F Plasmids Are the Major Carriers of Antibiotic Resistance Genes in Human-Associated Commensal *Escherichia coli*

Craig Stephens,[a,b] Tyler Arismendi,[a] Megan Wright,[a] Austin Hartman,[a] Andres Gonzalez,[a] Matthew Gill,[a] Mark Pandori,[c] David Hess[a]

aBiology Department, Santa Clara University, Santa Clara, California, USA
bPublic Health Program, Santa Clara University, Santa Clara, California, USA
cAlameda County Public Health Laboratory, San Leandro, California, USA

ABSTRACT The evolution and propagation of antibiotic resistance by bacterial pathogens are significant threats to global public health. Contemporary DNA sequencing tools were applied here to gain insight into carriage of antibiotic resistance genes in *Escherichia coli*, a ubiquitous commensal bacterium in the gut microbiome in humans and many animals, and a common pathogen. Draft genome sequences generated for a collection of 101 *E. coli* strains isolated from healthy undergraduate students showed that horizontally acquired antibiotic resistance genes accounted for most resistance phenotypes, the primary exception being resistance to quinolones due to chromosomal mutations. A subset of 29 diverse isolates carrying acquired resistance genes and 21 control isolates lacking such genes were further subjected to long-read DNA sequencing to enable complete or nearly complete genome assembly. Acquired resistance genes primarily resided on F plasmids (101/153 [67%]), with smaller numbers on chromosomes (30/153 [20%]), IncI complex plasmids (15/153 [10%]), and small mobilizable plasmids (5/153 [3%]). Nearly all resistance genes were found in the context of known transposable elements. Very few structurally conserved plasmids with antibiotic resistance genes were identified, with the exception of an ~90-kb F plasmid in sequence type 1193 (ST1193) isolates that appears to serve as a platform for resistance genes and may have virulence-related functions as well. Carriage of antibiotic resistance genes on transposable elements and mobile plasmids in commensal *E. coli* renders the resistome highly dynamic.

IMPORTANCE Rising antibiotic resistance in human-associated bacterial pathogens is a serious threat to our ability to treat many infectious diseases. It is critical to understand how acquired resistance genes move in and through bacteria associated with humans, particularly for species such as *Escherichia coli* that are very common in the human gut but can also be dangerous pathogens. This work combined two distinct DNA sequencing approaches to allow us to explore the genomes of *E. coli* from college students to show that the antibiotic resistance genes these bacteria have acquired are usually carried on a specific type of plasmid that is naturally transferrable to other *E. coli*, and likely to other related bacteria.

KEYWORDS *Escherichia coli*, F plasmid, antibiotics

The resistance of pathogenic bacteria to antibiotics is an ongoing threat to global public health (https://www.who.int/antimicrobial-resistance/global-action-plan/en/). The U.S. Centers for Disease Control and Prevention has designated certain antibiotic-resistant *Enterobacteriaceae* as a major public health hazard (https://www.cdc.gov/drugresistance/biggest_threats.html). The most well-known member of this family, *Escherichia coli*, is ubiquitous as an intestinal commensal in humans, but it can act as a diarrheagenic gastrointestinal tract pathogen (1) or as an extraintestinal pathogen

Address correspondence to Craig Stephens, cstephens@scu.edu.

College students studying E. coli from college students have found that antibiotic resistance is very common, and the resistance genes are usually carried on plasmids that can easily move between E. coli and related bacteria.

causing urinary tract infections (2) and sepsis (3). Common *E. coli* lineages causing either intestinal or extraintestinal disease are increasingly found to be resistant to multiple drugs (4–6). Antibiotic resistance in *E. coli* can arise by mutations in diverse targets or by acquisition of preexisting genes whose products target antibiotics for alteration or efflux (7, 8). Mobile resistance genes have the greatest potential for spread of antimicrobial resistance in the microbiome. The goal of the study presented here was to examine genes underlying antibiotic resistance phenotypes in *E. coli* by applying genome analysis tools capable of unambiguously assigning the responsible genes to a chromosome or plasmid. Further, we sought to identify the local context of resistance genes to assess their potential for mobility within the genome.

Although genomic analysis of *E. coli* has largely focused on isolates from pathogenic contexts, deeper analysis of the commensal *E. coli* population from which such isolates likely emerge will provide new insights into the genetic reservoir that they are drawing from (9–11). Conjugal plasmids are key vectors for disseminating this reservoir of genetic information (12). In commensal *E. coli*, F plasmids are the most common conjugal plasmids (13), and they were historically the first to be associated with transmissible antibiotic resistance ("R factors") (14). F plasmids have been prominent in the evolution of medically important lineages such as sequence type 131 (ST131) (15). However, non-F plasmids have also been implicated in the evolutionary dynamics of antibiotic resistance in *Enterobacteriaceae*, such as in recent work on the *mcr-1* gene (encoding colicin resistance) demonstrating that this gene is most often associated with X plasmids (16). Determining the structures of large bacterial plasmids is a significant challenge for DNA sequencing based on short read lengths (17), due to the high frequency of repetitive mobile elements typically residing on them. As a consequence, large plasmids have generally not been carefully analyzed outside of major pathogenic lineages. The advent of low-cost, long-read length sequencing methods is now lowering barriers to such analysis (18). We employed a combination of short-read and nanopore-based long-read sequencing methods to generate complete genome sequences that include all plasmids in complete form, allowing definitive assessment of the genomic context of resistance genes.

RESULTS

Isolation and characterization of commensal *E. coli*. A collection of 101 commensal *E. coli* isolates, obtained from healthy college students between 2014 and 2019, were phenotypically characterized for antibiotic resistance. The majority (56/101 [55%]) of the commensal isolates analyzed were phenotypically resistant to at least one of the following classes of antibiotics: β-lactams (36%), sulfonamides (35%), aminoglycosides (34%), trimethoprim (27%), tetracyclines (27%), quinolones (25%), macrolides (17%), or phenicols (3%). Over one-third of the isolates (37%) were multidrug resistant (MDR) (defined as resistant to three or more classes of antibiotics). These 101 isolates were subjected to short-read DNA sequencing to obtain draft-level genome assemblies adequate for resistance gene identification. After alleles were grouped together, 22 distinct acquired resistance genes were identified (Table 1), which accounted for over 85% of observed antibiotic resistance phenotypes. The primary exception was for quinolone resistance (25% of isolates), in which case known mutations in the chromosomal *gyrA* and *parC* genes (19, 20) were present in 23 out of 25 resistant isolates.

Commensal *E. coli* isolates were assessed for phylogenetic diversity by multilocus sequence typing (MLST) inferred from the draft genome assemblies. Among the 59 MLST types identified (data in Tables S1 and S2 in the supplemental material), ST95 (12 isolates), ST69 (8 isolates), and ST10 (7 isolates) were the most abundant. Representatives of all major *E. coli* phylogroups were present, with B2 constituting the largest set. Isolates from phylogroup D (primarily ST69 and ST38) were notable for a high frequency of multidrug resistance (13/15 isolates [87%]), significantly higher than that of the overall collection (37%, chi-square test, $P < 0.01$).

Acquired antibiotic resistance genes in bacteria are often carried on plasmids, so the presence of known replicons was assessed using PlasmidFinder (21). Based on these

TABLE 1 Acquired resistance genes identified in commensal *E. coli* isolates

Drug and resistance gene(s)	No. (%) identified in all isolates (*n* = 101)
β-Lactams	
bla_{TEM}	31 (31)[a]
bla_{CTX-M}	5 (5)[b]
bla_{SHV-1}	1 (1)
bla_{CMY}	2 (2)[c]
Aminoglycosides	
strA, *strB*	27, 27 (27)
aadA (3 alleles)	19 (19)[d]
aadB	1 (1)
aac(3)-IId	8 (8)
aph(3′)	1 (1)
sat	2 (2)
Sulfonamides	
sul1	18 (1)
sul2	29 (29)
Trimethoprim	
dfrA (5 alleles)	27 (27)[e]
Tetracyclines	
tetA	18 (18)
tetB	6 (6)
tetD	4 (4)
Macrolides	
mphA	17 (17)
mphB	1 (1)
Phenicols	
cmlA	2 (2)
cat	2 (2)
Quinolones	
qnrS	1 (1)

[a]bla_{TEM-1B} in 29 isolates, bla_{TEM-1C} in 1 isolate, and bla_{TEM-34} in 1 isolate.
[b]$bla_{CTX-M-14}$ in 3 isolates and $bla_{CTX-M-27}$ in 2 isolates.
[c]bla_{CMY-2} in 1 isolate and $bla_{CMY-M-14}$ in 1 isolate.
[d]*aadA1* in 5 isolates, *aadA2* in 3 isolates, and *aadA5* in 11 isolates.
[e]*dfrA1* in 4 isolates, *dfrA5* in 5 isolates, *dfrA7* in 2 isolates, *dfrA8* in 1 isolate, *dfrA12* in 3 isolates, and *dfrA17* in 12 isolates.

replicons, 79% of isolates were predicted to contain at least one large conjugal plasmid, with FIB (66% of isolates) and FII (62%) replicons being most frequent, followed by I-complex replicons (B/O, K, Z, and I1) in 16% of isolates (data in Tables S1 and S2). No other replicons were found in more than one of the draft genomes. The absence of conjugal plasmid replicons in an isolate was associated with pan-susceptibility (19/45 of pan-susceptible isolates lacked identifiable plasmid replicons versus 2/56 of antibiotic-resistant isolates, chi-square test, $P < 0.001$).

In assemblies based on short-read data, contigs containing antibiotic resistance genes, plasmid replicons, or genes encoding components of conjugal machinery were typically short (<20 kb) and linear (17). Limitations to assembly of short-read data were overcome by the addition of long-read sequencing, and integrating both types of read with hybrid assemblers (18, 22, 23). Using either Unicycler or Flye, assembly of a complete chromosome was achieved for 47 isolates, and for another 3 isolates, the chromosome was present in two to four large contigs. These genomes were therefore considered to be fully or nearly fully assembled and sufficient for assignment of antibiotic resistance genes to chromosomes or plasmids. The 50 genomes comprised representatives from six phylogroups (phylogroup A, 4 isolates; phylogroup B1, 6 isolates; phylogroup B2, 25 isolates; phylogroup D, 10 isolates; phylogroup E, 1 isolate;

phylogroup F, 4 isolates) and 33 MLST groups and included 29 isolates carrying acquired resistance genes (Table 2), and 21 lacking acquired resistance genes (Table S1).

Local context of antibiotic resistance genes. To better understand how acquired antibiotic resistance genes are mobilized in commensal *E. coli*, these genes in the completely assembled genomes were examined for surrounding mobile genetic elements such as insertion sequences (ISs), transposons (Tns), and integrons (Table 2) (24, 25). bla_{TEM-1} was always found in a Tn*2* transposable element, though a minority (6/18) resided in a full-length Tn*2* (~5 kb). In the majority of the partial Tn*2* elements, much of the Tn*2* sequence upstream of bla_{TEM-1} was replaced by IS*26*, reducing it to 1.2 to 1.6 kb. In these cases, a second IS element (*1A* or *CR2*) was located on the other flank of the partial Tn*2*. *tetA* was always found on Tn*1721*, and *tetB* on some form of Tn*10*. *strA* and *strB* were always located on Tn*5393*, usually with *sul2* immediately adjacent, followed by IS*26*, suggesting that this entire set moves as a unit. As with Tn*2*, only a few isolates carried complete versions of Tn*1721* or Tn*5393*, with the sizes of the residual elements varying. In four isolates, *sul2* was located in the context of IS*5075*/IS*CR2*, rather than adjacent to Tn*5393*. *mphA* was always found as part of the mobile three-gene cluster between IS*26* and IS*6100* fragments.

Eleven (38%) of the 29 isolates in Table 2 contained intact class 1 integrons (26) carrying one to three resistance genes, and three contained partial class 1 integrons. Of the 14 class 1 integrons observed, only two were located on chromosomes (SCU-164 and SCU-397). Twenty-five intact resistance genes were found as cassettes in class 1 integrons (intact or partial). Figure 1 shows the most common cassette configuration, with *dfrA17* (trimethoprim resistance) and *aadA5* (aminoglycoside resistance). Alleles of *dfrA* and *aadA* were found as cassettes; the only other intact cassette was *cmlA* (chloramphenicol resistance). *sul1* was present adjacent to the cassette regions of 11 intact class 1 integrons, but it was absent in three partial integrons and present in one partial integron lacking the cassette region. One isolate (SCU-105) contained an intact Tn*7*-associated class 2 integron on the chromosome with *dfrA1*, *satA1*, and *aadA1* cassettes. A second isolate (SCU-175) contained a partial class 2 integron with only fragments of Tn*7* in the adjacent sequence. *aadA* and *dfrA* genes were found only in the context of class 1 and 2 integrons. In total, 42 acquired resistance genes (27% of the total) were associated with class 1 or 2 integrons.

Plasmids and resistance genes. In the fully assembled genomes, plasmids partitioned into two general pools, designated here as "small" (1 to 13 kb; $n = 86$; mean size, 4.6 kb) and "large" plasmids (26 to 190 kb; $n = 63$; mean size, 103 kb). Figure 2 shows the size distribution of plasmids from the subset of antibiotic-resistant isolates (gray bars); the size distribution of plasmids from antibiotic-susceptible isolates was similar. The majority of the 63 large plasmids were associated with F replicons (49/63 [77%]), and in most cases, multiple subtypes of F replicons were found on the same plasmids. Ten plasmids had IncI complex replicons (Z, B/O, K, or I1). F- and I-complex replicons are typically associated with plasmids capable of conjugation, and the genes encoding components of the conjugal machinery typically take up 35 to 40 kb for both types of plasmids, although many of the F plasmids were missing at least 20% of the conjugation-associated genes (data not shown). None of the assembled plasmids contained replicons of multiple types, and only two putative plasmids had no identifiable replicons using PlasmidFinder (21). Plasmids encoding one or more antibiotic resistance genes were primarily from the large plasmid pool (Fig. 2), with two exceptions discussed more below. The fully assembled genomes from isolates containing acquired resistance genes ($n = 29$ isolates) had significantly more large plasmids per isolate (1.6 ± 0.6) than the genomes from isolates lacking acquired resistance genes ($n = 21$ isolates, 0.95 ± 0.79 large plasmids/isolate) [independent *t* test, $t(48) = 2.78$, $P = 0.0077$].

Nearly 80% (123/154 [78%]) of acquired antibiotic resistance genes in the fully assembled commensal *E. coli* genomes resided on plasmids (Table 2). Most plasmid-

TABLE 2 Plasmids and antibiotic resistance determinants in fully assembled commensal *E. coli* genomes

Phylogroup	MLST	Isolate	Resistance phenotype(s)[a]	Genome component (GenBank accession no.)	Size	Resistance genes (associated mobile elements)[b]	Plasmid replicon(s)[c]
A	10	SCU-103	AMP, CEF, AZM, STR, SXT, TET	pSCU-103-1 (CP054458)	139 kb	*aadA5-dfrA17* (class 1 integron cassette) plus *sul1* adjacent; *bla*$_{CTX-M-27}$ (IS*903C*, IS*Ecp1* flank); *strA-strB* (Tn*5393'*) plus *sul2* (Tn*5393'* flank); *mphA* (IS*6100*, IS*26* flank); *tetA* (Tn*1721'*)	F1A, F1B, FII
		SCU-118	AMP, STR, SXT, TET	pSCU-118-1 (CP051717)	85 kb	*aadA1-dfrA1* (class 1 integron cassette) plus *sul1* adjacent; *bla*$_{TEM-1B}$ (Tn*2'*); *mphB* (flanked by Tn*402*); *strA-strB* (Tn*5393'*) plus *sul2* (RSF1010-like); *tetA* (Tn*1721'*)	F1B'
B1	3695	SCU-106	STR, TET	pSCU-106-2 (CP053236)	112 kb	*strA-strB* (Tn*5393'*); *tetA* (Tn*1721* within Tn*5393'*)	F1B', FIC(II)
		SCU-308	AMP, STR, SXT	pSCU-308-1 (CP053282)	152 kb	*bla*$_{TEM-1B}$ (Tn*2'*) plus *strA-strB* (Tn*5393'*) and *sul2*; *dfrA5* (class 1 integron cassette fragment)	F1B', FII
B2	14	SCU-387	AMP, AZM	pSCU-387-2 (CP051690)	39 kb[d]	*bla*$_{TEM-1B}$ (Tn*1,2,3*-like'); *mphA*	FII
	73	SCU-112	AMP, CEF (int)	pSCU-112-1 (CP051726)	104 kb	*aadA1'* (class 1 integron cassette fragment); *bla*$_{SHV-1}$ (IS*26* flanking both sides)	F1B', FII, Col156
	91	SCU-121	TET	pSCU-121-1 (CP054329)	68 kb	*tetA* (Tn*1721'*)	FII
	95	SCU-108	AMP	pSCU-108-2 (CP051737)	72 kb	*bla*$_{TEM-1B}$ (Tn*2*)	FII
		SCU-123	AMP, STR, SUL	pSCU-123-2 (CP051713)	95 kb	*bla*$_{TEM-1C}$ (Tn*2c*); *strA-strB'* (Tn*5393'*) plus *sul2* (Tn*5393'* flank)	B/O/K/Z (B/O)
		SCU-306	AZM, SXT	pSCU-306-1 (CP053232)	129 kb[d]	*aadA2-dfrA12* (class 1 integron cassette) plus *sul1* adjacent; *mphA* (IS*6100*, IS*26*, 2 copies)	F1B, FII, Col156
	131	SCU-182	AMP, GEN	pSCU-182-1 (CP054376.1, CP054375.1, CP054374.1, CP054373.1)	168 kb[d]	*aac(3)-IId* (IS*26*, IS*10'*); *bla*$_{TEM-1B}$ (Tn*2'*)	F1A, F1B, FII, Col156
		SCU-481	AMP, AMC, AZM (int), CHL, SXT	pSCU-481-1 (JABLYB000000000.1)	144 kb	*aadB-aacC'-cmlA6* (class 1 integron cassette); *aadA5-dfrA17* (class 1 integron cassette) plus *sul1* adjacent; *bla*$_{TEM-34}$ (Tn*2*); *mphA* (IS*26*, IS*6100* flank)	F1A, F1B, FII
	144	SCU-125	STR, SXT	pSCU-125-2 (CP051702)	93 kb	*dfrA5* (class 1 integron cassette) plus *sul1* adjacent; *strA-strB* (Tn*5393'*, IS*CR2'*) plus *sul2* (Tn*5393'* flank)	B/O/K/Z (Z)
	357	SCU-124	AMP	pSCU-124-2 (CP051708)	73 kb	*bla*$_{TEM-1B}$ (Tn*2*)	FII
	1193	SCU-147	AMP, AZM, GEN, STR, SXT, TET	pSCU-147-1 (CP054326)	105 kb	*aac(3)-IId* (IS*10'*, IS*26* flank); *aadA5-dfrA17* (class 1 integron cassette fragment); *bla*$_{TEM-1B}$ (Tn*1,2,3*-like'); *mphA* (IS*6100*, IS*26* flank); *strA-strB* (Tn*5393'*) plus *sul2* (Tn*5393'* flank); *tetA* (Tn*1721'*)	F1A, F1B, Col156
		SCU-204	STR, SUL	pSCU-204-1 (CP054414.1)	88 kb	*strA-strB* (Tn*5393'*) plus *sul2* (RSF1010-like)	F1A, F1B, Col156
		SCU-390	AMP, STR, SUL	pSCU-390-1 (CP054321)	91 kb	*bla*$_{TEM-1B}$ (Tn*2'*); *strA-strB* (Tn*5393'*) plus *sul2* (RSF1010-like)	F1A, F1B', Col156
	2279	SCU-479	AMP, AMC, CEF, CHL, STR, SUL, TET	Chromosome (CP054317)	5.2 Mb	*bla*$_{CTX-M-14}$ (IS*Ecp1*); *bla*$_{CMY-121}$ (IS*Ecp1*); *bla*$_{TEM-1B}$ (IS*Ecp1*); *strA-strB* (Tn*5393'*) plus *sul2* (Tn*5393'* flank); *tetA* (Tn*1721'*)	
D	38	SCU-164	SXT, TET	Chromosome (CP054343)	5.4 Mb[d]	*dfrA7* (class 1 integron cassette) plus *sul1* adjacent; *sul2* (IS*5075*, IS*CR2*); *tetD* (flanked by IS*26* and Tn*2'*)	
		SCU-397	AMP, CEF, CHL, STR, SXT, TET	Chromosome (CP054828.1)	5.3 Mb[d]	*bla*$_{CTX-M-14}$ (2 copies, each between IS*ECP1* and IS*903B'*); *bla*$_{TEM-1B}$ (Tn*2'*); *dfrA7* (class 1 integron cassette) plus *sul1* adjacent; *strA-strB* (Tn*5393'*) plus *sul2* (Tn*5393'* flank); *tetD* (IS*26*, Tn*2'* flank); *catA1*	
		SCU-486	AMP, CEF, AZM, GEN, STR, SXT, TET	Chromosome (CP051749)	5.2 Mb	*bla*$_{CTX-M-14}$ (IS*903B'*, IS*Ecp1* flank); *strA-strB* (Tn*5393'*); *sul2* (IS*5075*, IS*CR2* flank); *tetD* (IS*26*, Tn*2'* flank); *bla*$_{TEM-1B}$ (Tn*2'*); *catA1* (IS*26* flank)	

(Continued on next page)

TABLE 2 (Continued)

Phylogroup	MLST	Isolate	Resistance phenotype(s)[a]	Genome component (GenBank accession no.)	Size	Resistance genes (associated mobile elements)[b]	Plasmid replicon(s)[c]
				pSCU-486-1 (CP051750)	84 kb	*aac(3)-IId* (IS*26*, IS*Kpn11-like'* flank); *bla*$_{CTX-M-14}$ (IS*903B'*, IS*Ecp1* flank); *dfrA5* (class 1 integron cassette) plus *sul1* adjacent; *mphA* (IS*6100*, IS*26* flank)	F1B', FII
	69	SCU-313	AMP, AZM, GEN, STR, SUL, TET	pSCU-313-1 (CP051695)	105 kb	*aac(3)-IId* (IS*26*, IS*10'*); *aadA5-dfrA17* (class 1 integron cassette) plus *sul1* adjacent; *bla*$_{TEM-1B}$ (Tn*1,2,3*-like'); *mphA* (IS*26*, IS*6100*); *strA-strB* (Tn*5393'*) plus sul2 (Tn*5393'* flank, IS*26*); *tetA* (Tn*1721'*)	F1A, F1B'
		SCU-482	AMP, AZM, STR, SXT	pSCU-482-1 (CP053248)	145 kb	*aadA5-dfrA17* (class 1 integron cassette) plus *sul1* adjacent; *bla*$_{TEM-1B}$ (Tn*2'*); *mphA* (IS*26*, IS*6100'* flank); *strA-strB* (Tn*5393'*) plus *sul2* (RSF1010-like)	F1B', FII, Col156
	106	SCU-318	AMP, STR, SUL, TET	pSCU-318-1 (CP051693)	105 kb	*bla*$_{TEM-1B}$ (Tn*2'*); *strA-strB* (Tn*5393'*) plus *sul2* (Tn*5393'* flank); *tetB* (Tn*10'*)	F1B, FII
	394	SCU-105	AMP, CEF, AZM, STR, SXT	Chromosome (CP051738)	5.2 Mb	*dfrA1-sat2-aadA1* (class 2 integron cassettes in Tn*7*)	
				pSCU-105-1 (CP051739)	173 kb[d]	*strA-strB* (Tn*5393'*); *sul2* (IC*R2'*, IS*5075'* flank)	F1B, FII
				pSCU-105-2 (CP051740)	9.7 kb	*bla*$_{TEM-1B}$ (Tn*2'*); *mphA* (IS*6100'*, IS*26'* flank)	
	963	SCU-109	AMP, AMC, CEF, GEN	Chromosome (CP051733)	5.0 Mb	*bla*$_{CMY-2}$	
				pSCU-109-1 (CP051734)	110 kb	*aac(3)-IId* (IS*26*, IS*10'* flank); *bla*$_{TEM-1B}$ (Tn*2'*)	F1B', FII, Col156
F	62	SCU-175	AMP, AZM, STR, SXT, TET	pSCU-175-1 (CP054380.1)	124 kb	*aadA1-dfrA1-sat2* (class 2 integron in Tn*7*); *mphA* (IS*26*, IS*6100* flank); *sul2* (IS*CR2'* flank); *tetB* (Tn*10'*)	B/O/K/Z (Z)
				pSCU-175-2 (CP054381.1)	72 kb	*bla*$_{TEM-1B}$ (Tn*2*)	FII
	379	SCU-172	AMP	pSCU-172-3 (CP054356)	76 kb	*bla*$_{TEM-1B}$ (Tn*2*)	FII
	648	SCU-120	AMP (int), CEF, AZM, STR, SXT, TET	pSCU-120-1 (CP054336)	143 kb	*aadA5-dfrA17* (class 1 integron cassette) plus *sul1* adjacent; *mphA* (IS*6100*, IS*26* flank); *tetA* (Tn*1721'*)	F1A, F1B', FII
				pSCU-120-3 (CP054338)	6.2 kb	*strA-strB* (Tn*5393'*) plus *sul2* (Tn*5393'* flank)	

[a]Abbreviations: AMP, ampicillin; AMC, amoxicillin-clavulanic acid; AZM, azithromycin; CEF, cephalothin; CHL, chloramphenicol; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; NOR, norfloxacin; STR, streptomycin; SUL, sulfamethoxazole alone; SXT, sulfamethoxazole-trimethoprim; TET, tetracycline; TMP, trimethoprim alone. "int" in parentheses indicates that the size of the zone of inhibition for the antibiotic met the manufacturer's criteria for "intermediate" resistance. Note that quinolone resistance is reported here only when due to an acquired gene; resistance due to chromosomal mutations is reported in Table S1 in the supplemental material.

[b]Identification of antibiotic resistance genes was done with ResFinder (48). A prime symbol indicates that the identified antibiotic resistance gene was incomplete (between 60 and 90% present). Mobile elements were identified using the Galileo Antimicrobial Resistance (GAMR) software (24). A prime symbol indicates that the transposable element was smaller than the published full version of the element. "flank" indicates that the resistance gene was not within the identified mobile element, but within 1 kb adjacent to it.

[c]Identification of plasmid replicons was done with PlasmidFinder (21). A prime symbol indicates that the identified replicon sequence was incomplete (between 60 and 90% present).

[d]Assembly was noncircular, suggesting gap of unknown size between ends.

borne resistance genes (103/123 [84%]) were on molecules containing at least one F replicon. Fifteen of the remainder were on three IncI complex plasmids in isolates SCU-123 (B/O replicon), SCU-125 (Z replicon), and SCU-175 (Z replicon). Finally, five were located on two small plasmids (pSCU-105-2 and pSCU-120-3). Only six isolates carried acquired resistance genes on their chromosome (Table 2), totaling 30 acquired resistance genes. The validity of assignment of resistance genes to plasmids or chromosomes was confirmed by electroporation of purified genomic DNA preparations into laboratory *E. coli*; predicted chromosomal loci never generated antibiotic-resistant electroporants (data not shown).
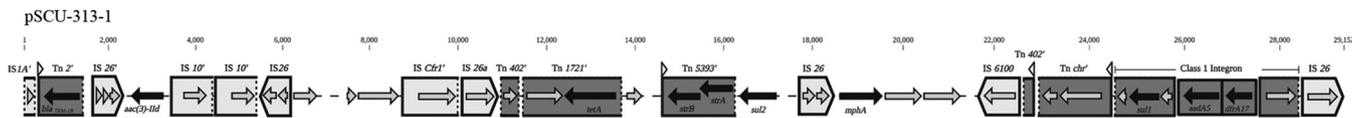
pSCU-313-1

FIG 1 Conserved cluster of antibiotic resistance genes, transposable elements, and a class I integron in pSCU-313-1. Transposable elements and resistance genes were identified using ResFinder (48) and GAMR (24) and visualized using BioRender. IS elements are indicated by light gray boxes, with their name above the box; transposons and the class 1 integron are indicated by dark gray boxes, with their name above. Conserved inverted repeats known to be associated with transposable element boundaries are indicated by triangles above the boundaries. Dashed lines indicate breakpoints (defined by sequence alignment) of interrupted elements; partial elements are indicated by a prime symbol following their name. Antibiotic resistance genes are indicated by black arrows, with their name underneath.

Small plasmids rarely contained antibiotic resistance genes, but there were two exceptions. pSCU-105-2 is a 9.7-kb plasmid containing bla_{TEM-1} and *mphA* (macrolide resistance), and pSCU-120-3 is a 6.2-kb plasmid containing *strA-strB* (streptomycin resistance) and *sul2* (sulfonamide resistance) (Fig. 3). Based on read coverage data, pSCU-105-2 was present at roughly 50 copies per chromosome equivalent, and pSCU-120-3 was present at ~8 copies per chromosome. These plasmids were readily transferred by electroporation into laboratory *E. coli* strains and stably maintained (data not shown). pSCU-105-2 is a ColE1-type plasmid in which the colicin E1 gene and associated immunity function have been replaced by a 6-kb mobile element comprised of Tn2 containing bla_{TEM-1} and a macrolide resistance locus (*mphA-mrx-mphR*). The pSCU-120-3 replication functions appear to reside in a 3-kb backbone found in several other plasmids (pSCU-103-4, pSCU-105-4, and pSCU-175-5), to which a 3-kb composite element is attached containing a partial Tn5393 (*strA-strB*) followed by *sul2*.

Multidrug-resistant isolates often exhibited regions in which multiple genetic elements (ISs, Tns, and/or integrons) aggregated into larger, potentially mobile units (27, 28). The largest conserved resistance locus found in the 29 fully assembled isolates was a 19-kb segment shared by plasmids in isolates SCU-103, SCU-147, and SCU-313 (Fig. 1). These plasmids (pSCU-103-1, pSCU-147-1, and pSCU-313-1) are otherwise not closely related, sharing only 50 to 60% of their contents, nor are the host *E. coli* closely related, coming from distinct MLSTs and phylogroups.

Only one complete plasmid with antibiotic resistance genes was found to be highly conserved in multiple isolates in this collection. Plasmids pSCU-204-1 (88 kb) and pSCU-390-1 (90.7 kb), both from ST1193 (B2) isolates, are 99.9% identical in nucleotide
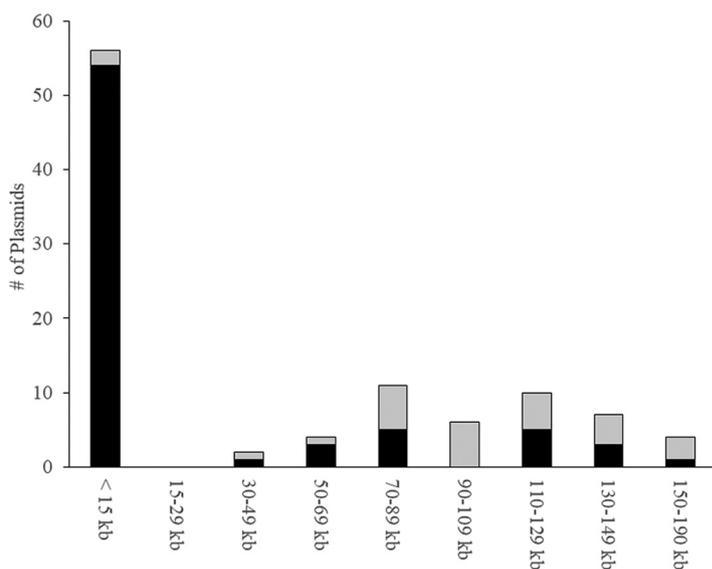

FIG 2 Acquired antibiotic resistance genes are primarily on large (>70-kb) plasmids in commensal *E. coli*. The *y* axis indicates the number of plasmids identified. Total (black and gray) bars indicate all plasmids in each size range from the genomes of antibiotic-resistant isolates (Table 2); black bars indicate only the plasmids that actually contained acquired antibiotic resistance genes.
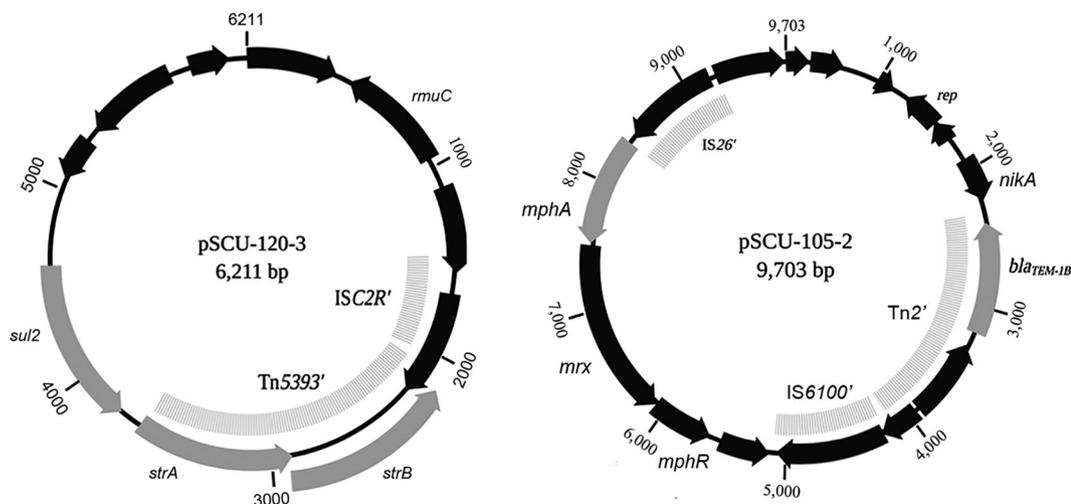
FIG 3 Rare small plasmids from commensal *E. coli* containing antibiotic resistance genes. Transposable elements and resistance genes were identified using ResFinder (48) and GAMR (24) and visualized using BioRender.

sequence over the 88-kb length of pSCU-204-1 (Fig. 4). These plasmids are in turn closely related to the 90-kb plasmid pC32_1 from *Shigella flexneri* strain C32 and an 88-kb pNMEC-075A plasmid from *E. coli* ST1193 strain MCJCHV-1. Differences between all these plasmids are focused in a 20-kb variable region containing antibiotic resistance genes, shown in Fig. 4B.

DISCUSSION

Only a small fraction of the thousands of *E. coli* genomes in the NCBI Genomes database are completely assembled. The work presented here was made possible by the development of affordable lab-scale long-read DNA sequencing (17, 18). This is essential for exploring the architecture of bacterial genomes, since chromosomes and plasmids are generally littered with repetitive transposable elements that preclude unambiguous assembly from short-read sequencing data. The goal of this study was to conclusively determine how antibiotic resistance genes are carried in commensal *E. coli*. To accomplish this, we generated 50 new, fully or near-fully assembled genomes using hybrid assemblers such as Unicycler and Flye (22, 23). As a caveat, we note that these assemblers employ distinct strategies that are affected differently by the quality and quantity of long- and short-read data (29). Flye has a higher residual error rate than Unicycler at the nucleotide level, so Unicycler assemblies were preferred for archiving in GenBank (37/50 chromosomes and nearly all plasmids from this project). However, when Flye was able to span gaps that Unicycler could not, the resulting assemblies were sufficient for the purposes of this project.

Previous population-based investigations of commensal *E. coli* plasmids and antibiotic resistance have relied on PCR to identify plasmid replicons (13, 30). These studies found replicon distributions similar to those reported here, with F replicons most abundant by far, followed by the I complex (B/O, K, Z, and I1). Whether particular replicon types were associated with higher frequencies of antibiotic resistance varied. Johnson et al. (30) found positive associations between FIA, FIA, and FIB replicons and several antibiotic resistance traits, and Marcadé et al. (31) found that bla_{TEM-1} genes are strongly associated with F replicons. On the other hand, Moran et al. (13) noted that only B/O replicons were significantly more abundant in antibiotic-resistant isolates. Using contemporary DNA sequencing methods, we determined that in the commensal *E. coli* we analyzed, 66% of acquired resistance genes were located on plasmids containing F replicons and 10% were on plasmids with I-complex replicons, compared to 19% residing on the chromosome.

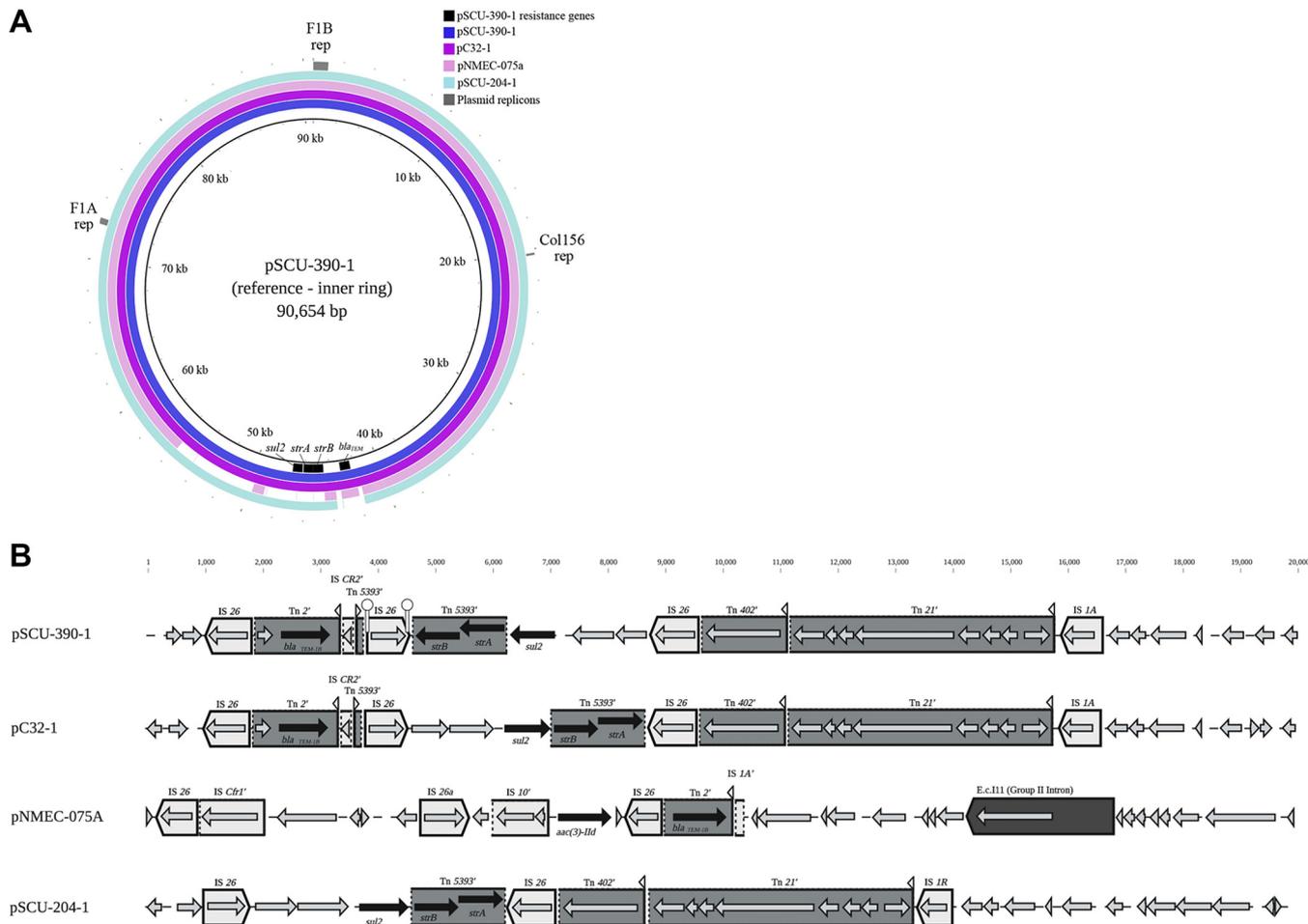For the most part, the large plasmids carrying antibiotic resistance genes were not

FIG 4 A large conserved plasmid carrying antibiotic resistance is found in ST1193 isolates. (A) Alignment of plasmids pSCU-390-1, pSCU-204-1, pNMEC-075A (GenBank accession no. CP030112.1), and pC32_1 (GenBank accession no. CP041619.1) using BRIG (BLAST Ring Image Generator) (50); numbering starts at the F1B replicon. Select genetic regions shared by all of the plasmids are indicated on the outside ring, including plasmid replicons identified by Plasmid Finder. Antibiotic resistance genes in plasmid pSCU-390-1 are annotated in the inner ring. The variable region from approximately 40 to 60 kb, indicated by gaps in the alignment, is shown in panel B. (B) Comparison of the variable regions located between 40 and 60 kb in the conserved ST1193 plasmids. Transposable elements and resistance genes were identified using ResFinder (48) and GAMR (24) and visualized using BioRender. IS elements are indicated by light gray boxes, with their name above the box; transposons and the class 1 integron are indicated by dark gray boxes, with their name above. Conserved inverted repeats known to be associated with transposable element boundaries are indicated by triangles above the boundaries. Dashed lines indicate breakpoints (defined by sequence alignment) of interrupted elements; partial elements are indicated by a prime symbol following their name. Antibiotic resistance genes are indicated by black arrows, with their name underneath.

highly conserved, perhaps due to their "cargo" (including antibiotic resistance genes) being in constant flux due to mobile elements. Plasmids pSCU-204-1 and pSCU-390-1from ST1193 isolates are intriguing exceptions, as their structures are very similar and align closely with a plasmid (pNMEC-075A) from the only other fully assembled ST1193 genome in GenBank (32), as well as with a plasmid (pC32_1) from a *Shigella flexneri* isolate. Johnson et al. (33) recently reported that, based on draft genome sequences, plasmids similar to pNMEC-075A are likely present in many *E. coli* ST1193 isolates. The ST1193 lineage is globally distributed, and it has emerged within the United States in the past decade as a significant extraintestinal pathogen (33). What key functions this conserved plasmid may provide to *E. coli* ST1193, other than serving as the primary platform for mobile antibiotic resistance genes, remain to be determined. Notably, this plasmid completely lacks the genes associated with the F-plasmid conjugal machinery, and yet its presence in a *Shigella* isolate suggests that it is still capable of horizontal transmission between cells.

Large, low-copy-number plasmids make up vastly less of the DNA content of *E. coli* cells than chromosomal DNA. Why most transposable elements carrying resistance

genes are located on these small fractions of the genome is unknown. Tn*7* is one of the few transposons known to have a preferred integration site on the chromosome, but it nevertheless has a strong preference for insertion into conjugal plasmids (34). Sequence-independent factors related to replication mechanism (as in the case of Tn*7*), topology, or methylation state may influence target preference, and in turn may be influenced by host factors. Five of the 10 (50%) complete genomes we assembled from phylogroup D isolates contained acquired resistance genes on their chromosomes, a much higher frequency than the collection as a whole (6/29 isolates with acquired resistance genes [21%]). The types of plasmids, resistance genes, and mobile elements observed in phylogroup D isolates did not appear to be distinct from those in the remainder of isolates with acquired resistance genes, but perhaps as-yet unidentified host factors in this lineage influence the distribution of mobile elements between plasmids and chromosomes.

Very few transposable elements were observed on small mobilizable plasmids, despite their diversity and apparent abundance. Transposition onto small plasmids can occur; indeed, pSCU-105-2 likely resulted from transposition of a 6-kb Tn*2* (bla_{TEM-1}) macrolide resistance module onto a ColE1 plasmid backbone. Numerous nearly identical homologs to pSCU-120-3 are found in GenBank, including p12579_4 from *E. coli* O55:H7 strain RM12579, an enteropathogenic strain isolated in California in 1974 (35), and pCERC2, identical in a commensal *E. coli* isolate from Australia in 2012 (36). The authors noted that these plasmids had likely been circulating globally in human-associated *E. coli* for decades, indicating their stability. Nevertheless, the low frequency of small plasmids carrying resistance genes in *E. coli* suggests that transpositional events involving small plasmids are generally inhibited, usually unstable, or are selected against. This may be fortunate for the human host, given that high-level expression of a resistance gene on a high-copy-number plasmid can potentiate a higher level (and in the case of β-lactams broader-spectrum) of antibiotic resistance (37, 38). The 9.7-kb pSCU-105-2 plasmid may illustrate this, as despite the plasmid-borne bla_{TEM-1} gene being wild type in sequence, SCU-105 displays an enhanced resistance to cephalosporins not seen in other isolates with this gene.

From the bacterial perspective, clustering of resistance genes on plasmids is advantageous for facilitating dramatic and simultaneous gains in resistance to multiple antibiotics. Nevertheless, the evolutionary dilemma of the "plasmid paradox" reflects the assumption that plasmid replication and maintenance costs exacted on the host are only offset under conditions where the plasmid explicitly provides a selective advantage, such as in the presence of antibiotics (39). Under such conditions, the plasmid is a symbiont; in their absence, the plasmid is a parasite. It should therefore be advantageous for resistance genes to move to the chromosome, where the host could benefit from them at a reduced cost. Recent experimental work on plasmid-host relationships (40) suggests that plasmid-host coevolution and compensatory mutations can reduce costs of plasmid maintenance and favor continued carriage of resistance genes and other genetic cargo on plasmid vectors. These findings have implications as well for the movement of such plasmids into new hosts (41); clearly there is much still to learn in this field.

Understanding the mobility of antibiotic resistance genes within genomes, within species, and within the microbiome at large can provide critical insights into trends in drug resistance among pathogens. The work presented here focuses on commensal *E. coli*, many of which can convert into opportunists causing extraintestinal infections (e.g., urinary tract infections [UTIs] or sepsis) (42). Almost half of the isolates examined here were from phylogroup B2, from which most extraintestinal pathogenic *E. coli* (ExPEC) strains derive (43, 44), and common ExPEC types represented among them included ST95 (12 isolates), ST1193 (4 isolates), ST73 (3 isolates), ST131 (3 isolates), and ST69 (phylogroup D, 8 isolates). The potential for F and other conjugal plasmids to facilitate acquisition of antibiotic resistance in *E. coli* and related species, including *Shigella*, *Klebsiella*, *Enterobacter*, *Salmonella*, and *Citrobacter*, will continue to be explored in future work.

MATERIALS AND METHODS

Strains and media. Commensal *E. coli* bacteria were obtained from self-administered rectal swabs by study participants (college students aged 19 to 22 years old) over a 6-year period from 2014 to 2019. The study protocol and informed consent documents were approved by the Human Subjects Research Committee at Santa Clara University. Swabs were streaked on CHROMagar Orientation agar (CHROMagar, Paris, France) (45) containing no antibiotics and incubated at 37°C for 16 to 24 h. Colonies were identified by color and restreaked for isolation. No more than one isolate per student was included in the data reported here. Isolates were identified to the species level by the API20E system (bioMérieux) and/or 16S rRNA sequencing. Isolates used in this work are described in Tables S1 and S2 in the supplemental material.

DNA sequencing, assembly, and analysis. Genomic DNA was prepared from broth-grown cultures using the Macherey Nagel microbial DNA isolation kit. DNA preparations were assessed by agarose gel electrophoresis, UV spectroscopy, and Qubit fluorometry. Library preparation and sequencing with the Illumina MiSeq platform followed the manufacturer's recommendations. 150-bp paired-end reads were trimmed based on length and quality using BBDUK (https://jgi.doe.gov/data-and-tools/bbtools/). *De novo* assembly of Illumina reads was done using the Geneious assembler (BioMatters LTD, Auckland, New Zealand). Long-read sequencing on the Oxford Nanopore MinION instrument followed the native genomic DNA barcoding sequencing protocol (protocol LSK108, Oxford Nanopore Technologies). MinION data were processed in MinKNOW (v. 3.6.5) using the Guppy basecaller (v.3.2.10), and demultiplexed by Epi2Me (Oxford Nanopore Technologies). Genome assemblies are described in Table S1 (isolates with completely or near-completely assembled genomes) and Table S2 (isolates with draft assemblies only). GenBank accession information is provided in Table 2 and Table S1; GenBank entries include metadata such as read coverage.

Assembly of MinION reads, combined with MiSeq reads, was done with the Unicycler (version 0.4.8) hybrid assembler (22). When genome assembly could not be achieved with Unicycler, Flye (version 2.6) (23) was applied to the same data. Following assembly with Flye, contigs were polished with Pilon (46) using short-read data. Unicycler assemblies were preferred, as Pilon polishing of Flye contigs leaves a significant residual error rate of 0.2 to 1%, but this did not interfere with the ultimate goal of localizing genes to plasmids or chromosomes.

Annotation was done by RAST v2 (47). Assembled genomes were analyzed using several tools from the Center for Genomic Epidemiology (https://cge.cbs.dtu.dk/services/), including ResFinder v3.2 (48) for identifying acquired antibiotic resistance genes and/or relevant mutations, MLST version 2.0 for multi-locus sequence typing (49), and PlasmidFinder version 1.3 for identification of plasmid replicons (21). IncI complex plasmids were differentiated into B/O, I, K, and Z subtypes by comparison to the *repA* sequences for the respective subtypes recommended by Moran et al. (13). Other mobile genetic elements were identified using the Galileo Antimicrobial Resistance (GAMR) software (ArcBio, Cambridge, MA, USA), which is derived from the Multiple Antibiotic Resistance Annotator (MARA) database (24).

Phenotypic testing. Antimicrobial susceptibility testing was performed by Kirby-Bauer disk diffusion assays, using guidelines from the manufacturer (Hardy Diagnostics). Antibiotics tested included β-lactams (ampicillin and cephalothin), aminoglycosides (gentamicin, kanamycin, and streptomycin), chloramphenicol, quinolones (nalidixic acid and norfloxacin), macrolides (azithromycin), tetracyclines, sulfonamides (sulfamethoxazole), and trimethoprim.

Analysis of plasmid mobilization of antibiotic resistance. Plasmid DNA was isolated from *E. coli* cultures using the ZR Plasmid Miniprep Classic kit (Zymo Research) and analyzed on 1% agarose gels. Because large plasmids are not recovered with high efficiency from plasmid preparations, both plasmid and genomic DNA samples were used for electroporation with commercial electrocompetent *E. coli* NEB5α (New England Biolabs). Colonies were selected on LB agar plus ampicillin (50 μg/ml), streptomycin (50 μg/ml), gentamicin (20 μg/ml), or oxytetracycline (10 μg/ml).

Data availability. All complete or nearly complete *E. coli* genome sequences described herein have been archived in GenBank as part of BioProject PRJNA624897. Individual GenBank accession numbers are provided in Table 2 and Table S1.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

TABLE S1, DOCX file, 0.1 MB.

TABLE S2, DOCX file, 0.02 MB.

REFERENCES

1. Levine MM. 1987. *Escherichia coli* that cause diarrhea: enterotoxigenic, enteropathogenic, enteroinvasive, enterohemorrhagic, and enteroadherent. J Infect Dis 155:377–389. https://doi.org/10.1093/infdis/155.3.377.

2. Foxman B. 2010. The epidemiology of urinary tract infection. Nat Rev Urol 7:653–660. https://doi.org/10.1038/nrurol.2010.190.

3. Russo TA, Johnson JR. 2003. Medical and economic impact of extraintestinal infections due to *Escherichia coli*: focus on an increasingly important endemic problem. Microbes Infect 5:449–456. https://doi.org/10.1016/S1286-4579(03)00049-2.

4. Phan MD, Forde BM, Peters KM, Sarkar S, Hancock S, Stanton-Cook M, Zakour NLB, Upton M, Beatson SA, Schembri MA. 2015. Molecular characterization of a multidrug resistance IncF plasmid from the globally disseminated *Escherichia coli* ST131 clone. PLoS One 10:e0122369. https://doi.org/10.1371/journal.pone.0122369.

5. Tchesnokova VL, Rechkina E, Larson L, Ferrier K, Weaver JL, Schroeder DW, She R, Butler-Wu SM, Aguero-Rosenfeld ME, Zerr D, Fang FC, Ralston J, Riddell K, Scholes D, Weissman S, Parker K, Spellberg B, Johnson JR, Sokurenko EV. 2019. Rapid and extensive expansion in the United States of a new multidrug-resistant *Escherichia coli* clonal group, sequence type 1193. Clin Infect Dis 68:334–337. https://doi.org/10.1093/cid/ciy525.

6. Guiral E, Quiles MG, Muñoz L, Moreno-Morales J, Alejo-Cancho I, Salvador P, Alvarez-Martinez MJ, Marco F, Vila J. 2018. Emergence of resistance to quinolones and β-lactam antibiotics in enteroaggregative and enterotoxigenic *Escherichia coli* causing traveler's diarrhea. Antimicrob Agents Chemother 63:e01745-18. https://doi.org/10.1128/AAC.01745-18.

7. Blair JM, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJ. 2015. Molecular mechanisms of antibiotic resistance. Nat Rev Microbiol 13:42–51. https://doi.org/10.1038/nrmicro3380.

8. Iredell J, Brown J, Tagg K. 2016. Antibiotic resistance in Enterobacteriaceae: mechanisms and clinical implications. BMJ 352:h6420. https://doi.org/10.1136/bmj.h6420.

9. Tenaillon O, Skurnik D, Picard B, Denamur E. 2010. The population genetics of commensal *Escherichia coli*. Nat Rev Microbiol 8:207–217. https://doi.org/10.1038/nrmicro2298.

10. Gordon DM, Cowling A. 2003. The distribution and genetic structure of *Escherichia coli* in Australian vertebrates: host and geographic effects. Microbiology 149:3575–3586. https://doi.org/10.1099/mic.0.26486-0.

11. Bailey JK, Pinyon JL, Anantham S, Hall RM. 2010. Commensal *Escherichia coli* of healthy humans: a reservoir for antibiotic-resistance determinants. J Med Microbiol 59:1331–1339. https://doi.org/10.1099/jmm.0.022475-0.

12. Amábile-Cuevas CF, Chicurel ME. 1992. Bacterial plasmids and gene flux. Cell 70:189–199. https://doi.org/10.1016/0092-8674(92)90095-t.

13. Moran RA, Anantham S, Pinyon JL, Hall RM. 2015. Plasmids in antibiotic susceptible and antibiotic resistant commensal *Escherichia coli* from healthy Australian adults. Plasmid 80:24–31. https://doi.org/10.1016/j.plasmid.2015.03.005.

14. Meynell EL, Meynell GG, Datta N. 1968. Phylogenetic relationships of drug-resistance factors and other transmissible bacterial plasmids. Bacteriol Rev 32:55–83. https://doi.org/10.1128/MMBR.32.1.55-83.1968.

15. Johnson TJ, Danzeisen JL, Youmans B, Case K, Llop K, Munoz-Aguayo J, Flores-Figueroa C, Aziz M, Stoesser N, Sokurenko E, Price LB, Johnson JR. 2016. Separate F-type plasmids have shaped the evolution of the H30 subclone of *Escherichia coli* sequence type 131. mSphere 1:e00121-16. https://doi.org/10.1128/mSphere.00121-16.

16. Matamoros S, Van Hattem JM, Arcilla MS, Willemse N, Melles DC, Penders J, Vinh TN, Hoa NT, de Jong MD, Schultsz C. 2017. Global phylogenetic analysis of *Escherichia coli* and plasmids carrying the *mcr-1* gene indicates bacterial diversity but plasmid restriction. Sci Rep 7:15364. https://doi.org/10.1038/s41598-017-15539-7.

17. Arredondo-Alonso S, Willems RJ, van Schaik W, Schürch AC. 2017. On the (im)possibility of reconstructing plasmids from whole-genome short-read sequencing data. Microb Genom 3:e000128. https://doi.org/10.1099/mgen.0.000128.

18. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. Microb Genom 3:e000132. . https://doi.org/10.1099/mgen.0.000132.

19. Yoshida H, Bogaki MA, Nakamura MI, Nakamura SH. 1990. Quinolone resistance-determining region in the DNA gyrase *gyrA* gene of *Escherichia coli*. Antimicrob Agents Chemother 34:1271–1272. https://doi.org/10.1128/aac.34.6.1271.

20. Heisig P. 1996. Genetic evidence for a role of *parC* mutations in development of high-level fluoroquinolone resistance in *Escherichia coli*. Antimicrob Agents Chemother 40:879–885. https://doi.org/10.1128/AAC.40.4.879.

21. Carattoli A, Zankari E, Garcìa-Fernandez A, Larsen MV, Lund O, Villa L, Aarestrup FM, 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.

22. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput Biol 13:e1005595. https://doi.org/10.1371/journal.pcbi.1005595.

23. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.

24. Partridge SR, Tsafnat G. 2018. Automated annotation of mobile antibiotic resistance in Gram-negative bacteria: the Multiple Antibiotic Resistance Annotator (MARA) and database. J Antimicrob Chemother 73:883–890. https://doi.org/10.1093/jac/dkx513.

25. Partridge SR, Kwong SM, Firth N, Jensen SO. 2018. Mobile genetic elements associated with antimicrobial resistance. Clin Microbiol Rev 31:e00088-17. https://doi.org/10.1128/CMR.00088-17.

26. Deng Y, Bao X, Ji L, Chen L, Liu J, Miao J, Chen D, Bian H, Li Y, Yu G. 2015. Resistance integrons: class 1, 2 and 3 integrons. Ann Clin Microbiol Antimicrob 14:45. https://doi.org/10.1186/s12941-015-0100-6.

27. Baquero F. 2004. From pieces to patterns: evolutionary engineering in bacterial pathogens. Nat Rev Microbiol 2:510–518. https://doi.org/10.1038/nrmicro909.

28. Stokes HW, Gillings MR. 2011. Gene flow, mobile genetic elements and the recruitment of antibiotic resistance genes into Gram-negative pathogens. FEMS Microbiol Rev 35:790–819. https://doi.org/10.1111/j.1574-6976.2011.00273.x.

29. De Maio N, Shaw LP, Hubbard A, George S, Sanderson ND, Swann J, Wick R, AbuOun M, Stubberfield E, Hoosdally SJ, Crook DW, Peto TEA, Sheppard AE, Bailey MJ, Read DS, Anjum MF, Walker AS, Stoesser N, REHAB consortium. 2019. Comparison of long-read sequencing technologies in the hybrid assembly of complex bacterial genomes. Microbial Genomics 5:e000294. https://doi.org/10.1099/mgen.0.000294.

30. Johnson TJ, Logue CM, Johnson JR, Kuskowski MA, Sherwood JS, Barnes HJ, DebRoy C, Wannemuehler YM, Obata-Yasuoka M, Spanjaard L, Nolan LK. 2012. Associations between multidrug resistance, plasmid content, and virulence potential among extraintestinal pathogenic and commensal *Escherichia coli* from humans and poultry. Foodborne Pathog Dis 9:37–46. https://doi.org/10.1089/fpd.2011.0961.

31. Marcadé G, Deschamps C, Boyd A, Gautier V, Picard B, Branger C, Denamur E, Arlet G. 2009. Replicon typing of plasmids in *Escherichia coli* producing extended-spectrum β-lactamases. J Antimicrob Chemother 63:67–71. https://doi.org/10.1093/jac/dkn428.

32. Nielsen DW, Ricker N, Barbieri NL, Wynn JL, Gómez-Duarte OG, Iqbal J, Nolan LK, Allen HK, Logue CM. 2018. Complete genome sequence of the multidrug-resistant neonatal meningitis *Escherichia coli* serotype O75:H5:K1 strain mcjchv-1 (NMEC-O75). Microbiol Resour Announc 7:e01043-18. https://doi.org/10.1128/MRA.01043-18.

33. Johnson TJ, Elnekave E, Miller EA, Munoz-Aguayo J, Figueroa CF, Johnston B, Nielson DW, Logue CM, Johnson JR. 2019. Phylogenomic analysis of extraintestinal pathogenic *Escherichia coli* sequence type 1193, an emerging multidrug-resistant clonal group. Antimicrob Agents Chemother 63:e01913-18. https://doi.org/10.1128/AAC.01913-18.

34. Peters JE, Craig NL. 2001. Tn7: smarter than we thought. Nat Rev Mol Cell Biol 2:806–814. https://doi.org/10.1038/35099006.

35. Kyle JL, Cummings CA, Parker CT, Quiñones B, Vatta P, Newton E, Huynh S, Swimley M, Degoricija L, Barker M, Fontanoz S, Nguyen K, Patel R, Fang R, Tebbs R, Petrauskene O, Furtado M, Mandrell RE. 2012. *Escherichia coli* serotype O55:H7 diversity supports parallel acquisition of bacteriophage at Shiga toxin phage insertion sites during evolution of the O157:H7 lineage. J Bacteriol 194:1885–1896. https://doi.org/10.1128/JB.00120-12.

36. Anantham S, Hall RM. 2012. pCERC1, a small, globally disseminated plasmid carrying the *dfrA14* cassette in the *strA* gene of the *sul2-strA-strB* gene cluster. Microb Drug Resist 18:364–371. https://doi.org/10.1089/mdr.2012.0008.

37. San Millan A, Escudero JA, Gifford DR, Mazel D, MacLean RC. 2016. Multi-

copy plasmids potentiate the evolution of antibiotic resistance in bacteria. Nat Ecol Evol 1:10–18. https://doi.org/10.1038/s41559-016-0010.

38. Harrison E, Brockhurst MA. 2012. Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends Microbiol 20:262–267. https://doi.org/10.1016/j.tim.2012.04.003.

39. MacLean RC, San Millan A. 2015. Microbial evolution: towards resolving the plasmid paradox. Curr Biol 25:R764–R767. https://doi.org/10.1016/j.cub.2015.07.006.

40. Harrison E, Guymer D, Spiers AJ, Paterson S, Brockhurst MA. 2015. Parallel compensatory evolution stabilizes plasmids across the parasitism-mutualism continuum. Curr Biol 25:2034–2039. https://doi.org/10.1016/j.cub.2015.06.024.

41. Porse A, Schønning K, Munck C, Sommer MO. 2016. Survival and evolution of a large multidrug resistance plasmid in new clinical bacterial hosts. Mol Biol Evol 33:2860–2873. https://doi.org/10.1093/molbev/msw163.

42. Leimbach A, Hacker J, Dobrindt U. 2013. *E. coli* as an all-rounder: the thin line between commensalism and pathogenicity, p 3–32. *In* Dobrindt U, Hacker JH, Svanborg C (ed), Between pathogenicity and commensalism. Springer, Berlin, Germany.

43. Zhang L, Foxman B, Marrs C. 2002. Both urinary and rectal *Escherichia coli* isolates are dominated by strains of phylogenetic group B2. J Clin Microbiol 40:3951–3955. https://doi.org/10.1128/jcm.40.11.3951-3955.2002.

44. Le Gall T, Clermont O, Gouriou S, Picard B, Nassif X, Denamur E, Tenaillon O. 2007. Extraintestinal virulence is a coincidental by-product of commensalism in B2 phylogenetic group *Escherichia coli* strains. Mol Biol Evol 24:2373–2384. https://doi.org/10.1093/molbev/msm172.

45. Samra Z, Heifetz M, Talmor J, Bain E, Bahar J. 1998. Evaluation of use of a new chromogenic agar in detection of urinary tract pathogens. J Clin Microbiol 36:990–994. https://doi.org/10.1128/JCM.36.4.990-994.1998.

46. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.

47. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.

48. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.

49. Larsen MV, Cosentino S, Rasmussen S, Friis C, Hasman H, Marvig RL, Jelsbak L, Sicheritz-Pontén T, Ussery DW, Aarestrup FM, Lund O. 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50:1355–1361. https://doi.org/10.1128/JCM.06094-11.

50. Alikhan NF, Petty NK, Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics 12:402. https://doi.org/10.1186/1471-2164-12-402.