Whole genome sequencing of diverse Shiga toxin-producing and non-producing \textit{Escherichia coli} strains reveals a variety of virulence and novel antibiotic resistance plasmids

Liliana Losada\textsuperscript{a}, Chitrita DebRoy\textsuperscript{b}, Diana Radune\textsuperscript{a}, Maria Kim\textsuperscript{a}, Ravi Sanka\textsuperscript{a}, Lauren Brinkac\textsuperscript{a}, Subhashinie Kariyawasam\textsuperscript{b}, Daniel Shelton\textsuperscript{c}, Pina M. Fratamico\textsuperscript{d}, Vivek Kapur\textsuperscript{b}, Peter C.H. Feng\textsuperscript{e,∗}

\textsuperscript{a} J. Craig Venter Institute, Rockville, MD, USA
\textsuperscript{b} Department of Veterinary and Biomedical Sciences, Pennsylvania State University, University Park, PA, USA
\textsuperscript{c} Environmental Microbial and Food Safety Laboratory, Agricultural Research Service, USDA, Beltsville, MD, USA
\textsuperscript{d} Eastern Regional Research Center, Agricultural Research Service, USDA, Wyndmoor, PA, USA
\textsuperscript{e} Division of Microbiology, FDA, College Park, MD, USA

\begin{abstract}

The genomes of a diverse set of \textit{Escherichia coli}, including many Shiga toxin-producing strains of various serotypes were determined. A total of 39 plasmids were identified among these strains, and many carried virulence or putative virulence genes of \textit{Shiga toxin-producing E. coli} strains, virulence genes for other pathogenic \textit{E. coli} groups, and some had combinations of these genes. Among the novel plasmids identified were eight that carried resistance genes to aminoglycosides, carbapenems, penicillins, cephalosporins, chloramphenicol, dihydrofolate reductase inhibitors, sulfonamides, tetracyclines and resistance to heavy metals. Two of the plasmids carried six of these resistance genes and two novel IncHI2 plasmids were also identified. The results of this study showed that plasmids carrying diverse resistance and virulence genes of various pathogenic \textit{E. coli} groups can be found in \textit{E. coli} strains and serotypes regardless of the isolate’s source and therefore, is consistent with the premise that these mobile elements carrying these traits may be broadly disseminated among \textit{E. coli}.

\end{abstract}

1. Introduction

Shiga-toxin producing \textit{Escherichia coli} (STEC) is characterized by the production of Shiga toxins (Stx), a family of related protein toxins encoded by lambdoid prophages with two major types designated Stx1 and Stx2 (Perna et al., 2001; Tarr et al., 2005). Among STEC, O157:H7 is the most important serotype that causes food-borne outbreaks worldwide (Mead and Griffin, 1998; Perna et al., 2001), but other STEC serotypes have also caused infections and severe human illnesses (Brooks et al., 2005). STEC strains that cause severe diseases often produce the intimin protein that enable the pathogen to adhere to intestinal epithelial cells (Mead and Griffin, 1998; Tarr et al., 2005), but they also carry a variety of plasmids, some of which encode for putative virulence factors and others, for antibiotic resistance (Johnson and Nolan, 2009). Among the best studied STEC plasmid is the large pO157 plasmid in O157:H7, which is approximately 60 MDa and 92 kb in size (Johnson and Nolan, 2009). Similar large plasmids are present in other STEC serotypes, and some of these, like the plasmid in serotypes O26:H11, O26:H30, O145 and O157:H7 Sakai have been sequenced to determine the presence or absence of specific genes, such as ehxA (enterohemolysin), espP (serine protease) and katP (catalase peroxidase) (Machino et al., 1998; Brunder et al., 1999, Fratamico et al., 2011; Yan et al., 2012). But few have examined the plasmid contents of other \textit{E. coli} strains and serotypes. Since plasmids are highly mobile genetic elements that can be transferred, we examined the overall plasmid content of various STEC and non-STEC \textit{E. coli} serotypes isolated from various sources and sequenced these plasmids to determine the diversity of genes that are found on these plasmids.

2. Materials and methods

The 26 bacterial strains used in this study are from 18 different O-serogroups and comprised of 22 STEC and four non-STEC strains. Of these, 21 strains were isolated from different hosts, including seven from cows, eight from humans, three from pigs, and one each from horse, rabbit, and goat. Four environmental strains were isolated from surface waters in Maryland (Feng et al., 2010) and one strain was isolated from a sample of contaminated food. The metadata on these strains are shown in Table 1. The flagellar (\textit{H}) type of some strains was identified, but others were not fully serotyped or were non-motile (NM) and so, the \textit{H} types are not shown (Table 1). All strains were whole genome sequenced using a combination of 8 kb paired-end 454 pyrosequencing.
and sequencing by synthesis with Illumina (http://gsc.jcvi.org). The fragment libraries obtained were assembled with the Celera Assembler software (Miller et al., 2008). Genomes were improved to high quality draft status using JCVI’s automated closure tools. Most of the larger plasmids were not assembled into a single contig, however scaffolding was possible by manual curation using paired-read status and read-depth. The physical mate-pair link and the difference in read depth was sufficient to distinguish among plasmids where multiple plasmids existed. Replicon typing was done by in silico PCR (CLC Genomics Workbench) using primers described by Johnson et al., 2007.

3. Results and discussion

The presence of plasmids in the strains were identified by contigs that showed typical plasmid characteristics such as (i) evidence of circularity as determined by mate-pair read status, (ii) annotations commonly associated with plasmids, and (iii) sequence similarity with other E. coli plasmids. We identified 39 new plasmids whose sequences have not been published previously. These ranged from 5–250 kb in length and were assembled into 1–50 contigs (Supplementary File 1). The number of plasmids present ranged from 0 to 4 per strain, with 14 strains carrying only 1 plasmid, and 4 strains with 4 plasmids. These plasmids were designated by using the prefix “p” followed by the strain name and in the case of strains with multiple plasmids, they were numbered from the largest to the smallest [e.g. pEC4.0967–1, pEC4.0967–2, etc.]. There were four strains that did not carry any plasmids; one was an O121:H19 strain and the other three were O157 non-H7 strains that were not STEC. These included strain 391-71 (14b), which was isolated from a human source.

### Table 1

<table>
<thead>
<tr>
<th>BioProject</th>
<th>Accession</th>
<th>Strain</th>
<th>Host</th>
<th>Serotype</th>
<th>No. Plasmids</th>
<th>Plasmid replication types</th>
<th>Virulence plasmid-endowed traits*</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRJN51085</td>
<td>AEZ000000000</td>
<td>1.2741</td>
<td>Cow</td>
<td>O2:H4</td>
<td>2</td>
<td>FIB; Frep</td>
<td>Elx, Sta, TraT</td>
</tr>
<tr>
<td>PRJN51087</td>
<td>AEZ000000000</td>
<td>97.0246</td>
<td>Cow</td>
<td>O5</td>
<td>4</td>
<td>FIB; 3 not classified</td>
<td>Elx</td>
</tr>
<tr>
<td>PRJN51089</td>
<td>AEZ000000000</td>
<td>5.0888</td>
<td>Cow</td>
<td>O8</td>
<td>1</td>
<td>FIB</td>
<td>Elx, Iuc, EspP, TraT</td>
</tr>
<tr>
<td>PRJN51091</td>
<td>AEZ000000000</td>
<td>97.0259</td>
<td>Cow</td>
<td>O11</td>
<td>2</td>
<td>FIA; not classified</td>
<td>Elx, KatP, TraT</td>
</tr>
<tr>
<td>PRJN51097</td>
<td>AEZ000000000</td>
<td>1.2264</td>
<td>Goat</td>
<td>O76</td>
<td>1</td>
<td>FIB</td>
<td>Elx, Sta, TraT</td>
</tr>
<tr>
<td>PRJN51101</td>
<td>AEZ000000000</td>
<td>96.0497</td>
<td>Human</td>
<td>O91</td>
<td>1</td>
<td>FIA</td>
<td>Elx, EspP, Pet, SigA, Sat</td>
</tr>
<tr>
<td>PRJN51103</td>
<td>AEZ000000000</td>
<td>99.0741</td>
<td>Food</td>
<td>O91</td>
<td>1</td>
<td>FIB</td>
<td>Iuc, TraT</td>
</tr>
<tr>
<td>PRJN51105</td>
<td>AEZ000000000</td>
<td>3.2608</td>
<td>Horse</td>
<td>0103:H2</td>
<td>1</td>
<td>FIB</td>
<td>Elx, TraT</td>
</tr>
<tr>
<td>PRJN51107</td>
<td>AEZ000000000</td>
<td>93.0624</td>
<td>Human</td>
<td>O103:H16</td>
<td>1</td>
<td>FIB</td>
<td>Elx, TraT</td>
</tr>
</tbody>
</table>
| PRJN51109  | AEZ000000000 | 4.2202 | Cow | O111     | 4            | FIB; B/O; 2 not classified | Elx, KatP, TraT |}

* 0157 non-H7 strains — not STEC.

**Virulence and putative virulence genes identified by BLAST to Virulence Factor Data Base (VFDB):** Elx: enterohemolysin (eoxCABD) operon, Esp: serine protease (espP), KatP: catalase peroxidase (katP); Iuc: aerobactin synthesis (iuc) operon; TraT: complement resistance; Pet: plasmid encoded enterotoxin of enteragggregative E. coli (EAEC); pAPEC: plasmid of avian pathogenic E. coli (APEC) which contains the virulence factors of APEC; Sat: secreted autotransporter toxin; Paa: porcine attaching and effacing associated factor; Sta: heat stable enterotoxin; SigA: secretory immunoglobulin A.
O26:H11 and O26:H30, the traT gene is on the large STEC plasmid that also carry ehhA (Fratamico et al., 2011). Our results suggest that this may be the case with other _E. coli_ serotypes as well, since 12/15 strains that had _traT_ also had _ehhA_.

Several strains were found to carry virulence genes from various pathogenic _E. coli_ groups. For example, a non-STE C strain carried the _ehhA_ gene that is most common in STEC, and three STEC strains had both the _ehhA_ gene and the stable toxin (ST) gene of enterotoxigenic _E. coli_ (ETEC). We also found another STEC strain that carried both the _ehhA_ and the plasmid-encoded toxin (Pet) gene of enteropathogenic _E. coli_ (EAEC). Among 8 of the newly-sequenced plasmids, there were a total of 29 antibiotic and heavy metal resistance genes also carried genes involved in repair of UV damage.

Two of the plasmids were identified to be of the H incompatibility group (IncHI2), which are known to play a role in the acquisition of antibiotic resistance and are also frequently associated with virulence in Proteobacteria (Johnson et al., 2006). The 2 IncHI2 plasmids were identified in STEC strains Ec4.0967, a rabbit isolate of O145:H2 serotype and C691–71 (14b), a human isolate of O128:H21 serotype. Johnson et al., 2006 reported that extraintestinal APEC can carry the pAPEC plasmids that are IncHI2 type and encode for resistance to as many as eight antimicrobial agents. Hence, we compared the sequences of our two plasmids to that of pAPEC-O1-R, an IncHI2 plasmid of 241.387 bp in size that was found in an APEC strain (Johnson et al., 2006). The pEc40967-1 was smaller (215,572 bp) with an overall GC content of 44.7%, which is less than that of the chromosome [49.5%] or that of pAPEC-O1-R [46%] (Johnson et al., 2006). Previous analysis of IncHI2 plasmids had suggested that antibiotic and heavy metal resistance can be acquired horizontally by these plasmids in dedicated hotspots, while the rest of the plasmid body is highly conserved (Johnson et al., 2006; Garcia-Fernandez and Catrall, 2010). Consistent with those reports, there were large areas of conservation and synteny between pAPEC-O1-R and pEc40967-1 that roughly covered 82% and 93% of each plasmid, respectively, and included all of the previously identified constant regions found in IncHI2 plasmids (Garcia-Fernandez and Catrall, 2010). The two regions that were not conserved between the plasmids were the region in pAPEC-O1-R that encodes for streptomycin resistance, which was absent in pEc40967-1 and conversely, the region encoding for arsenic resistance in pEc40967-1, which was absent in pAPEC-O1-R [Fig. 1]. The plasmid pEcC691-71 was assembled into 7 contigs (202,180 bp) and also shared a high degree of synteny with pAPEC-O1-R and pEc40967-1. But, pEcC691-71 lacked the arsenic resistance operon found in pAPEC4067-1 and also lacked the antibiotic resistance genes and the silver and copper resistance operons that were located in the same region in pAPEC-O1-R. Since all these resistance genes were localized within the same region on these plasmids, this region may be a hotspot for the insertion of resistance genes (Johnson et al., 2006).

### Table 2

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Antibiotic resistance</th>
<th>Heavy metal resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td>pEcB41</td>
<td>sul2</td>
<td>Hg</td>
</tr>
<tr>
<td>pEc40967-4</td>
<td>tetb</td>
<td>As/Te</td>
</tr>
<tr>
<td>pEc970246</td>
<td>tetc</td>
<td>Fe/Mg</td>
</tr>
<tr>
<td>pEc990741</td>
<td>ant3ia; aph3ia; sul3; tet</td>
<td>Fe/Mg</td>
</tr>
<tr>
<td>pEc970264</td>
<td>aph33ia; aph33ia; aph6id; sul2; tetb</td>
<td>Fe</td>
</tr>
<tr>
<td>pEc40522-4</td>
<td>tetb; dfra12; aph3ia; sul1; mubc</td>
<td>Hg</td>
</tr>
<tr>
<td>pEc23916</td>
<td>ant3ia; aph3ia; cml1; e1; sul3; tet</td>
<td>Cu/Hg</td>
</tr>
<tr>
<td>pEc91649</td>
<td>aph33ib; aph6id; b1; cmy2; cml3; e1; sul1; sul2; tet</td>
<td>Hg</td>
</tr>
</tbody>
</table>

* Gene designations according to ARDB (10): _tet_ and _tbr_; tetracycline; _aph_3<sub>ia</sub> Streptomycin; _ant_3<sub>ia</sub> Spectinomycin; _aph_3<sub>3ib</sub> Kanamycin; _sul_1 and _sul_2: Sulfonamides; _mubc_ Macrolides; _cml_1 and _cml_3: Chloramphenicol; _dfra12_: Trimethoprim; _b1, cmy2_: Carbencillin.

* Fe: Iron; Ni: Nickel; Hg: Mercury; As: Arsenic; Te: Tellurite; Cu: Copper.

Fig. 1. Nucleotide alignment of IncHI2 plasmids pEc40967-1, pAPEC-O1-R and pEcC691-71. Alignments were performed in Artemis Comparative Tool (Abbott et al., 2005) and the numbers on the strands represent nucleotide coordinates. Synteny and identity along plasmid backbones are shown with red and blue solid lines. The red lines signify alignment of positive strands and blue lines signify inversions. Areas absent of lines indicate ‘hotspots’, where different resistance genes are carried by different strains. The functions and resistance profiles labeled are: As, arsenic/arsenate resistance, Te, tellurite; Ag, silver efflux; Cu, copper efflux; class I integron encodes resistance to aminoglycosides and sulfonamide (Johnson et al., 2008).
et al., 2006). The two IncH2 plasmids we identified showed a lot of sequence similarities but also some differences, hence they were further characterized with the in silico double-locus sequence typing (DLST) method developed for IncH2 plasmids (Garcia-Fernandez and Carattoli, 2010). Analysis of the two loci in the conserved region used for typing showed that pEcG691-71 was very similar to plasmid R478 that was isolated from Serratia marcescens (Gilmour et al., 2004) and therefore, these belonged in the same pDSLST type of IncH2 plasmids. In contrast, the two same loci in pEcG691-71 were distinct from the others and pEcC691-71 may represent a new pDSLST type of IncH2 plasmids.

This study showed that genes encoding for resistance to antibiotics, heavy metal, UV light, and complement are commonly found on E. coli plasmids. The presence of these genes however, is not indicative that they are actually expressed. For example, the ehxA gene on the large plasmid of sorbitol-fermenting O157 STEC strains is not expressed and so enterohemolysin is not produced (Karch and Bielazewska, 2001). Similarly, E. coli strains can have plasmids with an intact promoter and genes that encode for resistance to streptomycin, β-lactamase, sulphonamides and tetracycline, but not all expressed these genes, presumably due to a chromosomal transcriptional control that silenced the expression of plasmid genes (Enne et al., 2006). However, the genetic evidence that a large diversity of virulence and resistance genes carried by plasmids of the different E. coli serotypes is consistent with the premise that plasmid transfer, probably catalyzed by mobile elements, including transposases, is common among E. coli and that these strains, regardless of sources of isolation, may carry plasmids with multiple resistance genes, as well as trait genetic markers of various pathogenic E. coli groups.

Nucleotide sequence accession numbers

The information on all plasmids used in this analysis and the E. coli Whole Genome Shotgun project and GenBank accession numbers are described in Table 1. The entire sequence of all plasmids is available at ftp://ftp.jcvi.org/STEC/plasmids/. Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.plasmid.2015.12.001.

Acknowledgments

This project has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services under contract number HHSN272200900007C. The authors also thank Mark Adams, Derrick Fouts and William Nierman for their review of this manuscript.

References