

Molecular Epidemiologic Identification of *Escherichia coli* Genes That Are Potentially Involved in Movement of the Organism from the Intestinal Tract to the Vagina and Bladder

Jingping Xie,^{1,2} Betsy Foxman,¹ Lixin Zhang,¹ and Carl F. Marrs^{1*}

Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, Michigan,¹ and Department of Pediatrics and Communicable Diseases, University of Michigan Medical School, Ann Arbor, Michigan²

Received 22 February 2006/Returned for modification 10 April 2006/Accepted 10 May 2006

A first step in urinary tract infection (UTI) pathogenesis in the otherwise healthy host is the movement of uropathogenic *Escherichia coli* from the intestinal tract to the urinary tract. We conducted a genomic subtraction to isolate genetic regions associated with this movement. A representative UTI isolate present in the rectum, vagina, and bladder of a woman with UTI was chosen as the tester; the driver was a phylogenetically distant rectal isolate (based on pulsed-field gel electrophoresis analysis) with a profile of uropathogenic virulence genes similar to that of the tester. Tester-specific regions identified by the subtraction were screened, using DNA dot blot hybridization, against a collection of 88 uropathogens isolated from the rectum, urine, and/or vagina of women with UTIs and 54 *E. coli* isolates from the same women that were found only in the rectum. Twelve genetic regions occurred more often in multisite isolates than in rectal site-only isolates. Eleven of these 12 genetic regions are homologous to regions in the sequenced uropathogenic *E. coli* CFT073 strain.

Urinary tract infection (UTI) is one of the most commonly acquired bacterial infections in ambulatory and hospitalized populations; 11% of all women 18 years and older in the United States have a UTI each year (9). Most UTIs among otherwise healthy women are caused by *Escherichia coli*. Certain O:K:H serotypes and virulence factors occur more frequently in urinary than in fecal isolates, suggesting that uropathogenic *E. coli* (UPEC) is different from normal bowel inhabitants (17).

One critical aspect leading to UTI is the ability of UPEC strains to move from the intestinal tract and establish themselves in the urinary tract. In some cases this movement may be facilitated by UPEC strains establishing themselves first in the vagina (36, 37). Although some of this movement may be mechanical, the ability to establish colonization in the vagina and bladder must reflect bacterial characteristics. However, little is known about what genes or factors present in UPEC isolates help them move from the intestinal tract to the vagina and bladder and establish themselves extraintestinally.

Genomic subtraction makes it possible to identify genomic differences among strains. Genomic subtraction has been successfully employed to identify novel virulence-associated genes in UPEC strains (16, 22). We used a molecular epidemiologic strategy for bacterial gene discovery that selects bacterial isolates for genomic subtraction based on epidemiologic information and bacterial characteristics and screens epidemiologically defined bacterial collections with the resultant gene fragments to determine their potential significance and possible function (35, 44).

Here we report on the use of genomic subtraction followed by epidemiologic screening to identify 12 new genetic regions

potentially involved in the spread of *E. coli* from the intestinal tract into the vagina and bladder.

MATERIALS AND METHODS

***E. coli* isolates used in epidemiologic screening.** We started with a collection of *E. coli* isolates obtained from 166 women with *E. coli* UTIs and 94 women without UTIs and their most recent male sex partners in a study of sharing of *E. coli* among heterosexual sex partners (10). Three completely sequenced strains, uropathogenic *E. coli* CFT073 (41), hemorrhagic *E. coli* O157:H7 EDL933 (27), and the lab *E. coli* K-12 strain MG1655 (3), were used as controls for the hybridization. Strain TOP10 (Invitrogen, San Diego, CA) was used as the host strain for recombinant clones.

PFGE. Purification, rare-cutter restriction, and pulsed-field gel electrophoresis (PFGE) of minimally sheared *E. coli* DNAs were performed as previously described (12). Briefly, electrophoresis of NotI-digested DNAs was done in a pulsed-field apparatus (DR III; Bio-Rad, Hercules, CA) in 1.3% SeaKem HGT agarose at 14°C with pulse ramping from 10 to 22 s for 14 h and from 55 s to 60 s for 8 h at a field strength of 6 V/cm. Gels were stained with Vistra green dye (Amersham Biosciences, Piscataway, NJ) and then scanned with a Storm phosphorimager (Molecular Dynamics, Sunnyvale, CA). The electrophoretic patterns were analyzed with BioNumeric software (Applied Maths, Kortrijk, Belgium).

Selection of strains for subtraction and screening. We used PFGE to define the identity of isolates from different sites within individual women and identified 102 women with UTIs and four women without UTIs who had the same *E. coli* strain colonizing the urine, rectum, and vagina. In one case a woman with a UTI had two separate PFGE-defined isolate types present at all three sites. Among women with an *E. coli* isolate with a single PFGE type present in all three sites, we also obtained *E. coli* isolates of different PFGE types from their rectal samples. Altogether, there were 381 isolates obtained from the 106 women. All 381 isolates had been screened previously for the presence of 13 virulence genes and assigned a virulence signature, i.e., a binary score based on the presence or absence of each virulence gene: type 1, P-pilus family of fimbriae (*pff*), S-fimbrial adhesion (*sfa*), aerobactin (*aer*), group II capsule (*kpsMT*), outer membrane protein T (*ompT*), hemolysin (*hly*), cytotoxic necrotizing factor 1 (*cnfI*), DR binding adhesions (*drb*), group III capsule (CAP III), and three subclasses of the P-pilus family of fimbriae—*papG_{AD}* (class II), *papG₉₆* (class I), and *prsG₉₆* (class III) (11, 12, 20). We grouped the 381 isolates according to virulence signatures; the five largest groups ranged in size from 19 to 43 isolates (group 1 = 1110111100010, *n* = 43; group 2 = 1000110000000, *n* = 26; group 3 = 1101110000100, *n* = 21; group 4 = 1001110000000, *n* = 20; group 5 = 1010111100000, *n* = 19), representing 129 of the isolates which were derived from 40 women with UTIs and three women without UTIs. As group 2 contains

* Corresponding author. Mailing address: Department of Epidemiology, 109 Observatory Street, Ann Arbor, MI 48109-2029. Phone: (734) 647-2407. Fax: (734) 764-3192. E-mail: cfmarrs@umich.edu.

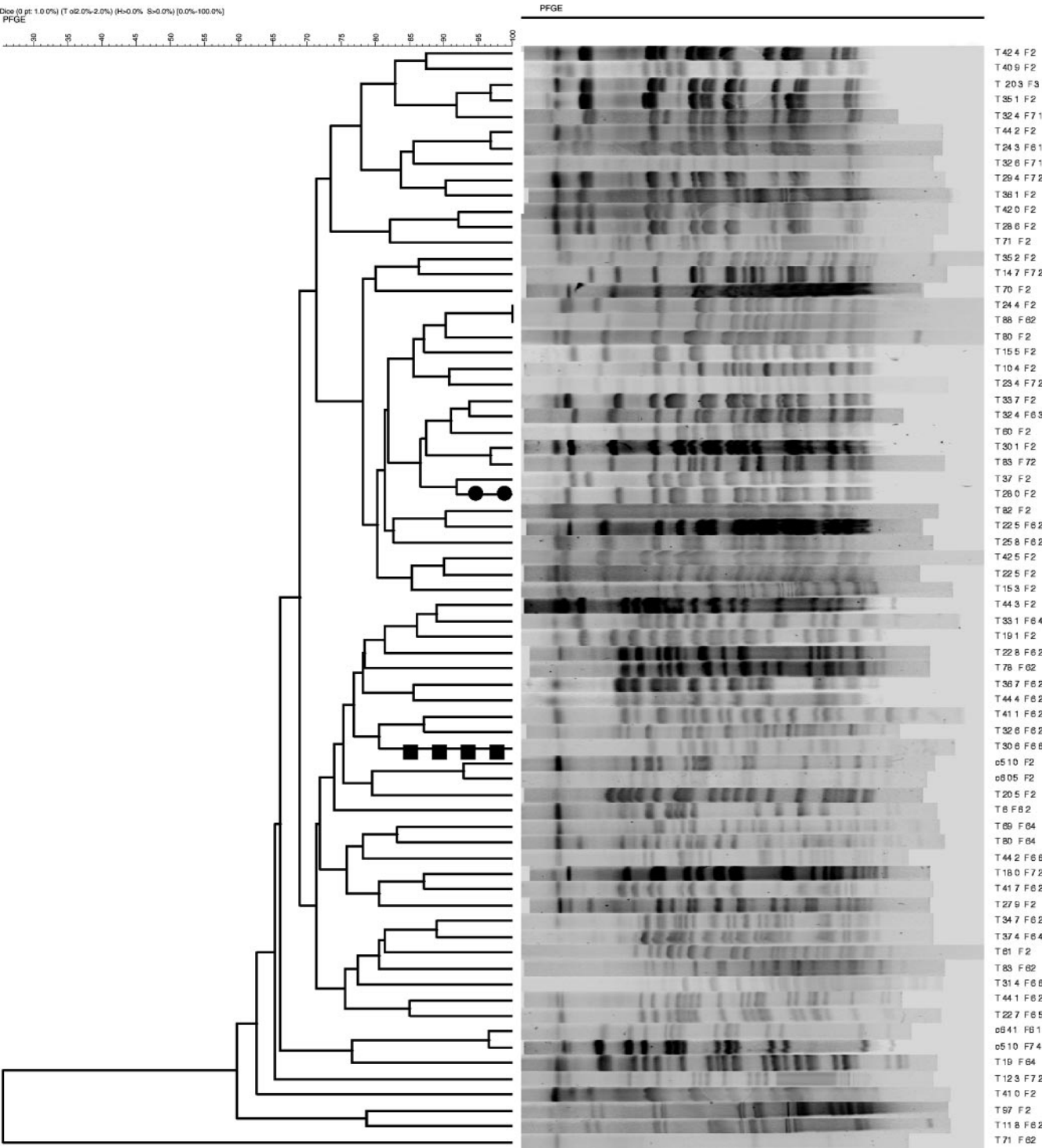


FIG. 1. Dendrogram of 43 *E. coli* isolates from the five most common virulence signatures plus 27 *E. coli* isolates found as unique virulence signatures. The filled circles show the positions of subtraction tester isolate T280F2, and the filled squares show the positions of subtraction driver isolate T306F66. The choice of the strains shown in this dendrogram is detailed in Materials and Methods.

the lowest number of known virulence genes, positive only for *fim*, *kpsMT*, and *ompT*, it had the greatest potential for identifying new virulence genes. In contrast, 27 isolates from 24 women with UTIs and one woman without a UTI had unique PFGE patterns and virulence signatures that were found only once. A dendrogram analysis of the 43 PFGE patterns seen from the five largest virulence signature groups plus the 27 patterns from the unique rectal *E. coli*

isolate is shown in Fig. 1. A representative shared isolate with a group 2 virulence signature, T280 F2, was chosen as the tester for genomic subtraction. For the subtraction driver, we selected isolate T306F66, a member of the group of unique rectal isolates with the closest match to the virulence signature of the tester (*fim*, *pff*, *aero*, *ompT*), which was phylogenetically distant based on the PFGE analysis (Fig. 1).

For epidemiologic screening, we chose 88 isolates from 85 women with UTIs (including the above-mentioned two separately shared isolates from one woman) and two without UTIs that were shared at all three sites based on PFGE patterns and 54 isolates from the original total transmission study set which were found as unique isolates by PFGE or virulence signature in the rectal flora of a woman who had at least one other isolate with a different PFGE or virulence signature pattern that was shared in at least two of the three sites.

Differential cloning by subtraction PCR. We used a commercial kit (Clontech PCR-Select bacterial genome subtraction kit; Palo Alto, CA.) to identify gene fragments specific to the tester strain through differential cloning. The genomic DNA of the driver (T306F66) was subtracted from that of the tester (T280F2) following the manufacturer's protocols to obtain tester-specific DNA. Briefly, genomic DNA was isolated from tester and driver strains, purified using phenol-chloroform extraction, and digested with *Rsa*I. Following purification of the digested DNA, the tester DNA was ligated with the adapter provided with the kit. The tester-specific DNA (sPCR) fragments were cloned into a pCR4-TOPO plasmid vector using the TOPO TA cloning kit (Invitrogen) and transformed into TOP10. The transformants were tested for inserts by a nested PCR using nested primers supplied with the genomic subtraction kit. For three of the clones, the resulting PCR products contained additional vector sequences, and for these clones an alternative approach to obtain regions lacking vector sequence was used. This involved a two-step process: first, amplification with M13 and T7 primers (35 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 2 min) followed by a nested PCR with primers provided with the subtraction kit (25 cycles of 94°C for 30 s, 68°C for 30 s, and 72°C for 2 min).

The sPCR products were spotted on nylon membranes and probed with fluorescence-labeled tester genomic DNA and driver genomic DNA using a commercially available kit (ECF random prime labeling and detection kit; Amersham). sPCR fragments that bound to both tester and driver DNA were removed from further analysis. Duplicate sPCR fragments were detected by cross-hybridization among all the tester-specific sPCR fragments, and only one of each was chosen for further analysis.

Data analysis. Differences in proportions of genes among collections were tested using the χ^2 test. Prevalence ratios were calculated as the ratio between the proportion with the gene in the collection of interest and the proportion with the gene in the unique strains isolated from the rectal flora (reference group). The magnitude of the associations of the combination sJX genetic regions (sJx) with previously known UTI virulence genes was estimated using the odds ratio and 95% confidence intervals, and the significance was determined using the χ^2 test. All statistical analyses were performed using SAS (v9.1). $P < 0.05$ was considered to be statistically significant.

Nucleotide sequence accession numbers. The GenBank accession numbers for the sPCR fragments are as follows: sJX208, DU098706; sJX113, DU098707; sJX128, DU098708; sJX129, DU098709; sJX13, DU098710; sJX150, DU098711; sJX198, DU098712; sJX204, DU098713; sJX206, DU098714; sJX210, DU098715; sJX76, DU098716; sJX77, DU098717; sJX80, DU098718; sJX83, DU098719.

RESULTS

All isolates were previously screened for the presence/absence of type1 pili (*fim*); P pili (*pff*); aerobactin (*aer*); group II capsule (*kpsMT*); cytotoxic necrotizing factor 1 (*cnfI*); the Dr family of adhesins (*drb*); hemolysin (*hly*); outer membrane protease T (*ompT*); three classes of P pili, *papG*_{J96} (class I), *papG*_{AD} (class II), and *prsG*_{J96} (class III); and S-fimbrial adhesin (*sfa*). After classifying isolates by virulence factors and genetic similarity, we selected a representative isolate from isolates found in the urine, vaginal cavity, and rectal flora of women that was positive only for *fim*, *kpsMT*, and *ompT* as the tester (T280F2), reasoning that it had the greatest potential for identifying new virulence genes (see Materials and Methods for details). As a subtraction driver isolate, we selected an isolate representative of isolates found only in the rectal flora (T306F66).

From the genomic subtraction of urine isolate T280F2 (tester) against the rectal isolate T306F66 (driver), we identified 185 sPCR fragments (Fig. 2). When separated by agarose gel electrophoresis, 60 of the sPCR products had multiple

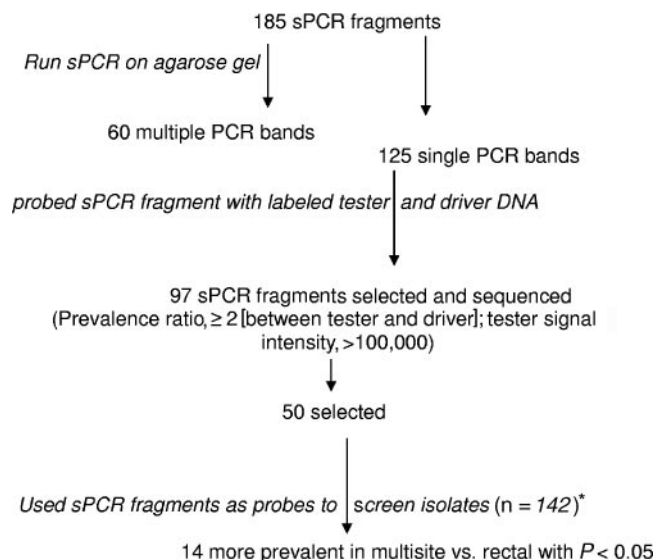


FIG. 2. Flow diagram showing how 185 sPCR fragments were analyzed to result in the discovery of 14 being significantly more common in *E. coli* isolates obtained from multiple sites than in those *E. coli* isolates uniquely found in the rectum. The asterisk shows that 88 *E. coli* isolates were present in all three sites (urine, vagina, and rectum), and there were 54 rectum-only isolates from women who also had a different *E. coli* isolate present in more than one site.

bands and were not considered further in the analysis. The remaining 125 sPCR fragments were each hybridized to labeled genomic DNA from the tester and driver strains to determine which were found in strain T280F2 but not in T306F66 (see Materials and Methods). Ninety-seven fragments were tester specific, and the genetic sequence was determined. As 97 was too large a number to reasonably screen, we selected one sPCR fragment representing each gene or operon, removing six sPCR fragments. We further reduced the number for screening by avoiding (i) genes that seemed to have housekeeping functions, (ii) genes present in the *E. coli* K-12 genome, (iii) already well characterized virulence genes, (iv) genes already found by one of our earlier subtractions that were already being analyzed (i.e., *evgS*), and (v) sPCR171, which had homology to a restriction-modification system. Finally, we preferentially chose sPCR fragments that were at least 350 bp, as larger probes work better technically in our dot blot hybridization system, although a few smaller sPCR fragments were tested. This left a total of 50 sPCR fragments. We used the 50 sPCR fragments to screen, using DNA dot blot hybridization, a collection of 88 uropathogens isolated from the rectum, urine, and vagina of women with UTIs and 54 *E. coli* isolates from the same women that were found only in the rectum. When isolates from multiple sites in a single woman were considered identical based on PFGE, only one of these isolates per woman was selected for screening. Figure 3A and B show a representative pair of dot blots for one of these probes. Figure 3C graphically represents the hybridization patterns for the duplicate blots. We used Southern blot analysis (Fig. 3D) to determine the appropriate cut point for classifying strains as containing or not containing sequences homologous to the sPCR probe. Similar analyses of all 50 sPCR probes identified

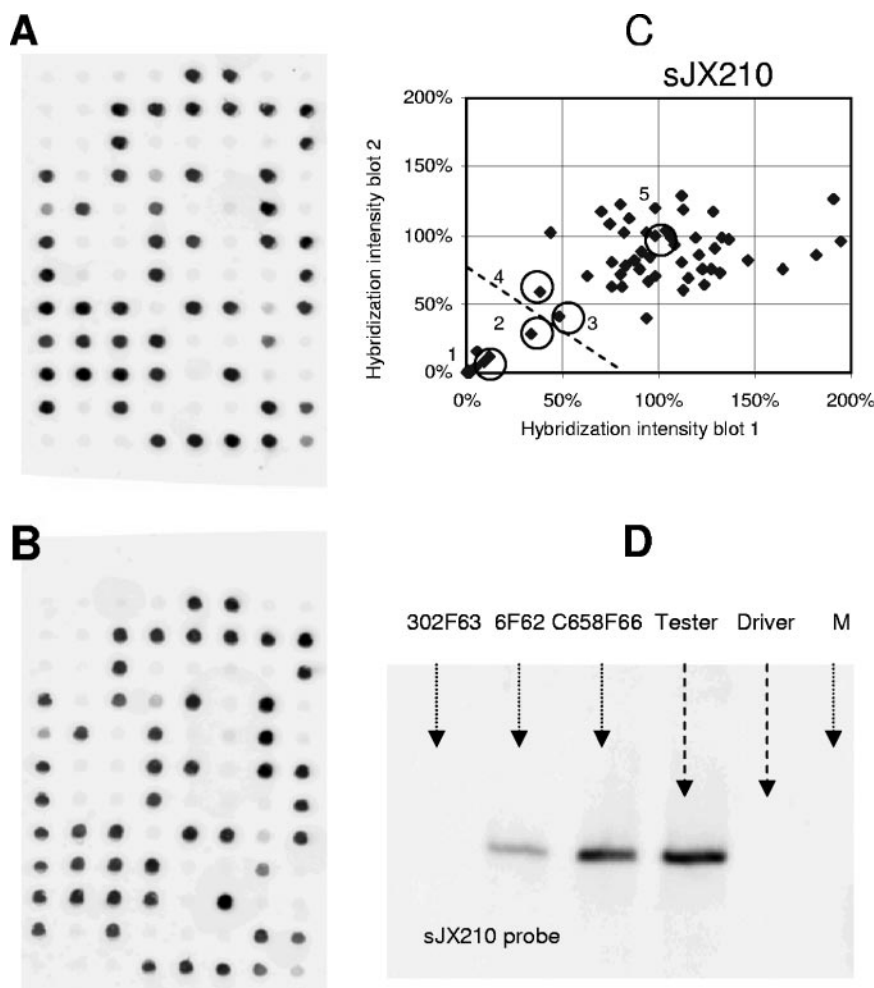


FIG. 3. (A and B) Duplicate dot blots (96 *E. coli* isolates) for probe sJX210. (C) Scatter plot of 96 *E. coli* isolates probed with sJX210 (A and B). Hybridization results are expressed as percentages of signal intensities of positive controls. Five isolates (see numbered circles in panel C) were chosen for Southern blot analysis: circle 1 surrounds duplicates of driver strain 306F66, and circle 2 (302F63), circle 3 (6F62), circle 4 (C658F66), and circle 5 surround duplicates of tester strain 280F2. (D) Southern blot assay using probe sJX210 with isolates listed; lane M contains DNA size markers.

14 sPCR fragments present significantly more often in isolates colonizing multiple sites than in rectum-only isolates (Table 1).

Thirteen of the 14 sPCR fragments present more often among isolates colonizing multiple sites were homologous to genetic regions in *E. coli* strain CFT073. Figure 4 shows their genomic locations in *E. coli* strain CFT073 and their positions relative to some previously identified uropathogenic virulence genes. While two of the sPCR fragments are near each other (the boundaries of sJX208 and sJX77 are only 870 bp apart), and two others, sJX113 and sJX76, are 5 kb apart but part of the same *auf* gene cluster, the nearest remaining combinations are at least 22 kb apart. None of the sPCR fragments were closer than 36 kb from known virulence genes (the distance between sJX198 and the *fim* gene cluster), and none were present within any of the three well-documented CFT073 pathogenicity islands, PAI II, between *danQ* and *yafV* (25); the *pap-hly* island at *pheV* (41); and the *pap* island at *pheU* (41).

Relatively little is known about most of the genetic regions in CFT073, and Table 1 lists the sizes of each sPCR fragment,

the CFT073 genes which they overlap, and the putative potential functions of those genes. sJX150 was the only sPCR fragment that was significantly more prevalent in the multiple-site collection than in the rectum-only collection that did not have a homologous sequence within CFT073, and it is part of an open reading frame with no known homologies to existing characterized proteins.

Many additional complete or partial *E. coli* genomic sequences are now available. We looked for sequence matches to the 14 sPCR fragments listed in Table 1 among the following genomic sequences: three UTI strains, CFT073 (41), F11 (GenBank accession no. NZ_AAJU000000000), and UTI89 (8); one K1 neonatal meningitis strain, R282 (39); three enteropathogenic *E. coli* strains, B171 (GenBank accession no. NZ_AAJX000000000), E110019 (GenBank accession no. NZ_AAJW000000000), and E22 (GenBank accession no. NZ_AAJV000000000); one enteroaggregative *E. coli* strain, 101-1 (GenBank accession no. NZ_AAMK000000000); two enterohemorrhagic *E. coli* strains, O157:H7 (O157Sakai) (14) and EDL933 (27); one enteroin-

TABLE 1. Prevalence of sJX subtraction candidates among 88 *E. coli* isolates from multiple sites and 54 unique rectal isolates, prevalence ratio, and putative functional homology^a

sPCR fragment ^b	Size (bp)	No. (%) of isolates		PR ^c	P value	Known CFT073 gene(s) to which homologous	Potential function(s)	Reference(s)
		Multisite (n = 88)	Rectum only (n = 54)					
sJX208	830	68 (77)	28 (52)	1.5	0.0014	C0021, C0022	C0021 = sugar hydrolase	3
sJX77	962	67 (76)	22 (41)	1.9	<0.001	C0022, C0023	Arylsulfatase, unknown	31, 41
sJX80	544	74 (84)	27 (50)	1.7	<0.001	C0168 (<i>yadL</i>), C0169 (<i>yadM</i>)	Fimbrial genes	27
sJX128	245	71 (81)	28 (52)	1.6	0.0003	Intergenic between C1819 and C1820 (<i>sitD</i>)	C1819 = prophage integrase, C1820 = iron transport	27, 41
sJX129	356	66 (75)	25 (46)	1.6	0.0005	C3752 (<i>yjiN</i>), C3753	Zinc alcohol dehydrogenase ureidoglycolate dehydrogenase	3, 21
sJX83	462	74 (84)	30 (56)	1.5	0.0002	C3775	Iron uptake receptor	27
sJX113	258	61 (69)	19 (35)	2.0	<0.001	C4207 (<i>aufG</i>)	Fimbrial adhesin	6
sJX76	400	69 (78)	20 (37)	2.1	<0.001	C4212 (<i>aufC</i>), C4213 (<i>aufB</i>)	Fimbrial usher and chaperone	6
sJX13	289	47 (53)	13 (24)	2.2	0.0006	C4839	Unknown	
sJX206	886	63 (72)	22 (41)	1.8	0.0002	C4894, C4895	Nucleoside-specific channel-forming protein; 5'-nucleotidase/2',3'-cyclic phosphodiesterase	21, 26
sJX204	196	68 (77)	21 (39)	2.0	<0.001	C5080 (<i>yddR</i>)	Nickel transporter	7
sJX210	187	54 (61)	20 (37)	1.7	0.004	C5352 (<i>yjgK</i>)	Unknown	
sJX198	907	55 (63)	20 (37)	1.7	0.0032	C5433, C5434	C ₄ -dicarboxylate transport; unknown	32, 40
sJX150	539	71 (81)	29 (54)	1.5	0.0006	Not present in CFT073	Unknown	
sJX ^d		64 (73)	12 (22)	3.3	<0.001			

^a Each multisite *E. coli* isolate was isolated from the rectal cavity, urine, and vaginal cavity of a woman, and the isolates were found to be identical using pulsed-field gel electrophoresis. One isolate of each PFGE pattern from each woman was included in this analysis. The 54 rectum-only *E. coli* isolates were isolated from the same women as were the multisite isolates but were found only in rectal samples. The subtraction probes were sJX from tester urinary isolate T280F2.

^b All sJX sequences are found in *E. coli* CFT073, except sJX150.

^c PR, prevalence ratio of probe positives in multisite isolates to those in rectum-only isolates.

^d sJX is the most common sPCR fragment in all 14 candidate sPCR signatures; it includes sJX77, sJX80, sJX83, sJX128, sJX129, sJX150, sJX204, and sJX208.

vasive *E. coli* strain, 53638 (GenBank accession no. NZ_AAKB000000000); two enterotoxigenic *E. coli* strains, B7A (GenBank accession no. NZ_AAJT000000000) and E24377A (GenBank accession no. NZ_AAJZ000000000); one human-colonizing strain, HS (GenBank accession no. NZ_AAJY000000000); and two laboratory strains, MG1655 (3) and W3110 (1, 15, 43). Table 2 shows what matches were found.

Many of the sJX genetic regions are highly prevalent, and their co-occurrence in a given *E. coli* isolate is also high (Table 3). If we look at those *E. coli* isolates that hybridized to all of the indicated eight sJX sPCRs (sJX77, sJX80, sJX83, sJX128, sJX129, sJX150, sJX204, and sJX208, combination labeled sJX), they were 3.3 times more likely to be present in *E. coli* strains isolated from multiple sites within a woman than in *E. coli* strains present only in the rectal flora of those same women. In each case where there was a significant co-occurrence between sJX and a previously known virulence gene, the odds ratio was substantially higher among multisite isolates than among rectum-only ones.

DISCUSSION

The successful movement of a particular *E. coli* bacterium from the intestinal tract to the vagina and/or bladder is likely the result of a complex combination of host behaviors, host susceptibility, and special attributes of the particular *E. coli* bacterium. Our goal was to identify *E. coli* genes that enable extraintestinal colonization. Using as our starting point a collection obtained in a study of sharing of *E. coli* isolates among heterosexual sex partners (10), we attempted to control for host behaviors and host susceptibility by limiting the comparison of our isolates found in multiple sites (rectum, vagina, and

bladder) within women to a set of *E. coli* rectum-only isolates that were present in those same women. Thus, we a priori controlled for unknown host behaviors and host susceptibilities that enable *E. coli* to colonize multiple sites.

Of the 14 resulting sPCR fragments found statistically significantly more often in *E. coli* isolates present in multiple sites than in those limited to the rectum, 13 had strong DNA homologies to sequences present in the genome of the completely sequenced pyelonephritis strain CFT073 (41) (Fig. 4). These 13 sPCR fragments appear to represent 11 different genetic regions, as sJX208 and sJX77 are located very close together and represent CFT073 genes C0021, C0022, and C0023, while sJX113 and sJX76 are found at different ends of the *auf* adhesin gene cluster (6).

These 12 new genetic regions (11 in CFT073 and 1 not in CFT073) are an interesting mixture of genes that seem to fit well into categories of known UTI virulence genes and those that are either less expected or of completely unknown function (Table 1). Adhesins make up a substantial portion of the demonstrated UPEC virulence functions, including type 1 pili (13), P fimbriae (19, 42), Dr family adhesins (24, 45), and S fimbriae (33). In addition to the *auf* gene cluster represented by sJX113 and sJX76, sJX80 contains portions of the *yadL* and *yadM* genes, which have been described as probable fimbrial genes (27). Another functional group well represented in known UPEC virulence factors is those involved in iron uptake, including the siderophore aerobactin (20, 38), synthesized by a set of *iuc* genes and transported by proteins encoded by *iut* genes; a siderophore receptor, *IroN*_{*E. coli*} (2, 18, 20, 29), and three additional genes with homology to iron uptake systems have been reported by Rasko et al. (28) to be present

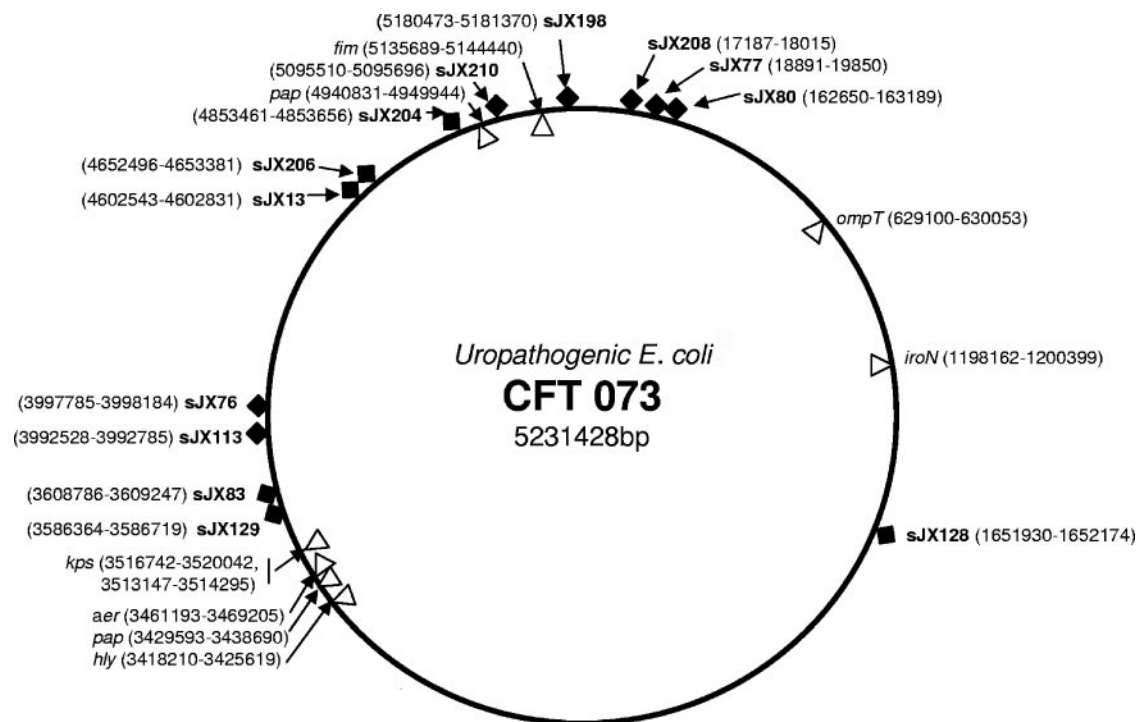


FIG. 4. Locations of 13 sJX subtraction sPCR fragments on the *E. coli* CFT073 map. Solid black boxes represent sPCR fragments present significantly more often in *E. coli* isolates found in the vagina, urine, and the rectum than in isolates found only in the rectum. Open triangles represent known virulence factors in CFT073.

more often in *E. coli* cystitis or pyelonephritis isolates than in fecal ones, and Parham et al. (25) have shown that these *fbp* genes are more common in prostatitis isolates than in cystitis or pyelonephritis isolates. Our study found two iron uptake genes that occurred more often in *E. coli* isolates present in multiple sites within a woman than in rectum-only isolates: sJX83, which is part of the C3775 gene encoding an outer membrane receptor involved in iron uptake (27), and sJX128,

which contains part of the *sitD* gene, which encodes the iron transport protein SitD (41).

Some of the probes associated with colonization in multiple sites within a woman overlap genes with homologies to genes with known functions but which have not been previously associated with UPEC virulence and whose possible roles in multisite colonization are unclear. These include sJX206, which overlaps with C4894 (*tsx*) and C4895, where the Tsx protein is a minor component of the *E. coli* outer membrane that is essential for the uptake of deoxynucleosides and nucleosides at submicromolar substrate concentrations (23), and C4895, another gene involved in nucleoside metabolism. Tsx also serves as a receptor for colicins (5) and bacteriophage (30). sJX204 covers a portion of the C5080 (*yddR*) gene, which has homology to *Yersinia* genes involved in nickel transport into bacteria. Interestingly, in pathogenic *Yersinia* species these uptake genes are adjacent to the urease gene cluster, and mutants defective in nickel uptake cannot make functional urease (31). sJX198 overlaps C5433, which has homology to C₄-dicarboxylate transport substrate binding genes, and C5434 encodes a hypothetical protein with no known homologies.

Eight of the 14 fragments occurred together in 73% of all isolates colonizing multiple sites but in only 22% of the rectum-only isolates. As the genes are spread throughout the CFT073 genome, this suggests that this combination of genes may be functionally linked. Further, the combination was strongly associated with several known UTI virulence factors, including *cnf1*, *prsG*₉₆, *hly*, *iroN*, *kpsMT*, and *sfa*, especially among the collection of isolates from multiple sites, suggesting

TABLE 2. Percent sequence matches of sJX subtraction fragments to *E. coli* genomic sequences^a

sPCR fragment	% Match to strain:						
	CFT073	F11	UTI89	R282 (K1)	B171	O157:H7	EDL933
sJX208	99	99	90	90	0	0	0
sJX77	99	99	99	99	0	0	0
sJX80	97	99	99	99	0	0	0
sJX128	98	98	98	98	0	0	0
sJX129	98	98	98	98	0	0	0
sJX83	99	99	99	99	0	98	98
sJX113	97	96	98	98	0	0	0
sJX76	100	100	100	100	0	0	0
sJX13	100	0	100	99	0	0	0
sJX206	99	99	99	99	0	0	0
sJX204	99	99	99	99	0	0	0
sJX210	100	100	100	100	0	0	0
sJX198	99	99	99	99	0	0	0
sJX150	0	0	0	0	100	0	0

^a Additional *E. coli* genomes 101-1, 53638, B7A, E110019, E22, E24377A, HS, MG1655 (K-12), and W3110 did not match any of the sPCR fragments. sJX fragments were identified from a genomic subtraction of urine isolate T280F2 (tester) against rectal isolate T306F66 (driver) (see text).

TABLE 3. Association of *sjx* (*sjx77*, -80, -83, -128, -129, -150, -204, and -208) with known UTI virulence genes among 88 *E. coli* isolates that include uropathogens isolated from urine (rectal, urine, and vaginal *E. coli* isolates identical by PFGE in a single woman) and 54 rectum-only isolates from the same individuals who had a different *E. coli* isolate present in multiple-site collections

Virulence factor(s) ^a (gene name)	Uropathogens isolated from urine (<i>n</i> = 88)		Rectum-only isolates (<i>n</i> = 54)		All <i>E. coli</i> isolates (<i>n</i> = 142)	
	OR ^b	<i>P</i> value	OR	<i>P</i> value	OR	<i>P</i> value
Aerobactin (<i>aer</i>)	1.0	0.9268	0.3	0.1649	0.8	0.5477
Capsule, group II (<i>kpsMT</i>)	6.9	0.0003	4.0	0.0382	8.5	<0.0001
Cytotoxic necrotizing factor 1 (<i>cnf1</i>)	23.0	<0.0001	2.6	0.3155	12.5	<0.0001
Dr family of adhesions (<i>drb</i>)	0.3	0.1953	1.8	0.6338	0.7	0.5727
Hemolysin (<i>hly</i>)	10.2	<0.0001	4.8	0.0406	10.4	<0.0001
IroN (<i>iroN</i>)	7.8	0.0002	2.0	0.3005	4.7	<0.0001
OmpT (<i>ompT</i>)	63.0	<0.0001	2.7	0.1623	16.9	<0.0001
P pilus family (<i>pff</i>)	3.2	0.0173	1.3	0.7500	3.3	0.0006
<i>papG_{AD}</i> (class II)	1.6	0.4004	4.3	0.0826	7.3	0.0073
<i>prsG₉₆</i> (class III)	16.8	0.0006	1.2	0.8896	9.0	<0.0001
S fimbriae (<i>sfa</i>)	29.6	<0.0001	6.5	0.0172	17.2	<0.0001

^a Known uropathogenic virulence factors.

^b The odds ratio (OR) is a measure of the strength of association between *sjx* and virulence factors. The further away that OR is from 1, the higher the association is between the two variables.

that extraintestinal movement is part of UPEC's armamentarium.

Since this is the first report of an association of these regions with UTI strains specifically, there is always a possibility that one or more of them occurred by chance alone in our analysis. However, the observation that most of these were uniquely present in the sequenced UTI and meningitis *E. coli* genomes compared to the other types of *E. coli* isolates that have been sequenced (Table 2) increases our confidence that these associations are real. It seems highly likely that our genomic subtraction has not found all of the genetic regions of interest in CFT073 or the other sequenced *E. coli* isolates that cause UTIs, as only 2 of the 12 new genetic regions were sampled twice in our sPCR set. The next useful approach will be to systematically do an "in silico" subtraction, where we determine all genes present in the majority of the UTI (and meningitis) strain genomes but absent in the majority of the commensal, laboratory, or diarrheal *E. coli* genomes. These genes can then be screened against large collections of UTI versus non-UTI isolates to determine those potentially important in uropathogenesis.

A limitation of the approach described here is that it does not find the more subtle differences between strains that will also be important in differentiating UPEC isolates from normal rectal *E. coli* isolates. Well-studied examples of these kinds of differences include phase variation of P pili (4) and type 1 fimbriae (13) and allelic sequence variations shown to be important in tissue specificity for P pili (42) and FimH (34).

The discovery of these 12 new genetic regions of *E. coli* that are more often found in *E. coli* isolates present in multiple sites within a woman than in rectum-only isolates is clearly just the

first step in determining what roles the genes that they represent may play in UPEC virulence. Additional association studies to look at their distributions in different UTI collections are needed. Ultimately, functional studies and animal virulence model studies may help further define the roles played by these genes.

ACKNOWLEDGMENTS

We thank Usha Srinivasan and Richard Bauer for their technical advice and Patricia Tallman for her assistance in the collection and maintenance of our *E. coli* collections.

REFERENCES

- Aiba, H., T. Baba, K. Hayashi, T. Inada, K. Isono, T. Itoh, H. Kasai, K. Kashimoto, S. Kimura, M. Kitakawa, M. Kitagawa, K. Makino, T. Miki, K. Mizobuchi, H. Mori, T. Mori, K. Motomura, S. Nakade, Y. Nakamura, H. Nashimoto, Y. Nishio, T. Oshima, N. Saito, G. Sampei, Y. Seki, S. Sivasundaram, H. Tagami, J. Takeda, K. Takemoto, Y. Takeuchi, C. Wada, Y. Yamamoto, and T. Horiuchi. 1996. A 570-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 28.0–40.1 min region on the linkage map. *DNA Res.* 3:363–377.
- Bauer, R. J., L. Zhang, B. F. Foxman, M. E. Jantunen, H. Saxen, and C. F. Marrs. 2002. Molecular epidemiology of three putative urinary tract infection virulence genes: *usp*, *iha*, *iroN*. *J. Infect. Dis.* 185:1521–1524.
- Blattner, F. R., G. Plunkett III, C. A. Bloch, N. T. Perna, V. Burland, M. Riley, J. Collado-Vides, J. D. Glasner, C. K. Rode, G. F. Mayhew, J. Gregor, N. W. Davis, H. A. Kirkpatrick, M. A. Goeden, D. J. Rose, B. Mau, and Y. Shao. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277:1453–1469.
- Blyn, L. B., B. A. Braaten, and D. A. Low. 1990. Regulation of *pap* pilin phase variation by a mechanism involving differential Dam methylation sites. *EMBO J.* 9:4045–4054.
- Bradley, D. E., and S. P. Howard. 1992. A new colicin that adsorbs to outer membrane protein Tsx but is dependent on the *tonB* instead of the *tolQ* membrane transport system. *J. Gen. Microbiol.* 138:2721–2724.
- Buckles, E. L., F. K. Bahrani-Mougeot, A. Molina, C. V. Lockatell, D. E. Johnson, C. B. Drachenberg, V. Burland, F. R. Blattner, and M. S. Donnenberg. 2004. Identification and characterization of a novel uropathogenic *Escherichia coli*-associated fimbrial gene cluster. *Infect. Immun.* 72:3890–3901.
- Chain, P. S. G., E. Carniel, F. W. Larimer, J. Lamerdin, P. O. Stoutland, W. M. Regala, A. M. Georgescu, L. M. Vergez, M. L. Land, V. L. Motin, R. R. Brubaker, J. Fowler, J. Hinnebusch, M. Marceau, C. Medigue, M. Simonet, V. Chenal-Francisque, B. Souza, D. Dacheux, J. M. Elliott, A. Derbise, L. J. Hauser, and E. Garcia. 2004. Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*. *Proc. Natl. Acad. Sci. USA* 101:13826–13831.
- Chen, S. W., C. S. Hung, J. Xu, C. S. Reigstad, V. Magrini, A. Sabo, D. Blasir, T. Bieri, R. R. Meyer, P. Ozersky, J. R. Armstrong, R. S. Fulton, J. P. Latreille, T. M. Hooton, E. R. Mardis, S. J. Hultgren, and J. I. Gordon. 2006. Identification of genes subject to positive selection in uropathogenic strains of *Escherichia coli*: a comparative genomics approach. *Proc. Natl. Acad. Sci. USA* 103:5977–5982.
- Foxman, B., R. Barlow, H. d'Arcy, B. Gillespie, and J. D. Sobel. 2000. Urinary tract infection: estimated incidence and associated costs. *Ann. Epidemiol.* 10:509–515.
- Foxman, B., S. Manning, P. Tallman, R. Bauer, L. Zhang, J. S. Koopman, B. Gillespie, J. D. Sobel, and C. F. Marrs. 2002. Uropathogenic *Escherichia coli* are more likely than commensal *E. coli* to be shared between heterosexual sex partners. *Am. J. Epidemiol.* 156:1133–1140.
- Foxman, B., L. Zhang, K. Palin, P. Tallman, and C. F. Marrs. 1995. Bacterial virulence characteristics of *Escherichia coli* isolates from first-time urinary tract infection. *J. Infect. Dis.* 171:1514–1521.
- Foxman, B., L. Zhang, P. Tallman, K. Palin, C. Rode, C. Bloch, B. Gillespie, and C. F. Marrs. 1995. Virulence characteristics of *Escherichia coli* causing first urinary tract infection predict risk of second infection. *J. Infect. Dis.* 172:1536–1541.
- Gunther, N. W., IV, J. A. Snyder, V. Lockatell, I. Blomfield, D. E. Johnson, and H. L. T. Mobley. 2002. Assessment of virulence of uropathogenic *Escherichia coli* type 1 fimbrial mutants in which the invertible element is phase-locked on or off. *Infect. Immun.* 70:3344–3354.
- Hayashi, T., K. Makino, M. Ohnishi, K. Kurokawa, K. Ishii, K. Yokoyama, C. G. Han, E. Ohtsubo, K. Nakayama, T. Murata, M. Tanaka, T. Tobe, T. Iida, H. Takami, T. Honda, C. Sasakawa, N. Ogasawara, T. Yasunaga, S. Kuhara, T. Shiba, M. Hattori, and H. Shinagawa. 2001. Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8:11–22.

15. Itoh, T., H. Aiba, T. Baba, K. Hayashi, T. Inada, K. Isono, H. Kasai, S. Kimura, M. Kitakawa, M. Kitagawa, K. Makino, T. Miki, K. Mizobuchi, H. Mori, T. Mori, K. Motomura, S. Nakade, Y. Nakamura, H. Nashimoto, Y. Nishio, T. Oshima, N. Saito, G. Sampei, Y. Seki, S. Sivasundaram, H. Tagami, J. Takeda, K. Takemoto, C. Wada, Y. Yamamoto, and T. Horiuchi. 1996. A 460-kb DNA sequence of the *Escherichia coli* K-12 genome corresponding to the 40.1–50.0 min region on the linkage map. *DNA Res.* 3:379–392.
16. Janke, B., J. Hacker, and G. Blum-Oehler. 2000. Genetic characterization of the uropathogenic *E. coli* strain 536—a subtractive hybridization analysis. *Adv. Exp. Med. Biol.* 485:53–56.
17. Johnson, J. R. 1991. Virulence factors in *Escherichia coli* urinary tract infection. *Clin. Microbiol. Rev.* 4:80–128.
18. Kanamaru, S., H. Kurazono, S. Ishitoya, A. Terai, T. Habuchi, M. Nakano, O. Ogawa, and S. Yamamoto. 2003. Distribution and genetic association of putative uropathogenic virulence factors *iroN*, *iha*, *kpsMT*, *ompT* and *usp* in *Escherichia coli* isolated from urinary tract infections in Japan. *J. Urol.* 170:2490–2493.
19. Manning, S. D., L. Zhang, B. Foxman, A. Spindler, P. Tallman, and C. F. Marrs. 2001. Prevalence of known P-fimbrial G alleles in *Escherichia coli* and identification of a new adhesin class. *Clin. Diagn. Lab. Immunol.* 8:637–640.
20. Marrs, C. F., L. Zhang, P. Tallman, S. D. Manning, P. Somsel, R. Raz, R. Colodner, M. E. Jantunen, A. Siitonen, H. Saxen, and B. Foxman. 2002. Variations in ten putative uropathogen virulence genes among urinary, fecal and periurethral *Escherichia coli*. *J. Med. Microbiol.* 51:138–142.
21. McClelland, M., K. E. Sanderson, J. Spieth, S. W. Clifton, P. Latreille, L. Courtney, S. Porwollik, J. Ali, M. Dante, F. Du., S. Hou, D. Layman, S. Leonard, C. Nguyen, K. Scott, A. Holmes, N. Grewal, E. Mulvaney, E. Ryan, H. Sun, L. Florea, W. Miller, T. Stoneking, M. Nhan, R. Waterston, and R. K. Wilson. 2001. Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2. *Nature* 413:852–856.
22. Miyazaki, J., W. Ba-Thein, T. Kumao, H. Akaza, and H. Hayashi. 2002. Identification of a type III secretion system in uropathogenic *Escherichia coli*. *FEMS Microbiol. Lett.* 212:221–228.
23. Nieweg, A., and E. Bremer. 1997. The nucleoside-specific Tss channel from the outer membrane of *Salmonella typhimurium*, *Klebsiella pneumoniae*, and *Enterobacter aerogenes*: functional characterization and DNA sequence analysis of the *tsx* genes. *Microbiology* 143:603–615.
24. Nowicki, B., R. Selvarangan, and S. Nowicki. 2001. Family of *Escherichia coli* Dr adhesins: decay-accelerating factor receptor recognition and invasiveness. *J. Infect. Dis.* 183(Suppl. 1):S24–S27.
25. Parham, N. J., S. J. Pollard, R. R. Chaudhuri, S. A. Beatson, M. Desvaux, M. A. Russell, J. Ruiz, A. Fivian, J. Vila, and I. R. Henderson. 2005. Prevalence of pathogenicity island II_{CFT073} genes among extraintestinal clinical isolates of *Escherichia coli*. *J. Clin. Microbiol.* 43:2425–2434.
26. Parkhill, J., G. Dougan, K. D. James, N. R. Thomson, D. Pickard, J. Wain, C. Churcher, K. L. Mungall, S. D. Bentley, M. T. G. Holden, M. Sebahia, S. Baker, D. Basham, K. Brooks, T. Chillingworth, P. Connor, A. Cronin, P. Davis, R. M. Davies, L. Dowd, N. White, J. Farrar, T. Feltwell, N. Hamlin, A. Haque, T. T. Hien, S. Holroyd, K. Jagels, A. Krogh, T. S. Larsen, S. Leather, S. Moule, P. O'Gaora, C. Parry, M. Quail, K. Rutherford, M. Simmonds, J. Skelton, K. Stevens, S. Whitehead, and B. G. Barrell. 2001. Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi CT18. *Nature* 413:848–852.
27. Perna, N. T., G. Plunkett III, V. Burland, B. Mau, J. D. Glasner, D. J. Rose, G. F. Mayhew, P. S. Evans, J. Gregor, H. A. Kirkpatrick, G. Posfai, J. Hackett, S. Klink, A. Boutin, Y. Shao, L. Miller, E. J. Grotbeck, N. W. Davis, A. Lim, E. Dimalanta, K. Potamou, J. Apodaca, T. S. Anantharaman, J. Lin, G. Yen, D. C. Schwartz, R. A. Welch, and F. R. Blattner. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7. *Nature* 409:529–533.
28. Rasko, D. A., J. A. Phillips, X. Li, and H. L. T. Mobley. 2001. Identification of DNA sequences from a second pathogenicity island of uropathogenic *Escherichia coli* CFT073: probes specific for uropathogenic populations. *J. Infect. Dis.* 184:1041–1049.
29. 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.
30. Schneider, H., H. Fsihi, B. Kottwitz, B. Mygind, and E. Bremer. 1993. Identification of a segment of the *Escherichia coli* Tss protein that functions as a bacteriophage receptor area. *J. Bacteriol.* 175:2809–2817.
31. Sebbane, F., M. A. Mandrand-Berthelot, and M. Simonet. 2002. Genes encoding specific nickel transport systems flank the chromosomal urease locus of pathogenic *Yersinia*. *J. Bacteriol.* 184:5706–5713.
32. Shirai, T., A. Suzuki, T. Yamane, T. Ashida, T. Kobayashi, J. Hitomi, and S. Ito. 1997. High-resolution crystal structure of M-protease: phylogeny aided analysis of the high-alkaline adaptation mechanism. *Protein Eng.* 10:627–634.
33. Sokolowska-Kohler, W., G. Schonian, R. Bollmann, A. Schubert, J. Parschau, A. Seeberg, and W. Presber. 1997. Occurrence of S and F1C/S-related fimbrial determinants and their expression in *Escherichia coli* strains isolated from extraintestinal infections. *FEMS Immunol. Med. Microbiol.* 18:1–6.
34. Sokurenko, E. V., V. Chesnokova, D. E. Dykhuizen, I. Ofek, X. R. Wu, K. A. Krogfelt, C. Struve, M. A. Schembri, and D. L. Hasty. 1998. Pathogenic adaptation of *Escherichia coli* by natural variation of the FimH adhesin. *Proc. Natl. Acad. Sci. USA* 95:8922–8926.
35. Srinivasan, U., B. Foxman, and C. F. Marrs. 2003. Identification of a gene encoding heat-resistant agglutinin in *Escherichia coli* as a putative virulence factor in urinary tract infection. *J. Clin. Microbiol.* 41:285–289.
36. Stamey, T. A. 1987. Recurrent urinary tract infections in female patients: an overview of management and treatment. *Rev. Infect. Dis.* 9(Suppl. 2):S195–S208.
37. Stapleton, A. E., C. L. Fennell, D. M. Coder, C. L. Wobbe, P. L. Roberts, and W. E. Stamm. 2002. Precise and rapid assessment of *Escherichia coli* adherence to vaginal epithelial cells by flow cytometry. *Cytometry* 50:31–37.
38. Torres, A. G., P. Redford, R. A. Welch, and S. M. Payne. 2001. TonB-dependent systems of uropathogenic *Escherichia coli*: aerobactin and heme transport and TonB are required for virulence in the mouse. *Infect. Immun.* 69:6179–6185.
39. University of Wisconsin *E. coli* Genome Project. <http://www.genome.wisc.edu/>.
40. Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealon, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkuch, Y. H. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304:66–77.
41. Welch, R. A., V. Burland, G. Plunkett III, P. Redford, P. Roesch, D. Rasko, E. L. Buckles, S.-R. Liou, A. Boutin, J. Hackett, D. Stroud, G. F. Mayhew, D. J. Rose, S. Zhou, D. C. Schwartz, N. T. Perna, H. L. T. Mobley, M. S. Donnenberg, and F. R. Blattner. 2002. Extensive mosaic structure revealed by the complete genome sequence of uropathogenic *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* 99:17020–17024.
42. Wullt, B. 2003. The role of P fimbriae for *Escherichia coli* establishment and mucosal inflammation in the human urinary tract. *Int. J. Antimicrob. Agents* 21:605–621.
43. Yamamoto, Y., H. Aiba, T. Baba, K. Hayashi, T. Inada, K. Isono, T. Itoh, S. Kimura, M. Kitagawa, K. Makino, T. Miki, N. Mitsuhashi, K. Mizobuchi, H. Mori, S. Nakade, Y. Nakamura, H. Nashimoto, T. Oshima, S. Oyama, N. Saito, G. Sampei, Y. Satoh, S. Sivasundaram, H. Tagami, H. Takahashi, J. Takeda, K. Takemoto, K. Uehara, C. Wada, S. Yamagata, and T. Horiuchi. 1997. Construction of a contiguous 874-kb sequence of the *Escherichia coli* K-12 genome corresponding to 50.0–68.8 min on the linkage map and analysis of its sequence features. *DNA Res.* 4:91–113.
44. Zhang, L., B. Foxman, S. D. Manning, P. Tallman, and C. F. Marrs. 2000. Molecular epidemiologic approaches to virulence gene discovery in uropathogenic *Escherichia coli*. *Infect. Immun.* 68:2009–2015.
45. Zhang, L., B. Foxman, P. Tallman, E. Cladera, C. Le Bouguenec, and C. F. Marrs. 1997. Distribution of *drb* genes coding for Dr binding adhesins among uropathogenic and fecal *Escherichia coli* isolates and identification of new subtypes. *Infect. Immun.* 65:2011–2018.