INFECTION AND IMMUNITY, Feb. 1990, p. 471–479
0019-9567/90/020471-09$02.00/0

Vol. 58, No. 2

Restriction Fragment Length Polymorphisms among Uropathogenic *Escherichia coli* Isolates: *pap*-Related Sequences Compared with *rrn* Operons

MICHEL ARTHUR,[1]† ROBERT D. ARBEIT,[2] CHEUNG KIM,[1] P. BELTRAN,[3] H. CROWE,[1]‡ SUZANNE STEINBACH,[4] CRAIG CAMPANELLI,[1] RICHARD A. WILSON,[5] ROBERT K. SELANDER,[3] AND RICHARD GOLDSTEIN[1]*

Section of Molecular Genetics and Epidemiology, Maxwell Finland Laboratory for Infectious Diseases, Boston University Schools of Public Health and Medicine,[1] and Departments of Medicine[2] and Pediatrics,[4] Boston University School of Medicine, Boston, Massachusetts 02118, and Departments of Biology[3] and Veterinary Science,[5] Pennsylvania State University, University Park, Pennsylvania 16802

Received 26 July 1989/Accepted 23 October 1989

Among the adhesin-encoding virulence operons associated with uropathogenic *Escherichia coli*, only *pap* (pyelonephritis-associated pilus)-related gene clusters typically exhibit variation in their structure and chromosomal copy number. To access further such variability, we compared *pap* restriction fragment length polymorphisms (RFLPs) with those detected among rRNA (*rrn*) operons, which encode an essential host function unrelated to virulence. To place such findings in a phylogenetic perspective, the *E. coli* isolates were also characterized by using multilocus enzyme electrophoresis. Variation in the *rrn* RFLP profiles correlated with evolutionary divergence resolved by multilocus enzyme electrophoresis; isolates with identical *rrn* profiles represented the same or closely related electrophoretic types. In contrast, such isolates frequently had different *pap*-related RFLPs, indicating that these genetic variations have developed recently relative to the changes associated with essential *rrn* operons or metabolic enzymes. Despite such fluctuations, two lines of evidence indicate conditions under which the *pap*-related RFLPs can be stably maintained. First, for each of 20 patients with urosepsis, both the primary urinary tract isolate and the concurrent blood isolate were identical. Second, although obtained from different patients, some isolates representing the same electrophoretic type also had identical *pap*-related RFLPs. Thus, the genotypic diversity of this virulence adhesin operon was not generated during the course of acute infection or during laboratory manipulations. Since fecal *E. coli* isolates frequently carry chromosomally encoded *pap*-related gene clusters, these findings suggest that the intra- and interchromosomal recombination events generating the polymorphisms associated with the *pap*-related sequences likely occur among the *E. coli* of the commensal reservoir.

Escherichia coli cause the substantial majority of urinary tract infections (33). As a group, such uropathogenic isolates exhibit properties that are not prevalent among commensal *E. coli* isolates found in the intestinal flora (40). Prominent among these phenotypes is the expression of adhesins that mediate bacterial attachment to the human uroepithelium and thereby confer a selective advantage in ascending colonization of the upper urinary tract by allowing the bacteria to resist the cleansing effect of urine flow in the ureters (29).

pap (pyelonephritis-associated pili) and *prs* (*pap*-related sequence) are two structurally and functionally related operon clusters that have been cloned from *E. coli* J96 and shown to express adhesin pili (9, 21, 22). *pap* encodes a P-fimbrial adhesin that specifically recognizes the globoseries of glycolipids present on both human erythrocytes bearing the P blood group antigen and uroepithelial cells (19, 21, 22). *prs* encodes an F adhesin that preferentially binds to the Forsmann antigen, which is a major constituent of sheep erythrocyte membranes and is also present on the cells of the human renal pelvis (21, 22).

In a recent survey, we observed that coexpression of these adhesin-binding specificities was significantly more frequent among isolates causing pyelonephritis than among isolates causing cystitis or among isolates obtained from stool specimens (2). Further, *E. coli* isolates from different patients were highly diverse with respect to the structure and copy number of sequences related to the *pap* operon (2, 16, 17, 32).

The overall genetic diversity among *E. coli* populations has been extensively studied by using multilocus enzyme electrophoresis (MLEE) (35). In this approach, differences in the electrophoretic mobilities of metabolic enzymes are used to identify allelic variation in the corresponding chromosomal genes (34). Such allelic variation is, in general, selectively neutral or nearly so and, therefore, minimally subject to convergent changes. Further, statistical analyses suggest that the rate of recombination between these chromosomal genes is low (35). As such, quantitative analysis of the results of MLEE have been used to define phylogenetic relationships among isolates. These studies indicate that natural *E. coli* populations are characterized by considerable genetic diversity and that isolates are distributed among genetically diverging lineages.

The primary goals of the current study were (i) to compare the genetic polymorphisms of uropathogenic virulence factor operons with those resolved for the essential chromosomal loci encoding rRNA and (ii) to correlate the diversity observed for the virulence factors with the evolutionary lin-

* Corresponding author.
† Present address: Unité des Agents Antibactériens, Institut Pasteur, Paris 75724 Cedex 15, France.
‡ Present address: Division of Infectious Diseases, Hartford Hospital, Hartford, CT 06115.

TABLE 1. Clinical sources and epidemiologic characteristics of infecting *E. coli*

Patient isolate	ET	Diagnosis	Age (yr)	Sex	Source of isolate	Hospital[a]	Date
45	1	Cystitis	2	Female	Urine	CM	March 1983
464	2	Biliary sepsis	56	Male	Bile, blood	VA	July 1987
457	3	Urosepsis	52	Male	Urine, blood	BCH	July 1987
23	4	Pyelonephritis	0.9	Male	Urine	CM	December 1984
18	5	Pyelonephritis	1	Female	Urine	CM	December 1984
175	5	Urosepsis	49	Female	Urine, blood	BCH	January1985
17	6	Cystitis	5	Male	Urine	CM	October 1984
213	7	Urosepsis	20	Female	Urine, blood	BCH	January 1987
49	8	Cystitis	7	Female	Urine	CM	November 1983
36	9	Cystitis	4	Female	Urine	CM	October 1983
471	10	Urosepsis	66	Male	Urine, blood	VA	September 1987
14	11	Cystitis	2	Male	Urine	CM	September 1984
19	12	Cystitis	3	Female	Urine	CM	September 1984
442	13	Urosepsis	30	Female	Urine, blood	BCH	June 1987
176	14	Urosepsis	47	Female	Urine, blood	BCH	February 1985
180	15	Urosepsis	0.1	Male	Urine, blood	BCH	April 1985
452	16	Urosepsis	51	Female	Urine, blood	BCH	June 1987
140	17	Unknown	70	Male	Feces	BCH	June 1985
141	17	Unknown	84	Female	Feces	BCH	June 1985
181	18	Urosepsis	50	Male	Urine, blood	BCH	April 1985
447	18	Urosepsis	94	Female	Urine, blood	BCH	June 1987
454	18	Biliary sepsis	52	Female	Bile, blood	BCH	July 1987
178	19	Urosepsis	34	Female	Urine, blood	BCH	February 1985
179	19	Urosepsis	71	Male	Urine, blood	BCH	May 1985
336	19	Urosepsis	26	Female	Urine, blood	BCH	June 1987
459	20	Urosepsis	57	Female	Urine, blood	BCH	July 1987
473	20	Urosepsis	61	Male	Urine, blood	VA	October 1987
254	21	Urosepsis	30	Female	Urine, blood	BCH	February 1987
469	22	Urosepsis	57	Male	Urine, blood	VA	September 1987
467	23	Biliary sepsis	61	Male	Bile, blood	VA	September 1987
177	24	Urosepsis	55	Male	Urine, blood	BCH	November 1984
440	24	Urosepsis	30	Female	Urine, blood	BCH	June 1987
456	25	Urosepsis	42	Female	Urine, blood	BCH	July 1987

[a] CM, Cleveland Metropolitan Hospital; VA, Boston Veterans Administration Medical Center; BCH, Boston City Hospital.

eages among the isolates as detected by MLEE. As a prerequisite to this analysis, we established that the diversity observed among the virulence-associated operons is not generated in the course of in vivo infection or in vitro manipulations.

MATERIALS AND METHODS

Patient population and collection of isolates. Patients with *E. coli* bacteremia associated with acute urinary or biliary tract infection were identified through the microbiology laboratories of the Boston City Hospital and the Veterans Administration Medical Center, Boston, Mass. Medical records were reviewed to obtain relevant clinical data (Table 1). The 20 patients with bacteremic urosepsis (i) had *E. coli* isolated from concurrent blood and urine specimens, (ii) had $\geq 10^5$ *E. coli* per ml of urine, and (iii) had clinical signs and symptoms indicating upper urinary tract infection (33). These 20 patients were selected arbitrarily without reference to clinical features or the properties of their isolates; thus, they represent a random sample of patients with bacteremic pyelonephritis. The three patients with bacteremic biliary sepsis had *E. coli* isolated from cultures of the gall bladder at surgery and from blood drawn within 48 h preceding surgery. Ten *E. coli* isolates previously characterized with respect to the expression and structure of virulence factors (2, 3) were chosen as representatives of the structural diversity of *pap*-related sequences; isolates 140 and 141 had identical *pap* restriction profiles and were chosen for that reason. In total, primary-source (urinary tract, biliary tract, or feces) isolates from 33 epidemiologically unrelated patients were studied (Table 1). For each of the 23 bacteremic patients, blood isolates were also examined.

For each isolate, a subculture was grown from a single colony and stored in 22% glycerol at −80°C. All studies were performed on fresh cultures prepared at 37°C on Luria agar or in Luria broth.

MLEE and statistical methods. The electrophoretic mobilities of 13 water-soluble enzymes were analyzed (Table 2). Methods of preparing lysates, performing starch-gel electrophoresis, and demonstrating specific enzyme activity have been previously described (34).

Electromorphs (allozymes) of each enzyme were equated with alleles at the corresponding chromosomal gene locus. Distinctive combinations of alleles over the 13 loci assayed (multilocus genotypes) were designated as electrophoretic types (ETs) (7). Genetic diversity at a locus was calculated as $h = (1 - \Sigma x_i^2)(n/n - 1)$, where x_i is the frequency of the ith allele and n is the number of isolates (27). Mean genetic diversity (h) is the arithmetic average of h values for all loci. Genetic distance (D) between pairs of isolates was expressed as the proportion of enzyme loci at which different alleles were detected (mismatches). Clustering was performed from a matrix of genetic distances by the average-linkage method (37).

Phenotypic characterization of the isolates. Expression of P, F, type 1, and X adhesins was detected by agglutination

TABLE 2. Genetic diversity among isolates at 13 metabolic enzyme loci

Enzyme locus (symbol)	Genetic diversity[a] (no. of alleles)	
	Urosepsis isolates[b] (n = 20)	All isolates[b] (n = 33)
Mannitol 1-phosphate dehydrogenase	0.753 (4)	0.699 (4)
Adenylate kinase	0.521 (2)	0.511 (2)
Acid phosphatase	0.679 (5)	0.729 (7)
Malic acid enzyme	0.605 (4)	0.655 (4)
Phosphoglucose isomerase	0.416 (3)	0.517 (4)
Isocitrate dehydrogenase	0.484 (3)	0.511 (3)
Malate dehydrogenase	0.100 (2)	0.278 (4)
Phenylalanyl leucine peptidase	0.605 (4)	0.655 (4)
Shikimic dehydrogenase	0.553 (4)	0.619 (5)
Alcohol dehydrogenase	0.363 (4)	0.498 (5)
Leucylglycyl glycine peptidase	0.416 (3)	0.438 (4)
6-Phosphogluconate dehydrogenase	0.705 (6)	0.693 (7)
Leucine aminopeptidase	0.616 (4)	0.676 (4)

[a] Mean values were 0.534 (3.7) for urosepsis isolates and 0.575 (4.4) for all isolates.

[b] See Materials and Methods for details regarding the selection and characterization of these strains.

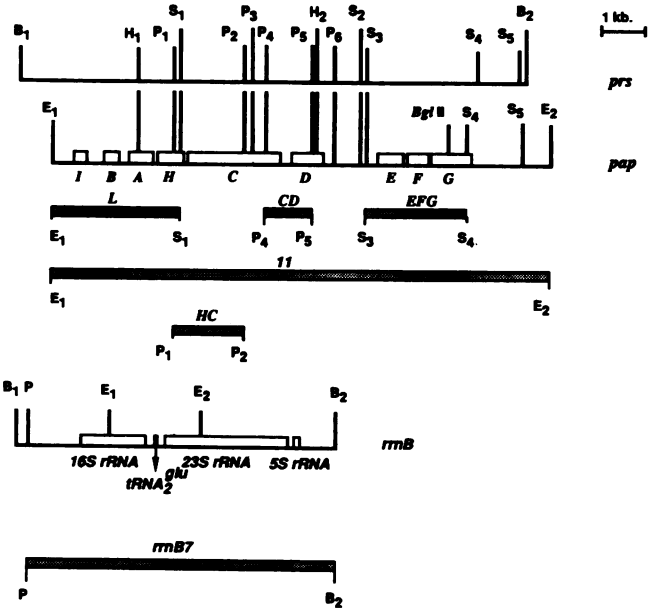

FIG. 1. Physical and genetic maps of the *pap* and *rrnB* operons indicating the positions of the DNA restriction fragments used to generate the probes. Detailed maps of the *pap* (P adhesin), *prs* (F adhesin), and *rrnB* (rRNA) operons have been presented (5, 6, 21, 22). *pap* and *prs* are entirely homologous operons; phenotypic specificity cannot be distinguished by restriction digests or available suboperon probes (2, 9, 21, 22). Restriction endonuclease cutting sites are denoted by vertical lines; genes are denoted by open horizontal boxes; and the DNA fragments used to generate the 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 after restriction enzyme designations refer to consecutive cutting sites.

assays as previously described (3). Isolates were considered to express a P adhesin if agglutination was positive for P1MM erythrocytes, positive for P1NN erythrocytes, and negative or comparatively weaker and slower for ppNN erythrocytes. All and only isolates expressing a P adhesin agglutinated latex beads coated with the minimal digalactoside receptor of the *E. coli* J96 P adhesin. Isolates were considered to express an F adhesin if hemagglutination was positive with sheep erythrocytes and to express an X adhesin if hemagglutination was positive with ppNN human erythrocytes. A type 1 adhesin was considered to be expressed if hemagglutination was positive with guinea pig erythrocytes; only this type of agglutination was inhibited by D-mannose.

The production of hemolysin was tested on Columbia agar supplemented with 4% sheep erythrocytes (3). Colonies surrounded by a clear halo after overnight culture at 37°C were defined as hemolysin positive.

Serotyping was performed at the *E. coli* Reference Laboratory, Pennsylvania State University, College Park, by standard techniques (31).

DNA hybridization. The hybridization procedures have been described previously in detail (2, 3). Briefly, plasmid DNA used to prepare the probes was obtained by alkaline lysis and purified by cesium chloride-ethidium bromide equilibrium density gradient centrifugation (23). The DNA restriction fragments used to generate the probes (Fig. 1) were separated by agarose gel electrophoresis. DNA from a slice of the gel containing the relevant restriction fragment was radiolabeled in the agarose by random oligo priming with [α-^{32}P]dCTP (800 Ci/mmol; Dupont, NEN Research Products, Boston, Mass.) (10, 11).

Colony hybridization membranes were prepared by in situ lysis of bacteria grown on nitrocellulose filters (BA-85; Schleicher & Schuell Co., Keene, N.H.) (23). Southern blot membranes (38) were prepared with DNA obtained by bacterial lysis in the presence of sodium lauryl sarcosine (15) and purified by cesium chloride-ethidium bromide equilibrium density gradient centrifugation. Restriction fragments were separated by electrophoresis in 0.8% agarose horizontal slab gels (25 by 16 by 0.7 cm) and transferred to Zeta-probe membranes (Bio-Rad Laboratories, Richmond, Calif.) as previously described (23). Between successive hybridizations, probes bound to the membranes were removed by alkaline treatment as recommended by the manufacturer. Hybridizations (42°C in the presence of 50% formamide) and washes were performed under stringent conditions as described previously (23).

RESULTS

Electrophoretic types and genetic diversity. Among the 33 primary source isolates, all 13 enzyme loci were polymorphic, with an average of 4.4 alleles per locus and a mean genetic diversity (h) of 0.575 (Table 2). A total of 25 distinctive multilocus ETs were defined (Fig. 2). For the 20 urinary isolates from patients with bacteremic urosepsis, the mean genetic diversity was 0.524. These 20 isolates were assigned to 15 ETs, three of which were represented by two isolates and one of which was represented by three isolates.

Cluster analysis of the ETs (Fig. 2) revealed 14 lineages (designated as clusters A to N) at a genetic distance of $\leq$0.3; that is, isolates representing a cluster had, on the average, more than 70% of alleles in common. There were four major lineages (designated as groups I to IV) at a distance of $\geq$0.6; that is, isolates representing different groups had, on the average, fewer than 40% of alleles in common. Urosepsis

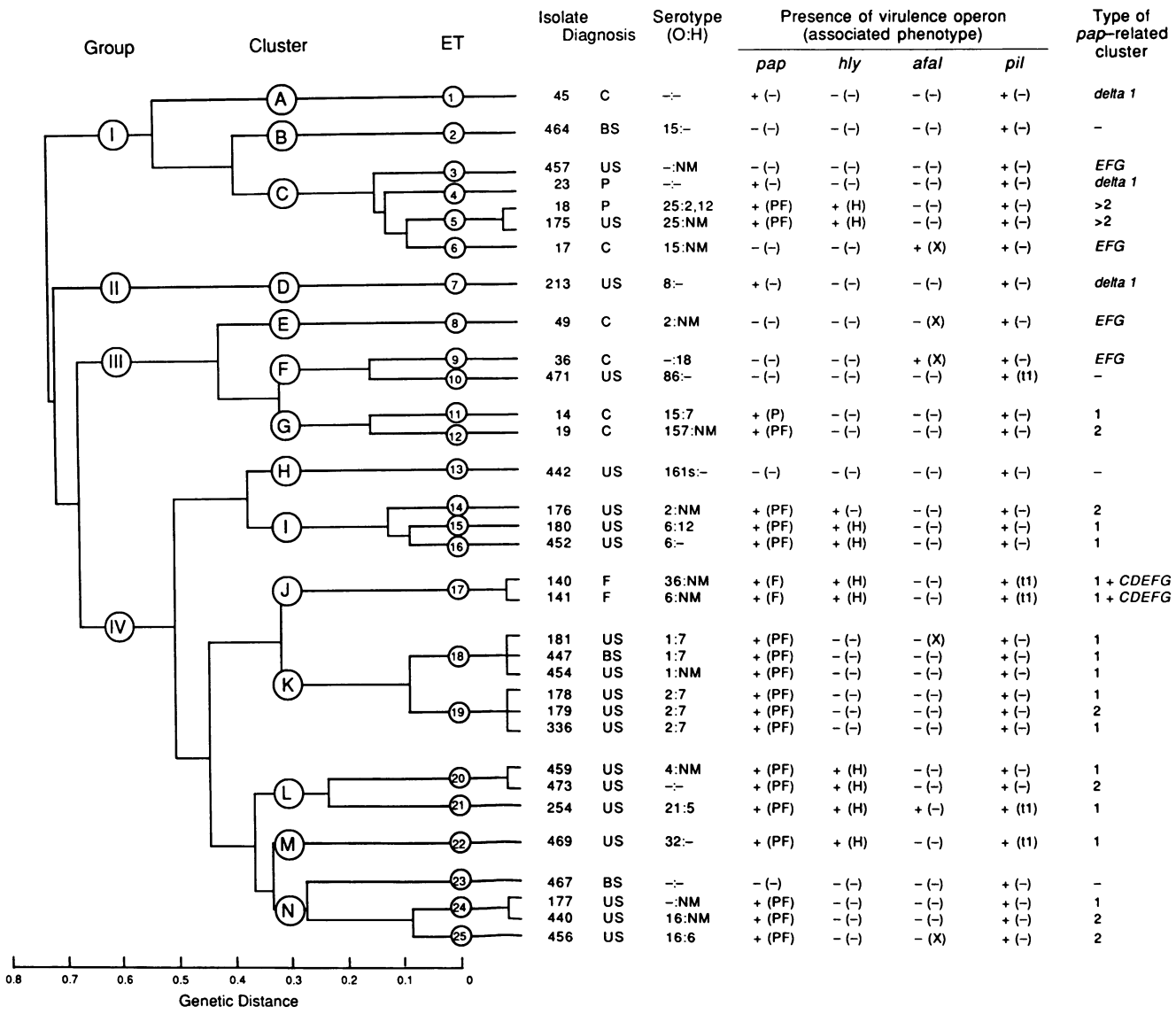

FIG. 2. Distribution of virulence factors among *E. coli* lineages. The dendogram depicts the phylogenetic relationships among the 33 primary source isolates as determined by MLEE (see Materials and Methods). Isolates with a genetic distance of 0.0 have the same combination of alleles over the 13 enzyme loci, i.e., the same ET, labeled 1 through 25. A distance of 0.08 corresponds to a single-locus difference. Lineages diverging within a genetic distance of 0.3 are referred to as clusters and labeled A through N. Lineages diverging within a genetic distance of 0.6 are referred to as groups and labeled I through IV. Diagnosis abbreviations: C, cystitis; US, urosepsis; P, pyelonephritis; BS, biliary sepsis; F, isolates obtained from stool specimen (unknown diagnosis; Table 1). For each isolate the presence of nucleotide sequences related to the *pap*, *hly*, *afaI*, and *pil* operons was detected by colony DNA hybridizations. For the *pap* operon a positive sign indicates a positive hybridization with the complete set of *pap* probes listed in Fig. 1. For each virulence factor, the associated phenotype is indicated in parenthesis. Abbreviations: (−), absence of phenotypic expression; *pap* (P), expression of the P adhesin-binding specificity; *pap* (F), expression of the F adhesin-binding specificity; *pap* (PF), expression of both adhesin-binding specificities; *hly* (h) production of hemolysin; *afaI* (X), expression of an X adhesin; *pil* (t1), expression of a type 1 adhesin (3). For each isolate the type and copy number of *pap*-related gene cluster as detected by Southern blot hybridization are listed (2). *pap*-related gene cluster abbreviations: *delta* 1, cryptic *pap*-related gene cluster lacking the conserved 1.0-kb *papCD* restriction fragments; −, negative hybridization with the *papHC*, *papCD*, and *papEFG* probes; *EFG*, cryptic *pap*-related gene cluster lacking homology to the essential *papHCD* region of the *pap* operon; >2, 5 copies of the *papHCD* region of the *pap* operon plus one copy of the *papEFG* region (2); 1, one complete copy of a functional *pap*-related gene cluster; 2, two complete copies of functional *pap*-related gene clusters; 1 + *CDEFG*, one complete copy of a functional *pap*-related gene cluster plus a genetic element related to the *papCD* and *papEFG* probes (2).

isolates were nonrandomly distributed among the lineages, with 16 of the 20 isolates representing ETs in group IV; moreover, 14 of these 16 isolates were in four clusters (I, K, L, and N). ET 18 and ET 19, which differed at only 1 of the 13 enzyme loci analyzed, were represented by a total of five urosepsis isolates.

Restriction polymorphisms associated with *rrn* operons. Polymorphism in the size of *Eco*RI restriction fragments carrying rRNA genes was studied by Southern blot hybridization (Fig. 3). The *rrnB7* probe was prepared from a 7.0-kilobase-pair restriction fragment that carries the entire *rrnB* operon of *E. coli* K-12 and flanking sequences (Fig. 1)

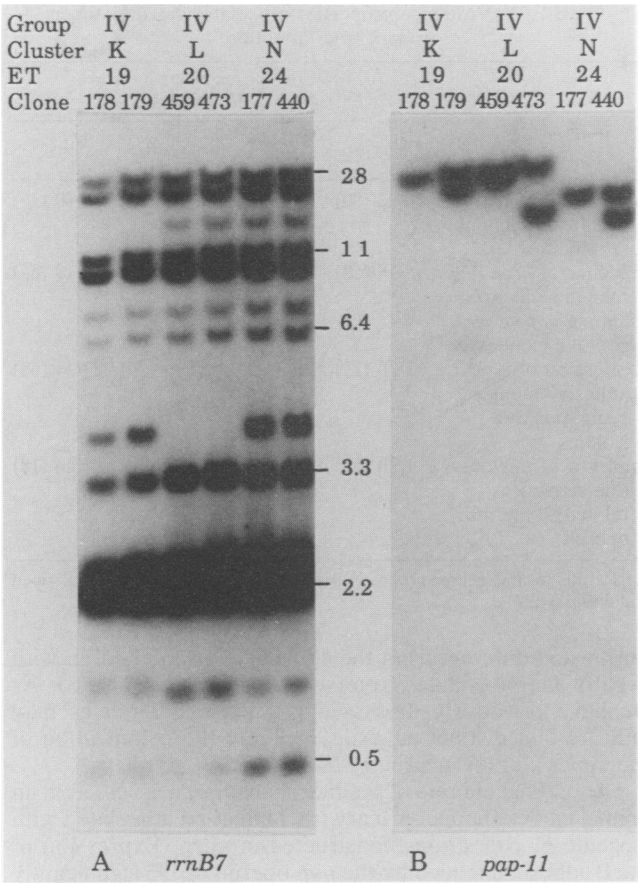

FIG. 3. Restriction fragment length polymorphisms among *E. coli* clinical isolates. For six primary source isolates obtained from six different patients, DNA samples were digested with *Eco*RI, electrophoresed in an agarose gel, and transferred to a membrane. (A) Hybridization with the *rrnB7* probe; (B) hybridization of the same membrane with the *pap-11* probe. For each *E. coli* isolate, the isolate number and lineage (group, cluster, and ET) are indicated (see text and Fig. 2). Between the two autoradiographs, the sizes of restriction fragments are given in kilobases.

(5, 6). Hence, restriction fragment length polymorphisms (RFLPs) may reflect variation in nucleotide sequence within or surrounding the ribosomal operons as well as variation in copy number of *rrn* operons (seven in *E coli* K-12) (4). All possible pairwise comparisons of *Eco*RI *rrn* restriction profiles among the 33 isolates were derived from electrophoresis of samples on the same gel, and all identities were confirmed by electrophoresis of samples in adjacent lanes.

All isolates assigned to the same ET had identical *Eco*RI *rrn* restriction profiles. In addition, identical restriction profiles were detected among isolates assigned to different but closely related ETs. This was observed for two pairs of ETs differing at one enzyme locus (ETs 18 and 19 and ETs 24 and 25), one pair of ETs differing at two enzyme loci (ETs 14 and 15), and one pair of ETs differing at three enzyme loci (ETs 20 and 21). In contrast, identical *rrn* restriction profiles were never detected between isolates of ETs assigned to different lineages. Overall, among the 33 isolates, there was slightly less diversity in *rrn* RFLPs (21 distinct *Eco*RI restriction profiles) than in multilocus genotypes (25 ETs over the 13 enzyme loci).

Distribution of *pap*-related sequences among *E. coli* lineages. We have previously reported in detail the use of Southern blot hybridizations and agglutination assays to establish relationships between the structure and the function of *pap*-related sequences (2, 3). Using this approach, we determined the distribution of this adhesin virulence factor among the 33 primary source isolates described here (Fig. 2).

Functional *pap*-related gene clusters were present in 22 of these 33 isolates expressing P, F, or both binding specificities. All 22 isolates possessed nucleotide sequences related to the entire set of *pap* probes (*papL*, *papHC*, *papCD*, and *papEFG* probes; Fig. 1); the *papHC* and *papCD* probes detected *Pst*I fragments of 1.0 and 1.7 kilobases (kb), respectively, indicating the conservation of four internal *Pst*I sites (*Pst*I sites 1 to 4; Fig. 1) (2). All three groups (I, III, and IV) that were represented by more than one isolate included both isolates with functional *pap*-related gene clusters and isolates lacking homology to the *papHC*, *papCD*, and *pap EFG* probes. Both types of isolates were detected within cluster N.

The number of gene clusters related to the entire *pap* operon was estimated by determining the number of *Eco*RI fragments detected by the entire set of *pap* probes as previously described (2). Different numbers of copies of a *pap*-related gene cluster were detected among isolates assigned to the same ET. This was the case for three of the six ETs that were represented by more than one isolate (ETs 19, 20, and 24; Fig. 2). A variable copy number of *pap*-related gene clusters was detected in six of the eight clusters that were represented by multiple ETs (Fig. 2).

papΔ1 gene clusters were defined by the absence of the conserved 1.0-kb *Pst*I fragment homologous to the *papCD* probe and the presence of a 0.8-kb fragment homologous to *papCD*. Isolates in this class were detected by the *papL*, *papHC*, *papCD*, and *papEFG* probes, although they expressed neither P nor F adhesin-binding specificities. As previously discussed (2), a small (ca. 0.2-kb) deletion or a point mutation creating a new *Pst*I site can account for these restriction profiles. Changes of this type may also account for the absence of adhesin expression, since mutations in the *papC* or the *papD* gene abolish expression of the P adhesin encoded by the J96 *pap* operon (28). Such cryptic *pap*-related gene clusters were detected in three isolates that were assigned to three different ETs (ETs 1, 4, and 7) distributed among three different clusters (A, C, and D) and two groups (I and II).

papEFG gene clusters were identified among four isolates that were detected only by the *papL* and *papEFG* probes. These isolates expressed neither P nor F adhesin-binding specificities. Such cryptic *pap*-related gene clusters were present in two of the five isolates that constitute cluster C (group I), in the single isolate of cluster E (group II), and in one of the two isolates which constitute cluster F (group III).

Nucleotide sequences related to the *afaI* operon. The distribution of nucleotide sequences related to the *afaI* operon, which encodes an X adhesin in *E. coli* KS52 (20), was studied by colony hybridization with the *afaIC* probe as described previously (3). Nucleotide sequences related to the *afaI* operon were detected in three isolates belonging to three different lineage groups (I, III, and IV) (Fig. 2).

Production of hemolysin. The distribution of nucleotide sequences related to the *hly* operon, which encodes the production of hemolysin in *E. coli* J96 (12), was studied by colony hybridization with an *hlyA* probe (3). All hemolytic isolates were detected by the *hlyA* probe. In addition, one probe-positive isolate (no. 176) was negative for hemolysin production. As found previously in our epidemiologic survey (3), nucleotide sequences related to the *hly* operon were

found only in a subset of the isolates detected by all the *pap* probes. These isolates included the two *pap* probe-positive isolates of cluster C (group I) and all *pap* probe-positive isolates of clusters I, J, L, and M (group IV) (Fig. 2). The *hlyA* probe did not hybridize with DNA from isolates assigned to group II or III.

Expression of type 1 adhesin. The distribution of nucleotide sequences related to the *pil* operon, which encodes a type 1 adhesin in *E. coli* J96 (24), was studied by colony hybridization with a *pilC* probe (3). The *pilC* probe detected all 33 isolates (Fig. 2). Expression of type 1 adhesin was infrequent (five isolates) and not associated with particular lineages.

Serotype. The O:H serotypes of these isolates are indicated in Fig. 2. Isolates of the same ET typically had the same serotype, and isolates representing divergent lineages typically had different O:H serotypes. Some O and H antigen specificities were detected in divergent lineages (e.g., O2 was found in clusters E, I, and K; O15 was found in clusters B, C, and G; H7 was found in clusters K and G).

Genotypic and phenotypic identities among isolates recovered from the same patient. Whereas isolates from 33 different patients were highly diverse, for each of the 23 bacteremic patients, the primary source isolate (20 from the urinary tract and 3 from the biliary tract) and the blood isolate were indistinguishable in the genotypic assays used in this study. In all cases, isolates from different sites of infection within the same patient were identical in (i) multilocus ET; (ii) *Eco*RI *rrn* restriction profiles as detected by the *rrnB7* probe; (iii) *Eco*RI and *Eco*RI-*Pst*I *pap* restriction profiles as detected by the entire set of five *pap* probes (Fig. 1); and (iv) hybridization class, as defined by colony hybridization with probes derived from the *pil*, *afaI*, and *hly* operons.

DISCUSSION

E. coli isolated from extraintestinal infections, are, as a group, different from those present in the commensal intestinal flora. First, bacterial phenotypes shown to mediate virulence are significantly more prevalent among *E. coli* isolates causing upper urinary tract infection than among isolates obtained from fecal specimens of healthy individuals (2, 40). Second, analysis of the genetic structure of *E. coli* populations by various techniques, including MLEE and serotyping, indicates that the majority of isolates that cause upper urinary tract infection represent a limited subset of the lineages found among the commensal flora (35).

Detailed genetic analyses of uropathogenic *E. coli* isolates have identified extensive diversity among chromosomal DNA sequences homologous to the *pap* adhesin pilus operon of strain J96 (2, 16, 17, 32). The first goal of the current study was to correlate the extent of the polymorphisms observed among these virulence-associated sequences with that observed for the operons that encode rRNA, an essential host function not associated with virulence. To place these analyses in the context of population genetics, the phylogenetic relationships among the *E. coli* isolates were estimated by MLEE. In addition, we sought to assess whether related phenotypic and/or genotypic variability developed during the course of in vivo infection or in vitro manipulation.

Distribution of virulence characteristics among the isolates causing urosepsis. The distribution of P, F, and X adhesins among the 20 urosepsis isolates in this study was compared with that found for urinary tract and fecal isolates previously analyzed in our laboratory (3). The urosepsis strains were similar to the 49 *E. coli* isolates from children with pyelonephritis and different from the 48 isolates from children with cystitis and 40 isolates from stool specimens (Table 3). As we have previously discussed (3), these differences may indicate that distinct adhesins facilitate the colonization of the upper and lower urinary tract.

TABLE 3. Virulence properties of isolates causing upper urinary tract infection

Virulence property	No. of isolates/total (%) with virulence property			
	Urosepsis	Pyelonephritis[a]	Cystitis[a]	Fecal[a]
Expression of P and F adhesin-binding specificities	16/20 (80)	32/49 (65)	11/48 (23)	6/40 (15)
Presence of *hly*-related sequences among *pap* probe-positive isolates	8/16 (50)	21/37 (57)	9/20 (45)	12/19 (63)
Expression of type 1 adhesin among *pil* probe-positive isolates	3/20 (15)	10/48 (21)	12/44 (27)	23/34 (68)
Presence of nucleotide sequences related to the *afaI* operon	1/20 (5)	13/49 (27)	24/48 (50)	7/40 (18)

[a] Results for these three groups of isolates have been previously described and discussed (2).

The pathogenic role of P adhesins has been assessed in an animal model of upper urinary tract infection inoculated with isogenic *E. coli* strains constructed in vitro. Expression of the P adhesin encoded by the *pap* operon of J96 significantly increased infectivity (29). Subsequent epidemiologic surveys of clinical isolates suggested that coexpression of P- and F-binding specificities encoded by *pap*-related sequences is critical in the pathogenesis of upper urinary tract infection (Table 3) (3). However, because 20% of upper urinary tract infections were due to isolates that lacked *pap*-related operons and did not express P or F adhesins (3), these observations do not rule out a role for other, as yet undefined, minor classes of adhesins.

The antigenic specificities and the distribution of the O:H serotypes among these isolates (Fig. 2) were similar to those previously observed for uropathogenic strains (7, 32). As previously described (8), the same O-antigen specificities were detected in divergent lineages, suggesting that these epitopes may be under selective pressure and subject to convergent evolution.

Population structure of isolates causing urosepsis. There are clear similarities in the structure of *E. coli* populations causing urosepsis (this report), pyelonephritis (7), or sepsis and meningitis in children (36). In the current study, although distantly related isolates were found to cause urosepsis, the majority (14 [70%] of 20) were relatively closely related, representing only four clusters (I, K, L, and N) within a single phylogenetic group (IV) (Fig. 2). If only enzyme loci common to this and previous studies are considered, estimates of genetic diversity among isolates causing urosepsis (this report), pyelonephritis (7, 35), and neonatal meningitis (36) are similar (data not shown). In contrast, estimates of genetic diversity among *E. coli* from stool specimens are significantly greater (35). Thus, urosepsis tends to be caused by isolates representing a limited set of related lineages. This perspective is in accord with a "clonal" concept of bacterial populations (1, 30).

Evolutionary divergence detected by *rrn* restriction profiles. MLEE and RFLP analyses detect different types of nucleotide sequence variations. MLEE specifically detects variations in the coding sequences that alter the electrophoretic mobility of the corresponding protein. The RFLPs detected in Southern blots may reflect either (i) nucleotide substitutions directly at the endonuclease restriction sites, which may be located in coding or noncoding sequences, or (ii) inversions, deletions, and insertions that alter the length of the DNA sequences between conserved restriction sites. Overall, MLEE defined a slightly larger number of distinct lineages than did the *rrn* RFLP studies and was therefore somewhat more sensitive to divergence.

Despite the inherent differences between MLEE and RFLP analyses, our results directly demonstrated an unexpectedly strong correlation between the *Eco*RI restriction polymorphisms associated with the *rrn* operons and the allelic variations detected by MLEE. Isolates representing the same ET consistently had the same restriction profile. Isolates representing different but closely related lineages (genetic distance, <0.3) occasionally had identical profiles, whereas distinct restriction profiles were observed for all pairs of more divergent isolates. Moreover, the genetic alterations detectable as allelic variants by MLEE and those resolved as restriction polymorphisms by Southern blot analyses occurred at strikingly similar rates. We therefore conclude that the RFLPs associated with the *rrn* operons, like the electrophoretic allelic variations in the metabolic enzymes, appear to be minimally subject to convergence and thus reflect divergent evolution.

Genetic variation and stability of virulence factors within and between lineages. In contrast to the strong concordance between the *rrn* restriction profiles and the lineages defined by MLEE, there was remarkable genetic diversity among the virulence adhesin operons of the isolates representing related lineages. There were isolates assigned to same phylogenetic cluster that were heterogeneous with respect to either the copy number of the type of *pap*-related genetic elements (in particular, cluster C) or to the presence of *afaI*-related sequences (Fig. 2). In particular, some isolates that were of the same ET and exhibited identical EcoRI *rrn* restriction profiles differed in the number of copies of *pap*-related gene clusters (Fig. 2).

There was also phylogenetic evidence indicating that the DNA polymorphisms associated with virulence operons are stably propagated. Five of six isolates assigned to the closely related ET 18 and ET 19 had identical *pap* restriction profiles. The complex restriction profiles of isolates assigned to ET 5 were also identical (Fig. 2), as were the profiles of the two isolates representing ET 17. Similarly, there was consistency in the distribution of hemolysin, with all but one cluster being homogeneous for the presence or absence of *hly*-related sequences. Clinical evidence likewise supported a model involving the stable propagation of virulence operon polymorphisms. For the 23 episodes of urinary and biliary tract sepsis, primary source and blood isolates from the same patient consistently had identical *pap*-related restriction profiles and the same phenotypic and genotypic pattern of virulence factors.

Taken as a whole, the results of this study begin to define the evolutionary context for the generation of the genetic diversity observed for adhesin virulence factors. The genetic alterations underlying the structural and functional diversity of the *pap*-related virulence operons are clearly occurring at an appreciably faster rate than changes that result either in the allelic variation of the essential chromosomal enzymes or in the restriction polymorphisms associated with the rRNA operons. On the other hand, variations detectable by detailed Southern blot analyses or phenotypic assays are not occurring either during the translocation of *E. coli* from the primary site of infection into the blood stream or during the laboratory manipulations required for these studies. Further, in some instances, virulence-associated properties have been stably maintained among phylogenetically related *E. coli* isolated from epidemiologically unrelated patients. Thus, our findings indicate that *pap*-related adhesin virulence factors can be stably inherited over relatively short periods of time, although they are often involved in DNA alterations occurring relatively recently in the evolution of the species.

Possible mechanisms generating genetic diversity among chromosomally encoded virulence factors. The particular types of genetic events that have generated the diversity and distribution of virulence factors observed among *E. coli* lineages cannot be identified by the approach used in this study. For example, the *pap*-related restriction profiles observed among the isolates representing group IV could be accounted for by the presence of a *pap*-related operon in the ancestor to those lineages and subsequent duplication and/or deletion of all or part of the *pap*-related sequences. Alternatively, members of these and other unrelated lineages may have acquired *pap*-related gene clusters by horizontal gene transfer. Similar alternative models have been proposed to account for the diversity and distribution of the RFLPs associated with the chromosomal loci encoding the capsular polysaccharide virulence factor of *Haemophilus influenzae* (26).

It has been hypothesized that *pap*-related sequences may be (or may have been) on transposons and are thereby transferred directly between different sites on the same chromosome or onto plasmid or phage vectors and then to the chromosome of a different isolate (32). Southern blot analyses of plasmids from other uropathogenic isolates failed to demonstrate plasmid carriage of *pap*-related operons (18). To our knowledge, there has been no direct experimental demonstration of the horizontal transfer of *pap*-related gene clusters. The spontaneous in vitro loss of hemolysin and mannose-resistant hemagglutination has been reported (13), as has the in vitro transfer of mannose-resistant hemagglutination between strains (14). However, in both of these reports, the underlying mechanism was not established; the receptor-binding specificity of the adhesin was not further characterized, and the presence of nucleotide sequences homologous to *pap* was not determined.

In our previous analysis of the polymorphisms associated with these virulence-associated operons (2), we emphasized the potential role of inter- and intrachromosomal recombination-based DNA rearrangements in generating such diversity. We continue to favor this mechanism. The results in the current report further indicate that such events occur neither in vivo during the course of acute infection nor in vitro during laboratory manipulations. Rather, based on previous observations that *E. coli* isolated from the feces commonly carry chromosomally encoded *pap*-related gene clusters (3), the recombinational events hypothesized to generate these polymorphisms are likely occurring within the intestinal reservoir.

Recently, Milkman and co-workers determined, for multiple *E. coli* isolates, extensive DNA sequence variations in the region of the *trp* operon (25, 39). Their results suggest that recombination of genomic segments between independent strains may be more common than previously appreci-

ated. Additional DNA sequence data and experiments with labeled *pap*-related operons for in vivo transfer studies will be required to establish an analogous role for recombination in the movement of virulence factors among pathogenic isolates.

ACKNOWLEDGMENTS

We are grateful to Catherine Squires, Noël Smith, Candice Johnson, Bernard Davis, Staffan Normark, Agnès Labigne-Roussel, David Low, Bruce Levin, Catherina Svanborg-Eden, Joel Maslow, Peter Rice, Laurent Gutmann, Kurt Stottmeier, Daniel Bernstein, Shiela Hull, Richard Hull, Roy Curtiss, Larry Gilbert, Heather Goldstein, and Howard Corwin for helpful discussions, encouragement, or gifts of bacterial strains and isolates. Manju Agarwal, Ilga Wohlrab, and Raminder Makkar 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 National Institutes of Health Biomedial Research Support Grants funding at Boston University School of Medicine and Public Health. M.A. was supported by a fellowship from the National Kidney Foundation. R.D.A. was supported by the Research Service of the Veterans Administration. S.S. was supported in part by funding from the Cystic Fibrosis Foundation, the Hood Foundation, and by National Institutes of Health Biomedical Research Support Grants funding from Boston City Hospital. R.K.S. and P.B. were supported by Public Health Service research grant AI 22144 from the National Institutes of Health.

LITERATURE CITED

1. **Achtman, M., and G. Pluschke.** 1986. Clonal analysis of descent and virulence among selected *Escherichia coli*. Annu. Rev. Microbiol. **40**:185–210.
2. **Arthur, M., C. Campanelli, R. D. Arbeit, C. Kim, S. Steinbach, C. E. Johnson, R. H. 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.
3. **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.
4. **Bachmann, B. J.** 1987. Linkage map of *Escherichia coli* K-12, edition 7., p. 807–876. *In* F. C. Neidhardt, K. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
5. **Berg, K. L., C. L. Squires, and C. Squires.** 1987. In vivo translation of a region within the *rrnB* 16S rRNA gene of *Escherichia coli*. J. Bacteriol. **169**:1691–1701.
6. **Brosius, J., A. Ullrich, M. A. Raker, A. Gray, T. J. Dull, R. R. Gutell, and H. F. Noller.** 1981. Construction and fine mapping of recombinant plasmids containing the *rrnB* ribosomal RNA operon of *E. coli*. Plasmid **6**:112–118.
7. **Caugant, D. A., B. R. Levin, G. Lidin-Janson, T. S. Whittam, C. Svanborg-Edén, and R. K. Selander.** 1983. Genetic diversity and relationships among strains of *Escherichia coli* in the intestine and those causing urinary tract infections. Prog. Allergy. **33**: 203–227.
8. **Caugant, D. A., B. R. Levin, I. Ørskov, F. Ørskov, C. Svanborg-Eden, and R. K. Selander.** 1985. Genetic diversity in relation to serotype in *Escherichia coli*. Infect. Immun. **49**:407–413.
9. **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.
10. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132**:6–13.
11. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267. (Addendum.)
12. **Felmee, T., S. Pellett, and R. A. Welch.** 1985. Nucleotide sequence of an *Escherichia coli* chromosomal hemolysin. J. Bacteriol. **163**:94–105.
13. **Hacker, J., S. Knapp, and W. Goebel.** 1983. Spontaneous deletions and flanking regions of the chromosomally inherited hemolysin determinant of an *Escherichia coli* O6 strain. J. Bacteriol. **154**:1145–1152.
14. **Hales, B. A., and S. G. B. Amyes.** 1986. The transfer of genes encoding production of mannose-resistant haemagglutination fimbriae from uropathogenic enterobacteria. J. Gen. Microbiol. **132**:2243–2247.
15. **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.
16. **Hull, S., S. Clegg, C. Svanborg-Edén, and R. Hull.** 1985. Multiple copies of genes in pyelonephritogenic *Escherichia coli* encoding adhesins binding globoseries glycolipid receptors. Infect. Immun. **47**:80–83.
17. **Hull, S. I., S. Beiler, and R. A. Hull.** 1988. Restriction fragment length polymorphism and multiple copies of DNA sequences homologous with probes for P-fimbriae and hemolysin genes among uropathogenic *Escherichia coli*. Can. J. Microbiol. **34**: 307–311.
18. **Johnson, J. R., S. L. Moseley, P. L. Roberts, and W. E. Stamm.** 1988. Aerobactin and other virulence factor genes among strains of *Escherichia coli* causing urosepsis: association with patient characteristics. Infect. Immun. **56**:405–412.
19. **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.
20. **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.
21. **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.
22. **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.
23. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
24. **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.
25. **Milkman, R., and A. Stoltzfus.** 1988. Molecular evolution of the *Escherichia coli* chromosome. II. Clonal segments. Genetics **120**:359–366.
26. **Musser, J. M., J. S. Kroll, E. R. Moxon, and R. K. Selander.** 1988. Clonal population structure of encapsulated *Haemophilus influenzae*. Infect. Immun. **56**:1837–1845.
27. **Nei, M.** 1978. Estimation of average heterozygosity and genetic distance from a small sample of individuals. Genetics **89**: 583–590.
28. **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.
29. **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. J. Clin. Invest. **75**:347–360.
30. **Ørskov, F., and I. Ørskov.** 1983. Summary of a workshop on the clone concept in epidemiology, taxonomy, and evolution of the

Enterobacteriacae and other bacteria. J. Infect. Dis. **148**:346–357.

31. **Ørskov, F., and I. Ørskov.** 1984. Serotyping of *Escherichia coli*, p. 43–112. *In* T. Bergan (ed.), Methods in microbiology. Academic Press, Inc., London.
32. **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.
33. **Rubin, R. H., N. E. Tolkoff-Rubin, and R. S. Cotran.** 1986. Urinary tract infection, pyelonephritis, and reflux nephropathy, p. 1058–1141. *In* B. M. Brenner and F. C. Rector, Jr. (ed.), The kidney, 3rd ed. The W. B. Saunders Co., Philadelphia.
34. **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.
35. **Selander, R. K., D. A. Caugant, and T. S. Whittam.** 1987. Genetic structure and variation in natural populations of *Escherichia coli*, p. 1625–1648. *In* F. C. Neidhardt, K. L. Ingraham, B. Magasanik, K. B. Low, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella typhimurium*: cellular and molecular biology. American Society for Microbiology, Washington, D.C.
36. **Selander, R. K., T. K. Korhonen, V. Vaisanen-Rhen, P. H. Williams, P. E. Pattison, and D. A. Caugant.** 1986. Genetic relationships and clonal structure of strains of *Escherichia coli* causing neonatal septicemia and meningitis. Infect. Immun. **52**:213–222.
37. **Sneath, P. H. A., and R. R. Sokal.** 1978. Numerical taxonomy. W. H. Freeman & Co., San Francisco.
38. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
39. **Stolzfus, A., J. F. Leslie, and R. Milkman.** 1988. Molecular evolution of the *Escherichia coli* chromosome. I. Analysis of structure and natural variation in a previously uncharacterized region between *trp* and *tonB*. Genetics **120**:345–358.
40. **Svanborg-Edén, C., and P. de Man.** 1987. Bacterial virulence in urinary tract infection. Infect. Dis. Clin. N. Am. **1**:731–750.