Ongoing Horizontal and Vertical Transmission of Virulence Genes and *papA* Alleles among *Escherichia coli* Blood Isolates from Patients with Diverse-Source Bacteremia

JAMES R. JOHNSON,^{1,2*} TIMOTHY T. O'BRYAN,^{1,2} MICHAEL KUSKOWSKI,^{3,4} AND JOEL N. MASLOW^{5,6}

Medical Service¹ and Geriatric Research, Education, and Clinical Center,³ Minneapolis VA Medical Center, and Departments of Medicine² and Psychiatry,⁴ University of Minnesota, Minneapolis, Minnesota, and Medical Service, Philadelphia VA Medical Center,⁵ and Department of Medicine, University of Pennsylvania,⁶ Philadelphia, Pennsylvania

Received 9 February 2001/Returned for modification 4 April 2001/Accepted 31 May 2001

The phylogenetic distributions of multiple putative virulence factors (VFs) and *papA* (P fimbrial structural subunit) alleles among 182 *Escherichia coli* blood isolates from patients with diverse-source bacteremia were defined. Phylogenetic correspondence among these strains, the *E. coli* Reference (ECOR) collection, and other collections of extraintestinal pathogenic *E. coli* (ExPEC) was assessed. Although among the 182 bacteremia isolates phylogenetic group B2 predominated, exhibited the greatest concentration of individual VFs, and contained the largest number of familiar virulent clones, other phylogenetic groups exhibited greater concentrations of certain VFs than did group B2 and included several additional virulent clones. Certain of the newly detected VF genes, e.g., *fyuA* (yersiniabactin; 76%) and *focG* (F1C fimbriae; 25%), were as prevalent or more prevalent than their more familiar traditional counterparts, e.g., *iut* (aerobactin; 57%) and *sfaS* (S fimbriae; 14%), thus possibly offering additional useful targets for preventive interventions. Considerable diversity of VF profiles was observed at every level within the phylogenetic tree, including even within individual lineages. This suggested that many different pathways can lead to extraintestinal virulence in *E. coli* and that the evolution of ExPEC, which involves extensive horizontal transmission of VFs and continuous remodeling of pathogenicity-associated islands, is a highly active, ongoing process.

The strains of *Escherichia coli* that cause extraintestinal infections such as urinary tract infection (UTI), meningitis, and bacteremia are distinct both from most intestinal commensal *E. coli* types and from diarrheagenic *E. coli* types (13, 57, 65, 70). These specialized extraintestinal pathogenic *E. coli* (ExPEC) strains (65) are thought to derive primarily from *E. coli* phylogenetic group B2 (as defined within the *E. coli* Reference [ECOR] collection by multilocus enzyme electrophoresis [MLEE]) (20, 54) and to acquire their unique pathogenicity from their distinctive virulence factors (VFs) (4, 10, 59, 60).

Putative VFs of ExPEC include diverse adhesins, toxins, polysaccharide coatings (including capsules and lipopolysaccharides), siderophores, serum resistance mechanisms, and invasins (21, 33, 70). Such VFs help the organisms colonize host surfaces, avoid and/or subvert host defense mechanisms, injure and/or invade host cells and tissues, and incite a noxious inflammatory response, thereby giving rise to clinical disease (12, 21, 70). The VF genes of ExPEC are thought to be primarily inherited vertically within evolutionary lineages but also to be transferred horizontally between lineages, in some instances on plasmids or on “pathogenicity-associated islands” (PAIs), which are gene blocks that contain multiple contiguous VF genes (2, 10, 16, 18, 36, 50, 61). Better understandings of the prevalence and evolutionary origins of the VFs of ExPEC and of the distinctive “virulent clones” that make up the ExPEC

population should hasten the development of the preventive measures that are sorely needed against these morbid and costly infections (42, 43, 47, 55, 72).

Recent methodological and epidemiological developments now permit the ready detection of a broad array of putative VFs of ExPEC (23, 27, 30, 33) and multiple F antigen-specific alleles of P fimbrial structural subunit gene *papA* (25, 32, 35), all of which are potential candidates for VF-specific preventive interventions such as vaccines (42, 43, 47, 55). We previously applied these developments to the analysis of a collection of urosepsis isolates from Seattle, Wash. (30, 33, 35). We undertook the present study to gain an understanding of the prevalence and phylogenetic distribution of such traits among diverse-source *E. coli* bacteremia isolates. Specifically, we analyzed a large collection of *E. coli* blood isolates from patients with diverse-source bacteremia within which phylogenetic relationships were previously established by MLEE and for which O:H serotypes and selected VF markers have been reported (22, 50). We sought to characterize this population with respect to the prevalence and phylogenetic distribution of a broad range of “newer” VFs and of the 12 recognized *papA* alleles and to correlate the phylogenetic structure of this population with that of the ECOR collection and other collections of ExPEC from the literature.

MATERIALS AND METHODS

Strains. The study population consisted of 182 of 187 previously published diverse-source *E. coli* bacteremia isolates from hospitalized or community-dwelling adults (50). (Five of the original 187 isolates were unavailable for analysis.) These strains were previously characterized with respect to genotype for *papHC* (P fimbriae), *papEFG* and *papG* (P fimbrial adhesin) alleles, *hly* (hemolysin), *sfa*

* Corresponding author. Mailing address: Infectious Diseases (111F), VA Medical Center, One Veterans Dr., Minneapolis, MN 55417. Phone: (612) 725-2000, ext. 4185. Fax: (612) 727-5995. E-mail: johns007@tc.umn.edu.

and/or *foc* (S and F1C fimbriae), *iut* (aerobactin system), *bma* (M fimbriae), *nfa* (nonfimbrial adhesins), and *afa* and/or *dra* (afimbrial and fimbrial Dr antigen-binding adhesins); pulsed-field gel electrophoresis type; mannose-resistant hemagglutination and hemolysin phenotypes; O:H serotype; ribotype; and MLEE type (22, 50). Members of the ECOR collection (54) were provided by Howard Ochman and the American Type Culture Collection (Manassas, Va.). Strains were stored in 20% glycerol at -70°C until use.

Virulence genotypes. In the present study the 182 bacteremia isolates were tested for 15 putative VF genes of ExPEC for which they had not previously been tested. Detection was by dot blot hybridization under stringent conditions, as previously described (27, 28, 33). Probes were generated and digoxigenin labeled using primers as previously described (28, 33). The VF genes encompassed five categories: adhesin, toxin, siderophore, capsule, and miscellaneous genes. The adhesin genes investigated were *papAH* (P fimbrial structural subunit), *sfaS* (S fimbrial adhesin), *focG* (the putative F1C fimbrial adhesin), and *iha* (iron-regulated gene homologue adhesin; a gene encoding a recently described putative adhesin found in ExPEC and *E. coli* O157:H7) (30, 74). The siderophore genes studied were *fyuA* (yersiniabactin) and *iroN* (a novel putative catecholate siderophore). The toxin genes studied were *cnfI* (cytotoxic necrotizing factor) and *cdtB* (cytolethal distending toxin). The capsule genes studied were *kpsM_{II}T_{II}* and *kpsM_{III}T_{III}* (group II and group III capsular polysaccharide synthesis, respectively). The miscellaneous VF genes were *cvaC* (colicin V; multifunctional serum resistance-associated plasmids), *traT* (serum resistance associated), *ibeA* (invasion of brain endothelium), and *malX*, a marker for a PAI from archetypal ExPEC strain CFT073 (18, 33, 36).

Strains also were tested by PCR for three VF gene regions, including *kpsM_{K1}T_{K1}* (K1 capsule synthesis), *kpsM_{II}T_{II}*, and *sfaS*, using primers and PCR conditions as previously described (33). Strains that were positive for *kpsM_{II}T_{II}* by dot blot hybridization but not by PCR were considered to be K2 capsule positive (33).

Strains that were positive for any *pap* element were tested for 12 alleles of *papA*, corresponding to the 11 established P fimbrial F types (F7-1, F7-2, and F8 to F16) plus the recently discovered F48 *papA* variant, by using a multiplex PCR-based assay, as previously described (23, 35).

Sequence determination of novel *papA*. The *papA* DNA sequence was determined as previously described (35) for the three strains which were positive for *papAH* by blotting and PCR but negative for a defined *papA* allele in the allele-specific *papA* assay. Predicted PapA peptides were aligned with reference PapA sequences for traditional F types F7-1, F7-2, and F8 to F16 plus the recently described F48 *papA* variant (35) and PapA from strain 536 (25, 32) by using CLUSTAL-W (75). An unrooted tree was inferred from the resulting PapA similarity matrix according to the NJ method (66) by using the application MEGA (40).

Phylogenetic comparison to the ECOR collection. To relate the known phylogenetic structure of the bacteremia isolates (as previously defined by MLEE) (50) to the known phylogenetic structure of the ECOR collection, a subset of the 182 strains representing each of the four major phylogenetic divisions of the bacteremia collection (i.e., clusters I, III, IV, and V) were categorized as to ECOR phylogenetic group based on random amplified polymorphic DNA (RAPD) analysis (3, 24, 26, 32, 34, 77). Fingerprints were generated in duplicate for each strain using (separately) arbitrary decamer primers 1247, 1254, and 1283 (3). Amplification conditions were as described previously (3) except that commercial PCR beads were used (3, 24, 26, 32, 34, 77). Each group of fingerprints from the three primers was combined in series to create a "virtual" composite fingerprint. Comparable fingerprints were generated in parallel for 10 ECOR control strains, including two from each of the five major ECOR phylogenetic groups as defined by Herzer et al., based on electrophoretic mobility polymorphisms (20). Each bacteremia isolate was inferred to derive from the phylogenetic group of the ECOR strain with which it exhibited the highest Pearson correlation coefficient in pair-wise comparisons based on analog densitometric scans of composite RAPD fingerprints (24, 26, 27, 32, 34).

Cluster analysis of VF profiles. For a subset of the bacteremia strains (i.e., those in electrophoretic type [ET] 47), a "tree" based on extended VF profiles was constructed according to the unweighted pair group method with averaging (UPGMA) (71) by using the application NT-SYSPc (Exeter Software, Setauket, N.Y.).

Statistical methods. Comparisons of proportions were tested using Fisher's exact test (two tailed). Comparisons of the prevalences of different traits within the same population were made using McNemar's test (15). A VF score was calculated for each strain as the sum of that for all VFs for which the strain was positive, with values for the four *pap* regions (*papA*, *papHC*, *papEFG*, and *papG*) and the three *sfa* and/or *foc* elements (*sfa* and/or *foc*, *sfaS*, and *focG*) proportionally adjusted to account for multiple detection of the same operon. Aggre-

gate VF scores for groups were compared using the Mann-Whitney U test. Correlations between VFs were tested using the phi coefficient, a chi-square-based measure of association for 2 by 2 tables. Because of multiple comparison, the threshold for statistical significance was a *P* value of <0.01 , with a *P* value of <0.05 considered to reflect borderline statistical significance.

Nucleotide sequence accession numbers. Newly determined *papA* sequences for strains CA033, CA039, and BOS020 were deposited in GenBank under accession no. AF332518, AF332519, and AF332520, respectively.

RESULTS

Prevalence of VFs and *papA* alleles. All of the newly assessed VFs were detected at least once in the population (Table 1). Among the adhesin genes, the recently described *iha*, although significantly less prevalent than *papA* or *papG* ($P < 0.001$; McNemar's test), was as prevalent as *sfa* and/or *foc* and was significantly more prevalent than the other mannose-resistant adhesin genes assessed, i.e., *afa* and/or *dra*, *nfa*, and *bma* ($P < 0.001$ for all comparisons; McNemar's test). *focG* was significantly more prevalent than the better-known sialosyl-binding adhesin gene *sfaS* ($P = 0.008$; McNemar's test). Although *hly* was the most prevalent toxin gene, *cnfI* and diarrhea-associated *cdtB* also were appreciably prevalent. Among the siderophore genes, *fyuA* (yersiniabactin), which only recently has been shown to occur in *E. coli* (33, 67), was significantly more prevalent than *iut* (aerobactin; $P < 0.001$; McNemar's test), and the prevalence of recently described *iroN* approached that of *iut*. Nearly 80% of the population exhibited group II capsule synthesis genes (*kpsM_{II}T_{II}*), with the K1 and the K2 variants accounting for 14 and 25% of *kpsM_{II}T_{II}*-positive strains, respectively. Group III capsules were uncommon. Meningitis-associated *ibeA* occurred in 12% of strains. The PAI marker gene *malX*, present in 71% of strains, was the second most prevalent trait after *kpsM_{II}T_{II}* (Table 1).

All 12 established *papA* alleles were detected at least once each in the population (Table 1). The recently described F48 *papA* variant (35) was the second most prevalent *papA* allele overall ($n = 20$), after the F10 variant ($n = 45$) (Table 1). Six strains were positive for *papA* by probe but negative for a defined *papA* allele in the F type-specific *papA* allele PCR assay. From five of these, full-length *papA* amplicons could be generated. Two of the *papA* amplicons repeatedly yielded uninterpretable sequence data; both were from strains known to have two copies of *pap* (not shown). Of the three amplicons that yielded interpretable sequence data, two proved to represent variants of the F11 *papA* allele. The third amplicon, from an isolate in ET 42, was distant from the 12 *papA* alleles included in the *papA* allele PCR assay but had 99.3% nucleotide identity with *papA* from archetypal pyelonephritis isolate 536 (25, 32).

Phylogenetic distribution of VFs. Projection of the VF and *papA* allele data onto the MLEE-based phylogenetic tree revealed diverse VF-specific patterns of distribution, which suggested both vertical and horizontal transmission of VFs, plus selective deletion of VFs from certain lineages (Fig. 1 and 2). Vertical inheritance of VFs within lineages was suggested by the significant concentration of specific VFs within particular phylogenetic groups, as was documented statistically at multiple hierarchical levels within the tree (Tables 1, 2, and 3). At the most basal level of the tree, i.e., that of major clusters (I to V), most VFs and *papA* alleles were significantly concentrated

TABLE 1. Distribution of VF genes in relation to phylogenetic cluster among 182 *E. coli* bacteremia isolates

Associated trait	No. (% of total)	Prevalence of associated trait (no. [%]) for phylogenetic group cluster ^a :			
		I (n = 21)	III (n = 119)	IV (n = 19)	V (n = 22)
<i>papA</i>	118 (65)	3 (14) ^d	96 (81) ^d	3 (16) ^d	16 (73)
F7-1	3 (2)	0	2 (2)	0	1 (5)
F7-2	14 (8)	0	8 (7)	0	6 (27) ^c
F8	7 (4)	0	5 (4)	0	2 (9)
F9	3 (2)	0	2 (2)	0	1 (5)
F10	45 (25)	1 (5) ^b	38 (32) ^c	0 ^c	6 (27)
F11	17 (9)	0	16 (13) ^c	1 (5)	0
F12	5 (3)	1 (5)	3 (3)	1 (5)	0
F13	9 (5)	0	9 (8) ^b	0	0
F14	12 (7)	0	12 (10) ^c	0	0
F15	3 (2)	0	3 (3)	0	0
F16	9 (5)	0	7 (6)	0	2 (9)
F48	20 (11)	1 (5)	18 (15) ^b	0	1 (5)
<i>papG</i>	101 (55)	1 (5) ^d	87 (73) ^d	2 (11) ^d	11 (50)
Allele II	58 (32)	0 ^d	45 (38) ^b	2 (11) ^b	11 (50)
Allele III	32 (18)	0	31 (26) ^d	0 ^b	0 ^b
Allele II + III	9 (5)	0	9 (8) ^b	0	0
Allele I + III	2 (1)	0	2 (2)	0	0
<i>sfa</i> and/or <i>foc</i>	66 (36)	0 ^d	66 (55) ^d	0 ^d	0 ^c
<i>sfaS</i>	25 (14)	0	25 (21) ^d	0	0 ^d
<i>focG</i>	45 (25)	1 (5) ^b	44 (37) ^d	0 ^c	0 ^c
<i>afa</i> and/or <i>dra</i>	15 (8)	6 (29) ^c	6 (5) ^b	0	3 (14)
<i>iha</i>	69 (38)	2 (10) ^c	50 (42)	2 (11) ^b	15 (68) ^c
<i>bma</i>	3 (2)	1 (5)	1 (1)	0	1 (5)
<i>nfa</i>	15 (8)	2 (10)	8 (7)	4 (21)	1 (5)
<i>hly</i>	82 (45)	0 ^d	75 (63) ^d	0 ^d	7 (32)
<i>cnfI</i>	50 (27)	1 (5) ^c	49 (41) ^d	0 ^c	0 ^{c,f}
<i>cdtB</i>	9 (5)	0	9 (8) ^b	1 (5)	0
<i>iroN</i>	83 (46)	2 (10) ^d	74 (62) ^d	4 (21) ^b	3 (14) ^c
<i>fyuA</i>	139 (76)	5 (24) ^d	114 (96) ^d	3 (14) ^d	17 (77)
<i>iut</i>	103 (57)	11 (52)	63 (53)	11 (52)	18 (82) ^b
<i>kpsM_{II}T_{II}</i>	144 (79)	10 (48) ^c	107 (90) ^d	7 (33) ^d	19 (82)
K1	36 (20)	2 (10)	29 (24) ^b	1 (5)	4 (18)
K2 ^e	20 (11)	2 (10)	10 (8)	0	8 (36) ^c
<i>kpsM_{III}T_{III}</i>	9 (5)	2 (10)	6 (5)	0	1 (5)
<i>ibeA</i>	21 (12)	1 (5)	19 (16) ^b	0	1 (5)
<i>cvaC</i>	12 (7)	0	10 (8)	1 (5)	1 (5)
<i>traT</i>	52 (29)	6 (29)	31 (26)	6 (29)	9 (41)
<i>PAI</i>	130 (71)	2 (10) ^d	113 (95) ^d	3 (14) ^d	12 (55)

^a Data for cluster II not shown (n = 1). Data for *papG* alleles II and III. *sfa* and/or *foc*, *afa* and/or *dra*, *hly*, *jut*, *nfa*, and *bma* are as reported elsewhere (20, 45).

^b $P < 0.05$, indicated group versus all other strains.

^c $P \leq 0.01$, indicated group versus all other strains.

^d $P \leq 0.001$, indicated group versus all other strains.

^e Strains were considered K2 positive if positive for *kpsM_{II}T_{II}* by blot but not by PCR.

within the most populous cluster, cluster III. Exceptions included *afa* and/or *dra* (associated with cluster I) and the F7-2 *papA* allele, *iha*, *iutA*, and the K2 *kpsMT* variant (all associated with cluster V) (Table 1). Aggregate VF scores were highest in cluster III (median, 8.0; $P < 0.001$ versus all other strains and versus cluster I, IV, or V). They were intermediate in cluster V (median, 6.0; $P > 0.10$ versus all other strains and $P < 0.001$ versus cluster I or IV), and lowest in clusters I and IV (median, 2.0; $P < 0.001$ versus all other strains; $P > 0.10$ for cluster I versus cluster IV).

A nonrandom phylogenetic distribution of VFs was apparent also at the level of the major subclusters (A to D) within cluster III (Table 2). Here, however, no consistent pattern as to which subcluster exhibited the greatest prevalence of specific traits emerged (Table 2). Similar findings were also made at the level of the four groups of individual lineages that made up

subcluster C, the most populous subcluster within cluster III (Table 3).

Exceptionally, particular VFs or *papA* alleles exhibited an "all or none" pattern of phylogenetic distribution, either occurring in all members of a phylogenetic group or being entirely absent from the group. Most traits, however, if present at all occurred in only a fraction of the members of any particular phylogenetic group, even if that group exhibited a significantly higher prevalence of the trait than did comparable groups at the same hierarchical level in the tree (Fig. 1 and 2; Tables 1 to 3). This phenomenon could be explained by selective gene deletion from members of an ancestrally VF-positive group and/or by horizontal entry of the trait into multiple smaller branches of the tree in the absence of a VF-positive common ancestor. Strongly favoring horizontal transfer in some instances was the markedly discontinuous distribution of certain

FIG. 2. Phylogenetic distribution of 12 F antigen-specific *papA* alleles (F types) among 182 *E. coli* bacteremia isolates. The MLEE-based tree is taken from Maslow et al. (50). Clusters I to V are labeled. The four major subclusters within cluster III, i.e., subclusters A to D, are in boldface. Data reflect the number of isolates within each lineage (Lin.) positive for the indicated *papA* allele. Minus signs, absence of positivity for the indicated allele; n, number of evaluable isolates in the lineage. (In addition to the data shown, strain BOS020 from lineage 42 was positive for the F536 *papA* allele according to comparative sequence analysis.)

TABLE 2. Phylogenetic distribution of VF genes within cluster III among 119 *E. coli* bacteremia isolates^a

Associated trait	No. (% of 119)	Prevalence of associated trait (no. [%]) for subcluster:			
		A (n = 8)	B (n = 24)	C (n = 62)	D (n = 24)
<i>papA</i>	96 (81)	2 (25) ^d	19 (79)	56 (90) ^c	19 (79)
F7-1	2 (2)	0	1 (4)	1 (2)	0
F7-2	8 (7)	0	0	8 (13) ^c	0
F8	5 (4)	0	0	0 ^b	5 (21) ^d
F9	2 (2)	0	2 (8) ^b	0	0
F10	38 (32)	0	5 (21)	22 (35)	11 (46)
F11	16 (14)	0	4 (17)	5 (8)	7 (29) ^b
F12	3 (3)	0	1 (4)	2 (3)	0
F13	9 (8)	0	5 (21) ^b	3 (5)	1 (4)
F14	12 (10)	0	6 (25) ^b	6 (10)	0
F15	3 (3)	0	1 (4)	2 (3)	0
F16	7 (6)	0	5 (21) ^c	2 (3)	0
F48	18 (15)	1 (13)	0 ^b	16 (26) ^d	1 (4)
<i>papG</i>	87 (73)	4 (3)	19 (16)	51 (43) ^b	13 (11) ^b
Allele II	45 (38)	3 (38)	11 (46)	19 (31)	12 (50)
Allele III	31 (26)	1 (13)	3 (13)	26 (42) ^d	1 (4) ^c
Allele II + III	9 (8)	0	3 (13)	6 (10)	0
Allele I + III	2 (2)	0	2 (8) ^b	0	0
<i>sfa</i> and/or <i>foc</i>	66 (56)	0 ^d	15 (63)	46 (74) ^d	5 (21) ^d
<i>sfaS</i>	25 (22)	0	3 (13)	20 (32) ^c	2 (8)
<i>focG</i>	44 (37)	0 ^b	11 (46)	29 (47) ^b	4 (17) ^b
<i>afa</i> and/or <i>dra</i>	6 (5)	0	3 (8)	0 ^c	3 (8)
<i>iha</i>	50 (42)	2 (25)	9 (38)	29 (47)	10 (42)
<i>bma</i>	1 (1)	0	1 (4)	0	0
<i>nfa</i>	8 (7)	0	0	6 (10)	2 (8)
<i>hly</i>	75 (64)	1 (13) ^c	20 (83) ^b	48 (77) ^d	6 (25) ^d
<i>cnfI</i>	49 (42)	1 (13)	10 (42)	37 (60) ^d	1 (4) ^d
<i>cdtB</i>	9 (8)	0	1 (4)	8 (13) ^b	0
<i>iroN</i>	74 (63)	4 (50)	18 (75)	44 (71)	8 (33) ^c
<i>fyuA</i>	114 (97)	8 (100)	21 (88)	62 (100) ^b	23 (96)
<i>iut</i>	63 (53)	5 (63)	10 (42)	36 (58)	12 (50)
<i>kpsM_{II}T_{II}</i>	107 (90)	6 (75)	18 (75) ^b	60 (97) ^b	22 (92)
K1	29 (25)	2 (25)	5 (21)	11 (18)	11 (46) ^b
K2 ^e	10 (8)	0	1 (4)	9 (15) ^b	0
<i>kpsM_{III}T_{III}</i>	6 (5)	1 (13)	5 (21) ^d	0 ^c	0
<i>cvaC</i>	10 (9)	2 (25)	1 (4)	5 (8)	1 (4)
<i>traT</i>	31 (26)	5 (63) ^b	5 (21)	14 (23)	7 (29)
<i>ibeA</i>	19 (15)	5 (63) ^c	3 (13)	5 (8) ^b	5 (21)
PAI	114 (96)	8 (100)	23 (96)	62 (100) ^b	20 (83) ^c

^a Data for *papG* alleles, *sfa* and/or *foc*, *afa* and/or *dra*, *hly*, *iut*, *bma*, and *nfa* are reported elsewhere (20, 45). One strain within cluster III fell outside the four subclusters. Thus, subclusters A to D account for only 118 strains, whereas the total in cluster III is 119 strains.

^b $P < 0.05$, indicated group versus all other strains.

^c $P \leq 0.01$, indicated group versus all other strains.

^d $P \leq 0.001$, indicated group versus all other strains.

^e Strains were considered K2 positive if positive for *kpsM_{II}T_{II}* by blot but not PCR.

VFs across the tree, with many intervening VF-negative lineages, a distribution pattern that would require multiple deletion events if not due to horizontal transfer (Fig. 1 and 2; Tables 1 to 3). Even more compelling evidence of horizontal transfer was provided by the distribution of *papA* alleles, since adjacent lineages commonly exhibited different alleles, whereas each shared its particular allele with phylogenetically distant lineages (Fig. 2; Table 3). Moreover, many individual lineages exhibited diverse *papA* alleles among their various members. These phenomena could not be explained by selective deletion alone and instead suggested horizontal exchange of *papA* alleles between lineages, possibly even independent of other *pap* elements.

Ordered diversity of VF genotypes within an individual lineage. Because of limiting numbers, statistical comparisons of aggregate VF data (e.g., Tables 1 to 3) could not be made at the level of the individual lineages (i.e., ETs) within the lineage

groups shown in Table 3, although diversity of VF profiles was still apparent even at this level (Fig. 1 and 2). However, ET 47, the single most populous lineage overall, provided an opportunity to analyze the continuing diversity of VF profiles even at the most differentiated extreme of the phylogenetic tree. The members of ET 47 exhibited seven distinct *papA* alleles and two *papG* allele configurations and were variably positive for most of the other traits detected within this lineage, with only *fyuA*, *kpsM_{II}T_{II}*, and the PAI marker being universally prevalent (Fig. 3). Analysis of the VF and *papA* allele data for ET 47 by UPGMA revealed that these traits were not randomly distributed (Fig. 3). Instead, they segregated in an arborizing pattern that could be explained most parsimoniously by the sequential acquisition or loss of specific markers by separate subclones within this lineage as the subclones progressively differentiated from a common ancestor (Fig. 3).

TABLE 3. Phylogenetic distribution of VF genes within subcluster IIIC among 62 *E. coli* bacteremia isolates^a

Associated trait	No. (% of 62)	Prevalence of associated trait (no. [%]) for lineages:			
		39 and 40 (n = 7)	41–44 (n = 15)	45–48 (n = 24)	49–53 (n = 15)
<i>papA</i>	56 (90)	7 (100)	15 (100)	19 (79) ^b	15 (100)
F7-1	1 (2)	1 (14)	0	0	0
F7-2	8 (13)	0	0	8 (33) ^d	0
F10	22 (35)	5 (71)	0 ^d	7 (29)	10 (67) ^c
F11	5 (9)	1 (14)	0	0	4 (27) ^c
F12	2 (3)	0	0	2 (8)	0
F13	3 (5)	1 (14)	0	2 (8)	0
F14	6 (10)	5 (71) ^d	0	1 (4)	0
F15	2 (3)	1 (14)	0	0	1 (7)
F16	2 (3)	1 (14)	0	1 (4)	0
F48	16 (26)	0	14 (93) ^d	0 ^d	2 (13)
<i>papG</i>	50 (81)	7 (100)	15 (100) ^b	14 (58) ^d	15 (100) ^b
Allele II	19 (31)	1 (14)	0 ^c	10 (42)	8 (53) ^b
Allele III	26 (42)	0 ^b	15 (100) ^d	4 (17) ^c	7 (47)
Allele II + III	6 (10)	6 (86) ^a	0	0	0
<i>sfa</i> and/or <i>foc</i>	46 (74)	5 (71)	13 (87)	17 (71)	11 (73)
<i>sfaS</i>	20 (32)	1 (14)	11 (73) ^d	1 (4) ^d	7 (47)
<i>focG</i>	29 (47)	5 (71)	3 (33) ^b	16 (67) ^b	5 (33)
<i>iha</i>	29 (47)	7 (100) ^c	1 (7) ^d	17 (71) ^c	4 (27)
<i>nfa</i>	6 (10)	0	4 (27) ^b	1 (4)	0
<i>hly</i>	48 (77)	6 (86)	15 (100) ^b	17 (71)	10 (67)
<i>cnfI</i>	37 (60)	6 (86)	15 (100) ^d	9 (38) ^c	7 (47)
<i>cdtB</i>	8 (13)	6 (86) ^d	0	2 (8)	0
<i>iroN</i>	44 (71)	5 (71)	10 (67)	14 (58)	14 (93) ^b
<i>fyuA</i>	62 (100)	7 (100)	15 (100)	24 (100)	15 (100)
<i>iut</i>	36 (58)	7 (100) ^b	0 ^d	20 (83) ^c	9 (60)
<i>kpsM_{II}T_{II}</i>	60 (97)	7 (100)	14 (93)	24 (100)	15 (100)
K1	11 (18)	0	2 (13)	1 (4) ^b	8 (53) ^d
K2 ^e	9 (15)	1 (14)	0	8 (33) ^d	0
<i>cvaC</i>	5 (8)	0	0	0	4 (27) ^c
<i>traT</i>	14 (23)	1 (14)	5 (33)	3 (13)	5 (33)
<i>ibeA</i>	5 (8)	0	0	1 (4)	4 (27) ^c
PAI	62 (100)	7 (100)	15 (100)	24 (100)	15 (100)

^a Data for *papG* alleles, *sfa* and/or *foc*, *afa* and/or *dra*, *hly*, *iut*, *bma*, and *nfa* are as reported elsewhere (20, 45). One strain within subcluster C (lineage 54) fell outside the four major groups of lineages. Thus, lineages 39 to 53 account for only 61 strains, whereas the total in subcluster C is 62 strains.

^b $P < 0.05$, indicated group versus all other strains.

^c $P \leq 0.01$, indicated group versus all other strains.

^d $P \leq 0.001$, indicated group versus all other strains.

^e Strains were considered K2 positive if positive for *kpsM_{II}T_{II}* by blot but not PCR.

Associations between individual VFs. Analysis of the total VF data set revealed many familiar patterns of association between VFs (23, 30, 33, 35). These included the positive associations among “J96-like” traits *sfa* and/or *foc*, *hly*, *cnfI*, and *papG* allele III, as a group, and among “CFT073-like” traits *iha*, *iut*, *papG* allele II, *kpsM_{II}T_{II}*, and the K2 *kpsMT* variant, as a group, together with the generally negative associations between these two groups (Fig. 4). Also of note were the negative associations of *afa* and/or *dra* and *nfa* with *sfa* and/or *foc*, the association of *ibeA* with the (meningitis-associated) K1 *kpsMT* variant, and the absence of an association between *iha* and the PAI marker gene *malX* despite their close physical proximity on the same PAI in archetypal strain CFT073 (18, 36).

The various *papA* alleles also exhibited multiple distinctive associations with specific VF genes (Fig. 5). Of note were the nonoverlapping groups of *papA* alleles associated with certain closely related VF genes such as *papG* allele II versus allele III, *sfaS* versus *focG*, and the various *kpsMT* variants (Fig. 5). Also of note were the shared associations of *hly* and *cnfI* with the F13, F14, and F48 *papA* alleles but the specific associations of

hly with the F7-2 *papA* allele and of *cnfI* with the F12 *papA* allele (Fig. 5).

Comparisons among the *papA* alleles themselves revealed associations between the F8 allele and both the F9 and F10 alleles ($P < 0.01$), between the F14 and F16 alleles ($P < 0.001$), and between the F9 and F13 alleles ($P < 0.001$). Many additional associations were noted at the level of $P < 0.05$ (not shown).

Correspondence of the present population with ECOR groups and with known virulent clones. Composite RAPD analysis revealed that clusters I, III, IV, and V of the present population (Fig. 1 and 2) correspond, respectively, with ECOR phylogenetic groups A, B2, B1, and D, as defined by MLEE (not shown). These findings, together with previously determined serotypes for the present population (50), the VF and *papA* data from the present study, and selective additional comparative RAPD analyses (not shown), allowed certain lineages to be presumptively linked with recognized virulent clones of ExPEC (Table 4). Of note, two of the distinct pathotypic groups within ET 47 (Fig. 3), one characterized by the F12 *papA* allele, *papG* allele III, and both *hly* and *cnf* and the

FIG. 3. Cluster analysis of VF and *papA* allele data for ET 47. Dendrogram (according to UPGMA) reflects similarity relationships based on VF and *papA* allele profiles among isolates from ET 47, which have indistinguishable genomic backgrounds according to MLEE. Strain CA039 was negative in the F PCR assay but contained a *papA* variant most similar to the F48 *papA* allele according to comparative sequence analysis.

other characterized by the F7-2 *papA* allele, *papG* allele II, and *hly* but not *cnf*, corresponded to two recognized virulence-associated clones within serogroup O6 (Table 4). Because of the multiple VF profile subdivisions within ET 47 (Fig. 3), ET 42, which was the second most populous lineage overall but which in comparison with ET 47 was pathotypically fairly homogeneous (Fig. 1 and 2), emerged as the most prevalent clonopathotype (not shown). This ET, one of whose members (BOS020) exhibited the recently described "F536" *papA* variant, putatively corresponds to an O6:K+;F48/F536-positive lineage, which is a prominent cause of UTI in dogs, cats, and

humans and which includes archetypal pyelonephritis isolate 536 (Table 4).

DISCUSSION

In the present study we assessed the phylogenetic distribution of multiple VF genes and *papA* alleles among *E. coli* isolates from patients with diverse-source bacteremia and defined the phylogenetic correspondence between these strains, the ECOR collection, and other collections of ExPEC. We found that although among the bacteremia isolates phyloge-

	Adhesins								Toxins			Siderophores		Capsules: <i>kpsMT</i>				Miscellaneous		
	<i>papG</i> allele	<i>sfa/</i>	<i>sfa/</i>	<i>sfa/</i>	<i>sfa/</i>	<i>sfa/</i>	<i>sfa/</i>	<i>sfa/</i>	<i>hly</i>	<i>cnf1</i>	<i>cdtB</i>	<i>fyuA</i>	<i>iut</i>	II	K1	"K2"	III	<i>cvaC</i>	<i>ibeA</i>	PAI
Adhesins	II	III	foc	sfaS	focG	dra	iha	nfa												
<i>papG</i> allele II	++	++	++	++	++	-	++	-	++	++	-	++	++	++	-	-	-	-	-	++
<i>papG</i> allele III	-	-	++	++	+	-	-	-	++	++	++	++	(+)	-	-	-	-	-	-	++
<i>sfa/foc</i>			++	++	++	(+)	-	(+)	++	++	-	++	+	+	-	(+)	-	-	-	++
<i>sfaS</i>						-	-	-	++	++	-	+	++	-	-	-	-	-	-	++
<i>focG</i>						-	+	-	++	++	+	++	-	-	-	-	-	-	-	++
<i>afa/dra</i>							-	-	-	-	-	-	+	-	+	-	-	-	-	-
<i>iha</i>								-	+	-	-	++	++	+	-	+	(+)	-	+	++
<i>nfa</i>								-	-	-	-	-	-	(+)	-	-	-	-	-	-
Toxins																				
<i>hly</i>									++	-	-	++	-	++	-	-	-	-	(+)	++
<i>cnf1</i>										+	++	(+)	-	-	-	-	-	-	-	++
<i>cdtB</i>										++	-	-	-	-	-	-	-	-	-	-
Siderophores																				
<i>fyuA</i>												-	++	-	-	-	-	-	-	++
<i>iut</i>													-	-	+	-	-	-	(+)	-
Capsules																				
<i>kpsMT</i> II														++	-	(++)	-	-	-	++
K1														-	-	-	+	++	-	-

FIG. 4. Correlations among VF genes for 182 *E. coli* bacteremia isolates. Only those VF genes that yielded at least one correlation at the $P \leq 0.01$ level are shown. Results for *papA* were similar to those for *papG*. Significance levels: -, $P > 0.01$; +, $P \leq 0.01$; ++, $P \leq 0.001$. (Parentheses denote negative associations.)

<i>papA</i> allele	Adhesins								Toxins			Siderophores		Capsule: <i>kpsMT</i>				Misc.
	<i>pap</i>				<i>sfa/</i>				<i>hly</i>	<i>cnf1</i>	<i>cdtB</i>	<i>fyuA</i>	<i>iut</i>	II	K1	"K2"	III	PAI
	A	G	allele II	allele III	<i>foc</i>	<i>sfaS</i>	<i>focG</i>	<i>iha</i>										
F7-1	-	-	-	-	-	-	-	-	++	-	-	-	-	-	-	-	-	-
F7-2	+	+	++	-	-	-	-	++	-	++	-	-	++	-	-	++	-	-
F8	-	-	+	-	-	-	-	++	-	-	-	-	-	-	-	-	-	-
F9	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	++	-
F10	++	-	-	-	+	-	-	++	-	-	-	++	++	+	-	-	-	++
F11	++	++	++	-	-	-	-	-	-	-	-	-	-	-	++	-	-	-
F12	-	-	-	+	-	-	-	-	++	-	+	-	-	-	-	-	-	-
F13	-	+	-	++	+	-	+	-	-	++	++	-	-	(+)	-	-	++	-
F14	+	++	-	++	-	-	++	+	-	++	+	++	-	-	-	-	-	-
F16	-	+	++	(+)	-	-	-	++	-	-	-	-	-	-	-	-	-	-
F48	++	++	(+)	++	++	++	-	-	-	++	++	-	-	(++)	-	-	-	-

FIG. 5. Correlations among F antigen-specific *papA* alleles and other VF genes for 182 *E. coli* bacteremia isolates. Only those *papA* alleles that yielded at least one correlation at the $P \leq 0.01$ level are shown. Significance levels: -, $P > 0.01$; +, $P \leq 0.01$; ++, $P \leq 0.001$. (Parentheses denote negative associations.)

netic group B2 predominated overall, other phylogenetic groups exhibited greater concentrations of certain VFs than did group B2 and included several additional recognized virulent clones. Moreover, certain of the newly detected VFs were as prevalent or more prevalent than their more familiar traditional counterparts, thus possibly constituting additional useful targets for preventive interventions. The considerable diversity of VF profiles observed at every level within the phylogenetic tree, including even within individual lineages, suggested that many different pathways can lead to extraintestinal virulence in *E. coli* and that the evolution of ExPEC, which involves extensive horizontal transmission of VFs and continuous remodeling of PAIs, is an active, ongoing process.

Our results are consistent with previous evidence that *E. coli* phylogenetic group B2 predominates among ExPEC and is the main repository (hence, presumably, the original source within the species) of many extraintestinal VFs (10, 23, 33, 45, 59, 60). Our data considerably enlarge the range of VFs known to be associated with phylogenetic group B2 in the context of diverse-source bacteremia (22, 50). However, they also demonstrate that certain VFs are significantly concentrated outside of group B2 in phylogenetic group A or D or are broadly distributed in the population without a focal concentration in any one phylogenetic group. This is consistent with recent findings derived from analysis of the ECOR collection and urosepsis isolates (23, 30, 33). The latter observations suggest that the evolutionary histories of the various VFs of ExPEC are more complex and diverse than can be accounted for by two-step models that focus primarily on phylogenetic group B2, as previously proposed based on an analysis of *pap*, *hly*, *sfa* and/or *foc*, and *kps* (10, 45).

The broad range of our VF gene screening allowed us to document an appreciable prevalence and/or nonrandom phylogenetic distribution for many newer VF genes, including *iha*, *fyuA*, *iroN*, *cnf*, *cdt*, and *ibeA*. Likewise, by discriminating between closely related variants of a particular VF gene or gene family we were able to detect disparate associations, prevalences, and phylogenetic backgrounds for the 12 *papA* alleles, *sfaS* versus *focG*, and the several *kpsMT* variants. This simultaneously broad and highly discriminating approach to VF genotyping provided the added advantage of generating distinctive extended VF "signatures" for individual strains and lineages. In many instances, with or without supplemental data these signatures allowed us to associate lineages from the

present population with known virulent clones from the literature (Table 4).

For example, we found that an O6:K+;F48/F536 clonal group that is associated with UTI in animals and humans (unpublished data; 25, 27, 34) and that includes archetypal pyelonephritis isolate 536 (8) was the single most prevalent clonopathotype in the study population. It accounted for ~7% of isolates overall, and its isolates outnumbered even those of the prominent O6:K2:H1 clonal group (~4%). The associated primary sources of bacteremia for these O6:K+;F48/F536 strains included UTI, pneumonitis, and intra-abdominal infections (49). The evident pathogenetic versatility of this clone with respect to host species, clinical syndrome, and anatomical site of infection illustrates the inappropriateness of the traditional designation of uropathogenic *E. coli* and supports instead a more inclusive rubric such as ExPEC (65). Consistent with this "generalist" concept, we identified within the study population several representatives of the O18:K1:H7 clonal group, which traditionally has been associated with neonatal meningitis (9, 38, 69) but which recently was shown to be a major contributor also to uncomplicated cystitis in women (24, 26, 41).

Correlation of the newly detected VFs, and especially the *papA* alleles, with the population's underlying phylogenetic structure suggested even more extensive horizontal transfer of VFs than was previously proposed for this population based on analysis of *pap*, *hly*, *sfa* and/or *foc*, *afa* and/or *dra*, and *iut* (50) or the three *papG* alleles (22). The evidence of horizontal transfer extended down to the most differentiated subdivisions of the phylogenetic tree, i.e., the individual ETs themselves, and beyond. This suggests that ExPEC strains are actively evolving with respect to VF profiles even today and hence represent a "moving target" for the pathotypic analyses that are needed to guide the development of VF-specific protective measures. Moreover, the ordered diversity of VF profiles observed within ET 47 (Fig. 3) suggests ongoing evolution of the genomic background at a level below the detection threshold of MLEE, which is insensitive to silent mutations and to peptide polymorphisms that do not alter the electrophoretic mobility of the particular markers analyzed (68, 79). We propose that in the instance of ET 47 the VF data, rather than being phylogenetically unreliable (as was previously suggested when diverse pathotypes were encountered within a single electrophoretic lineage) (69), actually supersede MLEE by revealing

TABLE 4. Proposed correspondence between specific lineages from present study and recognized virulent clones of ExPEC

ET lineage	Putative clonal group	Corresponding clonal groups from literature						Comments
		ECOR (23, 54)	Selander et al. (69)	Ochman and Selander (53)	Achtman et al. (1) pattern(s)	Cherifi et al. (11) clone	Whittam et al. (25, 80) ET	Vaisanen-Rhen et al. (76) clone(s)
29, 30	O16:K1			j,k/12	12		12 and 14	VI
34, 35	O4:K3/10:H5	59, 60	ET 18					IV
39	O2:K5/7:H1	57				2	21	Pylonephritis, sepsis Archetypal strains 196 and CP9 (29, 31) Urosepsis; usually <i>cdt</i> positive (33)
42	O6:K15/53:H31							Archetypal strain 536; UTI in dogs and cats; cystitis in humans (25, 27, 32, 34)
47	O6:K2:H1	56	Cluster B			1	1	V
47	O6:K13/K53:H1	56	Cluster B			1	1	VIII
51	O18:K1:H7	(61, 62) ^a	Cluster A	f,g,h,i/6,9	6 and 9		4 and 6	I and II
52, 61-62	O1/O2:K1:H7	61, 62	Cluster A	e/6,9	6 and 9		5 and 6	I and II
55, 56	O18:K5:H-/H7		Cluster E		11			VII
58, 59	O75:K5/100:H5	64			11			Usually <i>afa</i> and/or <i>dra</i> positive (39)
93	O7:K1:H-	38-41	Cluster H	1/3,15,16,m/3	3			Neonatal meningitis and sepsis
95-97	O15:K52:H1	(44) ^a						Epidemic and endemic UTI (56, 58, 62)
98-99	O1:K1:H-	35, 36		a/5	5			Neonatal meningitis and sepsis
100	O157:K+ :H+							UTI, urosepsis (33, 78)

^a Correspondence with indicated ECOR strains is approximate.

significant evolutionary diversity that MLEE is unable to detect. Formal testing of this hypothesis should be possible with the use of a phylotyping method more discriminating than MLEE, e.g., multilocus sequence typing (14, 45, 48, 63).

Although the strict clinical criteria for inclusion in the present study population (50) represent an epidemiological strength of the study, they do limit the generalizability of the findings to the larger *E. coli* population. The requirement that isolates must have caused bacteremia imposes a substantial “filter” that is predicted to bias the population not only toward the more virulent phylogenetic groups such as B2 and D (i.e., clusters III and V) but also toward the more virulent members within each group. This filtering effect would probably be strongest for intrinsically “low-virulence” groups such as A and B1 (i.e., clusters I and IV), most members of which may lack VFs and hence would be unable to cause bacteremia except in the presence of significant host compromise (23, 60). Consistent with this hypothesis, multiple VFs were numerically more prevalent within cluster I of the present population (group A equivalent) than among the group A ECOR strains (23). The genes for these included *afa* and/or *dra* (29 versus 0%), *iut* (52 versus 24%), *kpsM_{II}T_{II}* (48 versus 20%), *traT* (29 versus 16%), and *nfa*, *kps_{III}MT_{III}*, and *malX* (all 10 versus 0%). In contrast, only two VF genes were more prevalent among the group A ECOR strains, i.e., *iha* (28 versus 10%) and *fyuA* (52 versus 24%) (23). Similar trends were evident when cluster IV of the present population (group B1 equivalent) was compared with the group B1 ECOR strains (not shown) (23). Thus, broad phylogenetic conclusions must be tempered by recognition of the likely substantial influence of clinical context on the observed associations.

Previous studies have documented that PAIs from different ExPEC strains can differ considerably with respect to their constituent VF genes (9, 18, 19, 64, 73) and, for those VF genes that do consistently occur together in multiple PAIs, the specific genetic linkages between the VF genes (46). These largely anecdotal observations, together with our present population-based finding that even VFs known to be PAI associated in certain strains commonly exhibit divergent patterns of phylogenetic distribution both among lineages and even within a given lineage, strongly suggest that VF genes are highly mobile independent of PAIs and that PAIs are subject to continuous, ongoing remodeling. These data are consistent with a model in which PAIs, although perhaps occasionally subject to en bloc horizontal transfer (10, 16) or total deletion (5, 8, 19, 37), participate to a much greater extent in the horizontal transfer of VFs by providing genomic regions receptive to the insertion, retention, and release of individual VF genes.

The abundance of recombination-promoting elements within PAIs (16) presumably allows for the ready acquisition of any “wandering” VF genes that happen to enter the cell, plus ready deletion of previously integrated VF genes, including their donation to nearby plasmids, transposons, etc. The resulting rapid reassortment of VF genes within PAIs continuously generates novel combinations of VFs on which selective pressures can act. Those combinations of VFs that increase fitness in a particular niche or that broaden the range of niches in which the host cell can effectively compete are conserved and expand with the clone. However, just as the clonal frame of the bacterial host’s (presumably selection-neutral) genomic

background gradually breaks down over time from recombined genomic regions (17, 44, 51), so too is the host's highly selectable VF profile (i.e., the architecture of its PAIs) disrupted by newly acquired and deleted VF genes. According to this model, PAIs are not so much horizontally mobile and internally stable as they are horizontally stable but internally highly dynamic. Long-range VF mapping (7), comparative genomic analysis (6), and long-term culture experiments involving clonally mixed populations may allow direct testing of this model in the future.

In summary, we found that among *E. coli* blood isolates from patients with diverse-source bacteremia, phylogenetic group B2 predominated overall and accounted for most of the individual VFs and virulent clones. However, other phylogenetic groups exhibited greater concentrations of certain VFs than did group B2 and accounted for additional virulent clones. Certain newly detected VFs were as prevalent or more prevalent than were their more familiar counterparts and hence may constitute useful targets for future preventive interventions. The considerable diversity of VF profiles noted at every level within the phylogenetic tree suggests that many different pathways can lead to extraintestinal virulence in *E. coli* and that the evolution of ExPEC, which involves extensive horizontal transmission of VFs and continuous remodeling of PAIs, is a highly active and ongoing process.

ACKNOWLEDGMENTS

This work was supported by the Office of Research and Development, Medical Research Service, Department of Veterans Affairs (J.R.J. and J.N.M.), National Institutes of Health grant DK-47504 (J.R.J.), and National Research Initiative (NRI) Competitive Grants Program/United States Department of Agriculture grant 00-35212-9408 (J.R.J.).

Dave Prentiss (Minneapolis VAMC) helped prepare the figures. Ann Emery (Minneapolis VAMC) helped with manuscript preparation.

REFERENCES

- Achtman, M., A. Mercer, B. Kusecek, A. Pohl, M. Heuzenroeder, W. Aaronson, A. Sutton, and R. P. Silver. 1983. Six widespread bacterial clones among *Escherichia coli* K1 isolates. *Infect. Immun.* **39**:315-335.
- Arthur, M., R. D. Arbeit, C. Kim, P. Beltran, H. Crowe, S. Steinback, 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 *rm* operons. *Infect. Immun.* **58**:471-479.
- Berg, D. E., N. S. Akopyants, and D. Kersulyte. 1994. Fingerprinting microbial genomes using the RAPD or AP-PCR method. *Methods Mol. Cell. Biol.* **5**:13-24.
- Bingen, E., B. Picard, N. Brahimi, S. Mathy, P. Desjardins, J. Elion, and E. Denamur. 1998. Phylogenetic analysis of *Escherichia coli* strains causing neonatal meningitis suggests horizontal gene transfer from a predominant pool of highly virulent B2 group strains. *J. Infect. Dis.* **177**:642-650.
- Bloch, C. A., and C. K. Rode. 1996. Pathogenicity island evaluation in *Escherichia coli* K1 by crossing with laboratory strain K-12. *Infect. Immun.* **64**:3218-3223.
- Bloch, C. A., C. K. Rode, V. Obreque, and K. Y. Russell. 1994. Comparative genome mapping with mobile physical map landmarks. *J. Bacteriol.* **176**:7121-7125.
- Blum, G., M. Ott, A. Cross, and J. Hacker. 1991. Virulence determinants of *Escherichia coli* O6 extraintestinal isolates analysed by Southern hybridizations and DNA long range mapping techniques. *Microb. Pathog.* **10**:127-136.
- Blum, G., M. Ott, A. Lischewski, A. Ritter, H. Imrich, H. Tschape, and J. Hacker. 1994. Excision of large DNA regions termed pathogenicity islands from tRNA-specific loci in the chromosome of an *Escherichia coli* wild-type pathogen. *Infect. Immun.* **62**:606-614.
- Bonacorsi, S. P. P., O. Clermont, C. Tinsley, I. Le Gall, J.-C. Beaudoin, J. Elion, X. Nassif, and E. Bingen. 2000. Identification of regions of the *Escherichia coli* chromosome specific for neonatal meningitis-associated strains. *Infect. Immun.* **68**:2096-2101.
- Boyd, E. F., and D. L. Hartl. 1998. Chromosomal regions specific to pathogenic isolates of *Escherichia coli* have a phylogenetically clustered distribution. *J. Bacteriol.* **180**:1159-1165.
- Cherifi, A., M. Contrepois, B. Picard, P. Oullet, I. Orskov, F. Orskov, and J. De Rycke. 1991. Clonal relationships among *Escherichia coli* serogroup O6 isolates from human and animal infections. *FEMS Microbiol. Lett.* **80**:225-230.
- Donnenberg, M. S., and R. A. Welch. 1996. Virulence determinants of uropathogenic *Escherichia coli*, p. 135-174. In H. L. T. Mobley and J. W. Warren (ed.), *Urinary tract infections: molecular pathogenesis and clinical management*. ASM Press, Washington, D.C.
- Eisenstein, B. I., and G. W. Jones. 1988. The spectrum of infections and pathogenic mechanisms of *Escherichia coli*. *Adv. Intern. Med.* **33**:231-252.
- Enright, M. C., and B. C. Spratt. 1999. Multilocus sequence typing. *Trends Microbiol.* **7**:482-487.
- Fleiss, J. L. 1981. Statistical methods for rates and proportions, p. 112-137. John Wiley & Sons, New York, N.Y.
- Groisman, E. A., and H. Ochman. 1996. Pathogenicity islands: bacterial evolution in quantum leaps. *Cell.* **87**:791-794.
- Guttman, D. S., and D. E. Dykhuizen. 1994. Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. *Science* **266**:1380-1383.
- Guyer, D. M., J.-S. Kao, and H. L. T. Mobley. 1998. Genomic analysis of a pathogenicity island in uropathogenic *Escherichia coli* CFT073: distribution of homologous sequences among isolates from patients with pyelonephritis, cystitis, and catheter-associated bacteriuria and from fecal samples. *Infect. Immun.* **66**:4411-4417.
- Hacker, J., L. Bender, M. Ott, J. Wingender, B. Lund, R. Marre, and W. Goebel. 1990. Deletions of chromosomal regions coding for fimbriae and hemolysins occur in vitro and in vivo in various extraintestinal *Escherichia coli* isolates. *Microb. Pathog.* **8**:213-225.
- Herzer, P. J., S. Inouye, M. Inouye, and T. S. Whittam. 1990. Phylogenetic distribution of branched RNS-linked multicopy single-stranded DNA among natural isolates of *Escherichia coli*. *J. Bacteriol.* **172**:6175-6181.
- Johnson, J. R. 1991. Virulence factors in *Escherichia coli* urinary tract infection. *Clin. Microbiol. Rev.* **4**:80-128.
- Johnson, J. R., J. J. Brown, and J. N. Maslow. 1998. Clonal distribution of the three alleles of the Gal(α 1-4)Gal-specific adhesin gene *papG* among *Escherichia coli* strains from patients with bacteremia. *J. Infect. Dis.* **177**:651-661.
- Johnson, J. R., P. Delavari, M. Kuskowski, and A. L. Stell. 2001. Phylogenetic distribution of extraintestinal virulence-associated traits in *Escherichia coli*. *J. Infect. Dis.* **183**:78-88.
- Johnson, J. R., P. Delavari, and T. O'Bryan. 2001. *Escherichia coli* O18:K1:H7 isolates from acute cystitis and neonatal meningitis exhibit common phylogenetic origins and virulence factor profiles. *J. Infect. Dis.* **183**:425-434.
- Johnson, J. R., P. Delavari, A. L. Stell, T. S. Whittam, U. Carlino, and T. A. Russo. 2001. Molecular comparison of extraintestinal *Escherichia coli* isolates from the same electrophoretic lineages from humans and domestic animals. *J. Infect. Dis.* **183**:154-159.
- Johnson, J. R., T. T. O'Bryan, P. Delavari, M. Kuskowski, A. Stapleton, U. Carlino, and T. Russo. 2001. Clonal relationships and extended virulence genotypes among *Escherichia coli* isolates from women with a first or recurrent episode of cystitis. *J. Infect. Dis.* **183**:1508-1517.
- Johnson, J. R., T. T. O'Bryan, D. A. Low, G. Ling, P. Delavari, C. Fasching, T. A. Russo, U. Carlino, and A. L. Stell. 2000. Evidence of commonality between canine and human extraintestinal pathogenic *Escherichia coli* strains that express *papG* allele III. *Infect. Immun.* **68**:3327-3336.
- Johnson, J. R., T. A. Russo, J. J. Brown, and A. Stapleton. 1998. *papG* alleles of *Escherichia coli* strains causing first episode or recurrent acute cystitis in adult women. *J. Infect. Dis.* **177**:97-101.
- Johnson, J. R., T. A. Russo, F. Scheutz, J. J. Brown, L. Zhang, K. Palin, C. Rode, C. Bloch, C. F. Marrs, and B. Foxman. 1997. Discovery of disseminated J96-like strains of uropathogenic *Escherichia coli* O4:H5 containing genes for both *PapG₉₆* ("class I") and *PrsG₉₆* ("class III") Gal(α 1-4)Gal-binding adhesins. *J. Infect. Dis.* **175**:983-988.
- Johnson, J. R., T. A. Russo, P. I. Tarr, U. Carlino, S. S. Bilge, J. C. J. Vary, and A. L. Stell. 2000. Molecular epidemiological and phylogenetic associations of two novel putative virulence genes, *iha* and *iroN_{E. coli}*, among *Escherichia coli* isolates from patients with urosepsis. *Infect. Immun.* **68**:3040-3047.
- Johnson, J. R., A. E. Stapleton, T. A. Russo, F. S. Scheutz, J. J. Brown, and J. N. Maslow. 1997. Characteristics and prevalence within serogroup O4 of a "J96-like" clonal group of uropathogenic *Escherichia coli* O4:H5 containing the "class I" and "class III" alleles of *papG*. *Infect. Immun.* **65**:2153-2159.
- Johnson, J. R., A. Stell, and P. Delavari. 2001. Canine feces as a reservoir of extraintestinal pathogenic *Escherichia coli*. *Infect. Immun.* **69**:1306-1314.
- Johnson, J. R., and A. L. Stell. 2000. Extended virulence genotypes of *Escherichia coli* strains from patients with urosepsis in relation to phylogeny and host compromise. *J. Infect. Dis.* **181**:261-272.
- Johnson, J. R., A. L. Stell, P. Delavari, A. C. Murray, M. Kuskowski, and W. Gastra. 2001. Phylogenetic and pathotypic similarities between *Escherichia coli* isolates from urinary tract infections in dogs and extraintestinal infec-

- tions in humans. *J. Infect. Dis.* **183**:897–906.
35. Johnson, J. R., A. L. Stell, F. Scheutz, T. T. O'Bryan, T. A. Russo, U. B. Carlino, C. C. Fasching, J. Kavle, L. van Dijk, and W. Gaastra. 2000. Analysis of F antigen-specific *papA* alleles of extraintestinal pathogenic *Escherichia coli* using a novel multiplex PCR-based assay. *Infect. Immun.* **68**:1587–1599.
 36. Kao, J.-S., D. M. Stucker, J. W. Warren, and H. L. T. Mobley. 1997. Pathogenicity island sequences of pyelonephritogenic *Escherichia coli* CFT073 are associated with virulent uropathogenic strains. *Infect. Immun.* **65**:2812–2820.
 37. Knapp, S., J. Hacker, I. Then, D. Muller, and W. Goebel. 1984. Multiple copies of hemolysin genes and associated sequences in the chromosomes of uropathogenic *Escherichia coli* strains. *J. Bacteriol.* **159**:1027–1033.
 38. Korhonen, T. K., M. V. Valtonen, J. Parkkinen, V. Vaisanen-Rhen, J. Fine, F. Orskov, I. Orskov, S. B. Svenson, and P. H. Makela. 1985. Serotypes, hemolysin production, and receptor recognition of *Escherichia coli* strains associated with neonatal sepsis and meningitis. *Infect. Immun.* **48**:486–491.
 39. Korhonen, T. K., R. Virkola, V. Vaisanen-Rhen, and H. Holthofer. 1986. Binding of purified *Escherichia coli* O75X adhesin to frozen sections of human kidney. *FEMS Lett.* **35**:313–318.
 40. Kumar, S., K. Tamura, and M. Nei. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. The Pennsylvania State University, University Park, Pa.
 41. Kunin, C. M., T. H. Hua, C. Krishnan, L. Van Arsdale White, and J. Hacker. 1993. Isolation of a nicotinamide-requiring clone of *Escherichia coli* O18: K1:H7 from women with acute cystitis: resemblance to strains found in neonatal meningitis. *Clin. Infect. Dis.* **16**:412–416.
 42. Langermann, S., R. Mollby, J. E. Burlein, S. R. Palaszynski, C. G. Auguste, A. DeFusco, R. Strouse, M. A. Schenerman, S. J. Hultgren, S. Pinkner, J. Winberg, L. Guldevall, M. Soderhall, K. Ishikawa, S. Normark, and S. Koenig. 2000. Vaccination with FimH adhesin protects cynomolgus monkeys from colonization and infection by uropathogenic *Escherichia coli*. *J. Infect. Dis.* **181**:774–778.
 43. Langermann, S., S. Palaszynski, M. Barnhart, G. Auguste, J. S. Pinkner, J. Burlein, P. Barren, S. Koenig, S. Leath, C. H. Jones, and S. J. Hultgren. 1997. Prevention of mucosal *Escherichia coli* infection by FimH-adhesin-based systemic vaccination. *Science* **276**:607–611.
 44. Lawrence, J. G., and H. Ochman. 1998. Molecular archaeology of the *Escherichia coli* genome. *Evolution* **52**:9413–9417.
 45. Lecointre, G., L. Rachdi, P. Darlu, and E. Denamur. 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol. Biol. Evol.* **15**:1685–1695.
 46. Low, D., V. David, D. Lark, G. Schoolnik, and S. Falkow. 1984. Gene clusters governing the production of hemolysin and mannose-resistant hemagglutination are closely linked in *Escherichia coli* serotype O4 and O6 isolates from urinary tract infection. *Infect. Immun.* **43**:353–358.
 47. Lund, B., F. Lindberg, B.-I. Marklund, and S. Normark. 1988. Tip proteins of pili associated with pyelonephritis: new candidates for vaccine development. *Vaccine* **6**:110–112.
 48. Maiden, C. J., J. A. Bygraves, E. Feil, G. Morelli, J. E. Russell, R. Urwin, Q. Zhang, J. Zhou, K. Zurth, D. A. Caugant, I. M. Feavers, M. Achtman, and B. G. Spratt. 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc. Natl. Acad. Sci. USA* **95**:3140–3145.
 49. Maslow, J. N., M. E. Mulligan, K. S. Adams, J. C. Justis, and R. D. Arbeit. 1993. Bacterial adhesins and host factors: role in the development and outcome of *Escherichia coli* bacteremia. *Clin. Infect. Dis.* **17**:89–97.
 50. Maslow, J. N., T. S. Whittam, C. F. Gilks, R. A. Wilson, M. E. Mulligan, K. S. Adams, and R. D. Arbeit. 1995. Clonal relationships among bloodstream isolates of *Escherichia coli*. *Infect. Immun.* **63**:2409–2417.
 51. Milkman, R., and M. M. Bridges. 1990. Molecular evolution of the *Escherichia coli* chromosome. III. Clonal frames. *Genetics* **126**:505–517.
 52. Mobley, H. L. T., K. G. Jarvis, J. P. Elwood, D. I. Whittle, C. V. Lockatell, R. G. Russell, D. E. Johnson, M. S. Donnenberg, and J. W. Warren. 1993. Isogenic P-fimbrial deletion mutants of pyelonephritogenic *Escherichia coli*: the role of α Gal(1–4)- β Gal binding in virulence of a wild-type strain. *Mol. Microbiol.* **10**:143–155.
 53. Ochman, H., and R. K. Selander. 1984. Evidence for clonal population structure in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **81**:198–201.
 54. Ochman, H., and R. K. Selander. 1984. Standard reference strains of *Escherichia coli* from natural populations. *J. Bacteriol.* **157**:690–693.
 55. O'Hanley, P. 1996. Prospects for urinary tract infection vaccines, p. 405–425. In H. L. T. Mobley and J. W. Warren (ed.), *Urinary tract infections: molecular pathogenesis and clinical management*. ASM Press, Washington, D.C.
 56. Olesen, B., H. J. Kolmos, F. Orskov, and I. Orskov. 1995. A comparative study of nosocomial and community-acquired strains of *Escherichia coli* causing bacteraemia in a Danish university hospital. *J. Hosp. Infect.* **31**:295–305.
 57. Orskov, I., and F. Orskov. 1985. *Escherichia coli* in extra-intestinal infections. *J. Hyg. (London)* **95**:551–575.
 58. Phillips, I., S. Eykyn, A. King, W. R. Grandsden, B. Rowe, J. A. Frost, and R. J. Gross. 1988. Epidemic multiresistant *Escherichia coli* infection in West Lambeth health district. *Lancet* **i**:1038–1041.
 59. Picard, B., C. Journet-Mancy, N. Picard-Pasquier, and P. Goullet. 1993. Genetic structures of the B₂ and B₁ *Escherichia coli* strains responsible for extra-intestinal infections. *J. Gen. Microbiol.* **139**:3079–3088.
 60. Picard, B., J. Sevali Garcia, S. Gouriou, P. Duriez, N. Brahimi, E. Bingen, J. Elion, and E. Denamur. 1999. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection. *Infect. Immun.* **67**:546–553.
 61. Plos, K., S. I. Hull, R. A. Hull, B. R. Levin, I. Orskov, F. Orskov, 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.
 62. Prats, G., F. Navarro, B. Mirelis, D. Dalmay, N. Margall, P. Coll, A. Stell, and J. R. Johnson. 2000. *Escherichia coli* serotype O15:K52:H1 as a uropathogenic clone. *J. Clin. Microbiol.* **38**:201–209.
 63. Reid, S. D., C. J. Herbelin, A. C. Bumbaugh, R. K. Selander, and T. S. Whittam. 2000. Parallel evolution of virulence in pathogenic *Escherichia coli*. *Nature* **406**:64–67.
 64. Russo, T. A., U. B. Carlino, A. Mong, and S. T. Jodush. 1999. Identification of genes in an extraintestinal isolate of *Escherichia coli* with increased expression after exposure to human urine. *Infect. Immun.* **67**:5306–5314.
 65. Russo, T. A., and J. R. Johnson. 2000. A proposal for an inclusive designation for extraintestinal pathogenic *Escherichia coli*: ExPEC. *J. Infect. Dis.* **181**:1753–1754.
 66. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
 67. Schubert, S., A. Rakin, H. Karch, E. Carniel, and J. Heesemann. 1998. Prevalence of the “high-pathogenicity island” of *Yersinia* species among *Escherichia coli* strains that are pathogenic to humans. *Infect. Immun.* **66**:480–485.
 68. 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.
 69. Selander, R. K., T. K. Korhonen, V. Väisänen-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.
 70. Siitonen, A. 1994. What makes *Escherichia coli* pathogenic? *Ann. Med.* **26**:229–231.
 71. Sokal, R. R., and P. H. A. Sneath. 1963. *Principles of numerical taxonomy*. W. H. Freeman, San Francisco, Calif.
 72. Stapleton, A., and W. E. Stamm. 1997. Prevention of urinary tract infection. *Infect. Dis. Clin. N. Am.* **11**:719–733.
 73. Swenson, D. L., N. O. Bukanov, D. E. Berg, and R. A. Welch. 1996. Two pathogenicity islands in uropathogenic *Escherichia coli* J96: cosmid cloning and sample sequencing. *Infect. Immun.* **64**:3736–3743.
 74. Tarr, P. I., S. S. Bilge, J. C. Vary, Jr., S. Jelacic, R. L. Habeeb, T. R. Ward, M. R. Baylor, and T. E. Besser. 2000. Iha: a novel *Escherichia coli* O157:H7 adherence-conferring molecule encoded on a chromosomal region of conserved structure. *Infect. Immun.* **68**:1400–1407.
 75. Thompson, J. D., D. G. Higgins, and T. J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**:4673–4680.
 76. Vaisanen-Rhen, V., J. E. O. E. Vaisanen, A. Siitonen, I. Orskov, F. Orskov, S. B. Svenson, P. H. Makela, and T. Korhonen. 1984. P-fimbriated clones among uropathogenic *Escherichia coli* strains. *Infect. Immun.* **43**:149–155.
 77. Wang, G., T. S. Whittam, C. M. Berg, and D. E. Berg. 1993. RAPD (arbitrary primer) PCR is more sensitive than multilocus enzyme electrophoresis for distinguishing related bacterial strains. *Nucleic Acids Res.* **21**:5930–5933.
 78. Whittam, T. S., and R. A. Wilson. 1988. Genetic relationships among pathogenic *Escherichia coli* of serogroup O157. *Infect. Immun.* **56**:2467–2473.
 79. Whittam, T. S., H. Ochman, and R. K. Selander. 1983. Multilocus genetic structure in natural populations of *Escherichia coli*. *Popul. Biol.* **80**:1751–1755.
 80. Whittam, T. S., M. L. Wolfe, and R. A. Wilson. 1989. Genetic relationships among *Escherichia coli* isolates causing urinary tract infections in humans and animals. *Epidemiol. Infect.* **102**:37–46.