AMERICAN SOCIETY FOR MICROBIOLOGY | Antimicrobial Agents and Chemotherapy®

Characterization of a Large Antibiotic Resistance Plasmid Found in Enteropathogenic *Escherichia coli* Strain B171 and Its Relatedness to Plasmids of Diverse *E. coli* and *Shigella* Strains

Tracy H. Hazen,[a,b] Jane Michalski,[a] Sushma Nagaraj,[a] Iruka N. Okeke,[c] David A. Rasko[a,b]

Institute for Genome Sciences[a] and Department of Microbiology and Immunology,[b] University of Maryland School of Medicine, Baltimore, Maryland, USA; Faculty of Pharmacy, University of Ibadan, Ibadan, Nigeria[c]

ABSTRACT Enteropathogenic *Escherichia coli* (EPEC) is a leading cause of severe infantile diarrhea in developing countries. Previous research has focused on the diversity of the EPEC virulence plasmid, whereas less is known regarding the genetic content and distribution of antibiotic resistance plasmids carried by EPEC. A previous study demonstrated that in addition to the virulence plasmid, reference EPEC strain B171 harbors a second, larger plasmid that confers antibiotic resistance. To further understand the genetic diversity and dissemination of antibiotic resistance plasmids among EPEC strains, we describe the complete sequence of an antibiotic resistance plasmid from EPEC strain B171. The resistance plasmid, pB171_90, has a completed sequence length of 90,229 bp, a GC content of 54.55%, and carries protein-encoding genes involved in conjugative transfer, resistance to tetracycline (*tetA*), sulfonamides (*sull*), and mercury, as well as several virulence-associated genes, including the transcriptional regulator *hha* and the putative calcium sequestration inhibitor (*csi*). *In silico* detection of the pB171_90 genes among 4,798 publicly available *E. coli* genome assemblies indicates that the unique genes of pB171_90 (*csi* and *traI*) are primarily restricted to genomes identified as EPEC or enterotoxigenic *E. coli*. However, conserved regions of the pB171_90 plasmid containing genes involved in replication, stability, and antibiotic resistance were identified among diverse *E. coli* pathotypes. Interestingly, pB171_90 also exhibited significant similarity with a sequenced plasmid from *Shigella dysenteriae* type I. Our findings demonstrate the mosaic nature of EPEC antibiotic resistance plasmids and highlight the need for additional sequence-based characterization of antibiotic resistance plasmids harbored by pathogenic *E. coli*.

KEYWORDS EPEC, *Escherichia coli*, mosaic, plasmid

Received 15 May 2017 **Returned for modification** 19 June 2017 **Accepted** 26 June 2017

Accepted manuscript posted online 3 July 2017

Citation Hazen TH, Michalski J, Nagaraj S, Okeke IN, Rasko DA. 2017. Characterization of a large antibiotic resistance plasmid found in enteropathogenic *Escherichia coli* strain B171 and its relatedness to plasmids of diverse *E. coli* and *Shigella* strains. Antimicrob Agents Chemother 61:e00995-17. https://doi.org/10.1128/AAC.00995-17.

Address correspondence to David A. Rasko, drasko@som.umaryland.edu.

There has been an alarming increase in the frequency of antibiotic resistance in bacterial pathogens. Multidrug-resistant *Enterobacteriaceae* are of particular concern and include some groups of disease-causing *Escherichia coli* (1–3). *E. coli* strains have been identified that are resistant to all routinely prescribed antibiotics, including colistin, which is a last resort treatment option for carbapenem-resistant *Enterobacteriaceae* (4–7). The most widely reported antibiotic-resistant *E. coli* strains are those belonging to multilocus sequence type (MLST) 131 (ST131), which are usually identified as extraintestinal pathogenic *E. coli* (ExPEC) and often exhibit resistance to fluoroquinolones, as well as extended-spectrum cephalosporins (1, 8). *E. coli* can also serve as a reservoir of antibiotic resistance genes, including *mph*(A), which confers resistance to macrolides such as azithromycin (9). *E. coli* strains from healthy volunteers in different

geographic locations were also identified with resistance to antibiotics, including ampicillin, trimethoprim, and ciprofloxacin (10), and *E. coli* strains carrying tetracycline resistance genes were isolated from infants that had not received tetracycline (11). *E. coli* resistance genes and elements have been demonstrated to exist in fecal *E. coli* reservoirs and spread internationally for decades without detection (12).

Previous studies have also demonstrated that other pathotypes of disease-causing *E. coli*, such as the enteropathogenic *E. coli* (EPEC), can exhibit resistance to one or more antibiotics (13–17). One study demonstrated that EPEC strains from Brazil exhibited resistance to tetracycline, sulfonamides, ampicillin, chloramphenicol, and streptomycin (18). EPEC is one of six diarrheagenic pathotypes (EPEC, enterohemorrhagic *E. coli* [EHEC], enterotoxigenic *E. coli* [ETEC], enteroaggregative *E. coli* [EAEC], enteroinvasive *E. coli* [EIEC], and diffusely adherent *E. coli* [DAEC]) of disease-causing *E. coli* that cause human gastrointestinal illness (19–21). EPEC was also recently characterized in the Global Enteric Multisite Study as a leading cause of lethal diarrheal illness among infants in developing countries (22). Similar to the other *E. coli* pathotypes, EPEC can be molecularly classified by the presence of canonical virulence factors that are associated with mechanisms of pathogenesis or by phenotypes that are unique to each of the *E. coli* pathotypes (19–21). EPEC and EHEC are collectively called the attaching and effacing *E. coli* (AEEC) due to the chromosomally encoded locus of enterocyte effacement (LEE) pathogenicity island, which contains the intimin protein, involved in adherence, and a type III secretion system that delivers effector proteins into host cells (19, 20, 21, 23, 24). In addition to the LEE region, EHEC contains the Shiga toxin, which is carried by a phage and is absent from EPEC (20, 25). Meanwhile, EPEC strains are subclassified as typical EPEC (tEPEC) strains that contain the plasmid-encoded bundle-forming pilus (BFP) or atypical EPEC (aEPEC) strains that contain the LEE region but do not have the Shiga toxin genes or BFP genes (21).

E. coli plasmids and phage carry many of the virulence genes that contribute to differences in pathogenesis, and these virulence genes are used to distinguish among the different *E. coli* pathotypes (19, 20, 26). Many of the virulence plasmids and virulence-associated phage of *E. coli* have been sequenced and described in detail for the *E. coli* pathotypes, including EHEC, EPEC, ETEC, and EAEC (26–33). The EPEC adherence factor (EAF) virulence plasmid, which carries genes encoding the BFP of tEPEC, has been sequenced and described for several reference strains, including EPEC strains E2348/69 (29, 34) and B171 (28). In a previous study, we sequenced and compared the EAF plasmids from genomically diverse EPEC strains, demonstrating considerable differences in the gene content and stability of this canonical tEPEC virulence plasmid (35).

In contrast to our knowledge of the diversity of EPEC virulence plasmids, much less is known about the sequence diversity and distribution of EPEC antibiotic resistance plasmids. A large multidrug resistance plasmid from EPEC strain MB80 was previously characterized using subtractive hybridization (36). The *traI* and *traC* genes of the conjugative transfer system were absent from the virulence plasmid of EPEC reference strain E2348/69; however, they were detected in other EPEC strains, including the EPEC reference strain B171 (30, 36, 37). While portions of the MB80 plasmid were sequenced (36) and the draft genome sequence of EPEC strain B171 has been generated (30), the complete sequence of the resistance plasmid from these EPEC strains has not been previously described. Thus, to increase our understanding of EPEC antibiotic resistance plasmids, we sequenced the large, ~90-kb antibiotic and heavy metal resistance plasmid, pB171_90, from the EPEC reference strain B171, which was previously determined to confer resistance to tetracycline, streptomycin, chloramphenicol, and sulfonamides (30, 36, 37). We describe the distribution of the pB171_90 genes among a publicly available collection of 4,798 *E. coli* genome assemblies in GenBank, which includes *E. coli* of diverse pathotypes as well as commensal *E. coli* that lack canonical pathotype-specific virulence genes. We also investigated whether the pB171_90 plasmid genes may have been disseminated to other species of enteric bacteria. Thus, the current study provides a comprehensive description of the genetic content of an

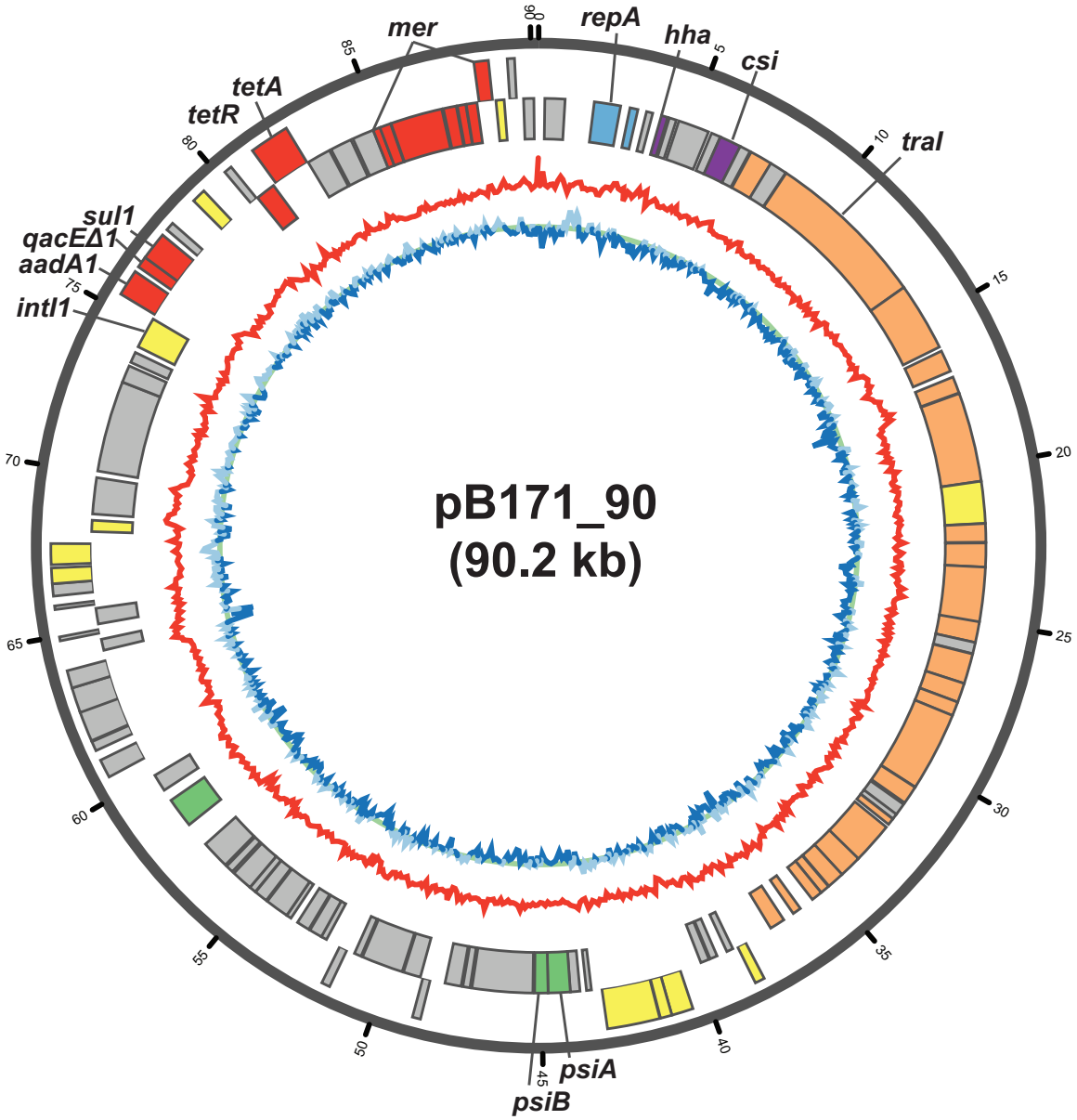

FIG 1 Circular plot of pB171_90. The outermost track of the plot contains the protein-coding genes on the positive strand, while the adjacent track contains the protein-coding genes on the negative strand. The predicted protein functions are indicated by colors as follows: replication (blue), virulence (purple), antibiotic resistance (red), conjugative transfer (orange), plasmid stability (green), transposition (yellow), and unknown (gray). The red line is the GC content along a 100-bp sliding window. The blue line is the GC skew along a 100-bp sliding window, with light blue indicating a positive skew and dark blue indicating a negative skew. The circular plot was generated using Circos 0.69-3 (80).

antibiotic resistance plasmid from a tEPEC reference strain as well as the distribution of the plasmid genes among a diverse collection of *E. coli* and other enteric pathogens.

RESULTS

Sequence characterization of the large antibiotic resistance plasmid, pB171_90, from EPEC strain B171. In the current study, we describe the complete sequence of the large antibiotic resistance plasmid from EPEC strain B171 to gain additional insight into its genetic content and potential role in antibiotic resistance and/or pathogenesis. We have designated this plasmid pB171_90 (GenBank accession number CP021212), as it is from EPEC strain B171 and has a sequence length of 90,229 bp. Plasmid pB171_90 has a GC content of 54.55% and contains 108 predicted protein-coding genes (Fig. 1). The predicted functions of the pB171_90 protein-coding genes include conjugative transfer,

plasmid stability, and antibiotic resistance (Table 1; Fig. 1). Among the pB171_90 antibiotic resistance genes are a spectinomycin-streptomycin resistance gene (*aadA1*, pB171_90_88) (GenBank accession number NG_052027.1) (38), *tetA* (pB171_90_95) encoding a tetracycline efflux pump, and *tetR* (pB171_90_94) encoding the *tetA* repressor (39–41), a multidrug efflux pump (*qacEΔ1*, pB171_90_89) (GenBank accession number L06822.4) (42), and a dihydropteroate synthase gene, which can confer resistance to sulfonamides (*sul1*, pB171_90_90) (GenBank accession number NC_006671.1) (43) (Table 1). The genes for spectinomycin-streptomycin resistance (*aadA1*), sulfonamide resistance (*sul1*), and the *qacEΔ1* efflux pump form a class 1 integron that includes a class 1 integrase gene, *intI1* (GenBank accession number NC_007100.1), and is located within a region (coordinates 74,154 to 77,947) that has 100% nucleotide identity to a previously described Tn*21* transposon from *E. coli* (GenBank accession number X12870.1) (44). Also carried on pB171_90 is a *mer* operon (pB171_90_99 to pB171_90_105), which contains all of the *mer* genes conferring narrow-spectrum mercury resistance, but the plasmid is lacking *merB* and *merG*, which are involved in organomercury detoxification and broad-spectrum mercury resistance (45).

The resistance plasmid, pB171_90, also carries two genes that were present on the EPEC resistance plasmid and not on the EPEC virulence plasmid and were previously suggested to have a role in pathogenesis (36). One potential pathogenicity gene is *csi*, a putative calcium sequestration inhibitor (36), whose function in virulence is poorly understood. The second virulence-associated gene of pB171_90, *hha*, encodes a transcriptional regulator of known virulence genes, including a hemolysin (46), genes involved in adherence and biofilm formation (47, 48), and genes involved in type III secretion (49). The *hha* gene has been identified on the chromosome and a virulence plasmid of *E. coli* O157:H7 Shiga toxin-producing *E. coli* (STEC) strains (50).

While sequencing the resistance plasmid of *E. coli* strain B171 using the Pacific Biosciences (PacBio) platform, we also took the opportunity to resequence the B171 EAF plasmid (here referred to as the virulence plasmid). The original sequence of the pB171 virulence plasmid was previously assembled from Sanger-sequenced plasmid subclones produced by different groups nearly 2 decades ago (28). To distinguish the PacBio-sequenced version of the plasmid from the original version, which is designated pB171 (GenBank accession number AB024946.1) (28), we have named the newly PacBio-sequenced version of the B171 virulence plasmid pB171_69 (GenBank accession number CP021211). The PacBio-sequenced virulence plasmid, pB171_69, has a GC content of 46.02% and a sequence length of 68,814 bp. The pB171_69 plasmid sequence is nearly identical to the original virulence plasmid, pB171, which has a GC content of 46% and a sequence length of 68,817 bp (28). Comparison of the two virulence plasmid sequences demonstrated that there were 21 single nucleotide polymorphisms (SNPs) and seven insertions or deletions in pB171_69 compared to the original version of the plasmid (see Table S2 in the supplemental material). These differences were identified in only 7 of the 80 previously described protein-coding genes (Table S2) (28). Overall, this comparative analysis demonstrates that the B171 virulence plasmids that were sequenced nearly 2 decades apart using different sequencing technologies are remarkably similar. Further investigation is necessary to determine whether these sequence differences are caused by the sequencing technology or whether they represent genetic changes that have been acquired by different laboratory strains of B171 over time and may influence any phenotypes of B171.

Distribution of the pB171_90 genes among diverse *E. coli* genomes. To determine the distribution of the pB171_90 genes among diverse *E. coli* genomes, we used large-scale BLAST score ratio (LS-BSR) analysis to conduct an *in silico* screen of 4,798 *E. coli* genome assemblies available in GenBank as of November 2016 for the presence of pB171_90 genes. LS-BSR analysis demonstrated that some of the genes, such as pB171_90_77 encoding a peptide deformylase, were present in nearly all of the *E. coli* genomes (4,764 of 4,798 genomes, 99%) (Table 1). Meanwhile, genes that were previously described as being unique to the B171 strain and a few other related EPEC

TABLE 1 Annotation of the EPEC strain B171 resistance plasmid, pB171_90

Gene ID or CDS[b]	Gene	Predicted protein function	Coding strand	Coordinates (bp)		No. of *E. coli* genomes (%)[a]
				Start	Stop	
pB171_90_1		CAAX protease self-immunity family protein	−	850	197	956 (19.92)
pB171_90_2	*repA*	IncFII family plasmid replication initiator RepA	−	2,649	1,792	3,214 (66.99)
pB171_90_3	*repB*	Replication regulatory RepB family protein	−	3,214	2,954	435 (9.07)
pB171_90_4		Hok/Gef family protein	−	3,704	3,498	1,359 (28.32)
pB171_90_5	*hha*	Hemolysin expression-modulating protein Hha	−	4,202	3,993	2,209 (46.04)
pB171_90_6		Conserved hypothetical protein	−	4,494	4,258	33 (0.69)
pB171_90_7		Conserved hypothetical protein	−	5,594	4,581	57 (1.19)
pB171_90_8		Putative gp46	−	6,009	5,701	58 (1.21)
pB171_90_9	*csi*	Putative Csi protein	−	6,716	6,006	31 (0.65)
pB171_90_10		Conserved hypothetical protein	−	7,066	6,713	31 (0.65)
pB171_90_11	*traX*	Type F conjugative transfer system pilin acetylase TraX	−	7,839	7,108	147 (3.06)
pB171_90_12		Conserved hypothetical protein	−	8,402	7,836	151 (3.15)
pB171_90_13	*traI*	Conjugative transfer relaxase protein TraI	−	13,717	8,420	134 (2.79)
pB171_90_14	*traD*	Type IV conjugative transfer system coupling protein TraD	−	15,948	13,717	135 (2.81)
pB171_90_15	*traT*	Enterobacterial TraT complement resistance family protein	−	16,802	16,092	3,017 (62.88)
pB171_90_16	*traS*	Putative protein TraS	−	17,549	17,004	71 (1.48)
pB171_90_17	*traG*	TraG-like, N-terminal region family protein	−	20,474	17,571	134 (2.79)
pB171_90_18		Conjugative relaxosome accessory transposon family protein	−	21,837	20,476	135 (2.81)
pB171_90_19	*trbB*	Type F conjugative transfer system pilin assembly thiol-disulfide isomerase TrbB	−	22,445	21,843	135 (2.81)
pB171_90_20	*traF*	Type F conjugative transfer system pilin assembly protein TraF	−	23,248	22,469	134 (2.79)
pB171_90_21	*traN*	Type F conjugative transfer system mating-pair stabilization protein TraN	−	25,059	23,245	133 (2.77)
pB171_90_22	*trbC*	Type F conjugative transfer system pilin assembly protein TrbC	−	25,693	25,013	132 (2.75)
pB171_90_23		Conserved hypothetical protein	−	26,058	25,690	113 (2.36)
pB171_90_24	*traU*	TraU family protein	−	27,055	26,069	2,118 (44.14)
pB171_90_25	*traW*	Type F conjugative transfer system protein TraW	−	27,687	27,052	138 (2.88)
pB171_90_26	*trbI*	Type F conjugative transfer system protein TrbI	−	28,151	27,687	132 (2.75)
pB171_90_27	*traC*	Type IV secretion system protein TraC	−	30,739	28,148	134 (2.79)
pB171_90_28	*traV*	Type IV conjugative transfer system protein TraV	−	31,349	30,750	137 (2.86)
pB171_90_29		Conserved hypothetical protein	−	31,819	31,409	99 (2.06)
pB171_90_30		Conserved hypothetical protein	−	31,988	31,803	102 (2.13)
pB171_90_31	*traR*	Prokaryotic DksA/TraR C4-type zinc finger family protein	−	32,216	31,992	140 (2.92)
pB171_90_32	*trbI*	Bacterial conjugation TrbI-like family protein	−	33,692	32,328	136 (2.83)
pB171_90_33	*traK*	Type F conjugative transfer system secretin TraK	−	34,419	33,679	138 (2.88)
pB171_90_34	*traE*	Type IV conjugative transfer system protein TraE	−	34,972	34,409	167 (3.48)
pB171_90_35	*traL*	Type IV conjugative transfer system protein TraL	−	35,297	34,992	178 (3.71)
pB171_90_36	*traA*	Type IV conjugative transfer system pilin TraA	−	35,658	35,299	193 (4.02)
pB171_90_37	*traJ*	Putative TraJ protein	−	36,419	36,138	97 (2.02)
pB171_90_38	*traM*	Relaxosome TraM domain protein	−	37,283	36,783	71 (1.48)
pB171_90_39		Integrase core domain protein	+	38,211	38,504	29 (0.60)
pB171_90_40		Putative membrane protein	−	38,880	38,665	257 (5.36)
pB171_90_41		Conserved hypothetical protein	−	39,463	39,221	363 (7.57)
pB171_90_42		Conserved hypothetical protein	−	39,786	39,505	2,115 (44.08)
pB171_90_43		Transposase family protein	+	40,493	41,143	497 (10.36)
pB171_90_44		Putative transposase	+	41,143	41,490	4,145 (86.39)
pB171_90_45		Transposase C of IS*166* homeodomain protein	+	41,510	43,081	1,225 (25.53)
pB171_90_46		Hok/Gef family protein	−	43,577	43,416	2,991 (62.34)
pB171_90_47		Conserved hypothetical protein	−	44,099	43,785	559 (11.65)
pB171_90_48	*psiA*	PsiA family protein	−	44,815	44,096	2,803 (58.42)
pB171_90_49	*psiB*	Protein PsiB	−	45,246	44,812	3,155 (65.76)
pB171_90_50		ParB-like nuclease domain protein	−	47,268	45,301	2,661 (55.46)
pB171_90_51		Conserved hypothetical protein	−	47,563	47,330	3,119 (65.01)
pB171_90_52	*ssbF*	Plasmid-derived single-stranded DNA-binding protein	−	48,147	47,620	3,080 (64.19)
pB171_90_53		Putative YddA protein	+	48,630	48,875	1,310 (27.30)
pB171_90_54		Methyltransferase small domain protein	−	49,524	48,961	2,546 (53.06)
pB171_90_55		Conserved hypothetical protein	−	50,930	49,569	2,554 (53.23)
pB171_90_56		Putative conserved predicted protein	−	51,212	50,982	2,622 (54.65)
pB171_90_57		Conserved hypothetical protein	+	51,475	51,732	2,354 (49.06)
pB171_90_58		Putative YchA	−	52,404	52,213	3,245 (67.63)
pB171_90_59		Conserved hypothetical protein	−	52,823	52,401	3,210 (66.90)
pB171_90_60		Antirestriction family protein	−	53,295	52,870	3,241 (67.55)
pB171_90_61		Putative YciA	−	53,705	53,544	2,014 (41.98)
pB171_90_62		Putative YfdA	−	54,478	53,708	2,911 (60.67)

(Continued on next page)

TABLE 1 (Continued)

Gene ID or CDS[b]	Gene	Predicted protein function	Coding strand	Coordinates (bp)		No. of *E. coli* genomes (%)[a]
				Start	Stop	
pB171_90_63		Conserved hypothetical protein	−	54,957	54,523	3,120 (65.03)
pB171_90_64		Conserved hypothetical protein	−	55,192	54,971	3,039 (63.34)
pB171_90_65		DNA methylase family protein	−	55,876	55,193	3,149 (65.63)
pB171_90_66		Putative YfbA	−	56,203	55,952	1,906 (39.72)
pB171_90_67		Conserved hypothetical protein	−	57,187	56,261	2,675 (55.75)
pB171_90_68	*parM*	Plasmid segregation protein ParM	−	58,977	57,997	385 (8.02)
pB171_90_69		Conserved hypothetical protein	−	59,932	59,366	347 (7.23)
pB171_90_70		Bacterial regulatory, TetR family protein	+	60,580	61,137	144 (3)
pB171_90_71		Conserved hypothetical protein	+	61,446	61,736	142 (2.96)
pB171_90_72		Conserved hypothetical protein	+	61,768	62,694	143 (2.98)
pB171_90_73		Conserved hypothetical protein	+	62,663	63,421	141 (2.94)
pB171_90_74		DoxX family protein	+	63,423	63,950	141 (2.94)
pB171_90_75		Conserved hypothetical protein	−	64,645	64,283	123 (2.56)
pB171_90_76		Hypothetical protein	+	64,852	64,971	395 (8.23)
pB171_90_77	*def*	Peptide deformylase	−	65,655	65,146	4,764 (99.29)
pB171_90_78		Hypothetical protein	+	65,785	65,895	144 (3)
pB171_90_79		Conserved hypothetical protein	+	66,183	66,548	1,079 (22.49)
pB171_90_80		Transposase zinc-binding domain protein	+	66,548	67,015	1,440 (30.01)
pB171_90_81		Transposase family protein	+	67,109	67,735	906 (18.88)
pB171_90_82		Transposase family protein	−	68,490	68,125	3,488 (72.70)
pB171_90_83		Conserved hypothetical protein	−	69,867	68,692	71 (1.48)
pB171_90_84		Conserved hypothetical protein	−	73,048	70,082	797 (16.61)
pB171_90_85		Resolvase, N-terminal domain protein	−	73,611	73,051	865 (18.03)
pB171_90_86		Conserved hypothetical protein	−	74,087	73,737	794 (16.55)
pB171_90_87	*intI1*	Integron integrase family protein	−	75,303	74,290	1,093 (22.78)
pB171_90_88	*aadA1*	Streptomycin 3″-adenylyltransferase	+	75,452	76,243	569 (11.86)
pB171_90_89	*qacEΔ1*	Small multidrug resistance family protein	+	76,407	76,754	957 (19.95)
pB171_90_90	*sul1*	Dihydropteroate synthase	+	76,748	77,587	942 (19.63)
pB171_90_91		Putative *Achromobacter xylosoxidans* subsp. *denitrificans* plasmid pAX22, partial DNA for integron In*70*	+	77,784	78,041	345 (7.19)
pB171_90_92		Transposase DDE domain protein	+	78,944	79,366	57 (1.19)
pB171_90_93		Putative relaxase/helicase	+	80,191	80,433	897 (18.70)
pB171_90_94	*tetR*	Tetracycline repressor protein class B from transposon Tn*10*	−	81,115	80,465	956 (19.92)
pB171_90_95	*tetA*	Tetracycline resistance protein, class C	+	81,221	82,420	1,008 (21.01)
pB171_90_96		Multidrug resistance efflux transporter family protein	−	83,234	82,452	841 (17.53)
pB171_90_97		Helix-turn-helix domain protein	−	84,004	83,297	809 (16.86)
pB171_90_98		EAL domain protein	−	84,786	84,079	702 (14.63)
pB171_90_99	*merE*	MerE family protein	−	85,019	84,783	715 (14.90)
pB171_90_100	*merD*	Mercuric resistance transcriptional repressor protein MerD	−	85,378	85,016	694 (14.46)
pB171_90_101	*merA*	Mercuric reductase	−	87,090	85,396	683 (14.24)
pB171_90_102	*merC*	MerC mercury resistance family protein	−	87,564	87,142	685 (14.28)
pB171_90_103	*merP*	Mercuric transport protein periplasmic component	−	87,875	87,600	720 (15.01)
pB171_90_104	*merT*	MerT mercuric transport family protein	−	88,239	87,889	733 (15.28)
pB171_90_105	*merR*	Hg(II)-responsive transcriptional regulator	+	88,311	88,745	757 (15.78)
pB171_90_106		Putative transposase	−	89,118	88,852	3,807 (79.35)
pB171_90_107		Hypothetical protein	+	89,301	89,516	3,793 (79.05)
pB171_90_108	*pemK*	mRNA interferase PemK	−	90,074	89,742	1,073 (22.36)

[a]The number and percentage of publicly available *E. coli* genomes that contained each gene with a BSR of ≥0.7. The total number of *E. coli* genome assemblies that were available in GenBank as of November 2016 was 4,798.
[b]CDS, protein-coding genes.

strains had a narrow distribution among the 4,798 *E. coli* genomes analyzed. The *csi* gene encoding a putative calcium sequestration inhibitor (GenBank accession number Y08258) (36) was, prior to this study, identified as unique in the B171 strain, and the *traI* gene encoding a conjugative transfer relaxase protein had only been reported from EPEC resistance plasmids (36). In this study, a much broader and sequence-based analysis identified the *csi* gene with significant similarity (BSR ≥ 0.7) in 31 of the 4,798 *E. coli* genomes (0.65%), including B171 (Table 2). Unlike the very narrow distribution of the *csi* gene, the *traI* gene was detected in more but still relatively few *E. coli* genomes (134 genomes total; 2.79% of the *E. coli* genomes) that did not contain *csi* (Table 2). Also, both *csi* and *traI* were significantly more prevalent among the AEEC genomes, such as the EPEC, compared to the non-AEEC genomes (chi-squared *P* value < 0.05)

TABLE 2 *In silico* detection of select pB171_90 genes in AEEC genomes versus other *E. coli*

Gene	No. of genomes (%)[a]			*P* value (non-AEEC vs AEEC)[e]
	All *E. coli*[b]	Non-AEEC[c]	AEEC[d]	
repA	3,214 (66.99)	2,225 (65.89)	989 (69.6)	0.01379
hha	2,209 (46.04)	1,306 (38.67)	903 (63.55)	<2.2*e*−16
csi	31 (0.65)	9 (0.27)	22 (1.55)	1.16*e*−06
traI	134 (2.79)	83 (2.46)	51 (3.59)	0.03796
tetR	956 (19.92)	780 (23.1)	176 (12.39)	<2.2*e*−16
tetA	1,008 (21.01)	806 (23.87)	202 (14.22)	9.04*e*−14
sulI	942 (19.63)	802 (23.75)	140 (9.85)	<2.2*e*−16
mer (≥4 genes)[f]	719 (14.99)	594 (17.59)	125 (8.8)	9.44*e*−15

[a]The total number of genomes that contained each of the pB171_90 genes listed. The number in parentheses is the percentage of genomes with each gene.
[b]The total number of *E. coli* genomes analyzed was 4,798. These were all of the *E. coli* assemblies available in GenBank as of November 2016.
[c]The total number of non-AEEC genomes analyzed was 3,377.
[d]The total number of AEEC genomes analyzed was 1,421. The AEEC genomes were identified by *in silico* analysis as those that had the LEE gene *escV*.
[e]The numbers of non-AEEC versus AEEC genomes that contained each gene were compared using the chi-square test.
[f]The number of genomes that contained ≥4 of the 7 total genes in the pB171_90 *mer* operon.

(Table 2). All of the *E. coli* genomes that contained *csi* also contained the *traI* gene, except for the genome of EPEC strain 302014, which contained *csi* but not *traI* (see Table S1 in the supplemental material). Interestingly, while several of the pB171_90 genes were significantly more associated with AEEC genomes (IncFII *repA*, *hha*, *csi*, and *traI* genes), the antibiotic resistance genes (*tetA*, *sul1*, and *mer* genes) were more frequently associated with the non-AEEC genomes (Table 2). These results highlight the mosaic nature of these plasmids, with the resistance genes having a different distribution from other genes, such as *csi* and *hha*, that are unique to pB171_90 and related plasmids.

We further investigated the genome-wide diversity of the *E. coli* strains that contain the unique genes of pB171_90 (*csi* and *traI*) by constructing a whole-genome phylogeny. Phylogenomic analysis of the *csi*- and/or *traI*-containing genomes with a collection of diverse *E. coli* and *Shigella* reference genomes demonstrated that the *csi*- and/or *traI*-containing *E. coli* genomes occurred in five of the six *E. coli* phylogroups (A, B1, B2, D, and E) (Fig. 2). Overall, the genomes analyzed in the phylogenomic analysis grouped together into clades by pathotype (Fig. 2). The *csi* gene, which was among the pB171_90 genes that had the narrowest distribution, was identified in genomes belonging to four phylogroups, but these genomes were concentrated in multiple clades or doublets of ETEC or EPEC genomes (Fig. 2). EPEC strain B171 was present in a subclade of the EPEC2 lineage (51) that also includes other *csi*-containing EPEC genomes (Fig. 2). Interestingly, among the STEC genomes that contained the unique pB171_90 genes *csi* and *traI* were two strains (K9107F and K9_45F) that were phylogenomically related to genomes of O157:H7 strains and were determined by molecular serotyping to have the O157:H7 serotype (Fig. 2; Table S1). The findings from the phylogenomic analysis demonstrate that *csi* and *traI* have been acquired by genomically diverse *E. coli* and are not restricted to EPEC strains that are closely related to strain B171.

In silico detection of the 108 protein-coding genes of pB171_90 in the *csi*- and/or *traI*-containing genomes demonstrated that nearly all of the pB171_90 genes were present with significant similarity (BSR ≥ 0.7) in EPEC genomes 102550 and 103447 (Fig. 3). In addition, the genome of ETEC strain TW10722 contained nearly all of the pB171_90 genes but was lacking several genes from the *mer* operon (Fig. 3). The frequencies of genes in the different plasmid regions (putative virulence genes *hha* and *csi*, conjugative transfer genes, and resistance genes) were demonstrated in the clustered heatmap (Fig. 3). Since all but one of the genomes analyzed contain the *traI*

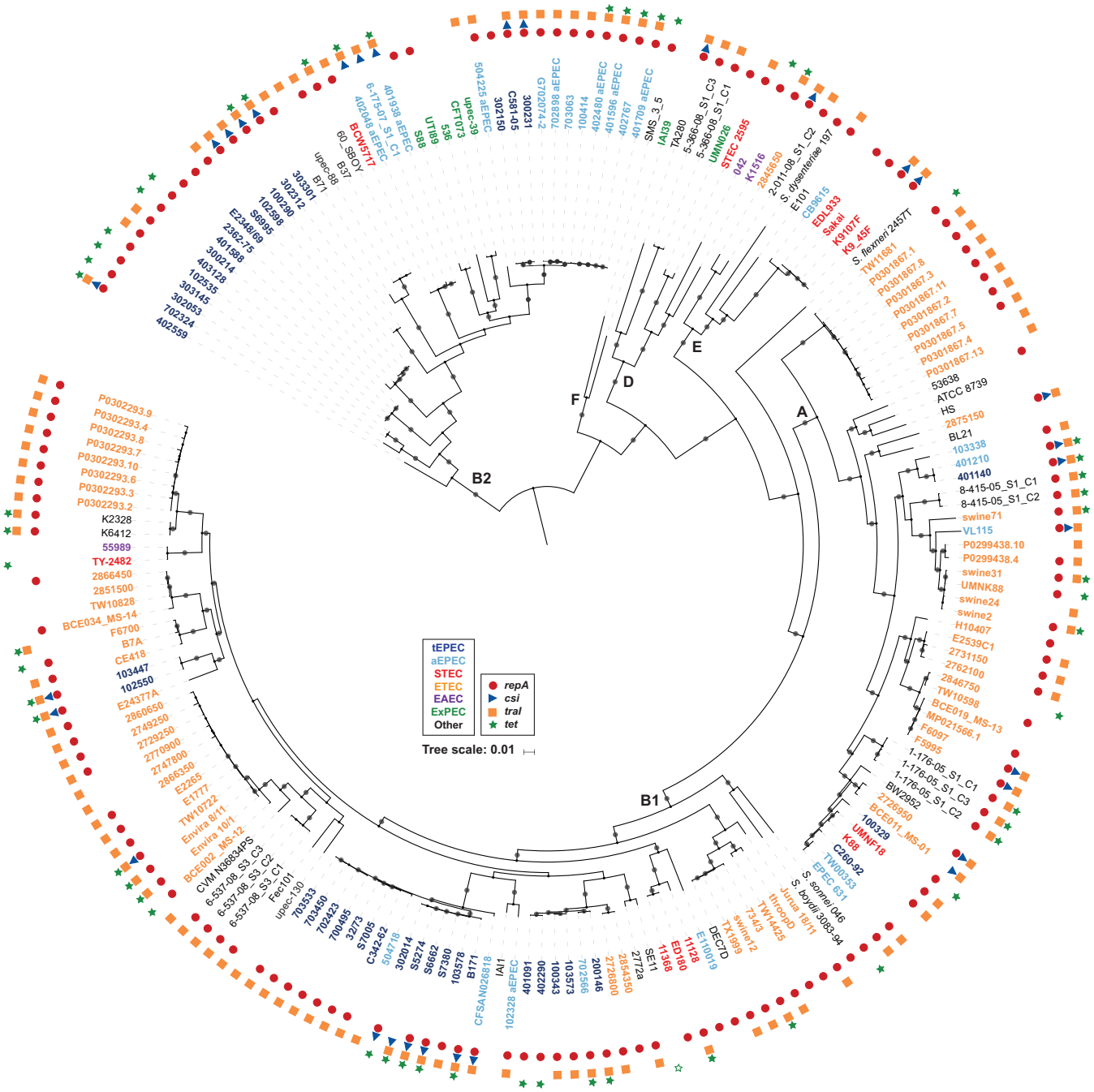

FIG 2 Phylogenomic analysis of *E. coli* genomes that carry genes with similarity to those of pB171_90. A SNP-based phylogeny was generated using ISG as previously described (35, 70). There were 149,150 conserved SNP sites identified among all of the *E. coli* genomes analyzed relative to the reference *E. coli* strain IAI39 (GenBank accession number NC_011750.1) that were used to construct a maximum-likelihood phylogeny with 100 bootstrap values using RAxML v.7.2.8 (71). The phylogeny was midpoint rooted and labeled using iTOL v.3 (72). Bootstrap values of ≥80 are designated by a gray circle. The scale bar indicates the approximate distance of 0.01 nucleotide substitutions per site. The *E. coli* phylogroups B1, A, B2, D, E, and F are denoted by bold letters. Symbols designate the presence of pB171_90 genes in each genome with *tetA* and *tetR* both represented by green stars (both genes, solid green star; one gene, open green star). Colors of the genome labels indicate the presumptive pathotype that was determined by *in silico* detection of canonical virulence genes.

gene, it was anticipated that these genomes would also contain most of the other genes involved in conjugative transfer (Fig. 3, green rectangle). The *csi*-containing genomes grouped together into several clusters (Fig. 3, yellow boxes); however, only two of the clusters also contained some or all of the resistance genes (Fig. 3, red boxes). The resistance genes were present mostly in genomes of three different clusters in the heatmap, of which two of the clusters also contained the *csi* and *hha*

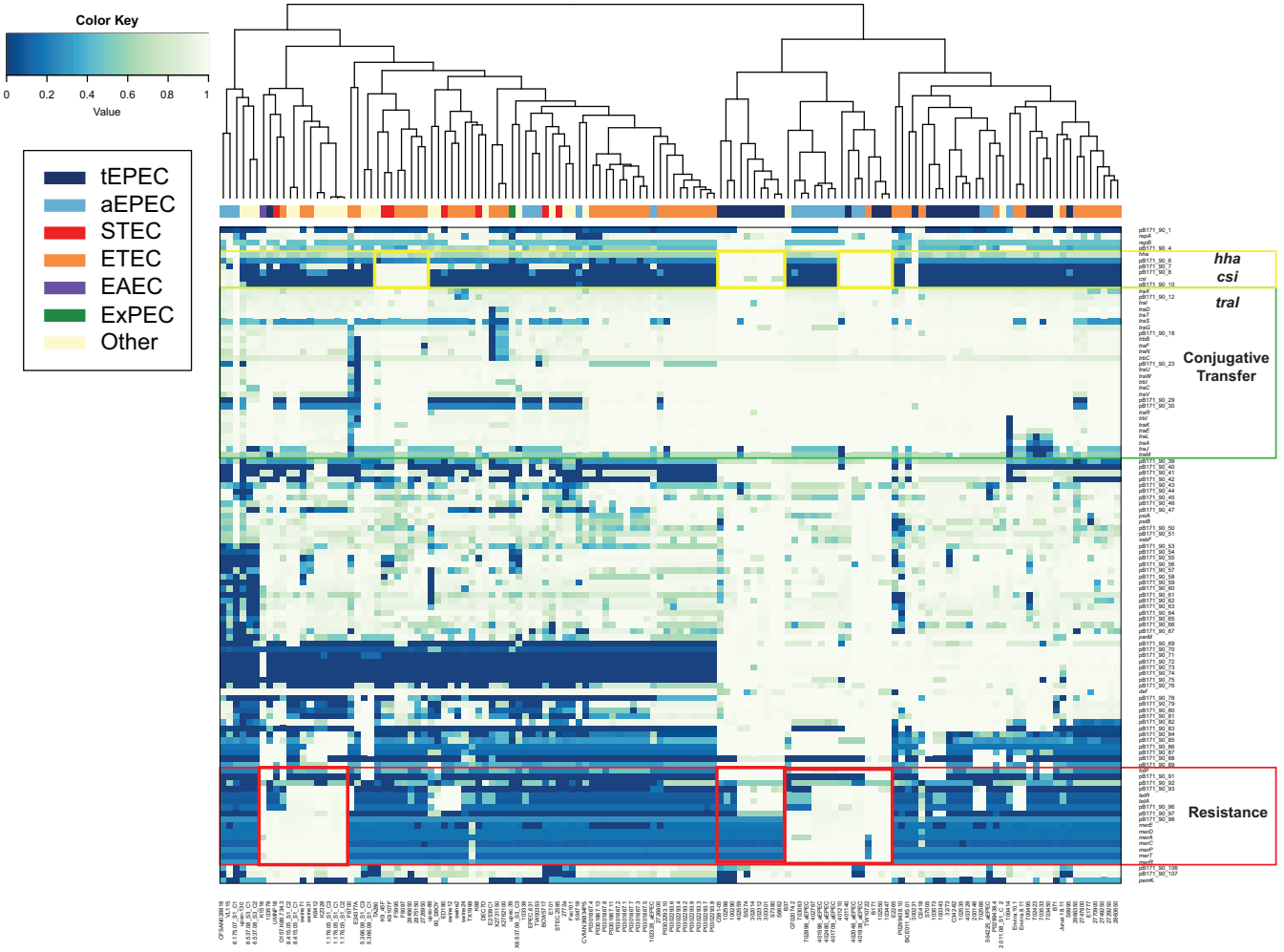

FIG 3 *In silico* detection of pB171_90 genes in diverse *E. coli* genomes. The presence or absence of all predicted protein-coding genes of pB171_90 in the 134 *E. coli* genomes that contained the pB171_90 genes *csi* and/or *tral* is indicated by the BSR values in the heatmap (see Table S1 in the supplemental material). The clustered heatmap was generated using the heatmap2 function of gplots v.3.0.1 in R v.3.3.2. The genomes were clustered using the complete linkage method with Euclidean distance estimation. The pB171_90 genes are represented by the rows, while each column represents a different *E. coli* genome. The predicted pathotype of each *E. coli* genome is indicated by the color of the rectangle at the top of the heatmap (see legend for detail).

genes while the third cluster that did not (Fig. 3, red boxes). These results demonstrate that the unique genes of the pB171_90 plasmid (*csi* and *hha*) do not travel exclusively with the antibiotic resistance genes of the pB171_90 plasmid, highlighting the mosaic nature of pB171_90 and related plasmids.

Phylogenetic analysis of the *csi* nucleotide sequences demonstrated that *csi* sequences from all but three of the genomes (aEPEC strains VL115 and CFSAN026818 and *E. coli* strain TA280) were present in two clades that we have designated clades I and II (Fig. 4). Clade I contained all of the *csi* sequences from tEPEC genomes, four of the *csi* sequences from aEPEC genomes, and one *csi* sequence from an ETEC genome (Fig. 4). The genomes of clade I belonged to phylogroups A, B1, and B2 and had at least 10 different MLSTs, while clade II contained *csi* sequences from six out of seven (85.7%) of the ETEC genomes and both of the STEC genomes (Fig. 4). Clade I genomes belonging to phylogroups A and E had four different MLSTs (Fig. 4). Thus, the *csi* genes exhibit the greatest similarity among genomes belonging to the same *E. coli* pathotype rather than from genomes of the same *E. coli* phylogroup.

Phylogenetic analysis of the *tral* gene sequences demonstrated that there was greater pathotype specificity than phylogroup specificity (Fig. 5). There were clades of *tral* sequences from ETEC genomes of different phylogroups (A or B1) and other clades that contained nearly all EPEC genomes also from multiple phylogroups (A, B1, and B2).

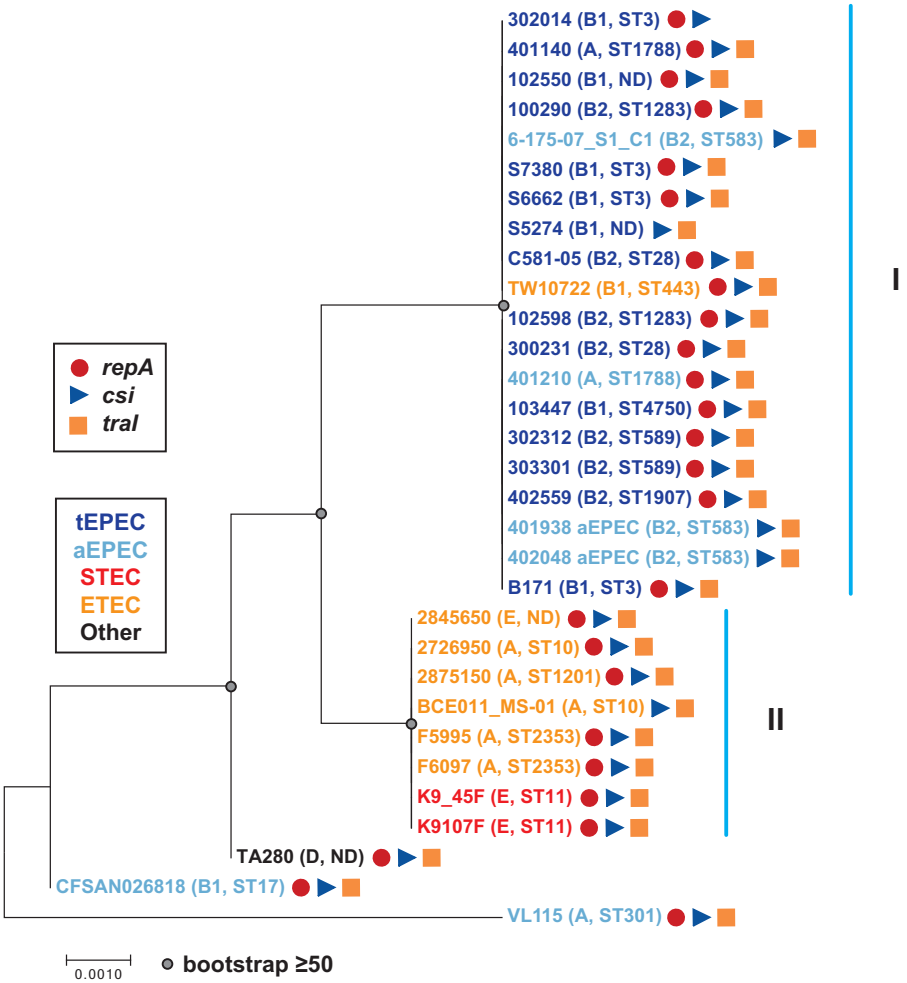

FIG 4 Phylogenetic analysis of *csi*. The *csi* nucleotide sequences of pB171_90 and other *E. coli* genomes were aligned using ClustalW. The alignment was used to construct a maximum-likelihood phylogeny with the Kimura 2-parameter model and 1,000 bootstraps using MEGA7 (79). The scale bar represents the approximate distance of 0.001 nucleotide substitutions per site. Bootstrap values of ≥50 are indicated by a gray circle. The predicted *E. coli* pathotypes are indicated by the color of their genome label (see legend), and the phylogroup and MLST of each genome are denoted in parentheses.

The genomes that formed clades I and II in the *csi* phylogeny formed similar clades in the *tral* phylogeny (Fig. 5). The consistent sequence similarity of the *csi* and *tral* genes occurring in the same genomes, along with the close proximity of *csi* (pB171_90_9) and *tral* (pB171_90_13) in the pB171_90 plasmid (Table 1), suggests that the *csi* and *tral* genes travel together.

The replication protein-coding gene of pB171_90, *repA* (pB171_90_2), was also significantly more associated with the AEEC genomes (69%) than with the non-AEEC genomes (65%), with a chi-squared *P* value of 0.013 (Table 2). The pB171_90 *repA* gene is most similar to the IncFIIA *repA* gene that has been identified in previously sequenced *E. coli* virulence plasmids, including plasmids from AEEC genomes (26, 35). Based on the IncF replicon typing scheme of Villa et al., the pB171 virulence plasmid belongs to the F13:A^{-}:B4 IncF replicon type (52), while the resistance plasmid pB171_90 belongs to the F79:A^{-}:B^{-} replicon type as determined by plasmid multilocus sequence typing (pMLST) analysis (53). Alignment of the *repA* nucleotide sequences from plasmids pB171_90 and pB171_69 indicated that they have 95% nucleotide identity (816/858 identical bases). Thus, both plasmids have related, but distinct, plasmid replicon types. In a previous study investigating the diversity of AEEC plasmids, we also identified multiple IncFII replicon types in the same genome (35). In contrast to the

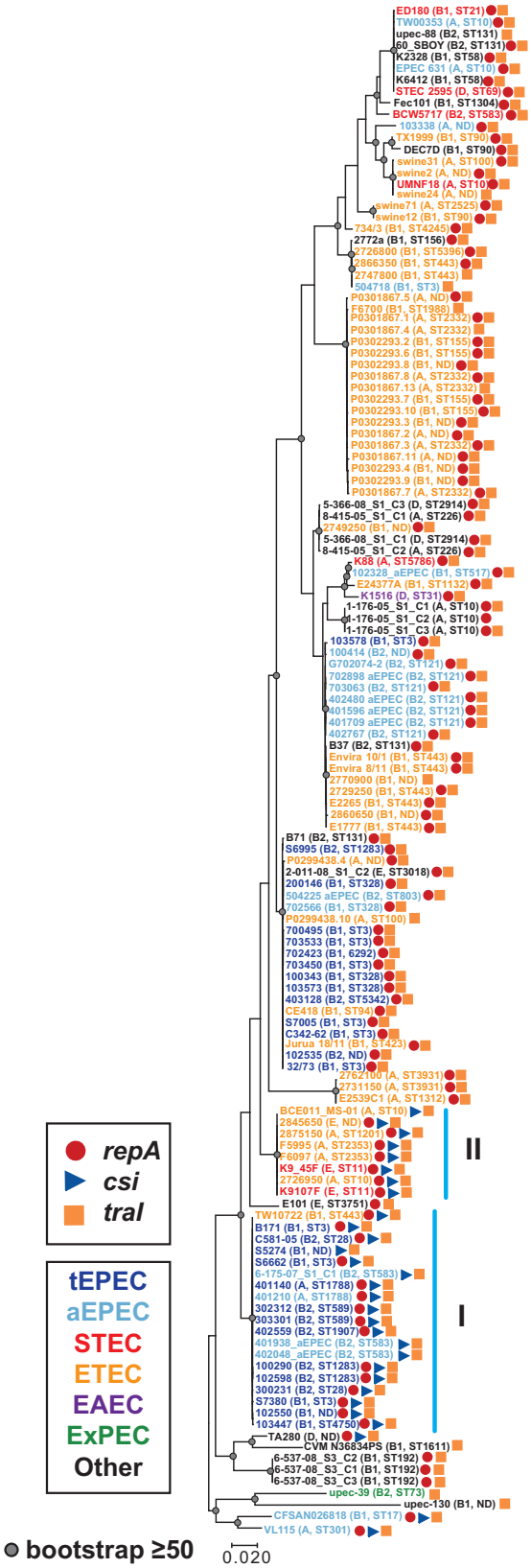

FIG 5 Phylogenetic analysis of *tral*. The *tral* nucleotide sequences of pB171_90 and other *E. coli* strains were aligned using ClustalW. The alignment was used to construct a maximum-likelihood phylogeny with the Kimura 2-parameter model and 1,000 bootstraps using MEGA7 (79). The scale bar represents the approximate distance of 0.02 nucleotide substitutions per site. Bootstrap values of ≥50 are indicated by a gray circle. The predicted *E. coli* pathotypes are indicated by the color of their genome label (see legend), and the phylogroup and MLST of each genome is denoted in parentheses. The blue vertical lines indicate the two clades identified in the *csi* phylogeny (Fig. 4).

formation of clades I and II by the *csi*-containing genomes in both the *csi* and *traI* phylogenies, phylogenetic analysis of *repA* sequences demonstrated that the similarity of the *repA* gene sequences differs among some of the *csi*-containing genomes. The *repA* sequences from genomes in clade I of the *csi* and *traI* phylogenies formed a similar clade, while the genomes from clade II did not form a clade and were instead distributed throughout the phylogeny (see Fig. S1 in the supplemental material). Interestingly, the genomes of clade I were primarily tEPEC and aEPEC genomes with one ETEC genome, while the genomes of clade II were ETEC and STEC (Fig. 4 and 5; Fig. S1). This suggests that the *csi* and *traI* genes were likely acquired as part of the same plasmid in the EPEC strains, whereas the *csi* and *traI* genes of the STEC and ETEC genomes are associated with plasmids that contain a different IncFII replicon type.

Another gene of interest on pB171_90 is the virulence gene transcriptional regulator, *hha*, which was identified in 46% of the *E. coli* genomes and was significantly more associated with AEEC genomes over non-AEEC genomes (chi-squared *P* value $<$ 2.2*e*$-$16) (Tables 1 and 2). Similar to the *repA* gene, phylogenetic analysis of the *hha* genes identified in the *csi* and/or *traI*-containing genomes demonstrated that the clade I genomes from the *csi* phylogeny also grouped together in the *hha* phylogeny, while the clade II genomes were distributed throughout the phylogeny (see Fig. S2 in the supplemental material). Interestingly, the *hha* genes identified in the *csi* and/or *traI*-containing genomes were more similar to the *hha* genes from plasmid pB171_90 and the *E. coli* O157:H7 plasmid pO157 from strain EDL933 (Fig. S2). These findings suggest that the *hha* genes identified in the clade I genomes may be located on plasmids containing the *repA*, *csi*, and *traI* genes, demonstrating that plasmids with similarity to pB171_90 exist among genomically diverse EPEC strains (Fig. S2).

Identification of pB171_90 genes in the genomes of other human enteric pathogens. We further investigated whether unique genes of pB171_90, such as *csi*, were present in other species of bacteria by using BLASTP to compare the predicted amino acid sequence of the Csi peptide to the nonredundant (nr) protein sequence database of GenBank. This query identified 16 genome or plasmid sequences of enteric pathogens other than *E. coli* that contained a region encoding a protein with similarity to Csi (Fig. 6; Table S1). Among these non-*E. coli csi*-containing genomes were a *Shigella dysenteriae* type I plasmid, pCAR10 (GenBank accession number KT754161.1) (54), and whole-genome sequences of six *Shigella sonnei*, two *Citrobacter freundii*, one *Klebsiella pneumoniae*, and six *Salmonella enterica* strains (Fig. 6).

Phylogenetic analysis of the *csi* genes from the *E. coli* genomes compared with the *csi* genes that were identified in other human enteric pathogens demonstrated that the *csi* genes formed three different phylogenetic groups (see Fig. S3 in the supplemental material). Group 1 contained all of the *csi* genes from the *E. coli* genomes as well as those from the *Shigella* genomes (*S. dysenteriae* and *S. sonnei*) and the single sequence from a *Klebsiella* genome (Fig. S3). Group 2 contained only the *csi* sequences from *Salmonella* genomes, while group 3 was made up of the *csi* sequences from the *Citrobacter* genomes (Fig. S3). The *csi* phylogenetic analysis demonstrates that there is species specificity among the *csi* genes.

In silico detection of the 108 protein-coding genes of pB171_90 in each of the *csi*-containing enteric species genomes demonstrated that $>$50% of the plasmid genes were present with significant similarity (BSR $\geq$ 0.7) in the *S. dysenteriae* type I plasmid pCAR10 and also in the *K. pneumoniae* genome and two of the *S. sonnei* genomes (Fig. 6). The Tn*21* transposon of pB171_90 that contains the *aadA1*, *sul1*, and *qacE*Δ*1*, was also detected with 100% nucleotide identity in sequenced plasmids and genomes from other species, including several plasmids from *S. dysenteriae* (p69-3818, GenBank accession number KT754167.1; p92-9000, GenBank accession number KT754166.1; and p80-547, GenBank accession number KT754160.1).

Comparison of the predicted protein-coding genes of the *S. dysenteriae* type I plasmid, pCAR10 (GenBank accession number KT754161.1) (54), with those of pB171_90, highlighted the significant similarity of these two antibiotic resistance carrying plasmids both in terms of gene content and order and also their levels of nucleotide similarity

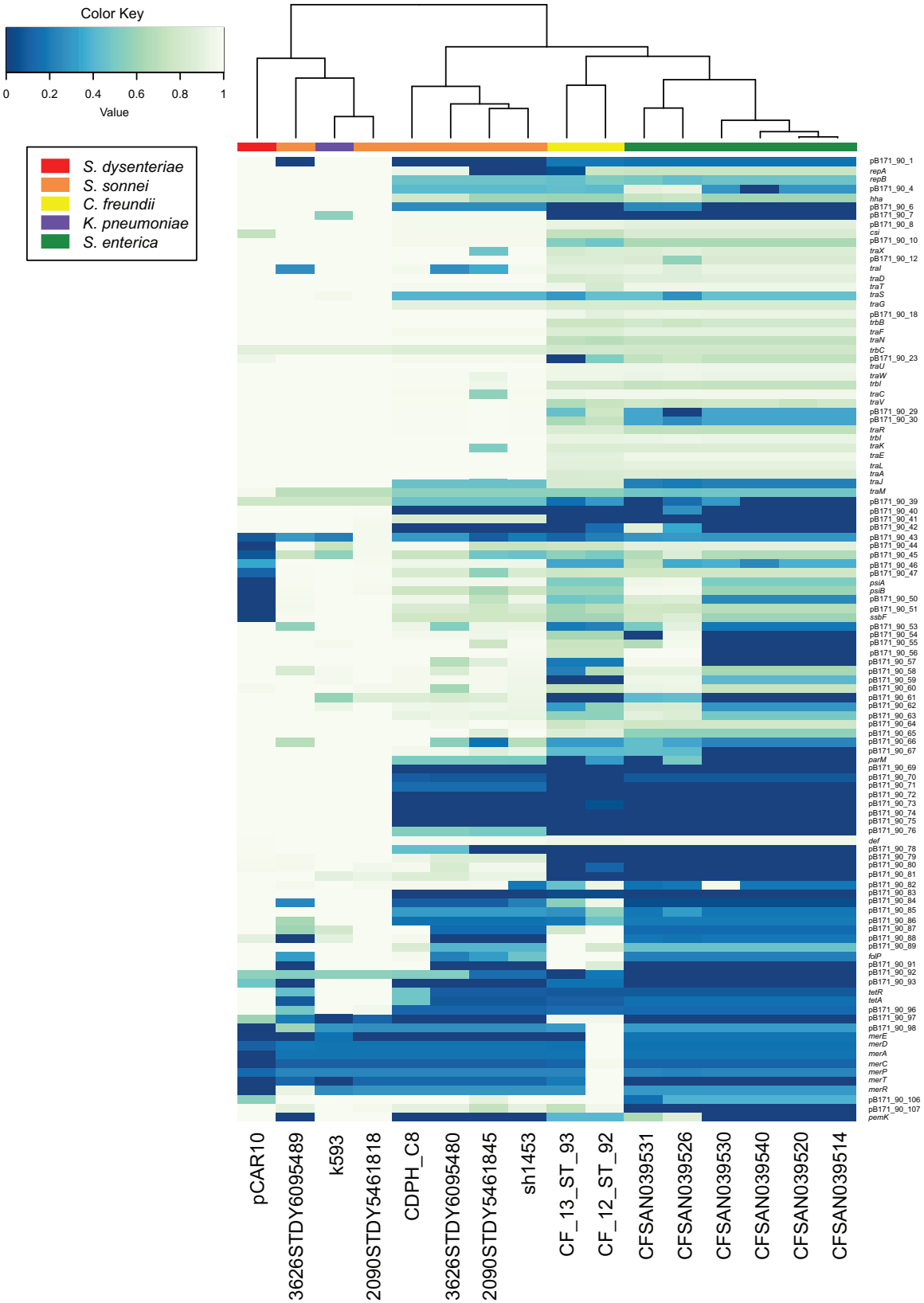

FIG 6 *In silico* detection of pB171_90 genes in other human enteric pathogens. The predicted protein-coding genes of pB171_90 were identified in additional human enteric pathogens that contained the pB171_90 gene *csi* and are listed in Table S1 in the supplemental material. The clustered heatmap was generated using the heatmap2 function of gplots v.3.0.1 in R v.3.3.2. The genomes were clustered using the complete linkage method with Euclidean distance estimation. The pB171_90 genes are represented by the rows, while each column represents a different genome. The species are designated by the color of the rectangle at the top of the heatmap (see legend).

(Fig. 6; see also Fig. S4 in the supplemental material). Interestingly, the replication protein-coding gene, the *csi* gene, and the conjugative transfer genes were highly conserved (BSR $\geq$ 0.7) between the two plasmids (Fig. S4). However, some genes, such as those involved in plasmid stability and the antibiotic resistance genes, differed between pB171_90 and pCAR10 (Fig. S4). Of the antibiotic resistance genes identified on pB171_90, pCAR10 also carries the tetracycline resistance gene *tetA*, the *tetA* repressor *tetR*, and the sulfonamide resistance gene *sul1* (Table 1; Fig. S4). However, pCAR10 contains additional antibiotic resistance genes that are missing from pB171_90, including a predicted bla_{TEM} β-lactamase gene and a dihydrofolate reductase gene, *dfrA15* (Table 1; Fig. S4). The pCAR10 plasmid also contains a small transporter designated *emrE*, which confers resistance to ethidium bromide, and has 100% nucleotide identity to the small multidrug resistance protein-coding gene *qacEΔ1* (pB171_90_89) of pB171_90 (Table 1; Fig. S4). The similarity of pB171_90 genes to genes identified in other species indicates that EPEC resistance plasmids may be transferred between EPEC and other human pathogens, including *S. dysenteriae* type I.

DISCUSSION

An increased prevalence of antibiotic resistance among certain groups of disease-causing *E. coli*, and the identification of *E. coli* as a reservoir of antibiotic resistance for routinely prescribed antibiotics (1, 9, 10, 55), highlights the importance of investigating the diversity of plasmids involved in disseminating resistance genes among *E. coli*. The pB171_90 plasmid belongs to the IncFII plasmid family, which counts among its members many of the virulence plasmids of diverse pathotypes of *E. coli*, including the virulence plasmids of EPEC (26). IncFII plasmids have also been described that carry antibiotic resistance genes among *E. coli* and other members of the family *Enterobacteriaceae* (26, 52) and have been reported to carry extended-spectrum β-lactamases (ESBLs), such as $bla_{CTX\text{-}M\text{-}15}$ (8, 56, 57). To provide insight into the diversity of resistance plasmids among EPEC strains, we provide the complete sequence and description of the antibiotic and heavy metal resistance plasmid, pB171_90, from the *E. coli* reference strain B171.

By investigating the presence of the pB171_90 genes among a large collection of *E. coli* genomes, we determined that genes such as *csi* and *traI* that were previously identified as unique to the B171 resistance plasmid (36) can be found in genomically diverse *E. coli* belonging to phylogroups A, B1, B2, D, and E (Fig. 2) as well as other enteric pathogens. Although the majority of the genomes that exhibited similarity to the genes of pB171_90 were identified as EPEC or ETEC, the pB171_90 genes were also identified in genomes of other diverse pathotypes (EAEC, STEC, ExPEC/uropathogenic *E. coli* [UPEC]) as well as *E. coli* genomes that didn't contain any of the pathotype-specific genes and may not be associated with illness (see Table S1 in the supplemental material). Interestingly, pB171_90 also contained a copy of the transcriptional regulator *hha*, which has been previously identified on the chromosome and plasmid of *E. coli* O157:H7 (46, 50). The *hha* gene was initially described as a regulator of hemolysin production but has since been described as a regulator of the LEE pathogenicity region and biofilm formation (47–49). Further studies are necessary to determine the role of *hha* in regulating virulence genes of EPEC strain B171. The pB171_90 plasmid also had significant similarity with genes of two O157:H7 *E. coli* genomes as well as the pCAR10 plasmid from a *S. dysenteriae* type I strain (54). This finding suggests that an exchange of resistance plasmids has occurred between EPEC and the phylogenomically related EHEC O157:H7 and *S. dysenteriae*, which cooccur together in *E. coli* phylogroup E.

Another notable feature of the pB171_90 resistance plasmid is the presence of a Tn*21* transposon containing a spectinomycin-streptomycin resistance gene, a sulfonamide resistance gene, a small efflux pump, and a class 1 integrase gene, *intI1*. In a study investigating the prevalence of class 1 integrons among *E. coli* and *Klebsiella* isolated from hospitalized patients, the class 1 integrons were detected in 49% of the *E. coli* strains and were significantly associated with resistance to several classes of antibiotics (58). Comparison of pB171_90 with the sequenced EPEC and EHEC resis-

tance plasmids pO111-CRL$_{115}$ (GenBank accession number KC340959) and pO26-CRL$_{125}$ (GenBank accession number KC340960) demonstrated that the tetracycline resistance and mercury resistance regions are conserved between these plasmids. Although plasmids pO111-CRL$_{115}$ and pO26-CRL$_{125}$ also contain a class 1 integron (59), the antibiotic resistance genes in this cassette differed from those of pB171_90, emphasizing the modular structure of *E. coli* resistance plasmids.

In summary, by characterizing the resistance plasmid of the EPEC reference strain B171, we provide further insight into the sequence diversity and host range of antibiotic and heavy metal resistance plasmids of *E. coli*. There were only a few publicly available *E. coli* genomes that contained nearly all of the resistance plasmid genes; however, genes from certain plasmid regions, including the antibiotic and heavy metal resistance genes, were present in diverse *E. coli* and other members of the family *Enterobacteriaceae*. Our study highlights the need for additional complete sequencing of resistance plasmids from EPEC and other pathotypes of *E. coli* to provide further insight into the dissemination and evolution of the *E. coli* resistance plasmids.

MATERIALS AND METHODS

Plasmid sequencing. Plasmid DNA of EPEC strain B171 (37) was isolated from 1 liter of culture grown in L broth using alkaline lysis extraction method followed by ethidium bromide-CsCl density gradient centrifugation as previously described (60). The plasmid band was isolated and rebanded (190,000 × *g*) overnight on a second identical gradient. Plasmid was siphoned using a 3-ml syringe and an 18-gauge needle. Ethidium bromide was removed using isopropanol-CsCl-saturated 10 mM Tris, 5 mM EDTA, and 10 mM NaCl (pH 8) and then dialyzed against three changes of 10 mM Tris and 1 mM EDTA (pH 8). Plasmid DNA was ethanol precipitated using 0.125 M NaCl (final) and dissolved in water prior to sequencing.

The purified plasmid DNA was sequenced using the Pacific Biosciences RS II platform with the P6C4 chemistry in a single flow cell. Pacific Biosciences sequencing was completed by generating a sequencing library from 5 to 20 kb using standard methods at the Institute for Genome Sciences Genomics Resource Center (http://www.igs.umaryland.edu/resources/grc/). A total of 115,336 reads were generated with an average read length of 7,036 bp and a maximum read length of 38,154 bp. PacBio raw data were corrected and assembled using HGAP (SMRTAnalysis 2.3.0) (61), Canu assembler v.1.2 (62), and Celera assembler v.8.2 (63) run with default parameters. The assemblies were assessed for inconsistencies and misassembly using Nucmer whole-genome alignments (64) and Circleator plots (GC skew) (65). A BLAST search against NCBI's complete nucleotide database was used to confirm that all of the contigs were plasmids (65). The Canu assembler generated the most complete assembly for each of the plasmids in this sample. Circular contigs were identified and Minimus2 was used for circularization (66). The final circularized assembly was polished using Quiver (SMRTAnalysis 2.3.0) (61).

***In silico* multilocus sequence typing and serotyping.** The seven loci (*adk*, *gyrB*, *fumC*, *icd*, *mdh*, *purA*, and *recA*) of the multilocus sequence typing (MLST) scheme developed by Wirth et al. (67) were identified in each of the genomes listed in Table S1 in the supplemental material. The allele sequences were used to query BIGSdb (68) to obtain the allele numbers and sequence type of each of the *E. coli* genomes analyzed.

The molecular serotype of each genome was determined using Serotype Finder v.1.1 (https://cge.cbs.dtu.dk/services/SerotypeFinder/) with the default settings of an 85% identity threshold and 60% minimum alignment length (69).

Phylogenomic analysis. The 134 *E. coli* genomes that contained genes with similarity to *csi* and/or *traI* of the B171 resistance plasmid were compared with 55 previously sequenced *E. coli* and *Shigella* genomes using the SNP-based *in silico* genotyper (ISG) (70) as previously described (35) (Table S1). There were 149,150 conserved SNP sites identified among all of the *E. coli* genomes analyzed relative to the *E. coli* reference strain IAI39 (GenBank accession number NC_011750.1) that were used to generate a maximum-likelihood phylogeny with 100 bootstrap values using RAxML v.7.2.8 (71). The phylogeny was constructed using the general time reversible (GTR) model of nucleotide substitution with the GAMMA model of rate heterogeneity and 100 bootstrap replicates. The phylogeny was then midpoint rooted and labeled using the interactive tree of life software (iTOL v.3) (72).

***In silico* detection of pB171_90 genes.** The genes of pB171_90 were predicted and annotated using an in-house annotation pipeline (73). These genes were then detected in a collection of 4,798 *E. coli* genome assemblies available in GenBank as of November 2016, using large-scale BLAST score ratio (LS-BSR) analysis as previously described (74, 75). The pB171_90 protein-coding genes were compared to each genome listed in Table S1 using TBLASTN (76) with composition-based adjustment turned off. The TBLASTN scores were used to generate a BSR indicating the detection of each gene cluster in each of the genomes (Table S1). The BSR was determined by dividing the score of a gene compared to a genome by the score of the gene compared to its own sequence. The genes that were identified with a BSR of ≥0.7 were considered present in the genomes analyzed. The numbers of genes that were present in AEEC compared to non-AEEC genomes was examined for statistical significance using the chi-squared test with Yates' continuity correction performed using R v.3.3.2. The clustered heatmap was generated

using the heatmap2 function of gplots v.3.0.1 in R v.3.3.2. The BSR values of the genomes were clustered in the heatmap using the complete linkage method with Euclidean distance estimation.

Pathotype-specific virulence genes were identified in each of the genomes by *in silico* analysis using LS-BSR as described above. The following canonical virulence genes, which have been previously used for the molecular-based classification of clinical *E. coli* strains to each of the pathotypes, were detected in each of the following *E. coli* genomes analyzed: tEPEC (LEE genes *eae* and *escV* and BFP genes *bfpA* and *bfpB*), aEPEC (LEE genes *eae* and *escV* only; no BFP or Shiga toxin genes), ETEC (*eltA* and *eltB* encoding the heat-labile toxin LT and/or *est* encoding the heat-stable toxin ST), EAEC (the transcriptional regulator *aggR*, *aatA* encoding an outermembrane protein, or *pic* and/or *pet* encoding autotransporters), STEC (Shiga toxin genes *stxA* and *stxB*), EIEC (*ipaB*, *mxi-spa*, and/or *ospB*, which are all involved in type III secretion), and ExPEC/UPEC (*papG* encoding a fimbrial adhesion precursor) (19–21, 77, 78).

Gene phylogenies. Individual gene phylogenies were constructed for the *repA*, *hha*, *csi*, and *traI* sequences that were detected by *in silico* analysis of each of the *E. coli* genomes analyzed (Table S1). A *csi* phylogeny was also constructed with the *csi* sequences from *E. coli* and other species of human enteric pathogens listed in Table S1. The nucleotide sequences of each gene were aligned using ClustalW, and maximum-likelihood phylogenies were constructed using the Kimura 2-parameter model and 1,000 bootstraps with MEGA7 (79).

Accession number(s). The assemblies of pB171_90 and pB171_68 are deposited in GenBank under the accession numbers CP021212 and CP021211, respectively.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/AAC.00995-17.

SUPPLEMENTAL FILE 1, PDF file, 0.9 MB.

ACKNOWLEDGMENT

This project was funded in part by federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under grant U19 AI110820.

REFERENCES

1. Johnson JR, Johnston B, Clabots C, Kuskowski MA, Castanheira M. 2010. *Escherichia coli* sequence type ST131 as the major cause of serious multidrug-resistant *E. coli* infections in the United States. Clin Infect Dis 51:286–294. https://doi.org/10.1086/653932.
2. Nordmann P, Naas T, Poirel L. 2011. Global spread of carbapenemase-producing *Enterobacteriaceae*. Emerg Infect Dis 17:1791–1798. https://doi.org/10.3201/eid1710.110655.
3. Pitout JD, Laupland KB. 2008. Extended-spectrum beta-lactamase-producing *Enterobacteriaceae*: an emerging public-health concern. Lancet Infect Dis 8:159–166. https://doi.org/10.1016/S1473-3099(08)70041-0.
4. Mediavilla JR, Patrawalla A, Chen L, Chavda KD, Mathema B, Vinnard C, Dever LL, Kreiswirth BN. 2016. Colistin- and carbapenem-resistant *Escherichia coli* harboring *mcr-1* and bla_{NDM-5}, causing a complicated urinary tract infection in a patient from the United States. mBio 7:e01191-16. https://doi.org/10.1128/mBio.01191-16.
5. Schwarz S, Johnson AP. 2016. Transferable resistance to colistin: a new but old threat. J Antimicrob Chemother 71:2066–2070. https://doi.org/10.1093/jac/dkw274.
6. McGann P, Snesrud E, Maybank R, Corey B, Ong AC, Clifford R, Hinkle M, Whitman T, Lesho E, Schaecher KE. 2016. *Escherichia coli* harboring *mcr-1* and bla_{CTX-M} on a novel IncF plasmid: first report of *mcr-1* in the United States. Antimicrob Agents Chemother 60:4420–4421. https://doi.org/10.1128/AAC.01103-16.
7. Yu H, Qu F, Shan B, Huang B, Jia W, Chen C, Li A, Miao M, Zhang X, Bao C, Xu Y, Chavda KD, Tang YW, Kreiswirth BN, Du H, Chen L. 2016. Detection of the *mcr-1* colistin resistance gene in carbapenem-resistant *Enterobacteriaceae* from different hospitals in China. Antimicrob Agents Chemother 60:5033–5035. https://doi.org/10.1128/AAC.00440-16.
8. Stoesser N, Sheppard AE, Pankhurst L, De Maio N, Moore CE, Sebra R, Turner P, Anson LW, Kasarskis A, Batty EM, Kos V, Wilson DJ, Phetsouvanh R, Wyllie D, Sokurenko E, Manges AR, Johnson TJ, Price LB, Peto TE, Johnson JR, Didelot X, Walker AS, Crook DW, Modernizing Medical Microbiology Informatics Group. 2016. Evolutionary history of the global emergence of the *Escherichia coli* epidemic clone ST131. mBio 7:e02162. https://doi.org/10.1128/mBio.02162-15.
9. Phuc Nguyen MC, Woerther PL, Bouvet M, Andremont A, Leclercq R, Canu A. 2009. *Escherichia coli* as reservoir for macrolide resistance genes. Emerg Infect Dis 15:1648–1650. https://doi.org/10.3201/eid1510.090696.
10. Erb A, Sturmer T, Marre R, Brenner H. 2007. Prevalence of antibiotic resistance in *Escherichia coli*: overview of geographical, temporal, and methodological variations. Eur J Clin Microbiol Infect Dis 26:83–90. https://doi.org/10.1007/s10096-006-0248-2.
11. Karami N, Nowrouzian F, Adlerberth I, Wold AE. 2006. Tetracycline resistance in *Escherichia coli* and persistence in the infantile colonic microbiota. Antimicrob Agents Chemother 50:156–161. https://doi.org/10.1128/AAC.50.1.156-161.2006.
12. Labar AS, Millman JS, Ruebush E, Opintan JA, Bishar RA, Aboderin AO, Newman MJ, Lamikanra A, Okeke IN. 2012. Regional dissemination of a trimethoprim-resistance gene cassette via a successful transposable element. PLoS One 7:e38142. https://doi.org/10.1371/journal.pone.0038142.
13. Nataro JP, Maher KO, Mackie P, Kaper JB. 1987. Characterization of plasmids encoding the adherence factor of enteropathogenic *Escherichia coli*. Infect Immun 55:2370–2377.
14. Gross RJ, Ward LR, Threlfall EJ, King H, Rowe B. 1982. Drug resistance among infantile enteropathogenic *Escherichia coli* strains isolated in the United Kingdom. Br Med J (Clin Res Ed) 285:472–473. https://doi.org/10.1136/bmj.285.6340.472.
15. Guerra B, Junker E, Schroeter A, Helmuth R, Guth BE, Beutin L. 2006. Phenotypic and genotypic characterization of antimicrobial resistance in *Escherichia coli* O111 isolates. J Antimicrob Chemother 57:1210–1214. https://doi.org/10.1093/jac/dkl127.
16. Moyenuddin M, Wachsmuth IK, Moseley SL, Bopp CA, Blake PA. 1989. Serotype, antimicrobial resistance, and adherence properties of *Escherichia coli* strains associated with outbreaks of diarrheal illness in children in the United States. J Clin Microbiol 27:2234–2239.
17. Vignoli R, Varela G, Mota MI, Cordeiro NF, Power P, Ingold E, Gadea P, Sirok A, Schelotto F, Ayala JA, Gutkind G. 2005. Enteropathogenic *Escherichia coli* strains carrying genes encoding the PER-2 and TEM-116 extended-spectrum beta-lactamases isolated from children with diarrhea in Uruguay. J Clin Microbiol 43:2940–2943. https://doi.org/10.1128/JCM.43.6.2940-2943.2005.
18. Scaletsky IC, Souza TB, Aranda KR, Okeke IN. 2010. Genetic elements

associated with antimicrobial resistance in enteropathogenic *Escherichia coli* (EPEC) from Brazil. BMC Microbiol 10:25. https://doi.org/10.1186/1471-2180-10-25.
19. Nataro JP, Kaper JB. 1998. Diarrheagenic *Escherichia coli*. Clin Microbiol Rev 11:142–201.
20. Kaper JB, Nataro JP, Mobley HL. 2004. Pathogenic *Escherichia coli*. Nat Rev Microbiol 2:123–140. https://doi.org/10.1038/nrmicro818.
21. Nisa S, Scanlon KM, Donnenberg MS. 2013. Enteropathogenic *Escherichia coli*, p 75–119. *In* Donnenberg MS (ed), *Escherichia coli* pathotypes and principles of pathogenesis, 2nd ed. Academic Press, San Diego, CA.
22. Kotloff KL, Nataro JP, Blackwelder WC, Nasrin D, Farag TH, Panchalingam S, Wu Y, Sow SO, Sur D, Breiman RF, Faruque AS, Zaidi AK, Saha D, Alonso PL, Tamboura B, Sanogo D, Onwuchekwa U, Manna B, Ramamurthy T, Kanungo S, Ochieng JB, Omore R, Oundo JO, Hossain A, Das SK, Ahmed S, Qureshi S, Quadri F, Adegbola RA, Antonio M, Hossain MJ, Akinsola A, Mandomando I, Nhampossa T, Acacio S, Biswas K, O'Reilly CE, Mintz ED, Berkeley LY, Muhsen K, Sommerfelt H, Robins-Browne RM, Levine MM. 2013. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter study, GEMS): a prospective, case-control study. Lancet 382:209–222. https://doi.org/10.1016/S0140-6736(13)60844-2.
23. McDaniel TK, Jarvis KG, Donnenberg MS, Kaper JB. 1995. A genetic locus of enterocyte effacement conserved among diverse enterobacterial pathogens. Proc Natl Acad Sci U S A 92:1664–1668. https://doi.org/10.1073/pnas.92.5.1664.
24. McDaniel TK, Kaper JB. 1997. A cloned pathogenicity island from enteropathogenic *Escherichia coli* confers the attaching and effacing phenotype on *E. coli* K-12. Mol Microbiol 23:399–407. https://doi.org/10.1046/j.1365-2958.1997.2311591.x.
25. Tarr PI, Gordon CA, Chandler WL. 2005. Shiga-toxin-producing *Escherichia coli* and haemolytic uraemic syndrome. Lancet 365:1073–1086. https://doi.org/10.1016/S0140-6736(05)71144-2.
26. Johnson TJ, Nolan LK. 2009. Pathogenomics of the virulence plasmids of *Escherichia coli*. Microbiol Mol Biol Rev 73:750–774. https://doi.org/10.1128/MMBR.00015-09.
27. Wang P, Xiong Y, Lan R, Ye C, Wang H, Ren J, Jing H, Wang Y, Zhou Z, Cui Z, Zhao H, Chen Y, Jin D, Bai X, Zhao A, Wang Y, Zhang S, Sun H, Li J, Wang T, Wang L, Xu J. 2011. pO157_Sal, a novel conjugative plasmid detected in outbreak isolates of *Escherichia coli* O157:H7. J Clin Microbiol 49:1594–1597. https://doi.org/10.1128/JCM.02530-10.
28. Tobe T, Hayashi T, Han CG, Schoolnik GK, Ohtsubo E, Sasakawa C. 1999. Complete DNA sequence and structural analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid. Infect Immun 67:5455–5462.
29. Brinkley C, Burland V, Keller R, Rose DJ, Boutin AT, Klink SA, Blattner FR, Kaper JB. 2006. Nucleotide sequence analysis of the enteropathogenic *Escherichia coli* adherence factor plasmid pMAR7. Infect Immun 74:5408–5413. https://doi.org/10.1128/IAI.01840-05.
30. Rasko DA, Rosovitz MJ, Myers GS, Mongodin EF, Fricke WF, Gajer P, Crabtree J, Sebaihia M, Thomson NR, Chaudhuri R, Henderson IR, Sperandio V, Ravel J. 2008. The pangenome structure of *Escherichia coli*: comparative genomic analysis of *E. coli* commensal and pathogenic isolates. J Bacteriol 190:6881–6893. https://doi.org/10.1128/JB.00619-08.
31. Crossman LC, Chaudhuri RR, Beatson SA, Wells TJ, Desvaux M, Cunningham AF, Petty NK, Mahon V, Brinkley C, Hobman JL, Savarino SJ, Turner SM, Pallen MJ, Penn CW, Parkhill J, Turner AK, Johnson TJ, Thomson NR, Smith SG, Henderson IR. 2010. A commensal gone bad: complete genome sequence of the prototypical enterotoxigenic *Escherichia coli* strain H10407. J Bacteriol 192:5822–5831. https://doi.org/10.1128/JB.00710-10.
32. Burland V, Shao Y, Perna NT, Plunkett G, Sofia HJ, Blattner FR. 1998. The complete DNA sequence and analysis of the large virulence plasmid of *Escherichia coli* O157:H7. Nucleic Acids Res 26:4196–4204. https://doi.org/10.1093/nar/26.18.4196.
33. Chaudhuri RR, Sebaihia M, Hobman JL, Webber MA, Leyton DL, Goldberg MD, Cunningham AF, Scott-Tucker A, Ferguson PR, Thomas CM, Frankel G, Tang CM, Dudley EG, Roberts IS, Rasko DA, Pallen MJ, Parkhill J, Nataro JP, Thomson NR, Henderson IR. 2010. Complete genome sequence and comparative metabolic profiling of the prototypical enteroaggregative *Escherichia coli* strain 042. PLoS One 5:e8801. https://doi.org/10.1371/journal.pone.0008801.
34. Iguchi A, Thomson NR, Ogura Y, Saunders D, Ooka T, Henderson IR, Harris D, Asadulghani M, Kurokawa K, Dean P, Kenny B, Quail MA, Thurston S, Dougan G, Hayashi T, Parkhill J, Frankel G. 2009. Complete genome sequence and comparative genome analysis of enteropathogenic *Escherichia coli* O127:H6 strain E2348/69. J Bacteriol 191:347–354. https://doi.org/10.1128/JB.01238-08.
35. Hazen TH, Kaper JB, Nataro JP, Rasko DA. 2015. Comparative genomics provides insight into the diversity of the attaching and effacing *Escherichia coli* virulence plasmids. Infect Immun 83:4103–4117. https://doi.org/10.1128/IAI.00769-15.
36. Nwaneshiudu AI, Mucci T, Pickard DJ, Okeke IN. 2007. A second large plasmid encodes conjugative transfer and antimicrobial resistance in O119:H2 and some typical O111 enteropathogenic *Escherichia coli* strains. J Bacteriol 189:6074–6079. https://doi.org/10.1128/JB.00349-07.
37. Riley LW, Junio LN, Libaek LB, Schoolnik GK. 1987. Plasmid-encoded expression of lipopolysaccharide O-antigenic polysaccharide in enteropathogenic *Escherichia coli*. Infect Immun 55:2052–2056.
38. Liebert CA, Hall RM, Summers AO. 1999. Transposon Tn*21*, flagship of the floating genome. Microbiol Mol Biol Rev 63:507–522.
39. Smith LD, Bertrand KP. 1988. Mutations in the Tn*10 tet* repressor that interfere with induction. Location of the tetracycline-binding domain. J Mol Biol 203:949–959.
40. Aldema ML, McMurry LM, Walmsley AR, Levy SB. 1996. Purification of the Tn*10*-specified tetracycline efflux antiporter TetA in a native state as a polyhistidine fusion protein. Mol Microbiol 19:187–195. https://doi.org/10.1046/j.1365-2958.1996.359886.x.
41. Roberts MC. 2005. Update on acquired tetracycline resistance genes. FEMS Microbiol Lett 245:195–203. https://doi.org/10.1016/j.femsle.2005.02.034.
42. Paulsen IT, Littlejohn TG, Radstrom P, Sundstrom L, Skold O, Swedberg G, Skurray RA. 1993. The 3′ conserved segment of integrons contains a gene associated with multidrug resistance to antiseptics and disinfectants. Antimicrob Agents Chemother 37:761–768. https://doi.org/10.1128/AAC.37.4.761.
43. Vedantam G, Guay GG, Austria NE, Doktor SZ, Nichols BP. 1998. Characterization of mutations contributing to sulfathiazole resistance in *Escherichia coli*. Antimicrob Agents Chemother 42:88–93.
44. Sundstrom L, Radstrom P, Swedberg G, Skold O. 1988. Site-specific recombination promotes linkage between trimethoprim- and sulfonamide resistance genes. Sequence characterization of *dhfrV* and *sulI* and a recombination active locus of Tn*21*. Mol Gen Genet 213:191–201.
45. Boyd ES, Barkay T. 2012. The mercury resistance operon: from an origin in a geothermal environment to an efficient detoxification machine. Front Microbiol 3:349. https://doi.org/10.3389/fmicb.2012.00349.
46. Nieto JM, Carmona M, Bolland S, Jubete Y, de la Cruz F, Juarez A. 1991. The *hha* gene modulates haemolysin expression in *Escherichia coli*. Mol Microbiol 5:1285–1293. https://doi.org/10.1111/j.1365-2958.1991.tb01902.x.
47. Sharma VK, Carlson SA, Casey TA. 2005. Hyperadherence of an *hha* mutant of *Escherichia coli* O157:H7 is correlated with enhanced expression of LEE-encoded adherence genes. FEMS Microbiol Lett 243:189–196. https://doi.org/10.1016/j.femsle.2004.12.003.
48. Sharma VK, Bearson BL. 2013. Hha controls *Escherichia coli* O157:H7 biofilm formation by differential regulation of global transcriptional regulators FlhDC and CsgD. Appl Environ Microbiol 79:2384–2396. https://doi.org/10.1128/AEM.02998-12.
49. Sharma VK, Zuerner RL. 2004. Role of *hha* and *ler* in transcriptional regulation of the *esp* operon of enterohemorrhagic *Escherichia coli* O157:H7. J Bacteriol 186:7290–7301. https://doi.org/10.1128/JB.186.21.7290-7301.2004.
50. Paytubi S, Dietrich M, Queiroz MH, Juarez A. 2013. Role of plasmid- and chromosomally encoded Hha proteins in modulation of gene expression in *E. coli* O157:H7. Plasmid 70:52–60. https://doi.org/10.1016/j.plasmid.2013.01.006.
51. Hazen TH, Sahl JW, Fraser CM, Donnenberg MS, Scheutz F, Rasko DA. 2013. Refining the pathovar paradigm via phylogenomics of the attaching and effacing *Escherichia coli*. Proc Natl Acad Sci U S A 110:12810–12815. https://doi.org/10.1073/pnas.1306836110.
52. Villa L, Garcia-Fernandez A, Fortini D, Carattoli A. 2010. Replicon sequence typing of IncF plasmids carrying virulence and resistance determinants. J Antimicrob Chemother 65:2518–2529. https://doi.org/10.1093/jac/dkq347.
53. Carattoli A, Zankari E, Garcia-Fernandez A, Voldby Larsen M, Lund O, Villa L, Moller Aarestrup F, Hasman H. 2014. *In silico* detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. Antimicrob Agents Chemother 58:3895–3903. https://doi.org/10.1128/AAC.02412-14.

54. Njamkepo E, Fawal N, Tran-Dien A, Hawkey J, Strockbine N, Jenkins C, Talukder KA, Bercion R, Kuleshov K, Kolinska R, Russell JE, Kaftyreva L, Accou-Demartin M, Karas A, Vandenberg O, Mather AE, Mason CJ, Page AJ, Ramamurthy T, Bizet C, Gamian A, Carle I, Sow AG, Bouchier C, Wester AL, Lejay-Collin M, Fonkoua MC, Hello SL, Blaser MJ, Jernberg C, Ruckly C, Merens A, Page AL, Aslett M, Roggentin P, Fruth A, Denamur E, Venkatesan M, Bercovier H, Bodhidatta L, Chiou CS, Clermont D, Colonna B, Egorova S, Pazhani GP, Ezernitchi AV, Guigon G, Harris SR, Izumiya H, Korzeniowska-Kowal A, et al. 2016. Global phylogeography and evolutionary history of *Shigella dysenteriae* type 1. Nat Microbiol 1:16027. https://doi.org/10.1038/nmicrobiol.2016.27.
55. Johnson JR, Johnston B, Clabots C, Kuskowski MA, Pendyala S, Debroy C, Nowicki B, Rice J. 2010. *Escherichia coli* sequence type ST131 as an emerging fluoroquinolone-resistant uropathogen among renal transplant recipients. Antimicrob Agents Chemother 54:546–550. https://doi.org/10.1128/AAC.01089-09.
56. Carattoli A. 2009. Resistance plasmid families in *Enterobacteriaceae*. Antimicrob Agents Chemother 53:2227–2238. https://doi.org/10.1128/AAC.01707-08.
57. Novais A, Viana D, Baquero F, Martinez-Botas J, Canton R, Coque TM. 2012. Contribution of IncFII and broad-host IncA/C and IncN plasmids to the local expansion and diversification of phylogroup B2 *Escherichia coli* ST131 clones carrying $bla_{CTX\text{-}M\text{-}15}$ and *qnrS1* genes. Antimicrob Agents Chemother 56:2763–2766. https://doi.org/10.1128/AAC.06001-11.
58. Rao AN, Barlow M, Clark LA, Boring JR, III, Tenover FC, McGowan JE, Jr. 2006. Class 1 integrons in resistant *Escherichia coli* and *Klebsiella* spp., US hospitals. Emerg Infect Dis 12:1011–1014. https://doi.org/10.3201/eid1206.051596.
59. Venturini C, Hassan KA, Roy Chowdhury P, Paulsen IT, Walker MJ, Djordjevic SP. 2013. Sequences of two related multiple antibiotic resistance virulence plasmids sharing a unique IS*26*-related molecular signature isolated from different *Escherichia coli* pathotypes from different hosts. PLoS One 8:e78862. https://doi.org/10.1371/journal.pone.0078862.
60. Sambrook JF, Fritsch EF, Maniatis T. 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
61. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
62. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 15 March 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27:722–736. https://doi.org/10.1101/gr.215087.116.
63. Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 33:623–630. https://doi.org/10.1038/nbt.3238.
64. Delcher AL, Salzberg SL, Phillippy AM. 2003. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinformatics Chapter 10:Unit 10.13.
65. Crabtree J, Agrawal S, Mahurkar A, Myers GS, Rasko DA, White O. 2014. Circleator: flexible circular visualization of genome-associated data with BioPerl and SVG. Bioinformatics 30:3125–3127. https://doi.org/10.1093/bioinformatics/btu505.
66. Sommer DD, Delcher AL, Salzberg SL, Pop M. 2007. Minimus: a fast, lightweight genome assembler. BMC Bioinformatics 8:64. https://doi.org/10.1186/1471-2105-8-64.
67. Wirth T, Falush D, Lan R, Colles F, Mensa P, Wieler LH, Karch H, Reeves PR, Maiden MC, Ochman H, Achtman M. 2006. Sex and virulence in *Escherichia coli*: an evolutionary perspective. Mol Microbiol 60:1136–1151. https://doi.org/10.1111/j.1365-2958.2006.05172.x.
68. Jolley KA, Maiden MC. 2010. BIGSdb: Scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics 11:595. https://doi.org/10.1186/1471-2105-11-595.
69. Joensen KG, Tetzschner AM, Iguchi A, Aarestrup FM, Scheutz F. 2015. Rapid and easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J Clin Microbiol 53:2410–2426. https://doi.org/10.1128/JCM.00008-15.
70. Sahl JW, Beckstrom-Sternberg SM, Babic-Sternberg JS, Gillece JD, Hepp CM, Auerbach RK, Tembe W, Wagner DM, Keim PS, Pearson T. 2015. The *in silico* genotyper (ISG): an open-source pipeline to rapidly identify and annotate nucleotide variants for comparative genomics applications. bioRxiv https://doi.org/10.1101/015578.
71. Stamatakis A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. https://doi.org/10.1093/bioinformatics/btl446.
72. Letunic I, Bork P. 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44:W242–W245. https://doi.org/10.1093/nar/gkw290.
73. Galens K, Orvis J, Daugherty S, Creasy HH, Angiuoli S, White O, Wortman J, Mahurkar A, Giglio MG. 2011. The IGS standard operating procedure for automated prokaryotic annotation. Stand Genomic Sci 4:244–251. https://doi.org/10.4056/sigs.1223234.
74. Sahl JW, Caporaso JG, Rasko DA, Keim P. 2014. The large-scale blast score ratio (LS-BSR) pipeline: a method to rapidly compare genetic content between bacterial genomes. PeerJ 2:e332. https://doi.org/10.7717/peerj.332.
75. Hazen TH, Donnenberg MS, Panchalingam S, Antonio M, Hossain A, Mandomando I, Ochieng JB, Ramamurthy T, Tamboura B, Qureshi S, Quadri F, Zaidi A, Kotloff KL, Levine MM, Barry EM, Kaper JB, Rasko DA, Nataro JP. 2016. Genomic diversity of EPEC associated with clinical presentations of differing severity. Nature Microbiol 1:15014. https://doi.org/10.1038/nmicrobiol.2015.14.
76. Gertz EM, Yu YK, Agarwala R, Schaffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. BMC Biol 4:41. https://doi.org/10.1186/1741-7007-4-41.
77. Hazen TH, Leonard SR, Lampel KA, Lacher DW, Maurelli AT, Rasko DA. 2016. Investigating the relatedness of enteroinvasive *Escherichia coli* to other *E. coli* and *Shigella* isolates by using comparative genomics. Infect Immun 84:2362–2371. https://doi.org/10.1128/IAI.00350-16.
78. Nataro JP, Steiner T, Guerrant RL. 1998. Enteroaggregative *Escherichia coli*. Emerg Infect Dis 4:251–261. https://doi.org/10.3201/eid0402.980212.
79. Kumar S, Stecher G, Tamura K. 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.
80. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. 2009. Circos: an information aesthetic for comparative genomics. Genome Res 19:1639–1645. https://doi.org/10.1101/gr.092759.109.