) 4.9 (*prs*)
B (I)	4	0	3	1	0	6.8	4.9
C (I)	2	0	2	0	0	6.8, 3.5	4.9, 2.9
D (I)	7	2	0	3	2	2.6	2.6
E (I)	5	1	0	4	0	2.6	1.8
F (I)	1	0	0	1	0	2.6, 1.6, 1.1	2.6, 1.6, 1.1
G (I)	1	0	0	1	0	2.6	2.6, 2.4
H (I)	1	0	0	1	0	2.6, 5.0	2.6, 5.0
I (I)	1	0	0	1	0	2.6, 2.3	2.6, 1.9
J (II)	3	0	0	0	3	2.3	1.9

[a] Hybridization classes as defined in Table 1.

These results indicate that the *Bgl*II site present in the *papG* gene is not conserved in the gene cluster encoding the P adhesin in this isolate. They also suggest that the J96 *prs* operon and the two gene clusters associated with the coexpression of P and F adhesins in this isolate share an extended homology beyond the adhesin gene. The apparent conservation of the *Eco*RI and *Pst*I-*Eco*RI *prs* junction fragments was also detected in gene clusters of five isolates expressing only an F adhesin that had a single *pap*-related gene cluster (Fig. 4, lanes 2a and 2b, and group B in Table 2) or multiple copies of *papEFG*-related sequences (Fig. 4, lanes 1a and 1b, and group C in Table 2).

In 16 other clinical isolates that expressed P, F, P and F, or neither P nor F adhesins, the *papEFG* probe detected 2.6-kb *Pst*I restriction fragments that comigrated with the 2.6-kb *Pst*I-*Bgl*II fragment of J96 carrying the *papG* gene (Fig. 4, lanes 5 to 8, and groups D to I in Table 2). Thus, in this group there exists an additional *Pst*I fragment that is not found in *pap* or *prs*. In strain J96 the 2.6-kb *Pst*I-*Bgl*II fragment is defined by a *Pst*I site located downstream of the *papD* gene and by the *Bgl*II site in the middle of the *papG* open reading frame. The additional *Pst*I site appears likely to be located within the *papG*-related adhesin gene of these clinical isolates based on (i) the conservation of *Pst*I sites found in the *papHCD* region (see above), and (ii) comigration of this new *Pst*I fragment of the isolates with the *Pst*I-*Bgl*II fragment of J96 (Fig. 4, lanes 4 to 8). The presence of this *Pst*I site did not correlate with the expression of a P versus an F adhesin (see, in particular, group D in Table 2).

DISCUSSION

Southern blot hybridization with a set of four *pap* probes revealed an unexpected number of restriction site polymorphisms. A total of 20 of the 22 class I isolates and 2 of the 3 class II isolates displayed different restriction profiles. These polymorphisms involved DNA sequences both within and surrounding *pap*-related gene clusters and a variable copy number of the actual *pap*-related gene clusters.

Conservation of internal *Pst*I-sites in the *papHCD* region of functional *pap*-related gene clusters. In *E. coli* J96 and all 20 class I clinical isolates expressing *pap*-related adhesins (Table 1), the *papHC* probe detected a single band of 1.7 kb, indicating the conservation of *Pst*I sites in single and multiple *pap*-related gene clusters (Fig. 3, lanes 3, 6, and 9). These bands had identical electrophoretic mobilities (Fig. 2). Also conserved were the two *Pst*I sites which delineate the 1.0-kb restriction fragment used to generate the *papCD* probe in the *pap* operon. This conclusion was based on the detection of 1.0-kb fragments with identical electrophoretic mobilities in all 20 class I isolates (Fig. 2 and 3). Thus, nucleotide sequences in the *papHCD* region of functional *pap*-related gene clusters appear to be highly conserved and to share the same organization as the J96 *pap* and *prs* operons. In contrast, the conserved 1.0-kb fragment was absent from the *Pst*I restriction profiles of the two class I isolates that did not express either a P or an F adhesin (Fig. 2, lanes 3 and 12). In these two isolates, the *papCD* probe detected a novel 0.8-kb *Pst*I fragment (data not shown). The presence of a small (ca. 0.2 kb) deletion or a point mutation creating a new *Pst*I site can account both for these restriction profiles and the absence of *pap*-related adhesin expression since mutations in the *papC* or the *papD* gene have been shown to abolish expression of the P adhesin encoded by the J96 *pap* operon (19).

Variability in the *papEFG* region of *pap*-related gene clusters. In contrast to the conservation of *pap*-related sequences in the *papHCD* region (see above), the *papEFG* probe detected restriction fragments with variable sizes (Table 2, and Fig. 2 and 4). Although some fragments with identical electrophoretic mobilities were often detected in several strains, 10 groups of isolates were defined based on distinct combinations of restriction fragments (Table 2). Thus, the high polymorphism found in *pap*-related restriction profiles results not only from the variability of individual gene clusters but also from the presence of different combinations of distinct gene clusters. At least three classes of *papG*-related adhesin genes were detected among the 22 class I isolates and strain J96 based on the presence or absence of *Pst*I and *Bgl*II restriction sites. (i) There was an adhesin gene cleaved only by *Bgl*II and carried by junction fragments of 6.8 kb (*Pst*I) and 4.9 kb (*Pst*I-*Eco*RI). This was found for the J96 *papG* gene, while it was detected in none of the 22 class I isolates (Fig. 2 and 4, and Table 2).

(ii) There were adhesin genes that were not cleaved by either enzyme but that were carried by junction fragments of 6.8 kb (*Pst*I) and 4.9 kb (*Pst*I-*Eco*RI). This was the case for the strain J96 *prsG* gene. In addition, restriction fragments with identical electrophoretic mobilities were detected in six class I isolates, suggesting the presence of gene clusters similar to *prs* which shared extended homology beyond the adhesin gene (Fig. 2 and 4 and groups B and C in Table 2). This may reveal an identical chromosomal location. Alternatively, the dissemination of a larger genetic element that includes *pap*-related sequences but that extends beyond the adhesin gene could account for these results. It should be noted that two gene clusters found in the same isolate gave identical junction fragments, suggesting the presence of two copies of such an element (Fig. 3, lanes 1 to 3, and Fig. 4, lanes 3a and 3b). This isolate expressed both P and F adhesins, indicating that this type of gene cluster is associated with expression of adhesins with distinct binding specificities.

(iii) There were adhesin genes that could be distinguished from the *papG* and *prsG* genes based on the presence of an internal *Pst*I site. This was the case for 16 isolates expressing different *pap*-related adhesins (groups D to I in Table 2). This site appeared to occupy the position of the *Bgl*II site of the J96 *papG* gene based on comigration of *Pst*I and *Pst*I-*Bgl*II restriction fragments (Fig. 4, lanes 4 to 8). In agreement with this hypothesis, nucleotide sequencing has recently revealed the presence of a *Pst*I site 4 base pairs upstream of the *Bgl*II site in the adhesin gene of the *pap*-related pap_{IA2} gene cluster encoding a P adhesin (15). Of note, an internal probe derived from the *papG*-related adhesin gene from the pap_{IA2} gene cluster detected 23 of 24 clinical isolates expressing a P adhesin, while a strain J96 *papG* probe detected only strain J96 (6). These results suggest that the $papG_{IA2}$ allele was the most frequently detected allele in both studies.

Our analysis demonstrates that the presence or absence of *Pst*I, *Eco*RI, or *Bgl*II sites within adhesin genes or the surrounding sequences is neither a conserved nor a discriminative feature of gene clusters encoding different binding specificities. The apparently close relationship found between particular gene clusters encoding distinct binding specificities (Table 2, and Fig. 2 and 4) suggests that a minor variation in the adhesin gene, such as a single amino acid substitution, may alter the binding specificity.

Copy number of *pap*-related gene clusters. The copy number of *pap*-related gene clusters was estimated by determining the number of *Eco*RI bands detected by the *papL*, *papHC*, *papCD*, and *papEFG* probes that together span almost the entire *pap* operon (Fig. 1). In strain J96 and 17 of

the 22 class I isolates, *pap*-related sequences were found to be organized in gene clusters that are likely to share the same organization as the J96 *pap* operon. One to three *Eco*RI bands were detected by the complete set of *pap* probes. In strains with multiple clusters, the *papHC* and *papCD* probes detected only the conserved 1.7- and 1.0-kb fragments in the *Pst*I and *Eco*RI-*Pst*I restriction digests, indicating the conservation of four internal *Pst*I sites in each cluster (Fig. 2, lanes 1 to 3). For 8 of these 17 class I isolates, the presence of a single gene cluster was associated with expression of a P adhesin only (three isolates), an F adhesin only (three isolates), and the null phenotype (two isolates). For the remaining nine isolates, expression of both P and F adhesins was associated with the presence of two (eight isolates) or three (one isolate) *pap*-related gene clusters. Thus, the presence of at least two copies of a *pap*-related gene cluster in these nine isolates indicates that two adhesins with distinct binding specificities may be encoded by two distinct operons, as has been found in strain J96 (14). Nonetheless, these results do not exclude the possibility that the same operon might encode an adhesin with overlapping specificities that can mediate bacterial binding to both the P human blood group and sheep erythrocyte antigens. In fact, based on results of a similar Southern blot analysis, we recently detected isolates from a patient with urosepsis that apparently had a single *pap*-related gene cluster while expressing both binding specificities (unpublished data).

In 5 of the 22 class I isolates, multiple *Eco*RI fragments were shown to carry nucleotide sequences related to the *pap* operon, although not all of the *Eco*RI bands were detected by the complete set of *pap* probes. The number of gene clusters with the same organization as the J96 *pap* operon could not be determined definitively, as these results imply either the presence of *Eco*RI sites within or between *pap*-related sequences or the presence of genetic elements related to only a part of the *pap* operon. For two isolates expressing only an F adhesin, the restriction profiles were consistent with the presence of a complete copy of a *pap*-related gene cluster plus an additional genetic element related only to the *papCD* and *papEFG* probes (Fig. 3, lanes 4 to 6). The restriction profiles of another isolate expressing both P and F adhesins indicated the presence of five clusters related to the *papL*, *papHC*, and *papCD* probes, while only a single *Eco*RI band was detected by the *papEFG* probe (Fig. 3, lanes 7 to 9). The restriction profiles of the two remaining class I isolates were consistent with the presence of two complete copies of *pap*-related gene clusters plus a genetic element related only to the *papEFG* probe (data not shown). Interestingly, the 2.3-kb *Pst*I and 1.9-kb *Eco*RI-*Pst*I fragments detected by the *papEFG* probes in one of these two isolates had the same electrophoretic mobilities as the *papEFG* probe-positive fragments of the three class II isolates (compare groups I and J in Table 2). Since class II isolates were not detected by the *papHC* and *papCD* probes, these observations strongly suggest that a genetic element related only to the *papEFG* probe can be found either alone (group J in Table 2) or with complete copies of *pap*-related gene clusters (group I in Table 2).

Copy number polymorphism of *pap*-related sequences: comparison with other virulence factors and significance. As described above, we resolved the existence of multiple *pap*-related gene clusters in the majority of *pap* probe-positive isolates (16 of 22 class I isolates). These findings are in dramatic contrast with the situation observed for other adhesin virulence factors of *E. coli*, such as for the *pil* (8, 18) and *afaI* (11) operons encoding the type 1 and X adhesins, respectively. Further characterization of the 30 clinical isolates described in this report revealed the presence of single gene clusters in all 28 *pil* probe-positive and 10 of 12 *afaI* probe-positive isolates (unpublished data). In agreement with these observations, Buchanan et al. (5) have described restriction profiles consistent with the presence of a single *pil*-related gene cluster in 51 of 60 *pil* probe-positive isolates. In the case of *afaI*-related sequences, Labigne-Roussel and Falkow (10) have detected the presence of two *afaI*-related gene clusters in only 5 of 15 *afaI* probe-positive isolates. In comparison with other adhesin operons, it thus appears that the high frequency of multiple *pap*-related genetic elements is a distinctive feature of *pap*-related sequences.

The unusual copy number polymorphism in *pap*-related sequences suggests the occurrence of possible recombination-based events involving duplication, deletion, or gene transfer. To investigate whether *pap*-related sequences are stably inherited in bacterial populations, we have recently used both multilocus gel electrophoresis (20) and Southern blot hybridization analysis to characterize isolates in blood and urine from patients with urosepsis, in addition to a subset of the isolates described in this study (M. Arthur, R. D. Arbeit, C. Kim, P. Beltran, H. Crowe, S. Steinbach, R. K. Selander, and R. Goldstein, manuscript in preparation). We found that the various *pap*-related genetic elements characterized in this study were not stably inherited among *E. coli* lineages. Rather, it appears that multiple and independent recombination events occurred in each of several genetically distant clones found to cause upper urinary tract infections.

Despite the occurrence of such recombination events, we also deciphered that *pap*-related genetic elements are inherited stably during infection and under laboratory conditions. This conclusion was based on our observation that no difference could be detected between *E. coli* isolated from blood versus that isolated from urine from the same patients. These findings therefore suggest that the polymorphisms described in this study reflect stably inherited differences within the limited time scale of an infection or laboratory procedures. These polymorphisms do, however, appear to result from recombinational events that occur at a much higher frequency than has been observed previously for essential chromosomal genes.

Based on comparisons of nucleotide sequences of three *pap*-related gene clusters, it has become apparent that variation of each major and minor pilus subunit type may occur independently (15). Recombination between *pap*-related sequences could account for these observations. The existence of multiple genetic elements, as resolved in this study, demonstrates that a wide variety of substrates exists for such recombination-based events in most clinical isolates. In addition to the analysis of Lund et al. (15), the detection of genetic elements related to only a part of the operon, as described in this report, further suggests that different regions of the *pap*-related gene cluster may be exchanged independently via recombinational events, thereby allowing for a constant flux in pilus adhesin specificity and antigenicity.

ACKNOWLEDGMENTS

We are grateful to Bernard Davis, Staffan Normark, Stanley Falkow, Agnès Labigne-Roussel, Catharina Svanborg-Edén, Peter Rice, Laurent Gutmann, Kurt Stottmeier, Daniel Bernstein, David Ozonoff, Shela Hull, Richard Hull, and Howard Corwin for encouragement or gifts of bacterial strains and clones. Manju Agarwall,

Vernard Coulter, and Ilga Wohlrab are acknowledged for technical assistance.

This study was supported by research grants DMB-8501039, DMB-8540693, DCB-8614076, and DMB-8841761 from the National Science Foundation (to R.G.), by a grant from the Community Technology Foundation of Boston University, and by funds from BRSG, National Institutes of Health, to the Boston University School of Medicine and Public Health. M.A. was supported by a fellowship from the Fonds National Suisse pour la Recherche Scientifique and from the National Kidney Foundation. R.D.A. was supported in part by the Research Service of the Veterans Administration and by funds from Biomedical Research Support Grants (BRSG), National Institutes of Health, to the Boston University School of Medicine. S.S. was supported in part by funding from the Hood Foundation and by funds from BRSG, National Institutes of Health, to the Boston University School of Medicine and Boston City Hospital. C.E.J. was supported by the Kidney Foundation of Ohio.

LITERATURE CITED

1. **Arthur, M., C. E. Johnson, R. H. Rubin, R. D. Arbeit, C. Campanelli, C. Kim, S. Steinbach, M. Agarwal, R. Wilkinson, and R. Goldstein.** 1989. Molecular epidemiology of adhesin and hemolysin virulence factors among uropathogenic *Escherichia coli*. Infect. Immun. **57**:303–313.
2. **Båga, M., M. Göransson, S. Normark, and B. E. Uhlin.** 1985. Transcriptional activation of a *pap* pilus virulence operon from uropathogenic *Escherichia coli*. EMBO J. **4**:3887–3893.
3. **Båga, M., M. Norgren, and S. Normark.** 1987. Biogenesis of *E. coli* pap pili: papH, a minor pilin subunit involved in cell anchoring and length modulation. Cell **49**:241–251.
4. **Båga, M., S. Normark, J. Hardy, P. O'Hanley, D. Lark, O. Olsson, G. Schoolnik, and S. Falkow.** 1984. Nucleotide sequence of the *papA* gene encoding the Pap pilus subunit of human uropathogenic *Escherichia coli*. J. Bacteriol. **157**:330–333.
5. **Buchanan, K., S. Falkow, R. A. Hull, and S. I. Hull.** 1985. Frequency among *Enterobacteriaceae* of the DNA sequences encoding type 1 pili. J. Bacteriol. **162**:799–803.
6. **Ekbäck, G., S. Mörner, B. Lund, and S. Normark.** 1986. Correlation of genes in the *pap* gene cluster to expression of globoside-specific adhesin by uropathogenic *Escherichia coli*. FEMS Microbiol. Lett. **34**:355–360.
7. **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.
8. **Hull, R. A., R. E. Gill, P. Hsu, B. H. Minshew, and S. Falkow.** 1981. Construction and expression of recombinant plasmids encoding type 1 or D-mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33**:933–938.
9. **Källenius, G., S. B. Svenson, R. Möllby, T. Korhonen, J. Windberg, B. Cedergren, I. Helin, and H. Hultberg.** 1982. Carbohydrate receptor structures recognized by uropathogenic *E. coli*. Scand. J. Infect. Dis. **33**(Suppl.):52–60.
10. **Labigne-Roussel, A., and S. Falkow.** 1988. Distribution and degree of heterogeneity of the afimbrial-adhesin-encoding operon (*afa*) among uropathogenic *Escherichia coli* isolates. Infect. Immun. **56**:640–648.
11. **Labigne-Roussel, A., M. A. Schmidt, W. Waltz, and S. Falkow.** 1985. Genetic organization of the afimbrial adhesin operon and nucleotide sequence from a uropathogenic *Escherichia coli* gene encoding an afimbrial adhesin. J. Bacteriol. **162**:1285–1292.
12. **Leffler, H., and C. Svanborg-Edén.** 1981. Glycolipid receptors for uropathogenic *Escherichia coli* on human erythrocytes and uroepithelial cells. Infect. Immun. **34**:920–929.
13. **Lindberg, F., B. Lund, L. Johansson, and S. Normark.** 1987. Localization of the receptor-binding protein adhesin at the tip of the bacterial pilus. Nature (London) **328**:84–87.
14. **Lund, B., F. Lindberg, B.-I. Marklund, and S. Normark.** 1987. The PapG protein is the α-D-galactopyranosyl-(1→4)-β-D-galactopyranose-binding adhesin of uropathogenic *Escherichia coli*. Proc. Natl. Acad. Sci. USA **84**:5898–5902.
15. **Lund, B., F. Lindberg, and S. Normark.** 1988. Structure and antigenic properties of the tip-located P pilus proteins of uropathogenic *Escherichia coli*. J. Bacteriol. **170**:1887–1894.
16. **Lund, B., B.-I. Marklund, N. Strömberg, 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.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Maurer, L., and P. E. Orndorff.** 1987. Identification and characterization of genes determining receptor binding and pilus length of *Escherichia coli* type 1 pili. J. Bacteriol. **169**:640–645.
19. **Norgren, M., S. Normark, D. Lark, P. O'Hanley, G. Schoolnik, S. Falkow, C. Svanborg-Edén, M. Båga, and B. E. Uhlin.** 1984. Mutations in *E. coli* cistrons affecting adhesion to human cells do not abolish pap pili fiber formation. EMBO J **3**:1159–1165.
20. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51**:873–884.
21. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
22. **Svanborg-Edén, C., and P. de Man.** 1987. Bacterial virulence in urinary tract infection. Infect. Dis. Clin. N. Am. **1**:731–750.