INFECTION AND IMMUNITY, June 1991, p. 2089–2096
0019-9567/91/062089-08$02.00/0

Vol. 59, No. 6

Frequency and Organization of *papA* Homologous DNA Sequences among Uropathogenic Digalactoside-Binding *Escherichia coli* Strains

KENNETH DENICH,[1] ABIE CRAIU,[1] HOPE RUGO,[1] GIRIJA MURALIDHAR,[1] AND PETER O'HANLEY[1,2,3]*

Department of Medicine, Division of Infectious Diseases,[2] and Department of Microbiology and Immunology,[1] Stanford University, Stanford, California 94305, and Veterans Administration Hospital, Palo Alto, California 94304[3]

Received 3 December 1990/Accepted 19 March 1991

The frequency of selected *papA* DNA sequences among 89 digalactoside-binding, uropathogenic *Escherichia coli* strains was evaluated with 12 different synthetic 15-base probes corresponding to *papA* genes from four digalactoside-binding piliated recombinant strains (HU849, 201B, 210B, and 200A). The *papA* probes encode amino acids which are common at the carboxy terminus of all strains, adjacent to the proximal portion of the intramolecular disulfide loop of strain 210B, or predicted to constitute the type-specific epitope for each of the four recombinant strains or other epitopes of strain HU849. The presence among the strains of DNA sequence homology to the *papA* probes was determined by in situ colony hybridization. Hybridization data suggest that there is a high frequency of homologous *papA* DNA sequences corresponding to selected regions of the *papA* gene from strain HU849 among the clinical strains. The following nucleotide locations which encode portions of the mature HU849 PapA are detected in a high percentage (42 to 70%) of clinical isolates: 208 to 222, 310 to 324, 478 to 492, 517 to 531, 553 to 567, and 679 to 693. These sequences encode portions of the predicted protective, immunogenic, and/or antigenic epitopes of this PapA. The data also indicate considerable heterogeneity of *papA* sequences among the strains, especially in the region of nucleotide bases corresponding to positions 391 to 418. These oligonucleotides encode the predicted PapA type-specific immunogenic dominant epitope. Determination of the extent of genetic variability in the *papA* gene among digalactoside-binding strains will require more extensive DNA sequencing of prototypic *papA* genes, additional hybridization studies employing other *papA* gene oligonucleotide probes, and assessment of the different *pap* operons and their copy number in each strain.

Adherence of *Escherichia coli* to uroepithelial cells is an important pathogenic step in the development of urinary tract infections. There are a number of adhesins expressed by uropathogenic *E. coli* strains (viz., Pap, mannose, AFA, Dr, S, and MN) which might mediate uroepithelial attachment (1, 7, 14, 24, 26, 29, 30, 38, 40, 41). However, pyelonephritogenic strains are characterized by a high frequency of pili associated with α-D-Galp-(1-4)-β-D-Galp (Gal-Gal) binding (1, 18, 27, 33, 38). The Gal-Gal-binding phenotype is considered crucial to the pathogenesis of nonobstructive, ascending urinary tract infections in the anatomically normal host (8, 9, 25). This is probably due to the absence of a natural host defense factor in urine to prevent Gal-Gal pili from binding to uroepithelial cells (25). In contrast, uromodulin (also termed Tamm-Horsfall protein) in urine abrogates adherence of mannose-binding bacteria to uroepithelial cells in vivo (28) or the binding of purified mannose pili to its receptor in vitro (25). Digalactoside-binding pili have been variably termed Pap pili (pili associated with pyelonephritis) or P pili because they bind to the P_1 blood group antigen, a globoside containing Gal-Gal that is present on human erythrocytes, uroepithelial cells, and a variety of other epithelial cells (12, 16, 17).

Many *pap* gene clusters have been cloned, and some have been extensively studied (2–4, 10, 11, 19–21, 23, 26, 31, 32, 33, 34, 39, 41–44). There are strains which simultaneously harbor multiple copies of different *pap* operons (3, 32). A *pap* operon encodes at least nine polypeptides involved in the biogenesis of receptor-binding pili (23, 39). In brief, the pili are heteropolymers composed of a major pilin subunit, PapA, and three minor tip-located proteins, PapE, PapF, and PapG (20, 23, 39). Expression of Pap pili and the adhesin requires synthesis of *papC* and *papD* gene products. PapC and PapD are required for the assembly of pilin subunits (A, E, F, and G) into a fiber at the cell surface (11, 23). Transcomplementation analysis has indicated that PapG is the Gal-Gal-specific adhesin, although PapF is also required for receptor specificity (20). The PapA protein is not required for specific binding; however, attempts to develop effective vaccines for immunoprophylaxis against *E. coli* urinary tract infections have employed this polypeptide since it is the most abundantly produced and immunogenic of the *pap* gene products (25, 31, 35). Anti-PapA immunoglobulin G (IgG) in the urine correlates with protection against subsequent colonization by homologous and heterologous PapA-piliated challenge strains (31, 35). The precise mechanism of protection afforded by anti-PapA IgG has not been elucidated. We have determined that specific PapA antibody does not prevent epithelial attachment or directly influence pilus expression, since the rate of spontaneous phase variation of Gal-Gal-binding wild-type *E. coli* isolates is not affected by culturing bacteria in broth or on agar in the presence of specific PapA antibody. We suspect that the antibody-coated bacteria are subjected to stress or impaired growth and/or more readily cleared as aggregates from the urinary tract.

The linear immunogenic and antigenic structures of the PapA pilin from the HU849 recombinant strain have been previously evaluated with polyclonal IgG and nine synthetic peptides corresponding to regions that span the major pilin sequence predicted to contain hydrophilic β turns (36).

* Corresponding author.

TABLE 1. Characteristics of control strains

E. coli strain	Origin	Binding status[a]	F-antigen status	PapA pilin[b]	Molecular mass (kDa) of PapA pilin[c]	Reference(s)
J96	Pyelonephritis	Gal-Gal, MS	F13	+	17.5	10, 29
HU849	Recombinant	Gal-Gal	F13	J96 DNA	17.5	10, 29, 31
SH48	Recombinant	MS		–		10, 29, 31
C1212[d]	Cystitis	Gal-Gal, MS	$F7_1$, $F7_2$	+	21.0, 17.0	26, 42
201B	Recombinant	Gal-Gal	$F7_1$	C1212 DNA	21.0	32
210B	Recombinant	Gal-Gal	$F7_2$	C1212 DNA	17.0	32
3669	Pyelonephritis	Gal-Gal, MS	F9	+	19.0	31,36
200A	Recombinant	Gal-Gal	F9	3669 DNA	19.0	32
HB101	K-12, laboratory	Negative		–		10

[a] Binding status refers to the ability of the strain to bind mannose or Gal-Gal latex beads (MS, mannose binding; Gal-Gal, digalactoside binding).

[b] PapA pilin refers to whether a strain expresses PapA pilin or the source of DNA to construct the recombinant digalactoside-binding strain.

[c] The molecular mass of PapA pilin was determined by the relative mobility of Coomassie blue-stained bands in 12.5% polyacrylamide separating gels when 25 to 50 μg of purified pili (26) were subjected to SDS-polyacrylamide gel electrophoresis (15).

[d] C1212 (O6:K2:H1) (29) was originally isolated from a woman with cystitis and first identified as strain AD111 by van den Bosch et al. (41). Unfortunately, we have inadvertently reported elsewhere (35) that strain C1212 is strain AD110, which is not correct (27b). AD110 strain is the same as strain C1210 (26, 27b).

There were at least five linear immunogenic epitopes contained within this 163-amino-acid PapA polypeptide. Peptides corresponding to amino acid residues 25 to 38 (R 25–38), R 38–50, and R 48–61, which jointly constitute the single intramolecular disulfide loop, and R 103–116 were bound at low titers. A prominent immunogenic epitope was identified by a peptide corresponding to R 65–75. In fact, this particular region (R 65–79), which is just distal to the disulfide loop of PapA pilins, constitutes the type-specific, immunogenic dominant epitope for this polypeptide family (36). The hypothesis that this portion constitutes a type-specific PapA epitope is supported by similar results from additional studies employing synthetic peptides corresponding to this region from 200A, 201B, and 210B recombinant digalactoside-binding strains (31) and specific PapA polyclonal IgG antibodies (24a). Also, two prominent antigenic epitopes for HU849 PapA pilin were localized to peptides corresponding to R 5–12 and R 93–104; two minor antigenic epitopes were localized to peptides corresponding to R 65–75 and R 119–131. None of the peptide antisera bound PapA pilins from heterologous Gal-Gal-binding strains, except anti-R 93–104 and anti-R 5–12 IgG (35, 36). Fifty-eight to 80% of Gal-Gal-binding strains were serologically cross-reactive to anti-R 93–104 or anti-R 5–12 IgG. Furthermore, evidence indicates that linear protective epitopes are located between R 5–12 and the type-specific immunogenic domain (R 65–75) for the HU849 PapA (35). Mice vaccinated with these peptides were protected from experimental *E. coli* pyelonephritis upon challenge with a homologous strain (35). (Immunogenic epitopes refer to a particular domain in the native protein that is recognized by the immune system and elicit antibodies able to bind synthetic peptides corresponding to this domain. Antigenic epitopes refer to domains that are recognized in the native protein by antibodies engendered by synthetic peptides corresponding to that region of the protein. Protective epitopes refer to synthetic peptide vaccines which protect BALB/c mice from experimental *E. coli* pyelonephritis [25].)

In the present study, we have determined the frequency of homologous *papA* DNA sequences among uropathogenic digalactoside-binding *E. coli* strains by employing synthetic oligonucleotides corresponding to specific structural or predicted β-turn determinants of four sequenced *papA* genes (4, 26a). Results from this study extend our understanding of the molecular epidemiology of the *papA* gene and provide a basis for future vaccine developments for immunoprophylaxis against *E. coli* urinary tract infections.

MATERIALS AND METHODS

Bacteria. A total of 89 digalactoside-binding *E. coli* strains were studied that were originally isolated over a 25-year period from the urine of patients living in geographically diverse locations in the United States. On the basis of rigorous radiological studies and clinical definitions (27), 37 isolates were classified as obtained from individuals with anatomically normal urinary tracts with a specific urinary tract infection. Eleven were from women with acute pyelonephritis; 16 were from women with acute cystitis; 5 were from women with recurrent cystitis; and 5 were from males with acute urethritis. These 37 isolates were also assessed for O serotype. They were distributed among 10 different O serotypes: 2, 4, 6, 7, 18, 23, 50, 75, 86, and 132. The remaining 52 isolates were from adults with undetermined urinary tract anatomy and not well-defined clinical syndromes of symptomatic urinary tract infection. Control strains that were employed in this study are listed in Table 1. HU849, 201B, 210B, and 200A are recombinant strains which harbor a *pap* operon (10, 31). They were grown and maintained on Luria agar supplemented with ampicillin (50 μg/ml).

Gal-Gal binding assays. Gal-Gal binding by bacteria was assessed by either latex agglutination or hemagglutination assays. Bacterial strains were grown on MacConkey agar at 37°C and harvested at 18 to 24 h into 0.1 M phosphate-buffered saline, pH 7.4 (PBS), at a concentration $\geq 10^8$ CFU/ml based on optical density. These bacteria were then employed in the assays. In the latex agglutination assay (26), synthetic α-D-Galp-(1-4)-β-D-Galp with an 8-methoxycarbonylactyl-glycoside linker was absorbed to latex beads (Chembiomed Ltd., Edmonton, Alberta, Canada). Equal volumes (50 μl) of bacteria were mixed with a 1% (vol/vol) latex suspension in PBS and assessed for slide agglutination. Lactose absorbed to latex was used as a negative control. In the other assay (26), the bacterial suspensions were first assessed for mannose-resistant hemagglutination with human blood group OP_1 erythrocytes at a final concentration of 2% (vol/vol) in PBS supplemented with 4% (wt/vol) D-mannose. Equal volumes (50 μl) of the bacterial and erythrocyte suspensions were mixed and assessed for hemagglutination.

TABLE 2. Synthetic *papA* oligonucleotides corresponding to β turns, intramolecular disulfide loops, and carboxy-terminal regions of four different *papA* genes

Probe	*papA* origin		Location of probe (nucleotide sequence/amino acid residues)	Oligonucleotide sequence (5'→3') and amino acid sequence (NH_2→COOH)	Comment
	Parent	Recombinant			
1	C1212	210B	679–693/163–167	ACC CTG ACT TAT CAG N L T Y Q	Common COOH-terminal sequence for all 4 *papA* genes
2	C1212	210B	252–266/21–24	A CCA TCT GGT ATT GA P C G I	Corresponds to first cysteine residue that forms a disulfide loop
3	C1212	210B	391–405/67–71	AAA AAA GCT GCT GGC K K A A G	Predicted type-specific epitope
4	C1212	201B	402–416/67–71	AAA CAG CTT CAA GGC K Q L Q G	Predicted type-specific epitope
5	3669	200A	404–418/65–69	AAA AAG GCA GCT ACA K K A A T	Predicted type-specific epitope
6	J96	HU849	394–408/67–71	AAA GGT GGT AAT GGC K G G N G	Predicted type-specific epitope
7	J96	HU849	208–222/5–9	CCA CAG GGG CAG GGT P Q G Q G	Corresponds to clonal antigenic and protective epitopes
8	J96	HU849	310–324/39–43	TCA AAA AGC TTC CTT S K S F L	Corresponds to immunogenic epitope within loop
9	J96	HU849	346–360/51–55	CCA ATG GAC TTA GAT P M D L D	Corresponds to immunogenic epitope within loop
10	J96	HU849	478–492/95–99	ACA AAT GGT GGT ACG T N G G T	Corresponds to a clonal antigenic epitope
11	J96	HU849	517–531/108–112	GCA GGT AAA AAC GTT A G K N V	Corresponds to an immunogenic epitope
12	J96	HU849	553–567/120–124	GAT GCT AAT ACC CTG D A N T L	Corresponds to an antigenic epitope

All mannose-resistant hemagglutination-positive bacteria were assessed for specific hapten inhibition of hemagglutination by mixing bacteria with Gal-Gal at a final concentration of 2% (wt/vol). After preincubation of the mixture for 1 h at room temperature, an equal volume of a 4% erythrocyte suspension with mannose was added and the mixture was assessed for agglutination.

Synthetic oligonucleotide probes. Twelve different 15-base oligonucleotides were synthesized (Table 2). The selection of oligonucleotides for synthesis corresponding to DNA segments of four known *papA* gene sequences (4, 26a) was based on one of three criteria. Probe 1 was selected because it represents a common DNA sequence at the carboxy terminus of the four *papA* genes. Probe 2 corresponded to a sequence from the *papA* gene from strain 210B that is adjacent to the proximal portion of the intramolecular disulfide loop. The presence of a cysteine-cysteine loop which involves R 22–61 is a discriminating feature of this polypeptide family (26). Probes 3 through 12 were selected based, in part, on their encoding predicted β turns according to the algorithm of Chou and Fasman (5). Probes 3 through 6 correspond to the selected four encoded amino acids assumed to constitute the type-specific, immunogenic dominant epitope for each of the four recombinant pili. The remaining six probes corresponded to the selected four encoded amino acids assumed to constitute either immunogenic or antigenic epitopes for the PapA protein from strain HU849 (36). The oligonucleotides were synthesized by cyanoethyl chemistry with an Applied Biosystems 380A DNA Synthesizer (Foster City, Calif.) and purified as recommended by the manufacturer. Purified oligonucleotides were 5' end labeled with [γ-^{32}P]ATP (3,000 Ci/mmol) and T4 polynucleotide kinase (22). The specific radioactivity of the probes was approximately 10^8 and 10^9 cpm/μg.

Colony hybridization. Bacterial colonies were replica plated onto nitrocellulose filters and screened by in situ hybridization. After 12 h of growth at 37°C, duplicate filters were made, denatured, neutralized, and baked at 80°C for 2 h as described by Grunstein and Wallis (6). The filters were washed three times in 3× SSC (standard saline citrate; 1× SSC is 0.15M NaCl plus 0.015 M sodium citrate [pH 8.0])–0.1% sodium dodecyl sulfate (SDS) for 5 min at room temperature and once in 3× SSC–0.1% SDS for 90 min at 50°C to remove bacterial debris. Filters were prehybridized for 12 h in 5× SSC–5× Denhardt's solution (1× Denhardt's solution is 0.02% bovine serum albumin, 0.02% Ficoll and 0.02% polyvinylpyrrolidone)–0.5% SDS–0.05% $Na_4P_2O_7 \cdot 10H_2O$–denatured salmon sperm (100 μg/ml) at 37°C. The filters were hybridized in this solution containing 10^6 cpm of labeled oligonucleotide per ml at 2 to 5°C below the calculated T_d (temperature of dissociation), at which a perfectly matched hybrid will be half disassociated. T_d was calculated as described by Suggs et al. (37): T_d = 4°C per GC base pair + 2°C per AT base pair. The temperature for hybridization was 37°C for all probes, except that for probe 8, which was carried out at 35°C. Filters were then washed four times for 10 min each at room temperature in 6× SSC–0.05% $Na_4P_2O_7 \cdot 10H_2O$ and once at 37°C for 30 min and exposed to X-ray film at −70°C for 12 h. After autoradiography, filters were routinely rewashed once at 45°C in 6× SSC–0.5% $Na_4P_2O_7 \cdot 10H_2O$ and exposed again to X-ray film as described above to assess the stringency of the hybridization. Additional washes above 45°C were performed to assess the specificity of the hybridization reaction. All in situ hybridizations under the different conditions of stringency were repeated at least two times as described to ensure consistency of results. Each filter contained at least one positive control (viz., the recombinant strain from which the probe was derived) and a standard negative control (strain HB101).

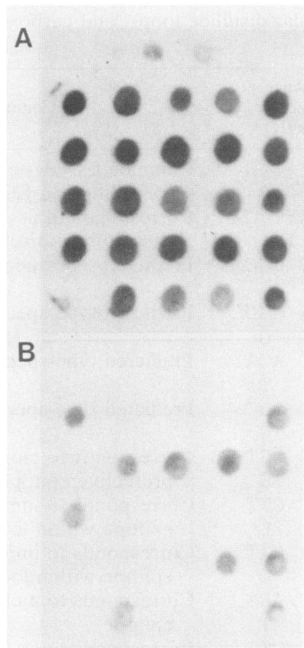

FIG. 1. Colony hybridization of 26 *E. coli* strains with probe 7 (including controls). (A) 5′-end-labeled ^{32}P-probe 7 was hybridized to the filter at 37°C overnight, washed with 6× SSC–0.5% $Na_4P_2O_7 \cdot 10H_2O$ at 37°C, and exposed to X-ray film. (B) The same filter was rewashed as described above at 45°C and again exposed to X-ray film.

TABLE 3. Hybridization of *E. coli* control strains with 12 different *papA* gene probes

E. coli strain[a]	Hybridization with oligonucleotide probe[b]											
	1	2	3	4	5	6	7	8	9	10	11	12
HB101	−	−	−	−	−	−	−	−	−	−	−	−
SH48	−	−	−	−	−	−	−	−	−	−	−	−
HU849	+	−	−	−	−	+	+	+	+	+	+	+
201B	+	−	−	+	−	−	−	−	−	−	−	−
210B	+	+	+	−	−	−	−	−	−	−	−	−
200A	+	−	−	−	+	−	−	−	−	−	−	−
J96	+	−	−	−	−	+	+	+	+	+	+	+
C1212	+	+	+	+	−	−	−	−	−	−	−	−
3669	+	−	−	−	+	−	−	−	−	−	−	−

[a] Strain HB101 is a laboratory K-12 derivative that lacks mannose-binding and *pap* operons. This strain was employed to harbor the mannose-binding pilus operon plasmid for recombinant strain SH48 and the *pap* operon for recombinant strains HU849, 201B, 210B, and 200A. SH48 and J96 contain homologous mannose-binding pilus operons. HU849 and J96 contain homologous *papA* genes; 201B and 210B *papA* genes are derived from the C1212; 200A and 3669 contain homologous *papA* genes.

[b] The nucleotide sequence and location of the 12 different synthetic *papA* oligonucleotide probes are shown in Table 2.

RESULTS

Assessment of positive hybridization. Eleven synthetic oligonucleotide probes corresponding to four different *papA* gene sequences (probes 2 to 12) and one probe corresponding to a common sequence at the carboxy terminus of these *papA* genes (probe 1) were employed in in situ colony hybridization (Table 2). All probes were 15 bases in length. In general, the pattern of colony hybridization for any given probe for a specific isolate was distinct and varied little from experiment to experiment. The intensity of a positive hybridization signal for any one probe among the clinical isolates exhibited variability: strong dark to gray (Fig. 1A). In addition, comparison of hybridization intensity between probes showed a similar degree of variability. In this report, we define a positive hybridization reaction as one that exhibited a signal following a 45°C wash and autoradiography (Fig. 1B). The only exception to this designation was the criterion employed for probe 8. We observed that the hybridization signal for probe 8 to its positive control was removed from the filter following a 45°C wash. Based on empiric observations, we defined a positive hybridization for probe 8 as one that exhibited a hybridization signal following a 37°C wash and autoradiography. All hybridization signals were removed from the filter after a 50°C wash, including those of the positive controls (data not shown). In addition, *E. coli* HB101 (the standard negative control) showed an occasional weak gray hybridization signal with some of the probes (probes 2 and 9 through 12) following the 37°C wash. There was, however, no detectable hybridization signal for any of the probes with strain HB101 following a 45°C wash. In contrast, *E. coli* SH48 (the other negative control) did not produce hybridization signals with any of probes following the 37°C wash. Table 3 summarizes the colony hybridization results for the control strains for all 12 *papA* probes. In each case, the oligonucleotide probe hybridized strongly with the positive control strain from which the probe sequences were derived. There was no hybridization signal detectable when probes which had a different *papA* DNA sequence were employed in reactions with control strains. All positive control strains, which share an identical carboxy-terminal sequence (represented by probe 1), hybridized as expected. Negative controls, strains HB101 and SH48, did not hybridize with any of the probes according to our definitions of a positive result.

Distribution of *papA* probes. Figure 2 summarizes the hybridization results for the 89 clinical isolates that exhibit Gal-Gal binding. All clinical isolates hybridized with probe 1. Probes 7, 8, 10, 11, and 12 also demonstrated a high frequency of homologous *papA* DNA sequences among the clinical isolates. These five probes are derived from the *papA* gene of strain HU849. Probes 7, 8, 10, 11, and 12 hybridized with 62 (70%), 46 (52%), 47 (53%), 44 (49%), and 37 (42%) of the 89 digalactoside-binding clinical isolates, respectively. In contrast, probes 3, 4, and 6 hybridized to only one strain each of the clinical isolates; probe 5 did not hybridize to any of the 89 clinical isolates.

Based on these data, Fig. 2 illustrates the 18 different patterns of hybridization (hybridization groups). We define a hybridization group as a group of clinical isolates that hybridize uniquely with a specific set of probes. For example, the 12 clinical isolates in hybridization group 16 each hybridized to probes 1, 2, 7, 8, 10, 11, and 12 (Fig. 2). No other group of clinical isolates had an identical pattern of hybridization. Excluding the common *papA* gene probe (probe 1), 75 (84%) of the 89 clinical isolates exhibited positive hybridization with at least one probe derived from the *papA* gene from the HU849 recombinant strain. In addition, 39 (44%) of the 89 strains hybridized with at least four unique probes corresponding to HU849 *papA* sequences. Only one clinical isolate in hybridization group 17 hybridized with all seven unique *papA* probes corresponding to strain HU849 sequences (Fig. 2). Nineteen (21%) of the 89 clinical isolates exhibited positive hybridization with at least one unique *papA* probe corresponding to 210B recombinant strain sequences (Fig. 2).

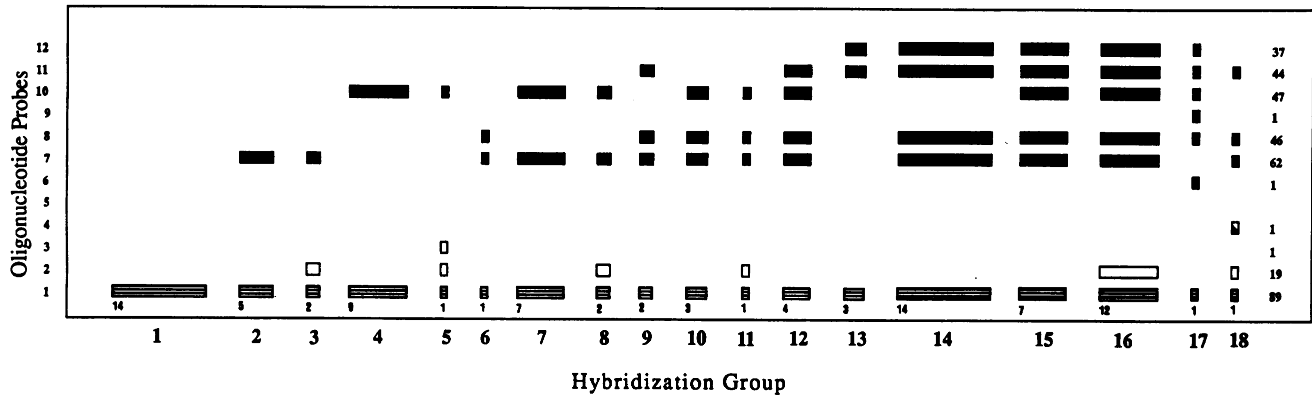

FIG. 2. Distribution of 89 digalactoside-binding *E. coli* strains, isolated from the urine of patients, into 18 hybridization groups. The numbers on the bottom indicate the hybridization group. The numbers on the far left indicate which oligonucleotide probe was used (Table 2). The numbers on the far right indicate the total number of isolates that hybridized with each oligonucleotide probe. The number under each box indicates the total number of isolates in a hybridization group (as described in the text) that hybridized with one or more oligonucleotide probes. The length of each box is proportional to the number under each box; probes derived from HU849 are indicated by a solid box; the probe derived from 201B is indicated by a black and white box; probes derived from 210B are indicated by an open box; and the common oligonucleotide sequence at the carboxyl terminus is indicated by a horizontally hatched box.

Table 4 lists the hybridization groups according to the clinical origins of each *E. coli* isolate. There was no obvious distribution for the hybridization groups among the 37 well-defined clinical isolates causing a specific urinary tract infection syndrome. In addition, when the hybridization results for these 37 well-defined clinical isolates were compared with their O serotypes, no correlation was observed (data not shown). It appears that the clinical isolates from symptomatic urinary tract infections are broadly distributed among the hybridization groups. It is noteworthy, however, that there is a high frequency of homologous *papA* DNA sequences among the 89 digalactoside-binding clinical isolates corresponding to probe 7. Nine (82%) of 11 pyelonephritis strains, 11 (69%) of 16 cystitis strains, and 38 (73%) of 52 uropathogenic strains hybridized to probe 7 (based on analysis of data from Table 4 and Fig. 2).

***papA* organization.** Table 5 indicates the percentage of clinical isolates which display identical hybridization patterns for each of the probes when two probes are compared. The percentage of isolates that gave identical hybridization patterns was calculated by adding the number of isolates positive for both probes and dividing by the total number of isolates positive for either probe. For example, 37 (84%) of 44 clinical isolates had an identical hybridization pattern when probes 11 and 12 were employed (Table 5). The percentage of clinical isolates that displayed an identical *papA* hybridization pattern with selected pairs of probes corresponding to the *papA* DNA sequences from strain HU849 ranged from 31 to 84% (Table 5). Identical hybridization patterns between combinations of the other probes (probes 2 through 6 and 9) were calculated and found to be considerably less (viz., <20%) (data not shown).

DISCUSSION

We examined the frequency of *papA* homologous DNA sequences among 89 digalactoside-binding *E. coli* isolates from patients with symptomatic urinary tract infections. We employed 12 different 15-base synthetic oligonucleotides that encode amino acid residues either predicted to correspond to β turns which constitute epitopes or to be at common locations within the mature PapA proteins encoded by four *papA* genes. The majority of *papA* gene probes were derived from strain HU849, corresponding to portions of its immunogenic, antigenic, and/or protective epitopes (35, 36). In this report, the colony hybridization data suggest a high frequency of homologous *papA* DNA sequences corresponding to selected regions from strain HU849. These data provide the impetus for more accurately determining the diversity of the *papA* gene among clinical isolates in order to develop vaccines against PapA. Because most clinical isolates contain two or more *pap* operons and because colony hybridization cannot distinguish whether the sequence homology of the multiple DNA probes used here is due to homologous sequences on separate operons in a given strain, we suspect that the genetic diversity of *papA* genes is much more extensive than we can estimate from colony hybridization data. Future hybridization studies which specifically evaluate the genetic diversity of the *papA* gene in clinical isolates must assess the number of different *pap* operons

TABLE 4. Distribution of *papA* hybridization groups among uropathogenic *E. coli* strains

Clinical syndrome	Total no. of strains	No. of strains in hybridization group:																	
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Acute pyelonephritis	11	1	3											1	5		1		
Acute cystitis	16	2	2	1	2			1					1		2	1	4		
Acute urethritis	5															2	3		
Recurrent cystitis	5	3						2											
Undefined	52	8		1	7	1	1	4	2	2	3	1	3	2	7	4	4	1	1

TABLE 5. Percentages of clinical isolates which display identical *papA* hybridization patterns between probes derived from the recombinant strain HU849[a]

Hybridization pattern with probe:	% of isolates with identical hybridization pattern with probe:			
	8	10	11	12
7	74	51	64	52
8		43	84	69
10			36	31
11				84

[a] The percentages of clinical isolates which displayed an identical *papA* hybridization pattern between probes is above the diagonal. An identical hybridization pattern is defined as the pattern observed when two probes hybridize to a given clinical isolate. The percentage was calculated by adding the number of isolates positive for both probes divided by the total number of isolates positive for either probe and multiplying by 100. The data were derived from Fig. 2.

harbored by each strain (3, 32). Future work will include screening strains by Southern blots of restriction fragment polymorphisms with *papA* probes.

The hybridization results reveal that among the clinical isolates there was a high frequency of *papA* homologous DNA sequences derived from the HU849 recombinant strain. This result may be due to probe selection bias since 7 to the 11 unique *papA* gene probes employed in this study were derived from the sequence of the *papA* gene of strain HU849. Excluding a probe representing a portion of the type-specific epitope (probe 6) and a probe corresponding to a portion of a minor immunogenic epitope within the disulfide loop (probe 9), the remaining five probes unique for HU849 *papA* DNA hybridized with 42 to 70% of the 89 uropathogenic, digalactoside-binding clinical isolates. This might not be surprising since other investigators (4, 13, 34, 42–44) have suggested that the *papA* genes of strain HU849 and other digalactoside-binding strains may have evolved from a common ancestral gene. The high prevalence of conserved regions in the *papA* genes supports the concept that broadly cross-protective PapA pilus vaccines can be developed. For example, probe 7 hybridized to 62 (70%) of 89 digalactoside-binding uropathogenic strains. Further sequence analysis of this region (R 5–12) in digalactoside-binding strains by genetic or protein sequencing methods might be worthwhile since it has already been demonstrated that this region constitutes both an antigenic and a protective epitope for one PapA against pyelonephritis (35). These data could provide a rational basis for new vaccine development.

We consider the absence of hybridization of any of the 11 unique *papA* gene probes to the clinical isolates to be due to *papA* DNA sequence divergence and not to the absence of a *papA* gene. This assumption is based on a number of characteristics of the strains. First, all the clinical isolates express digalactoside binding. Second, all the clinical isolates hybridized with probe 1, which corresponds to a conserved nucleotide sequence at the carboxyl terminus of the four *papA* genes. Third, the 89 clinical isolates hybridized to all 13 different restriction fragments derived from strain HU849, ranging in size from 500 to 1,800 bases, which span the entire *pap* operon, including restriction fragments within the *papA* gene (27a). Finally, the strong concordance of positive signals observed among the clinical isolates for a number of selected combinations of *papA* probes derived from HU849 sequences suggest the presence of a *papA* gene. We assume that sequence divergences as little as a single-base-pair difference in selected regions of the *papA* gene would account for the lack of hybridization observed among the clinical isolates, given the stringency of the hybridization conditions used in this study.

The hybridization data also suggest that selection pressure on the *papA* gene to diverge is not random. For example, probe 7 and 1 sequences, which are derived from regions corresponding to the amino or carboxy terminus, respectively, are not extensively divergent among clinical isolates. They are 70 to 100% homologous among the 89 digalactoside-binding clinical strains. These nonimmunogenic regions could be conserved either because there was no selective advantage for mutations in this region or because these regions are essential for pilus biogenesis (12, 34, 42–44). They might be involved in the translocation of PapA to the cell surface by interacting with accessory *pap* proteins (4, 13, 42–44). It is interesting that the COOH region of PapA pilins is absolutely conserved among strains which exhibit Gal-Gal binding and might likely be the location where PapD interaction occurs. Hultgren and coworkers (11) have shown that the PapG adhesin of uropathogenic *E. coli* contains separate regions for receptor binding and for interaction with other *pap* gene products. The receptor-binding domain was mapped to the amino-terminal half of the adhesin, whereas a hydrophobic preassembly domain which is necessary for the formation of a complex with PapD was located near the carboxy terminus. In the model of pilus assembly, they also suggest that a PapA-PapD complex might occur in the cytoplasm (11). It is possible that the conserved PapA portion at the carboxy terminus is involved with PapD interaction. Furthermore, 84% of the clinical isolates hybridized to a combination of probes 8 and 11 and probes 11 and 12 despite 193 and 22 bases intervening between these regions, respectively (Table 5). In contrast, there is much more discordance when other combinations of probes are examined in this manner (e.g., 36% concordance when probes 10 and 11 are employed despite only 25 bases intervening between these probes). We suggest that the observed variation in different regions cannot be explained by random point mutations or base substitutions alone. The large variations seen within a pair of probes are probably due to genetic recombination between *papA* genes. This assumption is in agreement with previous studies of restriction fragment polymorphisms with virulence-associated operons and hybridization studies employing large fragments of the *pap* operon of uropathogenic *E. coli* isolates (2, 3, 32, 34). They have shown that there is considerable diversity within *pap* cistrons which can be explained by inter- and intrachromosomal recombinations. Also, only 3 (3%) of the 89 clinical isolates hybridized to probes 3 through 6. These probes correspond to portions predicted to encode the putative PapA type-specific epitope (36). This result is not surprising since the highest degree of genetic diversity occurs in this region of the *papA* gene, based on the analysis of other *papA* genes sequenced by other groups (4, 42–44) and the four complete *papA* gene sequences of recombinant strains used in this study (26a). It also probably explains the serological diversity of Gal-Gal pili of *E. coli* strains. Furthermore, the fact that many of these strains might possess multiple *pap* gene copies suggests that the diversity is even more extensive.

In summary, results from this study extend our understanding of the molecular epidemiology of the *papA* gene among uropathogenic, digalactoside-binding *E. coli* strains (2, 3, 32, 34). Data indicate that the DNA sequence which encodes a portion of the protective epitope in the amino

terminus of one PapA (of strain HU849) is frequently harbored by uropathogenic strains. This provides further support that a human vaccine trial employing this epitope is warranted and might be expected to provide broad protection against *E. coli* pyelonephritis (35). However, other *papA* regions were not homologous to the *papA* gene of HU849, suggesting sequence divergence. The greatest genetic variability occurs in the region which encodes the predicted type-specific epitope of the mature PapA protein. It seems likely, therefore, that PapA vaccines composed of this epitope would be more likely to select for antigenic escape variants than would those made up of the more genetically conserved region at the amino terminus. To further assess the genetic variability of the *papA* gene among these strains, a number of other studies are required. They include Southern blot analysis of restriction fragment polymorphism, DNA sequencing performed on some or all of the strains, and the use of additional oligonucleotide probes corresponding to other *papA* genes (e.g., from strains 200A, 201B, and 210B). These studies will be required to phylogenetically characterize the *papA* gene among strains that cause urinary tract infections.

ACKNOWLEDGMENTS

We thank Kenneth Vosti for a number of strains from well-defined clinical episodes of urinary tract infection.

This work was supported by grants from the Veterans Administration and the National Institutes of Health (AI23435). Hope Rugo was supported, in part, by a National Institute of Health Training Grant (HL07100) to the University of California, San Francisco.

REFERENCES

1. **Archambaud, M., P. Courcoux, V. Ouin, G. Chabanon, and A. Labigne-Roussel.** 1988. Phenotypic and genotypic assays for the detection and identification of adhesins from pyelonephritic *Escherichia coli*. Ann. Inst. Pasteur Microbiol. **139:**557–573.
2. **Arthur, M., R. D. Arbeit, C. Kim, P. Beltran, H. Crowe, S. Steinbach, C. Campanelli, R. A. Wilson, R. K. Selander, and R. Goldstein.** 1990. Restriction fragment length polymorphisms among uropathogenic *Escherichia coli* isolates: *pap*-related sequences compared with *rrn* operons. Infect. Immun. **58:**471–479.
3. **Arthur, M., C. Campanelli, R. D. Arbeit, C. Kim, S. Steinbach, C. E. Johnson, R. Rubin, and R. Goldstein.** 1989. Structure and copy number of gene clusters related to the pap P-adhesin operon of uropathogenic *Escherichia coli*. Infect. Immun. **57:**314–321.
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. **Chou, P. Y., and G. D. Fasman.** 1978. Empirical predictions of protein conformation. Annu. Rev. Biochem. **47:**251–254.
6. **Grunstein, M., and J. Wallis.** 1979. Colony hybridization. Methods Enzymol. **68:**379–389.
7. **Hacker, J., A. Schrettenbuinner, G. Schroter, H. Duvel, G. Schmidt, and W. Goebel.** 1986. Characterization of *Escherichia coli* wild-type strains by means of agglutination with antisera raised against cloned P-, S-, and MS-fimbriae antigens, hemagglutination, serotyping and hemolysin production. Zentralbl. Bakteriol. Hyg. A **261:**219–231.
8. **Hagberg, L., I. Enberg, R. Freter, J. Lam, S. Olling, and C. Svanborg Edén.** 1983. Ascending, unobstructed urinary tract infection in mice caused by pyelonephritogenic *Escherichia coli* of human origin. Infect. Immun. **40:**273–283.
9. **Hagberg, L., R. Hull, S. Hull, S. Falkow, R. Freter, and C. Svanborg Edén.** 1983. Contribution of adhesion to bacterial persistence in the mouse urinary tract. Infect. Immun. **40:**265–272.
10. **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 mannose-resistant pili from a urinary tract infection *Escherichia coli* isolate. Infect. Immun. **33:**933–938.
11. **Hultgren, S., F. Lindberg, G. Magnusson, J. Kihlberg, J. Tennet, and S. Normark.** 1989. The Pap G adhesin of uropathogenic *E. coli* contains separate regions for receptor binding and for the incorporation into the pilus. Proc. Natl. Acad. Sci. USA **86:**4357–4361.
12. **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. Suppl. **33:**52–60.
13. **Klemm, P.** 1981. The complete amino acid sequence of the K88 antigen, a fimbrial protein from *Escherichia coli*. Eur. J. Biochem. **117:**617–627.
14. **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.
15. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227:**680–685.
16. **Leffler, H., and C. Svanborg-Edén.** 1980. Chemical identification of a glycosphingolipid receptor for *Escherichia coli* attaching to human urinary tract epithelial cells and agglutinating human erythrocytes. FEMS Microbiol. Lett. **8:**127–134.
17. **Leffler, H., and C. Svanborg-Edén.** 1981. Glycolipid receptors for uropathogenic *Escherichia coli* on human erythrocytes and uroepithelial cells. Infect. Immun. **34:**920–929.
18. **Lichodziejewska, M., R. Steadman, K. Jones, N. Topley, R. Mackenzie, and J. Williams.** 1989. Variable expression of P fimbriae in *Escherichia coli* urinary tract infection. Lancet **i:**1414–1418.
19. **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.
20. **Lund, B., F. Lindberg, B. I. Marklund, and S. Normark.** 1987. The *pap G* 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.
21. **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.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **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.
24. **Nowicki, B., A. Labigne, S. Moseley, R. Hull, S. Hull, and J. Moulds.** 1990. The Dr hemagglutinin, afimbrial adhesins AFA-I and AFA-III, and F1845 fimbriae of uropathogenic and diarrhea-associated *Escherichia coli* belong to a family of hemagglutinins with Dr receptor recognition. Infect. Immun. **58:**279–281.

24a. **O'Hanley, P.** Unpublished data.

25. **O'Hanley, P., D. Lark, S. Falkow, and G. Schoolnik.** 1985. Molecular basis of *Escherichia coli* colonization of the upper urinary tract in BALB/c mice: Gal-Gal pili immunization prevents *Escherichia coli* pyelonephritis in the BALB/c mouse model of human pyelonephritis. J. Clin. Invest. **75:**347–360.
26. **O'Hanley, P., D. Lark, S. Normark, S. Falkow, and G. Schoolnik.** 1983. Mannose-sensitive and Gal-Gal binding *Escherichia coli* pili from recombinant strains. J. Exp. Med. **158:**1713–1719.

26a. **O'Hanley, P., D. Low, J. Hardy, A. Craiu, and K. Denich.** Submitted for publication.

27. **O'Hanley, P., D. Low, I. Romero, D. Lark, K. Vosti, S. Falkow, and G. Schoolnik.** 1985. Gal-Gal binding and hemolysin phenotypes and genotypes associated with uropathogenic *Escherichia coli*. N. Engl. J. Med. **313:**414–420.

27a. **O'Hanley, P., H. Rugo, A. Schmidt, and G. Muralidhar.** Unpublished data.

27b. **Ørskov, I.** Personal communication.

28. **Ørskov, I., A. Ferenez, and F. Ørskov.** 1980. Tamm-Horsfall protein in the normal urinary tract that traps type-1 fimbriated *E. coli*. Lancet **i**:887.
29. **Ørskov, I., F. Ørskov, and A. Birch-Andersen.** 1980. Comparison of *Escherichia coli* fimbrial antigen F7 with type 1 fimbriae. Infect. Immun. **27**:657–666.
30. **Ott, M., J. Hacker, T. Schmoll, T. Harchau, T. K. Korhonen, and W. Goebel.** 1986. Analysis of the genetic determinants coding for the S-fimbrial adhesin (*sfa*) in different *Escherichia coli* strains causing meningitis or urinary tract infections. Infect. Immun. **54**:646–653.
31. **Pecha, B., D. Low, and P. O'Hanley.** 1989. Gal-Gal pili vaccines prevent pyelonephritis by piliated *Escherichia coli* in a murine model: single-component Gal-Gal pili vaccines prevent pyelonephritis by homologous and heterologous piliated *E. coli* strains. J. Clin. Invest. **83**:2102–2108.
32. **Plos, K., T. Carter, S. Hull, R. Hull, and C. Svanborg-Eden.** 1990. Frequency and organization of *pap* homologous DNA in relation to clinical origin of uropathogenic *Escherichia coli*. J. Infect. Dis. **161**:518–524.
33. **Plos, K., S. I. Hull, R. A. Hull, B. R. Levin, I. Ørskov, F. Ørskov, and C. Svanborg-Edén.** 1989. Distribution of the p-associated-pilus (*pap*) region among *Escherichia coli* from natural sources: evidence for horizontal gene transfer. Infect. Immun. **57**:1604–1611.
34. **Rhen, M., I. vanDie, V. Rhen, and H. Bergmans.** 1985. Comparison of the nucleotide sequences of the genes encoding the KS71A and $F7_1$ fimbrial antigens of uropathogenic *Escherichia coli*. Eur. J. Biochem. **151**:573–577.
35. **Schmidt, A., P. O'Hanley, D. Lark, and G. Schoolnik.** 1988. Synthetic peptides corresponding to protective epitopes of *Escherichia coli* digalactoside-binding pilin prevent infection in a murine pyelonephritis model. Proc. Natl. Acad. Sci. USA **85**:1247–1251.
36. **Schmidt, A., P. O'Hanley, and G. Schoolnik.** 1984. Gal-Gal pyelonephritis *E. coli* pili linear immunogenic and antigenic epitopes. J. Exp. Med. **161**:705–717.
37. **Suggs, S. V., R. Wallace, T. Hirase, E. H. Kawashima, and K. Itakura.** 1981. Use of synthetic oligonucleotides as hybridization probes: isolation of cloned cDNA sequences for human B_2-microglobin. Proc. Natl. Acad. Sci. USA **78**:6613–6617.
38. **Svanborg-Edén, C., and P. deMan.** 1987. Bacterial virulence in urinary tract infection. Infect. Dis. Clin. North Am. **1**:731–750.
39. **Uhlin, B. E., M. Norgren, M. Båga, and S. Normark.** 1985. Adhesion to human cells by *Escherichia coli* lacking the major subunit of digalactoside-specific pilus-adhesin. Proc. Natl. Acad. Sci. USA **82**:1800–1804.
40. **Vaisanen, V., T. Korhonen, M. Jakinen, G. Gahmberg, and C. Ehnholm.** 1982. Blood group M specific hemagglutinin in pyelonephritogenic *Escherichia coli*. Lancet **i**:1192.
41. **van den Bosch, J. F., U. Verbloom-Sohmer, P. Postma, J. deGraaff, and D. MacLaren.** 1980. Mannose-sensitive and mannose-resistant adherence to human uroepithelial cells and urinary virulence of *Escherichia coli*. Infect. Immun. **29**:226–233.
42. **van Die, I., and H. Bergmans.** 1984. Nucleotide sequence of the gene encoding the $F7_2$ fimbrial subunit of a uropathogenic *Escherichia coli* strain. Gene **32**:83–90.
43. **van Die, I., M. Dijksterhuis, H. deCock, W. Hoekstra, and H. Bergmans.** 1986. Structural variation of P-fimbriae from uropathogenic *Escherichia coli*, p. 39–46. *In* D. L. Lark (ed.), Protein-carbohydrate interactions in biological systems. Academic Press, Inc. (London), Ltd., London.
44. **van Die, I., W. Hoekstra, and H. Bergmans.** 1987. Analysis of the primary structure of P fimbrillins of uropathogenic *Escherichia coli*. Microb. Pathog. **3**:149–154.