JOURNAL OF CLINICAL MICROBIOLOGY, Nov. 2006, p. 4038–4048 Vol. 44, No. 11
0095-1137/06/$08.00+0 doi:10.1128/JCM.01172-06

The *Escherichia coli argW-dsdCXA* Genetic Island Is Highly Variable, and *E. coli* K1 Strains Commonly Possess Two Copies of *dsdCXA*

Rebecca L. Moritz and Rodney A. Welch*

Department of Medical Microbiology and Immunology, University of Wisconsin—Madison, Madison, Wisconsin 53706

Received 7 June 2006/Returned for modification 31 July 2006/Accepted 3 August 2006

The genome sequences of *Escherichia coli* pathotypes reveal extensive genetic variability in the *argW-dsdCXA* island. Interestingly, the archetype *E. coli* K1 neonatal meningitis strain, strain RS218, has two copies of the *dsdCXA* genes for D-serine utilization at the *argW* and *leuX* islands. Because the human brain contains D-serine, an epidemiological study emphasizing K1 isolates surveyed the *dsdCXA* copy number and function. Forty of 41 (97.5%) independent *E. coli* K1 isolates could utilize D-serine. Southern blot hybridization revealed physical variability within the *argW-dsdC* region, even among 22 *E. coli* O18:K1:H7 isolates. In addition, 30 of 41 K1 strains, including 21 of 22 O18:K1:H7 isolates, had two *dsdCXA* loci. Mutational analysis indicated that each of the *dsdA* genes is functional in a rifampin-resistant mutant of RS218, mutant E44. The high percentage of K1 strains that can use D-serine is in striking contrast to our previous observation that only 4 of 74 (5%) isolates in the diarrheagenic *E. coli* (DEC) collection have this activity. The genome sequence of diarrheagenic *E. coli* isolates indicates that the *csrRAKB* genes for sucrose utilization are often substituted for *dsdC* and a portion of *dsdX* present at the *argW-dsdCXA* island of extraintestinal isolates. Among DEC isolates there is a reciprocal pattern of sucrose fermentation versus D-serine utilization. The ability to use D-serine is a trait strongly selected for among *E. coli* K1 strains, which have the ability to infect a wide range of extraintestinal sites. Conversely, diarrheagenic *E. coli* pathotypes appear to have substituted sucrose for D-serine as a potential nutrient.

It is remarkable that aside from the normal intestinal commensal strains, there are genotypes of *Escherichia coli* that can cause different diseases in the intestine and others that infect and cause disease at extraintestinal sites. This diversity suggests that different genotypes of *E. coli* have the ability to respond successfully to different environments by expression of relevant colonization factors, immune avoidance mechanisms, and advantageous nutrient acquisition strategies. Comparison of the sequenced genomes of uropathogenic *E. coli* (UPEC) strain CFT073, laboratory K-12 strain MG1655, and enterohemorrhagic *E. coli* (EHEC) strain EDL933 showed that only 39.2% of the predicted proteins were common among all three strains (40). Aside from virulence- and colonization-related genes that are specific to the niches for each strain, these strains vary in some specific catabolic capabilities. Our laboratory has recently focused on the *argW-dsdCXA* chromosomal region that corresponds to the *E. coli* K-12 MG1655 53′ genetic map position. In this region, EDL933 possesses phage-like genes and *cscRAKB*, which permit non-phosphotransferase (PTS)-mediated sucrose utilization; and MG1655 and CFT073 have genes (*dsdCXA*) that encode the ability to use D-serine as a sole carbon and nitrogen source. Until recently, it was thought that naturally occurring D-amino acids were found only in bacterial cell walls and capsules. However, studies indicate that D-serine is found in the urine of humans at concentrations that range from 3 to 40 μg/ml (9). The D-serine deaminase (encoded by *dsdA*) converts D-serine to ammonia and pyruvate (6). *dsdX* and *dsdC* are a specific D-serine permease (2a) and a LysR-like transcriptional regulator, respectively, that control the expression of *dsdXC* and a *dsdXA* operon (33). In the absence of *dsdCXA*, D-serine concentrations greater than 50 μg/ml in growth medium are bacteriostatic for the growth of *E. coli* K-12 where D-serine is transported by CycA and inhibits L-serine and pantothenate synthesis (11, 12, 31). Our previous study demonstrated that urinary tract infection (UTI) and urosepsis *E. coli* isolates were more likely than isolates in the diarrheagenic *E. coli* (DEC) collection to catabolize D-serine as a sole carbon source (36). This suggests that *dsdCXA* provides a growth advantage for uropathogenic *E. coli* strains in the urinary tract. It was also demonstrated that although a CFT073 *dsdA* mutant had a growth defect during growth in human urine compared to the growth ability of its wild-type parent, the same mutant surprisingly demonstrated nearly a 1,000-fold advantage in its ability to colonize the bladders and kidneys of mice (36).

UPEC is one of two groups within the extraintestinal *E. coli* (ExPEC) pathotype which cause infections outside of the intestinal tract, such as UTIs, pneumonia, sepsis, and meningitis. UPEC strains are responsible for greater than 80% of all community-acquired urinary tract infections. Neonatal meningitis *E. coli* (NMEC) strains are the second major group among ExPEC isolates and are a leading cause of neonatal meningitis and sepsis (4, 7). Roughly 80% of NMEC isolates are of the K1 capsular serotype. NMEC strains, especially O18:K1:H7 strains, are notable because they are also a common cause of UTIs (27, 28, 32). Twenty percent of cases of neonatal meningitis are a complication of a UTI in the neonate (3, 8). Despite intensive research on specific ExPEC virulence factors, such as type 1 fimbriae and hemolysin, the patterns of gene expression among ExPEC isolates in their different niches

* Corresponding author. Mailing address: Medical School, University of Wisconsin, Madison, WI 53706. Phone: (608) 263-2700. Fax: (608) 262-8418. E-mail: rawelch@wisc.edu.

are only beginning to be understood. For example, together with the laboratories of H. L. Mobley and M. S. Donnenberg, we recently demonstrated the in vivo transcriptome of UPEC strain CFT073 shed in the urine of mice with experimental UTI (37). Compared to the transcriptome of CFT073 grown in complex laboratory medium, the bacteria infecting the urinary tract are starved for nitrogen and iron. Genes for peptide and amino acid utilization are up-regulated during infection. Sugars such as glucose, fructose, and sucrose are generally not found in urine, whereas peptides and amino acids are present. Overall, this suggests that during UTIs, amino acids such as D-serine are being used by UPEC isolates as carbon and nitrogen sources.

The original evidence that D-serine is present in mammals came from neurological studies that identified it as a neurotransmitter. D-Serine is found at nanomolar concentrations in the mammalian brain, where it acts as a potent and selective agonist at the *N*-methyl-D-aspartate-type excitatory amino acid receptor (20). NMEC strains have the unique ability among *E. coli* isolates to cross the blood-brain barrier (BBB) and cause meningitis in neonates (29). These strains cross the intestinal lumen to gain access to the bloodstream and to cause bacteremia and are present at levels of at least 10^3 CFU/ml (16). Upon reaching the BBB, NMEC strains rearrange the actin cytoskeleton of brain microvascular endothelial cells and invade the brain, causing disease (30). Aside from the requirement for expression of the K1 capsule and a few other factors, little is known about the virulence mechanisms and the pattern of bacterial gene expression involved in these events (30). Environmental conditions (pH, iron concentration, and osmolarity) have been shown to play a role in the ability of NMEC strains to invade brain microvascular endothelial cells (4). We hypothesize that NMEC strains use D-serine as a carbon and nitrogen source for growth in the brain or as an environmental signal that permits the expression of genes relevant to their location in the host. This hypothesis arose from our analyses of the genome sequence of the archetype K1 strain, strain RS218, which possesses two unlinked copies of *dsdCXA.* In order to examine these hypotheses, we assessed the ability of a collection of *E. coli* K1 isolates to utilize D-serine and compared the prevalence of intact copies of the *dsdCXA* locus by Southern blot hybridization analysis. The genetic island between the *argW* and *dsdCXA* loci was found to be a hypervariable region of the *E. coli* chromosome. Surprisingly, a majority of the K1 strains, including known NMEC strains, possess a second *dsdCXA* locus at the *leuX* island at position 97′. The two *dsdCXA* loci in strain E44, a spontaneous rifampin-resistant mutant of RS218, were both shown to be functional for the use of D-serine as a sole carbon source. Lastly, a survey of the ability of diarrheagenic *E. coli* isolates in the DEC collection reveals that unlike the extraintestinal *E. coli* strains, few DEC strains can use D-serine and instead have the ability to ferment sucrose.

MATERIALS AND METHODS

Bacterial strains and growth conditions. Table 1 describes the *E. coli* strains used in this study. The strains were acquired from several sources and varied in their clinical origins. Strains were grown in Luria-Bertani (LB) broth or agar at 37°C unless otherwise specified. Chloramphenicol (20 μg/ml) and kanamycin (50 μg/ml) were added to the broth and agar media when needed. To test for the ability to grow on D-serine as a sole source of carbon, morpholinepropanesulfonic acid (MOPS)–D-serine minimal medium was used, as described previously (36). Because nicotinamide auxotrophy frequently occurs among K1 strains, 5 μg/ml nicotinamide was included in the MOPS minimal medium (1, 27).

Sequence analysis. The unpublished *E. coli* RS218 sequence determined by the Genome Sequencing Center at the University of Wisconsin was used to compile *dsdCXA* regions at the *argW* and *leuX* genetic islands from the BLAST website (www.genome.wisc.edu/sequencing/rs218.htm) (2), and ORF Finder (41) from the National Center for Biotechnology Information was used to determine the identities of coding regions and the locations of predicted proteins. The RS218 closed sequence was later used to confirm our compilation and annotation (Guy Plunkett, personal communication). Available sequences (GenBank accession numbers) for CFT073 (AE14075), EDL933 (AE005174), MG1655 (U00096), pathogenicity-associated island II in UPEC strain 536 (PAI II_{536}) (AJ494981), *E. coli* HS (AAJY00000000), *E. coli* B171 (AAJX00000000), *E. coli* E24377A (AAJZ00000000), *E. coli* F11 (AAJU00000000), *E. coli* 53638 (AAKB00000000), *E. coli* B7A (AAJT00000000), *E. coli* E11019 (AAJW00000000), *E. coli* E22 (AAJV00000000), *Shigella flexneri* (AE14073), *Shigella dysenteriae* (CP000034), *Shigella boydii* (CP000036), *Shigella sonnei* (CP000038), and *E. coli* EC3132 (AF473544) (which contains *cscRAKB*) (23) were downloaded from GenBank. The genome sequence for *E. coli* O42 was obtained at the Wellcome Trust Sanger Institute website. The genome sequences of HS and E24377A are closed but unannotated at present. The sequences of B171, F11, 53638, B7A, E11019, and E22 are unclosed and unannotated.

Southern blot hydridizations. Genomic DNA was isolated from overnight cultures grown at 37°C in LB broth with aeration by using a Genomic DNA isolation kit (Promega). The DNA was digested with EcoRI and EcoRV for 25 min and then heat inactivated for 20 min at 80°C. The digested DNA was run on a 0.8% agarose gel, transferred to Hybond N+ membranes (Amersham Pharmacia Biotech Inc.), and UV cross-linked to the membrane. The appropriate DNA hybridization probe was made from the predicted coding region of *dsdC*, *dsdX*, *dsdA*, *ipuA*, or *ipuB* and labeled with α-^{32}P by using a Prime-It random primer labeling kit (Stratagene). Hybridization was done at 60°C overnight. The hybridization signals on the blots were visualized by using a Typhoon Imager 8600 phosphorimager.

PFGE analyses. The bacterial strains were grown in LB broth to an optical density at 600 nm of 0.85. The cells were embedded in agarose as described previously (21). Briefly, the cells were harvested, pelleted by centrifugation, and resuspended in PIV buffer (1 M NaCl, 10 mM Tris-HCl [pH 7.6]) and 1.6% InCert agarose (FMC BioProducts) in water and then placed in plug molds (Bio-Rad). The plugs were incubated overnight at 37°C with gentle shaking in EC lysis solution (6 nM Tris-HCl [pH 7.6], 1 M NaCl, 100 nM EDTA [pH 7.5], 0.5% Brij 58, 0.2% deoxycholate, 0.5% Sarkosyl, 1 mg/ml lysozyme, 20 μg/ml RNase). The EC lysis solution was removed and replaced with ESP solution (0.5 M EDTA [pH 9 to 9.5], 1% Sarkosyl, 50 μg/ml proteinase K), and the mixture was incubated at 50°C with gentle shaking overnight. The plugs were then washed with TE three times with shaking for 30 min and stored at 4°C until needed. The plugs were digested with 10 units of AvrII (New England Biolabs) for 3 to 18 h and then washed with TE for 30 min. Each digested plug was inserted and sealed into a well of a 1% pulsed-field gel electrophoresis (PFGE)-grade agarose gel (Bio-Rad) made with 0.5× TBE along with a PFGE lambda ladder (Bio-Rad). PFGE was performed with a CHEF Mapper XA system (Bio-Rad). The gels were run at 14°C with one of two cycles, depending on the desired separation size and length of the gel. For short-run gels, the initial switch was 2.16 s, the final switch was 54.17 s, the angle was 120°, and the gradient was 6.0 V/cm with linear ramping for a total run time of 22 h (19). For long-run gels, the initial switch was 2.16 s, the final switch was 1 min 4.41 s, the angle was 120°, and the gradient was 6.0 V/cm with linear ramping for a total run time of 26 h. The separated DNA fragments were transferred from the gels to Hybond N+ membranes (Amersham Pharmacia Biotech Inc.) by using 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and UV cross-linked to the membrane. The blot was analyzed with an α-^{32}P-labeled 858-bp DNA probe consisting of *argW* and the region 5′ to *argW* through to the last 708 bp of *yfdC*, a gene that is present in all *E. coli* strains sequenced. The blots were visualized as described above.

Construction of E44 *dsdA* mutants. The parent strain for all mutant constructions was *E. coli* strain E44, which is a spontaneous rifampin-resistant mutant derived from cerebrospinal fluid (CSF) isolate RS218 (O18:K1:H7) (1, 39). Single and double *dsdA* mutations were constructed by the λ-Red recombination method described by Datsenko and Wanner (14). Both E44 *dsdA* genes were deleted individually by using chloramphenicol (pKD3) and kanamycin (pKD4) resistance gene cassettes. Linear DNA was made by using the following PCR primers, which were identical to the 5′ and 3′ ends of *dsdA*: forward primer P0 (5′CCTGCTGCTATTTATCATCTAAGCGCAAAGAGACGTACTTGTGTA GGCTGGAGCTGCTTCG-3′), reverse primer P2 (5′-CACCCAGGGAAAGG

TABLE 1. D-Serine and sucrose phenotypes for *Escherichia coli* isolates

Strain (WAM no.)	Original name	Serotype	Source	Growth on D-serine	Sucrose utilization	Source (reference)
2659	B50	K1	CSF	−	+	J. Robbins, NIH
2660	B200	K1	CSF	+	−	J. Robbins, NIH
2661	GM84	K1	CSF	+	−	J. Robbins, NIH
2662	468624	K1	CSF	+	+	J. Robbins, NIH
2663	RS218	$O18_{ac}$:K1:H7	CSF	+	−	J. Robbins, NIH
2664	Marriot	K55	CSF	+	+	J. Robbins, NIH
2665	Cleveland	$O18_{ac}$:K1:H7	CSF	+	−	J. Robbins, NIH
2666	SLC	K1	Blood	+	−	J. Robbins, NIH
2667	Kansas City	K1	CSF	+	−	J. Robbins, NIH
2668	B82 Marriot	O4:K1:H7	CSF	+	−	J. Robbins, NIH
2937	N551	O?:K_{neg}:H5	Blood	+	−	J. Badger (13)
2938	E780	O16:K1	CSF	+	−	J. Badger (13)
2939	E407	O16:K1	CSF	+	−	J. Badger (13)
2940	E496	O16:K1:H7	CSF	+	−	J. Badger (13)
2941	E784	O16:K1	CSF	+	−	J. Badger (13)
2942	E912	O18:K1:H7	CSF	+	−	J. Badger (13)
2943	E221	O18:K1:H7	CSF	+	−	J. Badger (13)
2944	E253	O12:K1	Blood	+	−	J. Badger (13)
2945	E251	O12:K1:H−	CSF	+	−	J. Badger (13)
2946	EC10	O7:K1:H−	CSF	+	+	J. Badger (13)
2947	RS167	O16:K1:H6	CSF	+	−	J. Badger (13)
2948	E115	O18:K51	CSF	−	+	J. Badger (13)
2949	A63	O18:K5:H7	CSF	+	+	J. Badger (13)
2950	RS168	O1:K1:H−	CSF	+	−	J. Badger (13)
2951	E606	O18:K−	CSF	−	+	J. Badger (13)
2952	E334	?:K1	CSF	+	−	J. Badger (13)
2953	E308	O12:K1	CSF	+	−	J. Badger (13)
2954	E44	$O18_{ac}$:K1:H7	CSF	+	−	J. Badger (13)
3228	U60	O18:K1:H7	Urine	+	−	J. Johnson (27)
3229	U6	O18:K1:H7	Urine	+	−	J. Johnson (27)
3230	U74	O18:K1:H7	Urine	+	−	J. Johnson (27)
3231	U88	O18:K1:H7	Urine	+	−	J. Johnson (27)
3232	U91	O18:K1:H7	Urine	+	−	J. Johnson (27)
3233	S8	O15:nonK1	Stool	+	+	J. Johnson (27)
3234	S92	O25:nonK1	Stool	+	−	J. Johnson (27)
3235	U92	O18:K1:H7	Urine	+	−	J. Johnson (27)
3236	S39	O18	Stool	+	−	J. Johnson (27)
3237	S13	O18	Stool	+	−	J. Johnson (27)
3238	U33	O18:K1:H7	Urine	+	−	J. Johnson (27)
3239	U64	O18:K1:H7	Urine	+	−	J. Johnson (27)
3240	U13	O18:K1:H7	Urine	+	−	J. Johnson (27)
3241	S98	O18	Stool	+	−	J. Johnson (27)
3242	U39	O18:K1:H7	Urine	+	−	J. Johnson (27)
3243	U7	O18:K1:H7	Urine	+	−	J. Johnson (27)
3244	U94	O18:K1:H7	Urine	+	−	J. Johnson (27)
3245	U80	O18:K1:H7	Urine	+	−	J. Johnson (27)
3246	U45	O18:K1:H7	Urine	+	−	J. Johnson (27)
3247	S111	Ont	Stool	+	+	J. Johnson (27)
3248	U8	O18:K1:H7	Urine	+	−	J. Johnson (27)
3250	H16	O18:K1:H7	CSF	+	−	J. Johnson (27)
3251	C5	O18:K1:H7	CSF	+	−	J. Johnson (27)
3252	89-1449	O18:K1:H7	CSF	+	−	J. Johnson (27)
2645	NU14	O18:K1:H7	Cystitis	+	+	S. Hultgren
2039	536	O6:K15:H31	Pyelonephritis	+	+	J. Hacker
2267	CFT073	O6:K2:H1	Pyelonephritis	+	−	This lab
2625	MG1655	OR:H48	Fecal isolate	+	−	This lab
2392	EDL933	O157:H7	Diarrheal isolate	−	+	This lab
2265	CP9	O4:H5	Bacteremia	+	−	T. Russo
1218	J96	O4:H5	Pyelonephritis	+	−	This lab

ATGGCGATGCTGCGTTGAAACGTTACATATGAATATCCTCCTTAG), and reverse primer P2 (5′-CGTAAAAAGGGAGTCGATGTGGCAAAATCA TTAGTGCCCCTTACATATGAATATCCTCCTTAG). This amplified DNA was electroporated into the appropriate strain, the mutations were screened by PCR, the resistance cassette was removed, and the recombinase plasmid was cured by high-temperature growth as described previously (14). Complementation of the *dsdA* mutations was accomplished by first cloning the *dsdA* gene with the Shine and Delgarno sequence intact into pACYC177 (36). This plasmid, pWAM2682, was transformed into E44 *dsdA* double-knockout strain WAM3049, and its ability to grow on D-serine as a sole carbon source was confirmed.

Sucrose fermentation. Sucrose utilization was monitored and was indicated by the presence of red-pink colony formation following overnight, aerobic

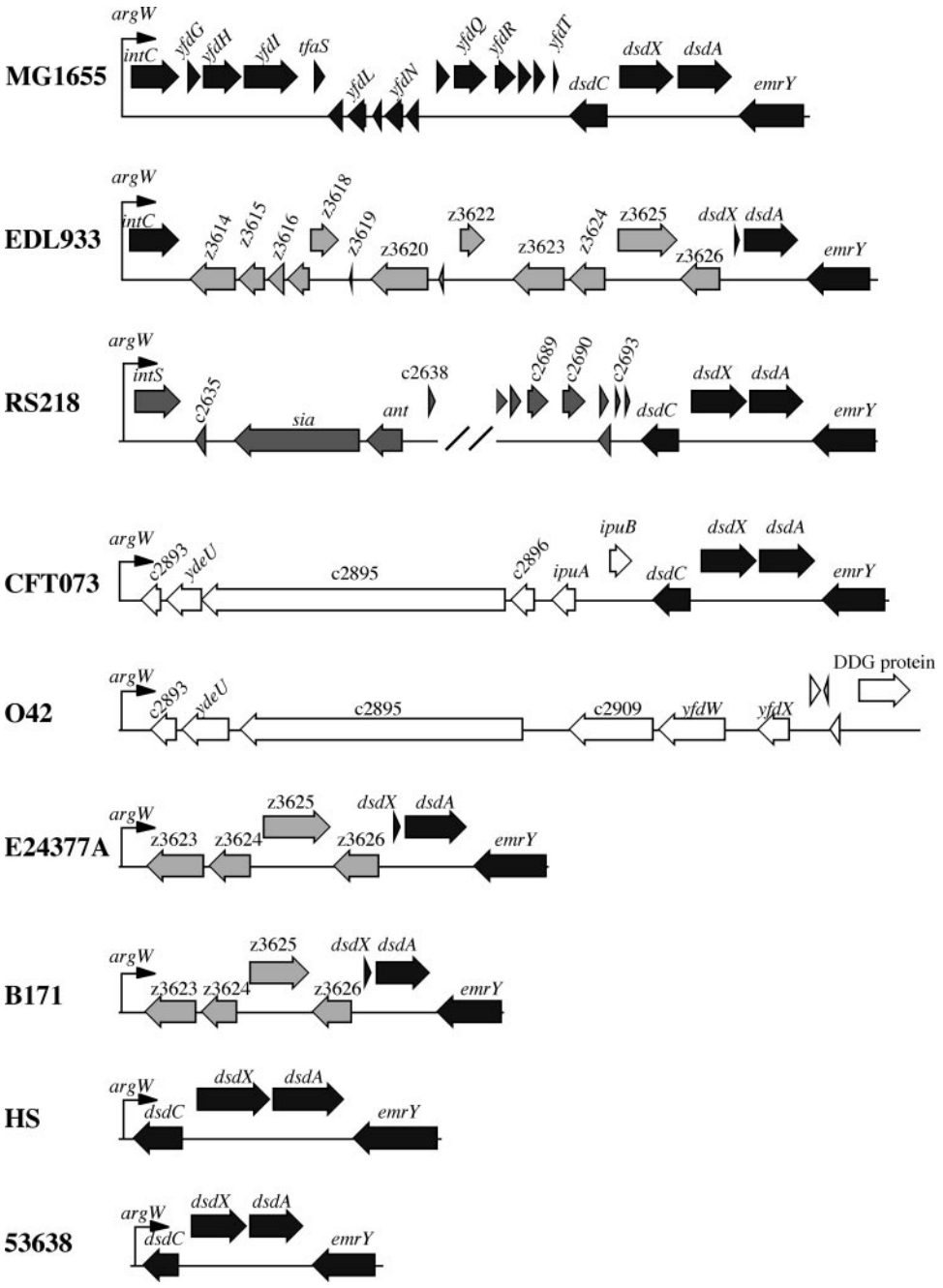

FIG. 1. Representation of *dsdCXA* loci *argW* genetic islands in the sequenced *Escherichia coli* genomes. Black arrows, genes found in MG1655; open arrows, genes present in CFT073; light gray arrows, genes found in EDL933; dark gray arrows, genes present in RS218; hatched marks in the map for RS218, 29,922 bp.

growth at 37°C on MacConkey agar-based medium (Difco) supplemented with 1% sucrose.

RESULTS

Sequence analysis of *E. coli argW-dsdCXA* regions. Our genome sequence comparison showed that *E. coli* strains CFT073, MG1655, and EDL933 all differ in genetic content at position 53′ in the *argW-dsdCXA* region (Fig. 1) (36). Specifically, the differences among the strains start from the 3′ end of the *argW* tRNA gene and continue through to the last 208 bases of *dsdX*. All sequenced strains contained the gene *yfdC* 5′ to *argW* and the gene *emrY* 3′ to *dsdA*. CFT073 contained 12,271 bp between *argW* and *dsdC*. This area consists of two FimBE-like recombinases, *ipuA* and *ipuB*, and hypothetical genes (10, 36). *E. coli* O157:H7 strain EDL933 had 14,124 bp between *argW* and the 208 bases of the 3′ end of *dsdX*. The EDL933 *argW-dsdX* region contained, in respective order 3′ to *argW*, an incomplete set of phage genes, hypothetical genes of unknown function, and the non-PTS-mediated sucrose utilization genes *csrRAKB* (23, 36). MG1655 had 10,311 bp of puta-

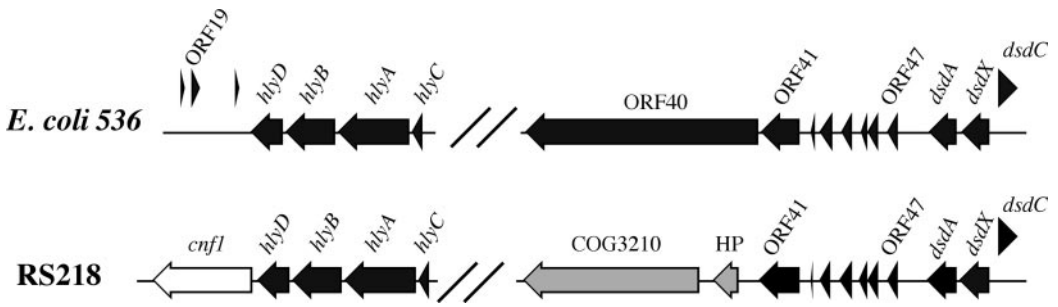

FIG. 2. Representation of *dsdCXA* loci at *leuX* genetic islands in ExPEC strains 536 and RS218. The genetic map for the strain 536 *leuX* genetic island covers approximately the 22,000- to 32,000-bp and 47,000- to 69,000-bp portions of the genetic island sequenced (17). The section presented for the strain RS218 *leuX* genetic island covers approximately the 42,000- to 54,000-bp and 69,000- to 90,000-bp genetic island sequenced. The tRNA *leuX* and *prf* genes are 5′ of the *hlyCABD* operon in the genetic islands of both strains. Black arrows, genes annotated in strain 536; open arrows, genes in RS218; gray arrows, open reading frames found in other strains. The internal hash marks cover ORFs 26 to 39, which are identical in both strains.

tive P22-like prophage genes and hypothetical genes, none of which encoded proteins similar to those in either of the two other strains. The original genetic and sequence characterization of the chromosomal *csrRAKB* genes of *E. coli* strain EC3132 by the laboratory of Lengeler identified that there was a disruption of the *dsdCXA* locus with the substitution of the non-PTS sucrose utilization locus (23). That region of strain EC3132 is similar to that of EDL933; however, the EDL933 phage-like genes were absent (data not shown).

Recently, the genomes of additional *E. coli* strains, including different pathotypes of typical and atypical enteropathogenic (EPEC), enterotoxigenic (ETEC), enteroinvasive (EIEC), and enteroaggregative (EAEC) strains, have been sequenced by TIGR and the Wellcome Trust Sanger Institute. Although these genome sequences are not fully closed or annotated, analysis of the *argW-dsdCXA* regions of those genomes showed additional variation in genetic content (Fig. 1). Because it has been argued that *Shigella* species are essentially pathotypes of *E. coli* (35), we included in the genome comparisons the *argW-dsdCXA* regions from *S. dysenteriae*, *S. sonnei*, *S. boydii*, and *S. flexneri*. Also included in the genome sequence comparisons was *E. coli* strain HS, a normal human fecal isolate of *E. coli*, which has been used as a negative disease control strain in oral challenges of human volunteers (18). Strain HS and strain 53638, an EIEC strain, had no apparent genes encoded between the 3′ end of *argW* and the intact *dsdCXA* locus. *dsdC* was 97 bp downstream of *argW* in HS and 110 bp downstream in strain 53638. The *argW-dsdCXA* regions of the three EPEC strains B171, E11019 (an atypical EPEC strain), and E22 (an EPEC-like rabbit diarrheal *E. coli* strain) were similar to the *argW-dsdCXA* region of EDL933; they possessed the sucrose utilization genes, a partial *dsdX* gene, and a full *dsdA* gene; but no hypothetical proteins or phage genes are present within a 440-bp region between *argW* and the *csrRAKB* locus. Interestingly, the *argW-dsdCXA* region in ETEC strain E24377A was very similar to those in the three EPEC strains, but it had a larger 1,690-bp region between *argW* and the *cscRAKB* genes. Strain O42, an EAEC strain, had the same sequence as CFT073 from *argW* through to the 3′ end of c2895. There then appeared to be a deletion of the entire *dsdCXA*, *emrYK*, and *evgAS* region in strain O42. It is recognized that the O42 genome sequence has not been closed, so the apparent deletion may be an assembly artifact.

The genome of the archetype NMEC strain, strain RS218, is being sequenced by the laboratory of F. R. Blattner at the University of Wisconsin. Sequence information available at the website for this laboratory, along with personal communications, confirmed that this strain has 40,587 bp of HK620-, Sf6-, and HK97-like phage genes inserted between *argW* and the *dsdCXA* locus (Fig. 1). The presence of the phage genes and the presence of a K1 capsular modification gene, *neuO*, were recently noted by the laboratory of E. R. Vimr (15). Our analysis of the RS218 genome sequence also found a second *dsdCXA* locus associated with the *leuX* genetic island at position 97′. This second *dsdCXA* locus is similar (with *dsdA* having 99% nucleotide identity) to one found on the *leuX* PAI II_{536}. PAI II_{536} is 102,200 bp in length and contains *hly* genes, *prf* genes, and a previously unannotated copy of *dsdCXA* (Fig. 2) (17). The RS218 version of this island is 122,709 bp and carries all of the same genes as PAI II_{536} plus the notable addition of *cnf1*. The RS218 and 536 *leuX* genetic islands differed between *prfI* and the 3′ end of *hlyD*, where RS218 had an additional 21,538 bp that included *cnf1* immediately downstream of the *hlyCABD* operon. Another major difference between the two *leuX* genetic islands was the lack of 536 open reading frames (ORFs) y1093 and y1094 in the RS218 island.

Two *E. coli* genomes had genetic similarities to RS218 in the *argW-dsdCXA* region. B7A, an ETEC isolate, had 45,108 bp of HK620- and Sf6-like phage genes and hypothetical genes within the *argW-dsdCXA* area that were similar to the RS218 phage genes. It is striking that the *argW-dsdCXA* area in strain B7A has the phage genes along with the sucrose utilization genes, a partial *dsdX* gene, and a full *dsdA* gene, similar to EDL933. UPEC cystitis isolate F11 has a *dsdCXA* locus similar to that of RS218 within a *leuX* genetic island at position 97′; yet at the *argW-dsdCXA* region, the arrangement is very similar to those of the three EPEC strains described above, with the sucrose utilization genes and the 3′ ends of *dsdX* and *dsdA* present.

Interestingly, none of the *Shigella* strains sequenced had an intact *dsdCXA* locus, and the sequences of the strains in the *argW-dsdC* area differed. *S. dysenteriae* was similar to EDL933,

TABLE 2. Phenotypes of diarrheal *E. coli* isolates (DEC collection)[a]

Strain	Serotype	Growth on D-serine	Sucrose fermentation	Strain	Serotype	Growth on D-serine	Sucrose fermentation
1A	O55:H6	−	+	8C	O111:HN	−	−
1B	O55:H6	−	+	8D	O111:H11	−	+
1C	O55:H6	−	+	8E	O111:H8	−	−
1D	O55:H6	−	+	9A	O26:H11	−	+
1E	O55:H6	−	+	9B	O26:HN	−	+
2A	O55:H6	−	+	9C	O26:HN	−	+
2B	O55:HN	−	+	9D	O26:H11	−	+
2C	O55:H6	−	+	9E	O26:H11	−	+
2D	O55:H6	−	+	10A	O26:H11	−	+
2E	O55:H6	−	+	10B	O26:H11	−	+
3A	O157:H7	−	+S	10C	O26:H11	−	+
3B	O157:H7	−	+S	10D	O26:H11	−	+
3C	O157:H7	−	+S	10E	O26:H11	−	+
3D	O157:H7	−	+S	10F	O15:HN	NT	NG
3E	O157:H7	−	+S	11A	O128:H2	−	+
3F	O157:HN	NT	NG	11B	O128:H2	−	+
4A	O157:H7	−	+S	11C	O45:H2	−	+
4B	O157:H7	−	+S	11D	O128:H2	−	+
4C	O157:H7	−	+S	11E	O128:H2	−	+
4D	O157:H7	−	+S	12A	O111:H2	−	+
4E	O157:H7	−	+S	12B	O111:H2	−	+
4F	O157:H7	M	M	12C	O111:HN	−	+
5A	O55:H7	−	+S	12D	O111:H2	−	+
5B	O55:H7	−	+S	12E	O111:HN	−	+
5C	O55:H7	−	+S	13A	O128:H7	−	+
5D	O55:H7	−	+S	13B	O128:H7	−	+
5E	O55:H7	−	+	13C	O128:H7	−	+
6A	O111:H21	+	−	13D	O128:H7	−	+
6B	O111:H12	−	+S	13E	O128:H47	−	+
6C	O111:H12	+	−	14A	O128:H21	−	+
6D	O111:H12	+	−	14B	O128:H21	−	+
6E	O111:HN	+	−*	14C	O128:H21	−	+
7A	O157:H43	−	+	14D	O128:HN	−	+
7B	O149:HN	M	M	14E	O128:H21	−	+
7C	O157:H43	−	+	15A	O111:H21	−	+
7D	O157:H43	−	+	15B	O111:H21	−	+
7E	O157:HN	−	+	15C	O111:H21	−	+
8A	O111:HN	−	+	15D	O111:H21	−	+
8B	O111:H8	−	−	15E	O111:H21	−	+

[a] +, positive phenotype; −, negative phenotype; +S, slow sucrose positive phenotype; NT, not tested; NG, no growth on MacConkey agar medium; *, poor colony formation on MacConkey agar medium; M, missing from collection.

with a partial *dsdX* gene and a full *dsdA* gene, but only the *cscA* and *cscR* genes were present. In place of *cscB* and *cscK*, which were directly upstream of the *cscA* gene in EDL933, were potential pseudogenes. There was 7,120 bp from the end of *argW* to the partial copy of *dsdX*. *S. flexneri* was also like EDL933, but without the phage genes, and had 5,368 bp in the *argW-dsdC* area. *S. boydii* had only *dsdA* and an unknown gene 407 bp after *argW*. *S. sonnei* had 51,748 bp in the *argW-dsdC* area. This area in *S sonnei* was similar to that in EDL933, with a partial *dsdX* gene, a full *dsdA* gene, and the sucrose utilization genes; but there were additional phage-like genes in *S. sonnei*.

Phenotypic survey of ability of *E. coli* strains to utilize D-serine and ferment sucrose. Because many K1 strains have the ability to cross the BBB and invade the meninges, where free D-serine may be available as a nutrient or an environmental signal, we examined the ability of K1 strains to use D-serine as a sole carbon source. The collection of strains tested is listed in Table 1. K1 strains that were neonatal blood or CSF isolates were examined. Additionally, we analyzed urine isolates, including K1 serotype isolates, such as O18:K1:H7, that have a broad ability to cause additional extraintestinal diseases, such as cystitis and pyelonephritis. We also included several well-characterized UPEC strains: strains NU14, CP9, 536, and J96. Because the genome sequence analysis showed that many diarrheagenic *E. coli* strains contain the *csrRAKB* genes, which encode the ability to ferment sucrose, in place of *dsdC* and the 5′ end of *dsdX*, the phenotypic survey also included isolates of the DEC collection (42). A total of 97.5% of the K1 isolates (40 of 41) isolates could grow with D-serine as a sole carbon source, suggesting the presence of an intact *dsdCXA* locus (Table 1). This rate of occurrence is considerably higher than that among the diarrheal strains tested in the DEC collection (4 of 74 strains) (Table 2) that could catabolize D-serine ($P = 0.001$, Fisher's exact test) (36).

The inverse relationship between D-serine and sucrose use was established when the strains in the collection were then compared for their sucrose fermentation phenotypes. Only four K1 strains fermented sucrose, with three of these (strains 468624, Marriot, and NU14) also being capable of using D-

serine (Table 1). B50 is the only K1 strain that could ferment sucrose but not use D-serine. Among the 76 available DEC strains, 2 failed to grow on MacConkey sucrose agar. Among the remaining 74 strains, 67 could ferment sucrose and 51 strains had strong fermentation-positive reactions following overnight growth (Table 2). There were 16 slow-sucrose-fermenting strains that displayed a light pink colony on the medium at 24 h. Interestingly, none of the DEC strains that could use sucrose could also use D-serine. Overall, the differences in sucrose phenotypes between the K1 and DEC strains were statistically significant ($P < 0.001$, Fisher's exact test).

Interestingly, the slow sucrose fermenters were clustered in DEC groups 3, 4, and 5 and represented the majority of the O157:H7 and O55:H7 isolates. The notable exception in this clustering was EDL933 (isolate 4F) in the DEC collection, which was a strong sucrose fermenter. In support of the genome sequence prediction, the four DEC isolates that could use D-serine but that could not ferment sucrose were within DEC group 6, which included O111:H21 EAEC strains. Among a group of eight EAEC strains, acquired from James Nataro at the University of Maryland, the D-serine and sucrose phenotypes were mixed, with five strains utilizing D-serine and three strains utilizing sucrose (strain data not shown).

Southern blot hybridization analysis for *ipuAB* and *dsdCXA* genes among K1 strains. Southern blot hybridization was used to examine the physical context of the region surrounding the *dsdCXA* genes among the ExPEC strains used in the phenotypic survey. The results are summarized in Table 3, and an example of a blot used in the analyses is shown in Fig. 3. The only K1 strain which could not grow on D-serine as a carbon source, strain B50, possessed an EcoRI-EcoRV fragment hydridization pattern similar to that of EDL933, where *dsdC* and the 5′ portion of *dsdX* were absent but *dsdA* was present. Two non-K1 strains, strains E115 and E606, isolated from CSF could not use D-serine, and they also had a *dsdCXA* hybridization fragment pattern similar to that of EDL933.

Four K1 strains, strains Kansas City, E253, E251, and E308, had *dsdX*- and *dsdA*-specific hybridization fragment patterns similar to those seen for MG1655, but with the *dsdC*-specific fragment being roughly 2,500 bp larger than that seen for MG1655. Strains E496, S92, CP9, and J96 had *dsdX*- and *dsdA*-specific hybridization fragment patterns similar to that of CFT073 but lacked *ipuA*- and *ipuB*-like sequences, which resulted in a *dsdC*-specific hybridization fragment different from that seen for CFT073. Eight strains had fragment patterns that were similar to the RS218 *argW*-associated *dsdCXA* fragment patterns, but three of the strains had larger *dsdC*-specific bands. One strain, RS168, had only a single copy of *dsdCXA* and had a hybridization pattern consistent with the version seen at the RS218 *leuX* genetic island.

Interestingly, 30 of the 41 K1 strains had two copies of *dsdCXA*-specific sequences. Two of these K1 strains, strains 468624 and EC10, also had the *fimBE*-like recombinases *ipuA* and *ipuB* upstream of *dsdC*, as was seen for CFT073. In these two instances, the *ipuA*- and *ipuB*-like sequences were found together on an EcoRI-EcoRV restriction fragment of the same size, suggesting that they are linked (10). These two strains also had a second *dsdCXA* locus that had a hybridization pattern similar to that of the RS218 *leuX* genetic island-associated copy.

For all 23 O18:K1:H7 isolates there was a *dsdCXA*-specific hybridization pattern similar to that of archetype strain RS218. There was a polymorphism involving the size of the *dsdA*-specific fragment among the O18:K1:H7 isolates. Three strains, strains E44 (a spontaneous rifampin-resistant RS218 mutant), NU14, and E912, had fragments similar in size to those predicted from the RS218 genome sequence. RS218 and 19 other O18:K1:H7 isolates had *dsdC*-specific fragments smaller than those predicted from the RS218 genome sequence. We suspect that there was either an error in the RS218 sequence or a rearrangement in this area during the strain's laboratory passage. Lastly, uropathogenic strain 536, studied extensively by the Hacker and Dobrindt laboratories, had a *dsdCXA*-specific hybridization pattern that was similar to that of cystitis isolate F11. They had the *dsdCXA* copy at the *leuX* genetic island and a second copy of *dsdX* and *dsdA*-like hybridization fragments similar to the *dsdXA*-like sequences for EHEC strain EDL933 at the *argW* genetic island.

PFGE analyses. PFGE and Southern blot hybridizations were used to examine possible clonal relationships and to confirm our observation of variability in the *argW-dsdCXA* region, especially among different K1 isolates. Examples of the PFGE AvrII digestion patterns for nine of the strains used in this analysis are shown in Fig. 4A. None of the AvrII fragment banding patterns were identical between E44, 891449, and NU14, which are all O18:K1:H7 isolates. In addition, when strains E44, 891449, and NU14 were probed for *argW*-like sequences, they showed hybridization fragments of dissimilar sizes. E912 and E221 were two other O18:K1:H7 isolates examined by PFGE; they had identical AvrII fragment patterns, with the exception of a single fragment. Interestingly, E912 and E221 had an *argW*-specific AvrII fragment different in size from those for the other three O18:K1:H7 isolates, isolates E44, 891449, and NU14. The sizes of the *argW*-specific AvrII fragments in the PFGE analyses are summarized in Table 3.

Strains CFT073 and ECOR56 are phylogenetically related, as determined on the basis of random amplified polymorphic DNA analysis (25), but they shared only four AvrII fragments of similar size (Fig. 4B). When these strains were probed for *argW*-like sequences, however, they shared a similar sized band of 125,000 bp. Strains CP9 and J96 are also thought to be closely related (25) but have different-sized *argW*-specific AvrII fragments (Fig. 4B).

Phenotypes of E44 *dsdA* mutants. Lastly, to assess the functionality of each copy of the duplicated *dsdCXA* genes that are commonly seen among the K1 strains, we constructed full deletions of each *dsdA* gene in strain E44. These were made individually (single-deletion mutants WAM3044 [E44 $dsdA_{argW}$] and WAM3047 [E44 $dsdA_{leuX}$]) and doubly (double-deletion mutant WAM3049 [E44 $dsdA_{argW,\ leuX}$]). Both single-deletion mutants, E44 $dsdA_{argW}$ and E44 $dsdA_{leuX}$, readily formed colonies after overnight incubation on MOPS–D-serine agar medium. Unlike the single-deletion strains, the mutant with a double *dsdA* deletion was unable to grow under these conditions. The $dsdA_{argW}$ gene was cloned into pACYC177 and transformed into E44 $dsdA_{argW,\ leuX}$ to produce strain WAM3180. WAM3180 was able to grow on MOPS--D-serine medium, thus utilizing D-serine as a sole carbon source and providing assurance that no second-site mutations in E44 $dsdA_{argW,\ leuX}$ were responsible for the growth defect (data not shown).

TABLE 3. Southern and PFGE hybridization results for *Escherichia coli* isolates

Strain (WAM no.)	Southern hybridization result[a]					Locus similarity to *E. coli* strains sequenced	Approximate size (bp) of *argW*-positive PFGE fragment
	ipuA	*ipuB*	*dsdC*	*dsdX*	*dsdA*		
2659	−	−	−	+	+	EDL933 53′[b]	475,000
2660	−	−	+	+	+	RS218 53′[b]	825,000
2661	−	−	++	++	++	RS218 53′, 97′[b]	925,000
2662	+	+	++	++	++	CFT073 53′,[b] RS218 97′[b]	500,000
2663	−	−	++	++	++	RS218 53′,[c] 97′	580,000
2664	+	+	++	++	++	CFT073 53′, RS218 97′[b]	290,000
2665	−	−	+	+	+	RS218 53′[b]	250,000
2666	−	−	++	++	++	RS218 53′, 97′[b]	380,000
2667	−	−	+	+	+	MG1655 53′[b]	>725,000
2668	−	−	++	++	++	CFT073 53′,[d] RS218 97′	500,000
2937	−	−	+	+	+	RS218 53′	415,000
2938	−	−	+	+	+	RS218 53′	825,000
2939	−	−	++	++	++	RS218 53′, 97′[b]	850,000
2940	−	−	+	+	+	CFT073 53′[b,d]	825,000
2941	−	−	+	+	+	RS218 53′	875,000
2942	−	−	++	++	++	RS218 53′, 97′	615,000
2943	−	−	++	++	++	RS218 53′, 97′	615,000
2944	−	−	+	+	+	MG1655 53′[b]	>840,000
2945	−	−	+	+	+	MG1655 53′[b]	>840,000
2946	+	+	++	++	++	CFT073 53′, RS218 97′[b]	750,000
2947	−	−	++	++	++	RS218 53′,[b] 97[b]	800,000
2948	−	−	−	+	+	EDL933 53′[b]	450,000
2949	−	−	++	++	++	CFT073 53′,[d] RS218 97′	500,000
2950	−	−	+	+	+	RS218 97′[b]	340,000
2951	−	−	−	+	+	EDL933 53′[b]	875,000
2952	−	−	+	+	+	RS218 53′[b]	825,000
2953	−	−	+	+	+	MG1655 53′[b]	>925,000
2954	−	−	++	++	++	RS218 53′, 97′	500,000
3228	−	−	++	++	++	RS218 53′,[c] 97′	NT[e]
3229	−	−	++	++	++	RS218 53′,[c] 97′	NT
3230	−	−	++	++	++	RS218 53′,[c] 97′	NT
3231	−	−	++	++	++	RS218 53′,[c] 97′	NT
3232	−	−	++	++	++	RS218 53′,[c] 97′	NT
3233	−	−	++	++	++	RS218 53′,[c] 97′?	NT
3234	−	−	+	+	+	CFT073 53′[b,d]	NT
3235	−	−	++	++	++	RS218 53′,[c] 97′	NT
3236	−	−	++	++	++	RS218 53′,[c] 97′	NT
3237	−	−	++	++	++	CFT073 53′, RS218 97′[b]	NT
3238	−	−	++	++	++	RS218 53′,[c] 97′	NT
3239	−	−	++	++	++	RS218 53′,[c] 97′	NT
3240	−	−	++	++	++	RS218 53′,[c] 97′	NT
3241	−	−	++	++	++	RS218 53′,[c] 97′	NT
3242	−	−	++	++	++	RS218 53′,[c] 97′	NT
3243	−	−	++	++	++	RS218 53′,[c] 97′	NT
3244	−	−	++	++	++	RS218 53′,[c] 97′	NT
3245	−	−	++	++	++	RS218 53′,[c] 97′	NT
3246	−	−	++	++	++	RS218 53′,[c] 97′	NT
3247	−	−	+	+	+	RS218 53′	NT
3248	−	−	++	++	++	RS218 53′,[c] 97′	NT
3250	−	−	++	++	++	RS218 53′,[c] 97′	NT
3251	−	−	++	++	++	RS218 53′,[c] 97′	NT
3252	−	−	++	++	++	RS218 53′,[c] 97′	300,000
2645	−	−	++	++	++	RS218 53′, 97′	500,000
2039	−	−	+	++	++	EDL933 53′, 536 97′	650,000
2267	+	+	+	+	+	CFT073 53′	125,000
2625	−	−	+	+	+	MG1655 53′	1,000,000
2392	−	−	−	+	+	EDL933 53′	560,000
2265	−	−	+	+	+	CFT073 53′[d]	275,000
1218	−	−	+	+	+	CFT073 53′[d]	600,000

[a] +, one copy; ++ two copies.
[b] Difference in restriction site 5′ to *dsdC* or 3′ of *dsdA*, resulting in a different-sized band.
[c] Different restriction site 5′ to *dsdC* compared to sizes predicted by sequence analysis.
[d] Similar to CFT073 but without *ipuA* and *ipuB*.
[e] NT, not tested by PFGE.

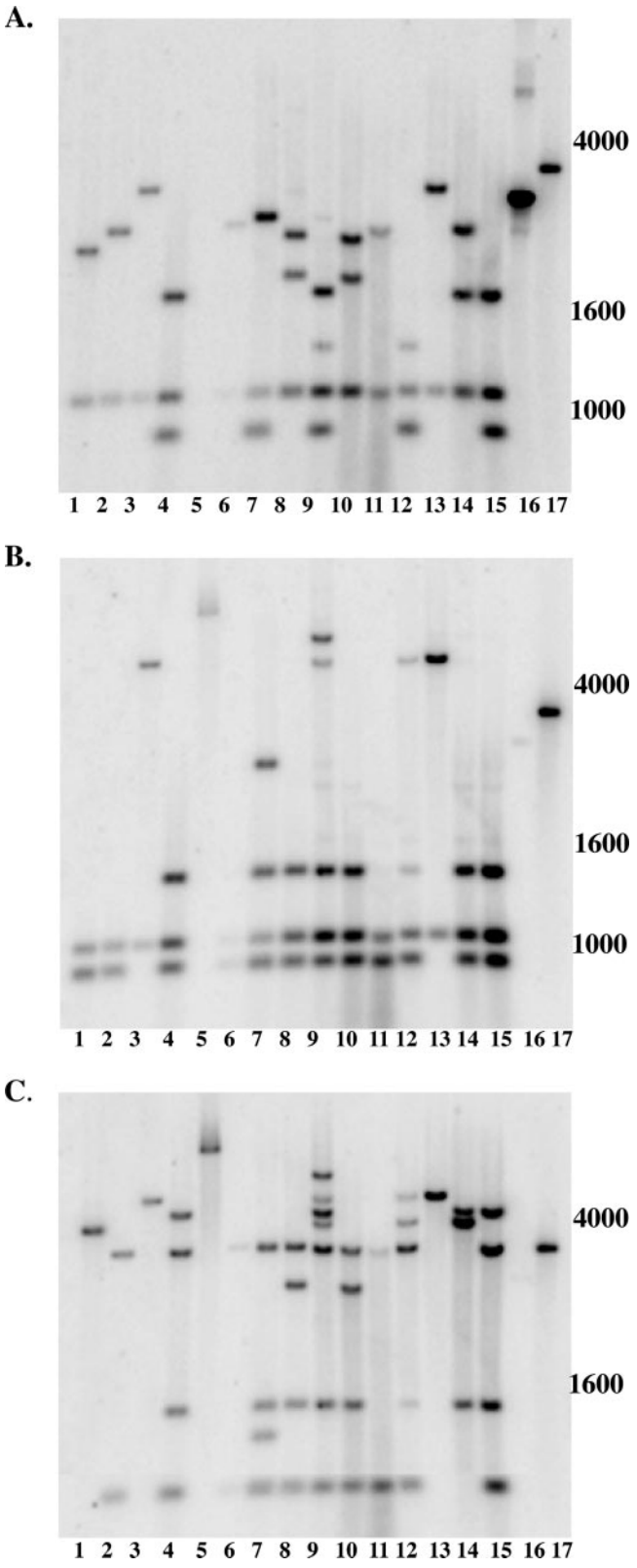

FIG. 3. Representative Southern blots of EcoRI- and EcoRV-digested genomic DNA. Lane 1, strain CFT073; lanes 2 to 4, strains WAM2952 to WAM2954; lanes 5 to 14, strains WAM2659 to WAM2668; lane 15, strain NU14; lane 16, positive control consisting of a PCR product from *ipuA-dsdC* made from CFT073; lane 17, positive control consisting of a PCR product of *dsdC-dsdA* in CFT073. (A) Hybridization to the *dsdC* probe; (B) hybridization to the *dsdX* probe; (C) hybridization to the *dsdA* probe.

DISCUSSION

In this study, we examined the physical context and phenotypes for genes found at the *E. coli argW-dsdCXA* genetic island. This region corresponds to the lambdoid bacteriophage PA-2 integration site at position 53′ on the *E. coli* K-12 MG1655 genetic map (34). The interest in this region comes from our studies that indicate that in archetype UPEC strain CFT073 genes are present within this genetic island that control the level of CFT073 bladder and kidney colonization of mice during experimentally induced UTIs (10, 36). An examination of the current collection of *E. coli* genome sequences supports our earlier observation that the genes for the use of D-serine, *dsdCXA*, are intact in ExPEC strains but that in diarrheal pathogens, such as EHEC strain EDL933, the *dsdCXA* genes are missing, in part due to a substitution with the sucrose utilization genes, *cscRAKB*. We analyzed a collection of *E. coli* isolates for their ability to utilize D-serine as a sole carbon source and to ferment sucrose. We found that there is generally a converse relationship for these traits; ExPEC strains often have the ability to use D-serine but not to ferment sucrose. EPEC, EHEC, and ETEC strains generally cannot use D-serine but can ferment sucrose. Even more striking was the observation that among 41 strains that possessed the K1 capsular antigen, 40 had the ability to use D-serine but only 3 strains could ferment sucrose. Because free D-serine is present in brain tissue and K1 strains are the most frequent bacterial causes of meningitis, we investigated this correlation further. Southern hybridization analyses indicated that 30 of the 41 K1 strains have two copies of the *dsdCXA* genes, just as was observed in the genome sequence of K1 archetype strain RS218. In vitro and in vivo studies that are examining the virulence effects of the mutants of E44 with single and double *dsdA* deletions are under way.

Our genetic analysis indicates each of the E44 *dsdCXA* loci is functional. However, the copies of the *dsdCXA* genes at the different island locations are not identical. The nucleotide sequences of the *dsdCXA* genes in the RS218 and the *leuX* PAI II_{536} islands are more similar to one another than to the copies of *dsdCXA* at the CFT073 and RS218 *argW* islands. It is possible that the copy of the *dsdCXA* genes at the RS218 *argW* island may be regulated differently than the copies at the CFT073 *argW* island or the *leuX*-associated copies because in the 211-bp intergenic promoter region between *dsdC* and *dsdX*, 5 nucleotide substitutions are unique to the RS218 *argW*-associated *dsdCXA* promoter region. One involves a substitution within one of two putative cyclic AMP receptor protein-binding sites, and the other involves the −35 promoter region for the *dsdC* gene (B. Haugen and R. Welch, unpublished observations).

The RS218 *argW-dsdCXA* genetic island has recently drawn the attention of the laboratory of E. R. Vimr. From the unannotated RS218 genome sequence, Vimr and colleagues identified a lambdoid-like phage integrated at the *argW* region (15). Linked to the prophage is a gene called *neuO*, which they

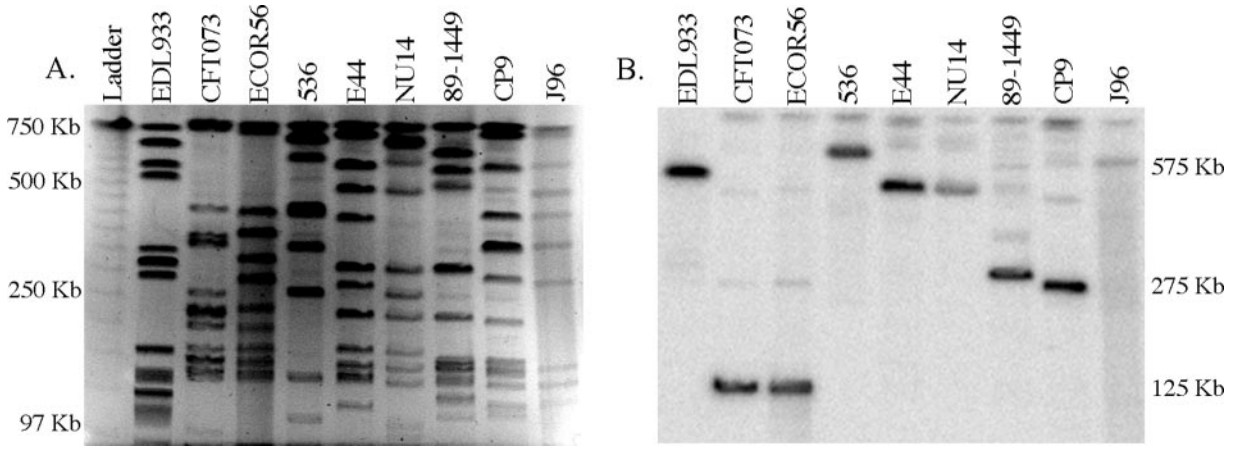

FIG. 4. PFGE characterization of the *Escherichia coli* isolates sequenced and phylogenetically related ExPEC strains. (A) PFGE profile with AvrII-digested genomic DNA; (B) hybridization with radiolabeled *argW*-specific probe. Strain designations are presented at the top of the gel.

found encodes an *O*-acetyltransferase necessary for the acetylation of the capsular polysialic acid (15). Interestingly, expression of *neuO* appears to undergo phase-switch control through a mechanism that involves slipped-strand DNA mispairing of a heptanucleotide repeat in its 5′ region (15). Although *neuO* is missing in non-K1 ExPEC strains like CFT073, FimBE-like recombinases *ipuA* and *ipuB* are encoded with the corresponding *argW-dsdC* CFT073 region. We recently showed that *ipuA* controls phase-switch control of the *fimS* element for expression of the type 1 fimbrial operon (10). These observations reinforce our model that ExPEC strains are successful at colonizing sequential niches during an ascending UTI because of a rich array of phase-switch-controlled gene expression states.

Southern blot hybridization analysis of the *dsdCXA* loci in NMEC and UPEC K1 strains, along with examination of the accumulated *E. coli* and *Shigella* genome sequences, showed that the region between *argW* and *dsdCXA* is hypervariable in size and content. The variability is evident even among O18:K1:H7 isolates that are considered to be of a common clonal descent (27). The variability is probably due to *argW* being an active site for the integration of lysogenic bacteriophages. The available sequences clearly show that complete phages or remnants of different phages occupy this region. We suggest that the *argW-dsdCXA* region provides an excellent site for the establishment of epidemiological markers that could differentiate not only different closely related K1 isolates but also, perhaps, isolates of the other *E. coli* pathotypes.

Our analysis also indicates that in many K1 strains the RS218 *leuX*-associated *dsdCXA* copy is present within an island that is very similar to the island in UPEC strain 536, PAI II_{536}, which has been studied extensively by Hacker, Dobrindt, and colleagues (17). Because of the general genetic content of *prf* fimbrial and *hlyCABD* virulence genes, this island has also been identified as PAI I_{C5} and PAI II_{J96} (5, 8, 22). The *prf* fimbria- and *hlyCABD*-containing island does vary in significant ways because the *dsdCXA*, *hra*, and *cnf-1* loci are not consistently present. In addition, the islands can be located at either a *leuX* or a *pheU* site in the ExPEC genomes (38). With a large amount of genomic sequence and epidemiological information becoming available, we predict that there will be increased confusion about the nomenclature of genetic islands. Currently, investigators number islands with no apparent logic other than the chronological order of their description. We propose that islands be given unique names that comprise their chromosomal location relative to the sequence of *E. coli* strain MG1655 and host strain number. In addition, the PAI nomenclature makes the assumption that the genetic islands play a role in pathogenesis. We propose that the phenotypically neutral abbreviation GI, for "genetic island," be used. Therefore the two islands that we have described here for RS218 would be named GI-RS218 *argW* and GI-RS218 *leuX*.

Lastly, our results permit some speculation about how different *E. coli* pathotypes evolved to take advantage of their nutritional opportunities and the possible global regulatory consequences for the use of those nutrients. EHEC, EPEC, and ETEC intestinal pathogens are not free-living, nor do they effectively colonize body sites outside of the mammalian intestine. Their typical human and mammalian hosts have diets rich in plant materials that contain sugars such as sucrose. Therefore, the predilection and utility of the intestinal pathogens to have the *cscRAKB* genes at the *argW-dsdCXA* genetic island are evident. Molecular evidence also suggests that the order of events for the intestinal pathogens was the loss of D-serine utilization and the acquisition of the sucrose-positive trait. However, for ExPEC strains, which are transient members of the human intestinal flora or members of the normal flora of carnivores, such as dogs (24, 26), the poor availability and scant need to use sucrose as a nutrient probably did not lead to a selective force. Because ExPEC strains occupy different sites outside of the intestine, it follows that they have evolved to take advantage of the nutrients available at those sites. Compared to the intestine, urine is a relatively carbohydrate-poor but peptide- and amino acid-rich environment. Our transcriptome studies with CFT073 grown in human urine or isolated from the urine of infected mice indicate that, in general, its carbohydrate transporters and catabolic pathways are poorly expressed relative to the expression of transporters of amino acids and peptides (37). At least one consequence of the carbohydrate versus amino acid nutritional choice is that the former would cause a classic state of catabolite repression, whereas the latter leads to induction of cyclic AMP

receptor protein-dependent gene expression. So, in the case of the enteric pathogenic *E. coli*, it can be hypothesized that their ability to break down sucrose to fructose and glucose would lead to the induction of strong state of catabolite repression. ExPEC strains, on the other hand, would accumulate amino acids such as D-serine instead of carbohydrates like sucrose and therefore would not be in a state of catabolite repression in their extraintestinal niche. We are currently testing the patterns of the nutritional capabilities of different *E. coli* pathotypes and how these differences affect the expression of their pathogenic factors.

ACKNOWLEDGMENTS

This work was supported by Public Health Service grant DK63250. We thank Brain Haugen, Shahaireen Pellett, Holly Hamilton, and Laura Walters for helpful discussions and editing of the manuscript. We also thank Alan Cross for digging through notebooks for specific information for the K1 strains that he originally described (13).

REFERENCES

1. **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.
2. **Altschul, S., F. W. Gish, W. Miller, E. W. Myers, and D. J. Lipman.** 1990. Basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2a. **Anfora, A. T., and R. A. Welch.** 2006. DsdX is the second D-serine transporter in uropathogenic *Escherichia coli* strain CFT073. J. Bacteriol. **188:**6622–6628.
3. **Bachur, R., and G. L. Caputo.** 1995. Bacteremia and meningitis among infants with urinary tract infections. Pediatr. Emerg. Care **11:**280–284.
4. **Badger, J. L., and K. S. Kim.** 1998. Environmental growth conditions influence the ability of *Escherichia coli* K1 to invade brain microvascular endothelial cells and confer serum resistance. Infect. Immun. **66:**5692–5697.
5. **Bidet, P., S. Bonacorsi, O. Clermont, C. De Montille, N. Brahimi, and E. Bingen.** 2005. Multiple insertional events, restricted by the genetic background, have led to acquisition of pathogenicity island IIJ96-like domains among *Escherichia coli* strains of different clinical origins. Infect. Immun. **73:**4081–4087.
6. **Bloom, F. R., and E. McFall.** 1975. Isolation and characterization of D-serine deaminase constitutive mutants by utilization of D-serine as sole carbon or nitrogen source. J. Bacteriol. **121:**1078–1084.
7. **Bonacorsi, S., and E. Bingen.** 2005. Molecular epidemiology of *Escherichia coli* causing neonatal meningitis. Int. J. Med. Microbiol. **295:**373–381.
8. **Bonacorsi, S., O. Clermont, V. Houdouin, C. Cordevant, N. Brahimi, A. Marecat, C. Tinsley, X. Nassif, M. Lange, and E. Bingen.** 2003. Molecular analysis and experimental virulence of French and North American *Escherichia coli* neonatal meningitis isolates: identification of a new virulent clone. J. Infect. Dis. **187:**1895–1906.
9. **Bruckner, J., S. Haasmann, and A. Friedrich.** 1994. Quantification of D-amino acids in human urine using GC-MS and HPLC. Amino Acids **6:**205–211.
10. **Bryan, A., P. Roesch, L. Davis, R. Moritz, S. Pellett, and R. A. Welch.** 2006. Regulation of type 1 fimbriae by unlinked FimB- and FimE-like recombinases in uropathogenic *Escherichia coli* strain CFT073. Infect. Immun. **74:**1072–1083.
11. **Cosloy, S. D., and E. McFall.** 1973. Metabolism of D-serine in *Escherichia coli* K-12: mechanism of growth inhibition. J. Bacteriol. **114:**685–694.
12. **Cosloy, S. D.** 1973. D-Serine transport system in *Escherichia coli* K-12. J. Bacteriol. **114:**679–684.
13. **Cross, A., S. P. Gemski, J. C. Sadoff, F. Orskov, and I. Orskov.** 1984. The importance of the K1 capsule in invasive infections caused by *Escherichia coli*. J. Infect. Dis. **149:**184–193.
14. **Datsenko, K. A., and B. L. Wanner.** 2000. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. Proc. Natl. Acad. Sci. USA **97:**6640–6645.
15. **Deszo, E. L., S. M. Steenbergen, D. I. Freedberg, and E. R. Vimr.** 2005. *Escherichia coli* K1 polysialic acid *O*-acetyltransferase gene, *neuO*, and the mechanism of capsule form variation involving a mobile contingency locus. Proc. Natl. Acad. Sci. USA **102:**5564–5569.
16. **Dietzman, D. E., G. W. Fischer, and F. D. Schoenknecht.** 1974. Neonatal *Escherichia coli* septicemia—bacterial counts in blood. J. Pediatr. **85:**128–130.
17. **Dobrindt, U., G. Blum-Oehler, G. Nagy, G. Schneider, A. Johann, G. Gottschalk, and J. Hacker.** 2002. Genetic structure and distribution of four pathogenicity islands (PAI I_{536} to PAI IV_{536}) of uropathogenic *Escherichia coli* strain 536. Infect. Immun. **70:**6365–6372.
18. **Formal, S. B., G. J. Dammin, E. H. Labrec, and H. Schneider.** 1958. Experimental *Shigella* infections: characteristics of a fatal infection produced in guinea pigs. J. Bacteriol. **75:**604–610.
19. **Gautom, R. K.** 1997. Rapid pulsed-field gel electrophoresis protocol for typing of *Escherichia coli* O157:H7 and other gram-negative organisms in 1 day. J. Clin. Microbiol. **35:**2977–2980.
20. **Hashimoto, A., and T. Oka.** 1997. Free D-aspartate and D-serine in the mammalian brain and periphery. Prog. Neurobiol. **52:**325–353.
21. **Heath, J. D., J. D. Perkins, B. Sharma, and G. M. Weinstock.** 1992. NotI genomic cleavage map of *Escherichia coli* K-12 strain MG1655. J. Bacteriol. **174:**558–567.
22. **Houdouin, V., S. Bonacorsi, N. Brahimi, O. Clermont, X. Nassif, and E. Bingen.** 2002. A uropathogenicity island contributes to the pathogenicity of *Escherichia coli* strains that cause neonatal meningitis. Infect. Immun. **70:**5865–5869.
23. **Jahreis, K., L. Bentler, J. Bockmann, S. Hans, A. Meyer, J. Siepelmeyer, and J. W. Lengeler.** 2002. Adaptation of sucrose metabolism in the *Escherichia coli* wild-type strain EC3132. J. Bacteriol. **184:**5307–5316.
24. **Johnson, J. R., A. L. Stell, and P. Delavari.** 2001. Canine feces as a reservoir of extraintestinal pathogenic *Escherichia coli*. Infect. Immun. **69:**1306–1314.
25. **Johnson, J. R., K. Owens, A. R. Manges, and L. W. Riley.** 2004. Rapid and specific detection of *Escherichia coli* clonal group A by gene-specific PCR. J. Clin. Microbiol. **42:**2618–2622.
26. **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 of the same electrophoretic lineages from humans and domestic animals. J. Infect. Dis. **183:**154–159.
27. **Johnson, J. R., P. Delavari, and T. T. O'Bryan.** 2001. *Escherichia coli* O18:K1:H7 isolates from patients with acute cystitis and neonatal meningitis exhibit common phylogenetic origins and virulence factor profiles. J. Infect. Dis. **183:**425–434.
28. **Kaijser, B.** 1973. Immunology of *Escherichia coli*: K antigen and its relation to urinary-tract infection. J. Infect. Dis. **127:**670–677.
29. **Kim, K. S.** 2001. *Escherichia coli* translocation at the blood-brain barrier. Infect. Immun. **69:**5217–5222.
30. **Kim, K. S.** 2002. Strategy of *Escherichia coli* for crossing the blood-brain barrier. J. Infect. Dis. **186**(Suppl. 2)**:**S220–S224.
31. **Maas, W. K., and B. D. Davis.** 1950. Pantothenate studies. I. Interference by D-serine and L-aspartic acid with pantothenate synthesis in *Escherichia coli.* J. Bacteriol. **60:**733–745.
32. **McCabe, W., R. B. Kaijser, S. Olling, M. Uwaydah, and L. A. Hanson.** 1978. *Escherichia coli* in bacteremia: K and O antigens and serum sensitivity of strains from adults and neonates. J. Infect. Dis. **138:**33–41.
33. **Nørregaard-Madsen, M., E. McFall, and P. Valentin-Hansen.** 1995. Organization and transcriptional regulation of the *Escherichia coli* K-12 D-serine tolerance locus. J. Bacteriol. **177:**6456–6461.
34. **Pugsley, A., P. D. Littmann-Louth, and C. A. Schnaitman.** 1979. Chromosomal location of the attachment site for the PA-2 prophage in *Escherichia coli* K-12. J. Virol. **29:**808–810.
35. **Pupo, G. M., D. K. Karaolis, R. Lan, and P. R. Reeves.** 1997. Evolutionary relationships among pathogenic and nonpathogenic *Escherichia coli* strains inferred from multilocus enzyme electrophoresis and *mdh* sequence studies. Infect. Immun. **65:**2685–2692.
36. **Roesch, P., L. P. Redford, S. Batchelet, R. L. Moritz, S. Pellett, B. J. Haugen, F. R. Blattner, and R. A. Welch.** 2003. Uropathogenic *Escherichia coli* use D-serine deaminase to modulate infection of the murine urinary tract. Mol. Microbiol. **49:**55–67.
37. **Snyder, J. A., B. J. Haugen, E. L. Buckles, C. V. Lockatell, D. E. Johnson, M. S. Donnenberg, R. A. Welch, and H. L. Mobley.** 2004. Transcriptome of uropathogenic *Escherichia coli* during urinary tract infection. Infect. Immun. **72:**6373–6381.
38. **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.
39. **Weiser, J. N., and E. C. Gotschlich.** 1991. Outer membrane protein A (OmpA) contributes to serum resistance and pathogenicity of *Escherichia coli* K-1. Infect. Immun. **59:**2252–2258.
40. **Welch, R., A. V. Burland, G. Plunkett, 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. 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.
41. **Wheeler, D., L. T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, L. Y. Geer, W. Helmberg, Y. Kapustin, D. L. Kenton, O. Khovayko, D. J. Lipman, T. L. Madden, D. R. Maglott, J. Ostell, K. D. Pruitt, G. D. Schuler, L. M. Schriml, E. Sequeira, S. T. Sherry, K. Sirotkin, A. Souvorov, G. Starchenko, T. O. Suzek, R. Tatusov, T. A. Tatusova, L. Wagner, and E. Yaschenko.** 2006. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res. **34:**D173–D180.
42. **Whittam, T. S., M. L. Wolfe, I. K. Wachsmuth, F. Orskov, I. Orskov, and R. A. Wilson.** 1993. Clonal relationships among *Escherichia coli* strains that cause hemorrhagic colitis and infantile diarrhea. Infect. Immun. **61:**1619–1629.