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 Hessa

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.
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
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; and phylogroup F, 2 isolates).

### 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*<sub>TEM</sub>        | 31 (31)<sup>a</sup>                        |
| *bla*<sub>CTX-M</sub>      | 5 (5)<sup>b</sup>                         |
| *bla*<sub>SHV-1</sub>      | 1 (1)                                      |
| *bla*<sub>CMY</sub>        | 2 (2)<sup>c</sup>                         |
| **Aminoglycosides**       |                                            |
| *strA, strB*               | 27, 27 (27)                                |
| *aadA (3 alleles)*         | 19 (19)<sup>a</sup>                       |
| *aadB*                     | 1 (1)                                      |
| *aac(3)-Idd*               | 8 (8)                                      |
| *aph(3*)*                  | 1 (1)                                      |
| *sat*                      | 2 (2)                                      |
| **Sulfonamides**           |                                            |
| *sul1*                     | 18 (1)                                     |
| *sul2*                     | 29 (29)                                    |
| **Trimethoprims**          |                                            |
| *dfrA (5 alleles)*         | 27 (27)<sup>a</sup>                       |
| **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)                                      |

<sup>a</sup>*bla*<sub>TEM-1B</sub> in 29 isolates, *bla*<sub>TEM-1C</sub> in 1 isolate, and *bla*<sub>TEM-34</sub> in 1 isolate.

<sup>b</sup>*bla*<sub>CTX-M-14</sub> in 3 isolates and *bla*<sub>CTX-M-27</sub> in 2 isolates.

<sup>c</sup>*bla*<sub>CMY-2</sub> in 1 isolate and *bla*<sub>CMY-3</sub> in 1 isolate.

<sup>d</sup>*aadA1* in 5 isolates, *aadA2* in 3 isolates, and *aadA5* in 11 isolates.

<sup>e</sup>*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.
shows the most common cassette configuration, contained intact class 1 integrons (4.6 kb) and “large” plasmids (26 to 190 kb; size distribution shown in Figure 1). Integrons are split into two general pools, designated here as “small” (1 to 13 kb; mean size, 6.7 kb) and “large” plasmids (26 to 190 kb; mean size, 103 kb). In total, 42 acquired resistance genes (27% of the total) were associated with class 1 or 2 integrons. In 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 cmrA (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 Tn7-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 Tn7 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 incl 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) | Genome component (GenBank accession no.) | Size | Resistance genes (associated mobile elements) | Plasmid replicon(s) |
|------------|------|---------|-------------------------|------------------------------------------|------|-----------------------------------------------|---------------------|
| A          | 10   | SCU-103 | AMP, CEF, AZM, STR, SXT, TET | pSCU-103-1 (CP054458) | 139 kb | adaAS-dfrA17 (class 1 integron cassette) plus sul1 adjacent; bla<sub>CTX-M-2</sub> (IS903C, ISEcp1 flank); strA-strB (Tn5393') plus sul2 (Tn5393' flank); mphA (IS6100, IS626 flank); tetA (Tn1721') | F1A, F1B, FII |
| B1         | 3695 | SCU-106 | STR, TET | pSCU-106-2 (CP053236) | 112 kb | strA-strB (Tn5393'), tetA (Tn1721 within Tn5393') | F1B', FIC(II) |
|            |      | SCU-308 | AMP, STR, SXT | pSCU-308-1 (CP053282) | 152 kb | bla<sub>TEM-1</sub> (Tn2') plus strA-strB (Tn5393') and sul2; dfrA5 (class 1 integron cassette fragment) | F1B', FII |
| B2         | 14   | SCU-387 | AMP, AZM (int) | pSCU-387-2 (CP051690) | 39 kb<sup>b</sup> | bla<sub>TEM-1</sub> (Tn, 1,2,3-like); mphA | FII |
|            | 73   | SCU-112 | AMP, CEF (int) | pSCU-112-1 (CP051726) | 104 kb | adaA' (class 1 integron cassette fragment); bl<sub>SHV-1</sub>, (IS526 flanking both sides) | F1B', FII, Col156 |
|            | 91   | SCU-121 | TET | pSCU-121-1 (CP054329) | 68 kb | tetA (Tn1721') | FII |
|            | 95   | SCU-108 | AMP | pSCU-108-2 (CP051737) | 72 kb | bla<sub>TEM-1</sub> (Tn2) | FII |
|            | SCU-123 | AMP, STR, SUL | pSCU-123-2 (CP051713) | 95 kb | bla<sub>TEC-1</sub> (Tn2c); strA-strB' (Tn5393') plus sul2 (Tn5393' flank) | B/O/K/Z (B/O) |
|            | SCU-306 | AZM, SXT | pSCU-306-1 (CP053232) | 129 kb<sup>b</sup> | adaA2-dfrA12 (class 1 integron cassette) plus sul1 adjacent; mphA (IS6100, IS626, 2 copies) | F1B, FII, Col156 |
|            | 131  | SCU-182 | AMP, GEN | pSCU-182-1 (CP054376.1, CP054375.1, CP054374.1, CP054373.1) | 168 kb<sup>b</sup> | adaC-ild (IS526, IS10); bl<sub>TEC-1</sub> (Tn2') | F1B, FII, Col156 |
|            |      | SCU-481 | AMP, AMC, AZM (int), CHL, SXT | pSCU-481-1 (JABLYBOO000000000.1) | 144 kb<sup>b</sup> | adaB-aacC-cmlA6 (class 1 integron cassette); adaA5-dfrA17 (class 1 integron cassette) plus sul1 adjacent; bl<sub>TEM-3</sub> (Tn2); mphA (IS62, IS6100 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 (Tn5393', ISCR2') plus sul2 (Tn5393' flank) | B/O/K/Z (Z) |
|            | 357  | SCU-124 | AMP | pSCU-124-2 (CP051708) | 73 kb | bl<sub>TEM-1</sub> (Tn2) | FII |
|            | 1193 | SCU-147 | AMP, AZM, GEN, STR, SXT, TET | pSCU-147-1 (CP054326) | 105 kb<sup>b</sup> | adaC-ild (IS510', IS526 flank); adaA5-dfrA17 (class 1 integron cassette fragment); bl<sub>TEM-1</sub> (Tn,1,2,3-like); mphA (IS6100, IS626 flank); strA-strB (Tn5393') plus sul2 (Tn5393' flank); tetA (Tn1721') | F1A, F1B, Col156 |
|            | SCU-204 | STR, SUL | pSCU-204-1 (CP054414.1) | 88 kb | strA-strB (Tn5393') plus sul2 (RSF1010-like) | F1A, F1B, Col156 |
|            | SCU-390 | AMP, STR, SUL | pSCU-390-1 (CP054321) | 91 kb | bl<sub>TEM-1</sub> (Tn2'); strA-strB (Tn5393') plus sul2 (RSF1010-like) | F1A, F1B, Col156 |
|            | 2279 | SCU-479 | AMP, AMC, CEF, CHL, STR, SUL, TET | Chromosome (CP054317) | 6.2 Mb<sup>b</sup> | bl<sub>CTX-M-14</sub> (ISEcp1); bl<sub>CP206-121</sub> (ISEcp1); bl<sub>TEM-1</sub=./B (IS5393)'; strA-strB (Tn5393') plus sul2 (Tn5393' flank); tetA (Tn1721') | F1A, F1B, Col156 |
| D          | 38   | SCU-164 | SXT, TET | Chromosome (CP054343) | 5.4 Mb<sup>b</sup> | dfrA7 (class 1 integron cassette) plus sul1 adjacent; sul2 (IS5075, ISCR2); tetD (flanked by IS26 and Tn2') | F1A, F1B, Col156 |
|            | SCU-397 | AMP, CEF, CHL, STR, SXT, TET | Chromosome (CP054828.1) | 5.3 Mb<sup>b</sup> | bl<sub>CTX-M-14</sub> (2 copies, each between ISEcp1 and IS903B'); bl<sub>TEM-1</sub> (Tn2'); dfrA7 (class 1 integron cassette) plus sul1 adjacent; strA-strB (Tn5393') plus sul2 (Tn5393' flank); tetD (IS26, Tn2' flank); catA1 | F1B, FII, Col156 |
|            | SCU-486 | AMP, CEF, AZM, GEN, STR, SXT, TET | Chromosome (CP051749) | 5.2 Mb | bl<sub>CTX-M-14</sub> (IS903B'; ISEcp1 flank); strA-strB (Tn5393'); sul2 (IS5075, ISCR2 flank); tetD (IS26, Tn2' flank); bl<sub>TEM-1</sub> (Tn2'); catA1 | F1A, F1B, Col156 |

(Continued on next page)
TABLE 2 (Continued)

| Phylogroup | MLST Isolate | Resistance phenotype(s)× | Genome component (GenBank accession no.) | Size | Resistance genes (associated mobile elements)×× | Plasmid replicon(s)×× |
|------------|--------------|--------------------------|-----------------------------------------|------|-----------------------------------------------|-----------------------|
| 69         | SCU-313 AMP, AZM, GEN, STR, SUL, TET | pSCU-313-1 (CP051695) | 105 kb | aac(3)-IId (IS26, IS107); aadA5-dfrA17 (class 1 integron cassette) plus sul1 adjacent; mphA (IS6100, IS26 flank) | F1A, F1B' | 
| 69         | SCU-482 AMP, AZM, STR, SXT | pSCU-482-1 (CP053248) | 145 kb | aadA5-dfrA17 (class 1 integron cassette) plus sul1 adjacent; blaTEM-1B (Tn2'; Tn2); mphA (IS6100', IS26' flank); strA-strB (Tn5393') plus sul2 (Tn5393' flank); tetA (Tn1721') | F1B', Fil | 
| 106        | SCU-318 AMP, STR, SUL, TET | pSCU-318-1 (CP051693) | 105 kb | blaTEM-1B (Tn2); strA-strB (Tn5393') plus sul2 (Tn5393' flank); tetB (Tn10') | F1B, Fil | 
| 394        | SCU-105 AMP, AZM, STR, SXT | Chromosome (CP051738) | 5.2 Mb | dfrA1-sat2-aadA1 (class 2 integron cassettes in Tn7) | 
| 963        | SCU-109 AMP, AMC, CEF, GEN | pSCU-109-1 (CP051737) | 110 kb | aac(3)-IId (IS26, IS107'); blaTEM-1B (Tn2'); mphA (IS6100', IS26' flank) | F1B', Fil, Col156 |
| F          | SCU-175 AMP, AZM, STR, SXT, TET | pSCU-175-1 (CP054380.1) | 124 kb | aadA1-dfrA1-sat2 (class 2 integron in Tn7); mphA (IS56, IS6100 flank); sul2 (ISCR7' flank); tetB (Tn10') | B/O/K/Z (Z) |
| 379        | SCU-172 AMP | pSCU-172-3 (CP054356) | 76 kb | blaTEM-1B (Tn2') | Fil |
| 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 (IS6100, IS26 flank); tetA (Tn1721') | F1A, F1B', Fil |
| 648        | SCU-120 AMP (int), CEF, AZM, STR, SXT, TET | pSCU-120-3 (CP054338) | 6.2 kb | strA-strB (Tn5393') plus sul2 (Tn5393' flank) | 

×Abbreviations: AMP, ampicillin; AMC, amoxicillin-clavulanic acid; AZM, azithromycin; CEF, cefalothin; CHL, chloramphenicol; GEN, gentamicin; KAN, kanamycin; NAL, nalidixic acid; NOR, norfloxacin; STR, streptomycin; SUL, sulfamethoxazole alone; SXT, sulfamethoxazole-trimethoprim; TEM, 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.

××Identification of antibiotic resistance genes was done with ResFinder (48). A prime symbol indicates that the identified antibiotic resistance gene was incomplete (less than 90% present). Mobile elements were identified using the Genome 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.

×××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).

××××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 remaining 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 electroprotants (data not shown).
Small plasmids rarely contained antibiotic resistance genes, but there were two exceptions. pSCU-105-2 is a 9.7-kb plasmid containing \( \text{bla}^{\text{TEM-1}} \) and \( \text{mphA} \) (macrolide resistance), and pSCU-120-3 is a 6.2-kb plasmid containing \( \text{strA-strB} \) (streptomycin resistance) and \( \text{sul2} \) (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 \( \text{Tn2} \) containing \( \text{bla}^{\text{TEM-1}} \) and a macrolide resistance locus (\( \text{mphA-mrx-mphR} \)). 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 \( \text{Tn5393} \) (\( \text{strA-strB} \)) followed by \( \text{sul2} \).

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
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 blaTEM-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
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-390-1 from 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, 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. Tn7 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 Tn7), 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 Tn2 (bla TEM-1) macrolide resistance module onto a ColE1 plasmid backbone. Numerous nearly identical homologs to pSCU-120-3 are found in GenBank, including pI2579.4 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 EpiZyme (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 multilocus sequence typing (49), and PlasmidFinder version 1.3 for identification of plasmid replicons (21). Incl complex plasmids were differentiated into B/O, I, K, and Z subtypes by comparison to the repA4 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.

ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health, grant R15AI130816-01A1. We thank the College of Arts and Sciences at Santa Clara University for supplemental funding.

We thank Tracy Ruscetti for ongoing feedback, Victoria Walton and Daryn Baker for technical support, Jim Grainger for assistance with figures, and the countless students in our undergraduate courses since 2012 who have participated in this project.
REFERENCES

1. Levine MM. 1987. Escherichia coli that cause diarrhea: enterotoxigenic, enteroaggregative, enteroinvasive, enterohemorrhagic, and enteropathogenic. 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/nurolo.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, Sarker S, Hancock S, Stanton-Cook M, Zakhour NLS, Upton M, Beaton 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, Muñoz L, Moreno-Morales J, Salvador P, Alvarez-Martinez MJ, Marco F, Vila J. 2018. Emergence of resistance to quinolones and β-lactam antibiotics in enterocaggregative and enterotoxigenic Escherichia coli causing traveler’s diarrhea. Antimicrob Agents Chemother 63:e00313-18. https://doi.org/10.1128/AAC.00313-18.

6. Blair JM, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJ. 2015. Molecular epidemiology of urinary tract infection. Nat Rev Urol 7:653–660. https://doi.org/10.1038/nrmicro2298.

7. Amábile-Cuevas CF, Chicurel ME. 1992. Bacterial plasmids and gene flux. Microbiol Rev 32:55–83. https://doi.org/10.1128/MBR.32.1.55-83.1986.

8. 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.

9. Bailey KK, Pinyon JL, Ananthan S, Hall RM. 2010. Commensal genetics of commensal Escherichia coli. Nat Rev Microbiol 8:207–217. https://doi.org/10.1038/nrmicro2298.

10. 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.

11. 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.

12. Partridge SR, Tsfat N. 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/10.1095/jac/dkh351.

13. Blazquez P. 1996. Genetic evidence for a role of parC mutations in develop-
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, Schanning 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, Tenailleau 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, Priester M, Abouelliel A, Sakhikumar 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.