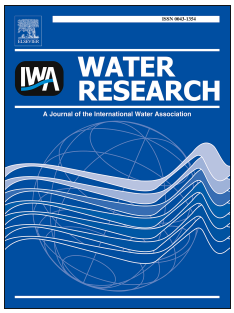
Journal Pre-proof



Characterization of water treatment-resistant and multidrug-resistant urinary pathogenic *Escherichia coli* in treated wastewater

Shuai Zhi, Paul Stothard, Graham Banting, Candis Scott, Kristin Huntley, Kanghee Ryu, Simon Otto, Nicholas Ashbolt, Sylvia Checkley, Tao Dong, Norma J. Ruecker, Norman F. Neumann

PII: S0043-1354(20)30364-X

DOI: https://doi.org/10.1016/j.watres.2020.115827

Reference: WR 115827

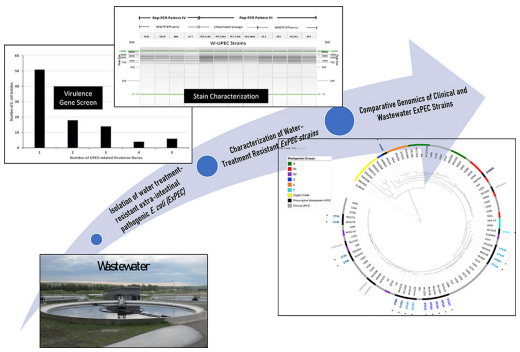
To appear in: *Water Research*

Received Date: 30 September 2019

Revised Date: 6 April 2020

Accepted Date: 11 April 2020

Please cite this article as: Zhi, S., Stothard, P., Banting, G., Scott, C., Huntley, K., Ryu, K., Otto, S., Ashbolt, N., Checkley, S., Dong, T., Ruecker, N.J., Neumann, N.F., Characterization of water treatment-resistant and multidrug-resistant urinary pathogenic *Escherichia coli* in treated wastewater, *Water Research* (2020), doi: https://doi.org/10.1016/j.watres.2020.115827.


Virulence
Gene Screen
Stain Characterization
Comparative Genomics of Clinical and
Wastewater ExPEC Strains
Characterization of Water-
Treatment Resistant ExPEC strains
Isolation of water treatment-
resistant extra-intestinal
pathogenic E. coli (ExPEC)
Wastewater

Characterization of Water Treatment-Resistant and Multidrug-Resistant Urinary Pathogenic *Escherichia coli* in Treated Wastewater

Shuai Zhi [a], Paul Stothard [b], Graham Banting [c], Candis Scott [c], Kristin Huntley [c], Kanghee Ryu [c], Simon Otto [c], Nicholas Ashbolt [c], Sylvia Checkley [d], Tao Dong [d], Norma J. Ruecker [e], and Norman F. Neumann [c*]

[a] School of Medicine, Ningbo University, Ningbo, China; [b] Faculty of Agricultural, Life and Environmental Sciences, University of Alberta, Edmonton, Alberta, Canada; [c] School of Public Health, University of Alberta, Edmonton, Alberta, Canada; [d] Faculty of Veterinary Medicine, University of Calgary, Calgary, Alberta, Canada; [e] City of Calgary, Water Quality Services, Calgary, Alberta, Canada.

* Corresponding Author: Norman F. Neumann, Ph.D.
357-E South Academic Building,
University of Alberta,
Edmonton, Alberta CANADA T6G 2G7
Telephone: 780-492-8502 Fax: 780-492-9070
Email: nfneuman@ualberta.ca

Abstract

A growing body of evidence has demonstrated that extraintestinal pathogenic *E. coli* (ExPEC), such as the urinary pathogenic *E. coli* (UPEC), are common constituents of treated wastewater, and therefore represent a potential public health risk. However, no single virulence gene, or set of virulence genes, can be used to conclusively identify this genetically diverse pathotype. As such we sought to identify and characterize the public health relevance of potential UPEC found in treated sewage/wastewater using a comparative genomics approach. Presumptive wastewater UPEC (W-UPEC) were initially identified by virulence gene screening against 5 virulence genes, and for which isolates containing ≥3 virulence genes were whole genome sequenced (n=24). Single nucleotide polymorphic (SNP) spanning tree analysis demonstrated that many of these wastewater UPEC (WUPEC) were virtually identical at the core genome (0.4 Mbp) when compared to clinical UPEC (C-UPEC) sequences obtained from NCBI, varying by as little as 1 SNP. Remarkably, at the whole genome level, W-UPEC isolates displayed >96% whole genome similarity to C-UPEC counterparts in NCBI, with one strain demonstrating 99.5% genome similarity to a particular C-UPEC strain. The W-UPEC populations were represented by sequence types (ST) known to be clinically important, including ST131, ST95, ST127 and ST640. Many of the W-UPEC carried the exact same complement of virulence genes as their most closely related C-UPEC strains. For example, O25b-ST131 W-UPEC strains possessed the same 80 virulence genes as their most closely related C-UPEC counterparts. Concerningly, W-UPEC strains also carried a plethora of antibiotic resistance genes, and O25b-ST131strains were designated as extended spectrum beta-lactamase (ESBL) producing *E. coli* by both genome profiling and phenotypic resistance testing. W-UPEC ST131 strains were found in the effluents of a single treatment plant at different times, as well as different wastewater treatment plants,

suggesting a differentially ability to survive wastewater treatment. Indeed, in sewage samples treated with chlorine doses sufficient for inducing a ~99.99% reduction in total *E. coli* levels, UPEC represented a significant proportion of the chlorine-resistant population. By contrast, no Shiga toxin-producing *E. coli* were observed in these chlorinated sewage libraries. Our results suggest that clinically-relevant UPEC exist in treated wastewater effluents and that they appear to be specifically adapted to survive wastewater treatment processes.

Keywords: urinary pathogenic *E. coli* (UPEC)**;** wastewater**;** treatment resistant; comparative genomics; extended spectrum beta-lactamase (ESBL); water quality.

Abbreviations:

E. coli, *Escherichia coli*; ExPEC, extraintestinal pathogenic *E. coli*; UPEC, urinary pathogenic *E. coli*; W-UPEC, wastewater UPEC; C-UPEC, clinical UPEC; ST, sequence types; ESBL, extended spectrum beta-lactamase; AMR, antimicrobial resistance; UTI, urinary tract infections; U.S. EPA, United States Environmental Protection Agency's; ATP, alternate test procedure; ProvLab, Alberta Provincial Laboratory for Public Health; WWTPs, wastewater treatment plants; LB, Luria-Bertani; MLST, multilocus sequencing typing; STEC, Shiga toxin-producing *E. coli*; gDNA, genomic DNA; CARD, Comprehensive Antibiotic Resistance Database; AST, Antibiotic Susceptibility Testing; EHEC, enterohemorrhagic *E. coli*.

1. Introduction

Drinking water treatment and waste sanitation represent the most important public health intervention strategies for control of infectious diseases in developed society. It is estimated that between 1900 and 1936 a reduction of 50% of infectious disease mortality rates in the U.S. could be attributed to the advent of drinking water treatment, accounting for a 75% decline in infant mortality and a 66% decline in childhood mortality (Cutler and Miller, 2005). An estimated economic return of 23:1 exists for clean water investments, and the 'cost-per-life-year-saved' has been estimated at $500 per individual (based on 2003 estimates), making water treatment the most cost-effective barrier to infectious disease prevention (Cutler and Miller, 2005). Consider the following – what would happen if microbes were able to breach this buttress of public health by evolving resistance to water treatment?

Indeed, several chlorine-specific transcription factors have been identified in *E. coli*, including *hypT*, *rclR* and *nemR* (Gray et al., 2013; Melnyk et al., 2015; Parker et al., 2013). Chlorine resistance and resistance to advanced oxidation products (AOPs) have been linked to the generalized stress response of *E. coli* (Cabiscol et al., 2000; Drazic et al., 2013; Du et al., 2015; Hillion and Antelmann 2015; Parker et al., 2013), as well as the production of heat shock proteins (Winter et al., 2008). It is estimated that at least 10% of the entire genome of *E. coli* is dedicated to the generalized stress-response (Landini et al., 2014), and this estimate does not include other adaptive stress responses such as the universal stress response, SOS response, and soxRS systems (Trastoy et al., 2018), which are involved in survival against nutrient deprivation, osmotic stress, and oxidative stress (including UV). The diversity of stress mechanisms in the microbial toolbox is astounding, and along with their ability to transfer genes horizontally,

microbes are empowered to evolve and adapt to virtually any environment. Collectively, this raises the public health specter that wastewater treatment may actually drive the natural co-selection and evolution of water treatment-resistance in microbes such as *E. coli*.

In fact, recent evidence suggests that some strains of *E. coli* are more resistant to water treatment than others, and include certain pathotypes such as extraintestinal pathogenic *E. coli* (ExPEC). Anastasi and colleagues (Anastasi et al., 2013), observed that various strains of *E. coli* differentially survived wastewater treatment processes (activated sludge, chlorination and UV irradiation), and that isolates surviving treatment were dominated by ExPEC pathotypes, such as urinary pathogenic *E. coli* (UPEC). In another study, Anastasia *et al.,* (Anastasi et al., 2010) observed that 59.5% of isolates surviving wastewater treatment possessed 1 or more virulence genes associated with UPEC (*papA*/*H*, *papE*/*F*, *papC*, *hlyA*, *cnf1*, and *iroN*). Based on virulence gene profiling, Adefisoye and Okoh (Adefisoye and Okoh, 2016) observed that >41.7% of *E. coli* isolates from treated wastewater effluents were considered as potential UPEC. Calhau *et al*. (2015), identified UPEC in wastewater effluents based on the presence of the virulence genes *iutA*, *papA*/*H*, and *sfa*, as well as various UPEC-associated pathogenicity islands, and observed a dominance of UPEC strains in hospital treated wastewaters, including sequence type (ST) 131 extended spectrum beta-lactamase (ESBL) producing strains. The observation that ExPEC dominates the population of *E. coli* in treated wastewater, led Paulshus and colleagues (2019) to suggest that these findings can only be explained by a 'higher-than-expected' level of infection in the community or that ExPEC naturally occur in wastewater. The common occurrence of emerging pathotypes of *E. coli*, such as ESBL-producing O25b-ST131strains in treated wastewater (Dolejska et al., 2011), elevates the concerns for public health.

UPEC are the most important cause of urinary tract infections (UTI) accounting for 65~75% of UTI cases (Flores-Mireles et al., 2015), resulting in >10 million physician visits each year in the U.S., and imposing an estimated $3.5 billion in healthcare costs (Flores-Mireles et al., 2015). Among the sequence types associated with clinical infections, the ST131 strain is of major public concern (Mathers et al., 2015). This sequence type has emerged on all continents and is associated with greater antibiotic resistance than other sequence types, and includes ESBL-producing strains (Mathers et al., 2015; Nicolas-Chanoine et al., 2014).

Unfortunately, no single virulence gene, or set of virulence genes, clearly define UPEC (Schreiber et al., 2017), raising questions about whether the strains observed in various wastewater studies outlined above may represent clinically-relevant UPEC. As such, our main objective was to determine whether the treatment resistant population of *E. coli* isolated from treated sewage/wastewater samples represented clinically-relevant UPECs. Collectively, our comparative whole genome data demonstrates that many of the water-treatment resistant strains found in treated wastewater represent clonally-identical and clinically-relevant UPEC strains, and include antibiotic resistant sequence types of global concern, such as the ST131 ESBL-producing UPEC.

2. Materials and methods

2.1 Bacterial strains

E. coli isolates for this study were obtained using two different approaches. One approach (Sample Set 1) employed the United States Environmental Protection Agency's (U.S. EPA) Alternate Test Procedure (ATP) (US EPA, 2010) for obtaining chlorine-stressed bacteria

from chlorinated sewage. This procedure incorporates a process step in which raw sewage is treated with chlorine bleach in the laboratory to reduce the viable *E. coli* concentration within the sewage by up ~ 4 $\log_{10}$, resulting in a chlorine-stressed and tolerant population. Briefly, raw sewage samples from 10 different sewage treatment plants in Alberta, Canada, were collected and sent to the Alberta Provincial Laboratory for Public Health (ProvLab) for analysis. Each raw sewage sample was treated with 3% sodium hypochlorite with a free chlorine residual of 0.3-0.5 ppm, and for a sufficient length of time to cause a ~4 $\log_{10}$ reduction in the culturable concentration of *E. coli*. Chlorine reactivity was neutralized by addition of a 10% solution of sodium thiosulfate. The ~4 $\log_{10}$ reduction in culturable *E. coli* levels was verified using parallel samples, with numbers estimated based on most probable numbers using a Colilert QuantiTray® system (IDEXX Laboratories, Inc.). Chlorine-treated wastewater samples were then used to inoculate either ColiTag® or lauryl trypticase broth/BCG media according to Method 9221.F in *Standard Methods for the Examination of Water and Wastewater* (American Public Health Association. et al., 1999). *E. coli* was isolated from ColiTag® or LTB/BG positive cultures by selective plating onto X-Gluc agar plates (Frampton et al., 1988) and incubating the plates at 44.5 ^{o}C for 24 h. Blue colonies were picked and streaked onto non-selective blood agar plates and incubated at 35 ^{o}C for 24 h. All presumptive wastewater *E. coli* isolates were confirmed as *E. coli* through comprehensive biochemical testing using a Vitek®2 Automated Bacterial Identification System (BioMerieux, St. Laurent, Quebec, Canada) according to the manufacturer's instructions. All isolates were stored in Tryptic Soy Broth (TSB) containing 50% skim milk at -80 ^{o}C. Individual isolates were subsequently thawed and cultured in TSB broth overnight at 37 ^{o}C prior to identification and characterization of UPEC isolates using molecular methods (see below).

158 The second approach for obtaining *E. coli* isolates (Sample Set 2) used partially-treated
159 and finished effluents from three municipal wastewater treatment plants (WWTPs) in the City of
160 Calgary as a source of treatment-resistant *E. coli*. Wastewater treatment configurations include
161 grit removal, primary clarification, activated sludge, secondary clarification and UV disinfection
162 (low or medium pressure at doses of 25-30mJ/cm^2 at peak flow), and with maximal treatment
163 capacities ranging from 140 ML/day to 1020 ML/day. Wastewater effluent samples were
164 processed by standard membrane filtration (100 mL) on X-Gluc agar plates, and blue isolates
165 were selected and inoculated into a 96-well plate containing 100uL 1X Luria-Bertani (LB) broth,
166 followed by incubation overnight without shaking at 37°C. For long-time storage of the plates,
167 50% LB-glycerol in a 1:1 ratio was added to each well and frozen at -80°C. A total of 1212
168 isolates were collected during the study, and a random collection of 261 isolates from this library
169 were selected for further analysis and confirmed as *E. coli* using a Vitek®2 Automated Bacterial
170 Identification System.
171
172 ***2.2 Analysis of UPEC-related virulence genes and molecular markers***
173 *E. coli* isolates were grown in TSB overnight at 37°C and their genomic DNA extracted from
174 the TSB culture using DNeasy Blood & Tissue kits (Qiagen, Toronto, Canada) according to the
175 manufacturer's instructions. *E. coli* isolates from Sample Set 1 (i.e., chlorinated sewage) were
176 analyzed by PCR against a panel of UPEC-related virulence genes and various other molecular
177 markers. A total of 376 chlorine-tolerant isolates from Sample Set 1 were tested against a panel
178 of two *E. coli*-related markers (*usp-IS30-flhDC* and *uidA*), two UPEC-related sequence type (ST)
179 markers (ST131 and O25b-ST131), five known UPEC-related virulence genes (described
180 below), and Shiga toxins 1 and 2 (*stx*1 and *stx*2). The *uspC-IS30-flhDC* marker was used to

exclude naturalized wastewater strains of *E. coli* from further analysis since these strains have been shown to dominate the population of treatment-resistant *E. coli* found in chlorinated sewage (Zhi et al., 2016). The *uidA* marker was used as a genetic confirmatory marker of all *E. coli* strains phenotypically-typed by the Vitek® Automated Bacterial Identification System. The presence of the ST131 and the O25b-ST131 markers were included in the panel since ST131 UPEC strains represent a dominantly emerging pandemic lineage of clinically-important UTI cases, and for which the O25b-ST131 sub-lineage is of particular concern (Clermont et al., 2009). The presence of five common UPEC-related virulence genes were also incorporated in the analysis, and included the genes *papC* (outer membrane usher protein), *sfa/foc* (S and F1C fimbriae), *fyuA* (outer membrane iron receptor), *chuA* (outer membrane heme receptor) and *iroN* (siderophore receptor). The *stx*1 and *stx*2 genes were also included in the analysis to determine if Shiga toxin-producing *E. coli* (STEC) were represented in the chlorine-tolerant *E. coli* population of treated sewage/wastewater. All primers used in this study are provided in **Table 1**. Purified genomic DNA (gDNA) was quantified using the Qubit fluorimeter (Thermo Fisher Scientific Inc.). All PCR reactions were performed on an ABI 2720 thermocycler (Applied Biosystems).

PCR protocols for the *uspC-IS30-flhDC* and *uidA* markers have been previously described by Zhi *et al*., (2016) and Taskin *et al*., (2011), respectively. For other PCR assays, the reactions consisted of 20 – 40 ng of gDNA template, 1X GoTaq mastermix (Promega), 10 µg/mL BSA and 500 nM of each primer. Cycling conditions for *pap*C and *sfa-foc* were: 95 ^{o}C, 2 min followed by 33 cycles of 30 s at 95 ^{o}C, 30 s at 63 ^{o}C and 45 s at 72 ^{o}C followed by a 7 min incubation at 72 ^{o}C. Cycling conditions for *fyuA*, *chuA* and *iroN* were: 95 ^{o}C for 2 min followed by 33 cycles of 30 s at 95 ^{o}C, 30 s at 63 ^{o}C and 1 min at 72 ^{o}C followed by a 7 min incubation at

72 °C. Cycling conditions for ST131marker were: 95 °C for 2 min followed by 35 cycles of 20 s at 95 °C, 20 s at 57 °C and 40 s at 72 °C followed by a 7 min incubation at 72 °C. Cycling conditions for O25b-ST131 marker were: 95 °C for 4 min followed by 30 cycles of 5 s at 94 °C, 10 s at 65 °C followed by a 5 min incubation at 72 °C. All PCR products were run on 1.5% agarose gels and photographed on an ImageQuant LAS 4000 (GE Healthcare Life Sciences).

In some experiments, clonal-relatedness between strains was determined by $(GTG)_5$ rep-PCR using the methods of Korvin *et al.,* (2014). Cycling conditions for rep-PCR experiments were: 95 °C for 2 min followed by 35 cycles of 1 min at 95 °C, 1 min at 50 °C, 8 min at 65 °C, followed by an 8 min incubation at 65 °C. The rep-PCR amplicons were run on a QIAxcel High Resolution Cartridge (Qiagen) using the OM1200 program with a 20 second injection time. A 15bp/10kb alignment marker and 250bp/8kb size marker were used for determining band sizes produced by rep-PCR.

2.3 Phenotypic characterization of antibiotic resistance in W-UPEC

Wastewater UPEC (W-UPEC) strains were phenotypically tested for antibiotic susceptibility using the VITEK®2 Antibiotic Susceptibility Testing (AST) cards: AST-N391 and AST-GN98 (BioMerieux, St. Laurent, Quebec, Canada) according to the manufacturer's instructions. Resistance to 27 individual antibiotics as well as an ESBL-specific combination were assessed in the analysis. Eight antibiotics were similar between the two VITEK®2 AST cards and acted as a quality control for resistance testing among different cards. For the purposes of this study, strains classified as having intermediate resistance or complete resistance were considered as resistant to a particular antibiotic. Previously isolated and purified presumptive W-UPEC isolates were streaked on tryptic soy agar plates and incubated for 24

hours at 37°C. After incubation, colonies were suspended in 3ml of 0.45% saline and adjusted to an optical density of 0.5-0.63 McFarland standard by DensiCHEK™Plus (BioMerieux, St. Laurent, Quebec, Canada). One-hundred forty-five µl of the adjusted suspensions (from each isolate) was transferred to a clear plastic test tube containing another 3 ml of 0.45% saline. The prepared tubes (2 for each isolate) were loaded into a VITEK® 2 Compact (BioMerieux) along with the AST-GN98 and AST-N391 cards (one card per suspension). Both positive and negative quality control samples were included in the analysis. Positive controls were *E. coli* ATCC 25912, *E. coli* ATCC 35218, and *Pseudomonas aeruginosa* ATCC 27853 (as per recommended) and the negative control was 0.45% saline.

2.4 Whole genome sequencing of presumptive wastewater UPEC (W-UPEC) strains and sequence assembly

Genomic DNA of each presumptive wastewater UPEC (W-UPEC) isolate from Sample Set 1, as identified from the PCR panel screen, was sent to Genome Quebec (Montreal, Canada) for sequencing using an Illumina HiSeq X platform (Illumina) with paired-end 150-nucleotide reads. Trimmomatic Version 0.38 was used to trim the low-quality reads of each genome with the following parameters: SLIDINGWINDOW=4:15, LEADING=3, TRAILING=3 MINLEN=36. *De novo* assembly was performed using SPAdes Version 3.9.1 (Bankevich et al., 2012) with '--careful' and '-k 21,33,55,77' options. The SPAdes-assembled contigs smaller than 1000 bp were removed from downstream analysis.

2.5 Phylogenetic analysis and multilocus sequencing typing (MLST) of W-UPEC strains

Whole genome phylogenetic analysis was performed using REALPHY 1.12 (Bertels et al., 2014). A maximum likelihood phylogenetic tree was generated based on the whole genome sequence (WGS) of presumptive W-UPEC strains from this study (n=24), 29 clinical UPEC (C-UPEC) strains, 6 cryptic *E. coli*, as well as 27 *E. coli* strains with known phylogroups (Sims and Kim, 2011). The phylogenetic tree was visualized using interactive Tree of Life (iTOL) (Letunic and Bork 2019) and was rooted against the *E. albertii* genome. Information pertaining to the bacterial strains used in this analysis can be found in **Supplementary Table S1**. MLST was also performed on all presumptive W-UPEC strains using MLST-2.0 with *Escherichia coli#1* scheme (Larsen et al., 2012). The serotype of W-UPEC strains and their genetically similar C-UPEC counterparts were predicted using SerotypeFinder 2.0 (Joensen et al., 2015).

2.6 Minimum spanning tree of W-UPEC and clinical UPEC (C-UPEC) strains based on core genome SNP analysis

Comparative core genome SNP analysis of the presumptive W-UPEC (n=24) and 328 C-UPEC strains obtained from NCBI databases (see **Supplementary Table S1** for list of isolates) was performed using REALPHY 1.12 (Bertels et al., 2014) to build a minimum spanning tree. In brief, whole genome sequences of all 352 strains were loaded into the local REALPHY input folder. One of the *E. coli* strains was randomly selected as a reference sequence. The genome sequences of all other strains were then mapped to the reference genome for SNP assessment. The core genome SNPs identified by REALPHY were analyzed by Phyloviz (Francisco et al., 2012) to generate a minimum spanning tree based on SNPs differences between the various strains. In addition, the core genome sequence alignments of all 352 strains were also used to generate a core genome SNP distance matrix using the R package Biostrings (Pagès et al., 2019).

2.7 Pairwise whole genome comparison of W-UPEC and clinical UPEC strains

Pairwise whole genome comparisons were performed on W-UPEC and closely related C-UPEC strains (as identified in minimum spanning tree and core genome similarity analysis described above). REALPHY 1.12 (Bertels et al., 2014) was used to determine the pairwise genome similarity with default parameters. The genome similarity between different strains was then calculated based on the core genome size result of the REALPHY analysis.

An additional pairwise whole genome comparison was performed among 30 *E. coli* O157:H7 strains (**Supplementary Table S1**) to determine the level of whole genome similarity that exists between strains of the same bacterial species that cause a similar pathology (i.e., gastrointestinal illness). Since *E. coli* O157:H7 strains are genetically more homogeneous than UPEC strains, the comparative genome analysis of *E. coli* O157:H7 strains established an upper threshold of whole genome intra-pathotype genetic variability that is associated with a particular disease phenotype. All *E. coli* O157:H7 strains were downloaded from the NCBI genome database. The upper median threshold of whole genome intra-pathotype variability in *E. coli* O157:H7 strains was then used as a threshold criterion for determining whole genome relatedness between W-UPEC and C-UPEC that would likely result in a similar pathology (i.e., urinary tract infection).

2.8 Whole genome analysis of virulence genes and antibiotic resistance genes in W-UPEC

W-UPEC strains and selected C-UPEC strains from NCBI were analyzed for virulence gene and antibiotic resistance gene composition. For the virulence gene search, a local blast database of W-UPEC and C-UPEC strains was built using BLAST 2.5.0 , after which ~2,600 virulence factors in the core dataset of the Virulence Factor Database (VFDB) (Chen et al., 2016) were

295 compared to the genome database to identify virulence gene matches in the UPEC genomes. The
296 minimum query coverage was set to 90% with a percent identity of ≥80%. Antibiotic resistance
297 genes in *E. coli* genomes were identified using the Resistance Gene Identifier (RGI) of the
298 Comprehensive Antibiotic Resistance Database (CARD) (Jia et al., 2017).
299
300
301 **3. Results**
302 ***3.1 Identification of presumptive W-UPEC from chlorine-treated sewage***
303 A library consisting of 376 *E. coli* isolates obtained from chlorinated sewage samples
304 (Sample Set 1) was screened by PCR against a panel of *E. coli* related markers, as well as UPEC
305 and STEC-related virulence genes. One-hundred forty-five (38%) isolates within this chlorine-
306 tolerant library contained the *uspC-IS30-flhDC* marker and were therefore considered as
307 naturalized wastewater strains of *E. coli*. Nine isolates that were biochemically-confirmed using
308 the automated VITEK®2 Automated Bacterial Identification system lacked the *uidA* gene.
309 Strains possessing the *uspC-IS30-flhDC* marker or lacking the *uidA* gene (154 of the original 376
310 isolates) were not included in the subsequent virulence gene analysis.
311 The remaining 222 chlorine-tolerant isolates were tested by PCR for the virulence genes
312 *papC*, *sfa/foc*, *fyuA*, *chuA*, *iroN*, *stx*1 and *stx*2 and were screened for ST131 and O25b-ST131
313 markers. None of the 222 chlorine-tolerant isolates possessed *stx*1 or *stx*2. Of the 222 isolates,
314 129 (or 34% of the original 376 chlorine-tolerant *E. coli*) did not possess any of the UPEC-
315 related virulence genes. Ninety-three isolates (25% of the original 376 chlorine-tolerant *E. coli*)
316 contained one or more UPEC-related virulence genes. Of these 93 isolates, 51 isolates possessed
317 one of the UPEC-related virulence genes, 18 possessed two virulence genes, 14 possessed three

318 virulence genes, 4 possessed four virulence genes, and 6 possessed five virulence genes. A
319 breakdown of the occurrence of each of these virulence genes is provided in **Figure 1**. Isolates
320 possessing at least 3 of the initial 5 virulence screen genes (*papC*, *sfa/foc*, *fyuA*, *chuA*, and *iroN*)
321 were subjected to whole genome sequencing (n=24 [or 6% of the original 376 chlorine-tolerant
322 isolates]). These 24 isolates were considered as presumptive W-UPEC. Five of these 24
323 presumptive W-UPEC isolates (WU 1155, 1030, 1266, 1036 and 1265) were also positive for the
324 ST131 PCR marker and all five were of the O25b-ST131 lineage.
325
326 ***3.2 Whole genome sequence analysis of presumptive W-UPEC from chlorinated sewage***
327 Genome characteristics of the 24 presumptive W-UPEC strains are summarised in **Table 2**.
328 Genome sizes ranged from 4.8 Mbp to 5.3 Mbp with an average G+C content of 50.5%. The
329 mean number of putative coding genes within these presumptive W-UPEC wastewater strains
330 was approximately 4762 and sequence coverage ranged from 332x to 569x.
331 Maximum likelihood phylogenetic analysis was used to compare whole genome sequences of
332 the 24 chlorine-resistant presumptive W-UPEC strains against 29 clinical UPEC (C-UPEC)
333 genomes from NCBI, as well *E. coli* strains with known phylogroups (Beatson et al., 2015;
334 Tenaillon et al., 2010; Touchon et al., 2009; Zhang and Lin 2012). Presumptive W-UPEC strains
335 scattered within phylogroups B1, B2, and D (**Figure 2**). In total, 21 of the W-UPEC strains
336 clustered within phylogroup B2, while only one and two strains grouped within phylogroups D
337 and B1, respectively. A similar phylogrouping pattern was also observed in clinical UPEC
338 strains with the majority of the clinical strains partitioning into phylogroup B2, followed by
339 phylogroup B1 and A.

Sequence types (ST) of the 24 chlorine-resistant presumptive W-UPEC was determined by MLST sequence analysis (**Figure 2**). Four strains were designated to unknown STs. For the remaining 20 strains, 11 STs were identified, among which the most predominant sequence type was ST131 (20.8%, 5/24), followed by ST95 (16.7%, 4/24), ST127 (8.3%, 2/24), and ST640 (8.3%, 2/24). The other STs (ST6331, ST349, ST80, ST625, ST538, ST357, ST117) were each represented by only one strain (4.2%, 1/24 each). Data also confirmed that the five ST131 strains identified by PCR were also designated as ST131 by whole genome sequencing analysis and also belonged to the O25b-ST131 sub-type, a current global pandemic strain (Can et al., 2015; Mathers et al., 2015; Nicolas-Chanoine et al., 2014). All of the chlorine-tolerant O25b-ST131 isolates originated from the same WWTP.

In total, the 24 presumptive W-UPEC strains originated from eight of ten different WWTPs surveyed across Alberta, Canada (**Figure 2**), suggesting that chlorine-tolerance among W-UPEC pathotypes is widely distributed. Remarkably, the chlorine-resistant population of presumptive W-UPECs from chlorinated sewage from one WWTP was shown to be comprised of 6 different sequence types, all of which are known to be clinically important, including the most common clinical UPEC sequence types ST131, ST95, and ST127 [**Figure 2**] (Toval et al., 2014; Yun et al., 2015). This suggests that a variety of W-UPEC strains appear to display tolerance to chlorine. Conversely, none of the chlorine-resistant strains represented in our chlorine-treated library were shown to contain Shiga toxin virulence genes (*stx*1 or *stx*2).

3.3 Genomic similarity between W-UPEC and C-UPEC strains

A core genome similarity comparison of the 24 presumptive W-UPEC isolates against 328 C-UPEC strains in NCBI was performed to further assess whether any of the chlorine-tolerant

363 W-UPEC strains actually represented clinically-important UPEC strains. A minimum spanning
364 tree was built based on core genome SNPs differences (**Supplementary Figure S1**) for
365 visualization of relatedness between W-UPEC and C-UPEC. Several W-UPEC strains clustered
366 closely with C-UPEC strains, differing by only a few SNPs across the core genome. For
367 example, the W-UPEC strain, WU1036, and the C-UPEC strain, U308, differed from each other
368 by only one SNP across a 0.4 Mbp core genome. This high genome similarity was also observed
369 among the W-UPEC strains, WU1025, and the C-UPEC strain U64 (4 SNPs differences). Other
370 W-UPEC strains, such as WU1157, differed from the most closely related C-UPEC strain, U60,
371 by only 4 SNPs. Similarly, the W-UPEC strain, WU1214, differed from the C-UPEC strain,
372 U341, by only 7 SNPs across the core genome.

373 In addition to the spanning tree analysis for determining core genome similarity, a core
374 genome SNP distance matrix of all 352 strains was generated (**Supplementary Table S2**). This
375 distance matrix was used to further characterize all pairs of W-UPEC and C-UPEC that had high
376 core genome similarity. A core genome SNP difference of 50 was used as a threshold to identify
377 genetic matches between W-UPEC and C-UPEC strains. As shown in **Table 3**, fifteen W-UPEC
378 strains were found to have C-UPEC matches that met this criterion. These fifteen W-UPEC
379 included the five ST131 isolates, the four ST95 isolates, the two ST127 isolates, one ST80
380 isolate, and two isolates with unknown sequence types.

381 Serotype prediction from genome sequences was subsequently done to evaluate similarity
382 between W-UPEC and their most closely related C-UPEC strains (as identified in the analysis
383 above) (**Table 4**). The predicted serotype of all W-UPEC strains matched the serotype
384 predictions of their most closely related C-UPEC counterparts (**Table 4**). All O25b-ST131 PCR
385 positive W-UPEC strains were predicted to have an O25:H4 serotype by whole genome

386 sequencing (**Table 4**) and matched the predictions for their mostly closely related clinical UPEC
387 counterparts.

388 Since UPEC are a genetically diverse group of pathogenic *E. coli*, we sought to further
389 evaluate and understand the overall intra-pathotype variation between chlorine-tolerant W-UPEC
390 and C-UPEC at the whole genome level. To root this comparison, we performed a pairwise
391 genome alignment against 30 clinical strains of *E. coli* 0157:H7 - a genetically more
392 homogeneous pathotype (STEC/EHEC) compared to UPEC. For this analysis, 435 paired
393 genome comparisons were performed across the 30 *E. coli* 0157:H7 genomes, and for which
394 whole genome similarity between any two representative diarrheal disease-causing *E. coli*
395 0157:H7 isolates ranged from 79.67% to 99.95%, with mean and median genetic similarity
396 estimates of 94.24% and 96.03%, respectively. We therefore selected the higher median
397 similarity value of 96.03% as the upper threshold of pathotype similarity to determine if any two
398 W-UPEC and C-UPEC strains were indeed closely related at the whole genome level and likely
399 to cause the same phenotypic disease pattern in humans. In total, 14 of the W-UPEC strains had
400 at least one closely related clinical UPEC strain that passed this criterion (**Table 3**). Whole
401 genome similarity of the W-UPEC and C-UPEC matches ranged from 96.05% to 99.49%,
402 suggesting an extremely high degree of genetic similarity at the whole genome level between
403 each of the W-UPEC and their most closed related C-UPEC strain. The genome of W-UPEC
404 strain WU1214 (an ST80 isolate) displayed the highest degree of similarity to its C-UPEC
405 counterpart (99.49%). Interestingly, and in agreement with the pairwise whole genome
406 comparison analysis, the O25b-ST131 PCR assay demonstrated that all five ST131 W-UPEC
407 strains belonged to an *E. coli* clonal group frequently isolated from urine samples of patients
408 with UTI (Clermont et al., 2009; Coque et al., 2008; Gibreel et al., 2012a; Vimont et al., 2012).

409

410 ***3.4 Comparative assessment of virulence gene composition in W-UPEC and C-UPEC***
411 ***strains***

412 The high degree of genetic relatedness between the W-UPEC strains and their C-UPEC
413 counterparts at the core and whole genome level was also reflected in virulence gene analysis.
414 Virulence gene composition was compared between each of the 14 W-UPEC strains identified
415 above and their most closely related C-UPEC counterparts (a maximum of four C-UPEC
416 representative strains were included for each pair). The results are provided in **Supplementary**
417 **Table S3.** W-UPEC strains WU965, WU1752, and WU1033 possessed the exact same
418 complement of virulence genes as their most closely related C-UPEC strains (**Supplementary**
419 **Table S3**). Based on whole genome searching, the five O25b-ST131 W-UPEC strains (WU1030,
420 WU1036, WU1155, WU1265, and WU1266) contained 78 different virulence genes that were
421 also found in their genetically related C-UPEC counterparts. Upon initial assessment using
422 curated whole genome data, the W-UPEC strains appeared to lack two virulence genes compared
423 to their most closely related C-UPEC strains, including the *iutA* gene (which encodes for
424 aerobactin receptor protein) and the *iucD* virulence gene (an aerobactin biosynthesis protein).
425 However, when these genes were blasted against raw genome data (including the short contigs),
426 the *iutA* gene and a partial *iucD* gene were found in short contig sequences, suggesting that the
427 five ST-131 W-UPEC strains actually possessed the same 80 virulence gene complement as the
428 C-UPEC strains. In addition to possessing these 80 virulence genes, the ST131 W-UPEC strains
429 also possessed the *sfaX* and *papX* virulence genes which were not found in their C-UPEC
430 counterparts. Interestingly, one W-UPEC ST131 strain (WU1266) also possessed the *gspC-M*
431 gene cluster associated with the Type 2 Secretion System, which was not observed in their most

closely related C-UPEC strains (albeit this gene cluster was observed in other W-UPEC and C-UPEC strains).

The additional effort of searching for virulence genes in raw genome data revealed the presence of the same complement of virulence genes between the majority of the W-UPEC strains and their most closely related C-UPEC counterparts (**Supplementary Table S3)**. This was true for all ST131 strains mentioned above, as well as W-UPEC strains WU1025, WU1151, and WU1214. In the case of the W-UPEC strain WU1214, although it carried all virulence genes similar to the related C-UPEC isolates (U57 and U341), the *papB* gene had only an 81% sequence identity. The exceptions to the virulence gene similarity analysis were: i) a W-UPEC strain (WU1157), which did not possess the *hlyCABD* gene operon as well as the *cnf1* gene compared to their most closely-related C-UPEC strains (U115, U203, U334, and U60); ii) a W-UPEC strain (WU1274), which did not possess the type 2 secretion system (T2SS) coding genes *gspCDEFGHIJKLM* compared to its closely related C-UPEC strains (U23, U67, U294, and U295); and c) a W-UPEC strain (WU3707), in which the *iroBCDEN* operon was absent when compared to its most closely related C-UPEC strains (U48, U139, U162, and U289).

3.5 Antibiotic resistance in W-UPEC strains

W-UPEC strains were also shown to possess a plethora of antibiotic resistance genes, as identified through the CARD database (Jia et al., 2017) (**Supplementary Table S4**). Sixty antibiotic resistance genes were identified in the ST131 W-UPEC isolates, and for which these isolates possessed the greatest number of antibiotic resistance genes compared to the other W-UPEC strains. Among the 60 antibiotic resistance genes identified, $bla_{CTX\text{-}M\text{-}15}$, $bla_{OXA\text{-}1}$, *ampC*, *aac(6')-Ib-cr*, *aac(3')-II*, *ptsI*, *parC* , *gyrA*, *catB3*, and *vagC* were only present in ST131 W-

455 UPEC strains and not the other W-UPEC strains. The presence of *bla*$_{CTX-M-15}$, *bla*$_{OXA-1}$, and the
456 *ampC* gene suggested that these isolates may represent ESBL-producing *E. coli*. The five ST131
457 W-UPEC strains also carried two genes encoding aminoglycoside-modifying enzymes (*AAC(3)-*
458 *IIb* and *AAC(6')-IIa*), representing the most common mechanisms for resistance against
459 aminoglycosides (Ramirez and Tolmasky, 2010). The five ST131 W-UPEC strains also
460 possessed four genes that confer resistance to fosfomycin, including *glpT*, *ptsI*, *uhpT,* and *mdtG*
461 (Ballestero-Tellez et al., 2017; Ohkoshi et al., 2017), among which *ptsI* was only found in the
462 five ST131 strains. Two additional antibiotic resistance genes, found only in the five ST131
463 strains, included *parC* and *gyrA* – genes involved in fluoroquinolone resistance (Johnning et al.,
464 2015). The chloramphenicol resistance gene *catB3* (Bunny et al., 1995) was also exclusively
465 present in the five ST131 strains.

466 The antibiotic resistance gene profiles of the remaining 19 W-UPEC isolates
467 (**Supplementary Table S4**) were quite similar to each other and contained upwards of 49
468 antibiotic resistance genes. These genes are known to confer resistance to various classes of
469 antibiotics including β-lactams, fluoroquinolones, fosfomycin, phenicol antibiotics, tetracycline,
470 ansamycins, macrolides, nitrofurantoins, aminoglycosides, aminocoumarin, peptide antibiotics,
471 and lincosamides.

472 Although the 24 W-UPEC from chlorine-treated sewage were shown to possess an
473 abundance of antibiotic resistance genes, phenotypically, only 11 of the isolates displayed
474 antibiotic resistance at clinical minimum inhibitory concentration breakpoints as determined by
475 antibiotic susceptibility testing using VITEK®2 AST-N391 and AST-GN98 cards (**Table 5**).
476 The five ST131 strains, were shown to be clinically-resistant to 16 antibiotics from five different
477 antimicrobial classes, including the penicillins and β-lactam/β-lactamase inhibitor combinations,

cephalosporins, aminoglycosides, and quinolones and fluoroquinolones. These isolates were phenotypically-classified as ESBLs. Antibiotic resistance in the other six W-UPEC were as follows: WU1157 (an ST127 strain) was only resistant to cefalexin (cephalosporins); WU3598 (an unknown ST strain) was resistant to ampicillin only; WU3165 (an ST117 strain) was resistant to the quinolone antibiotics enrofloxacin and marbofloxacin, and chloramphenicol (phenicols); WU2356 (an ST6331 strain) was resistant to chloramphenicol (phenicols); WU1630 (an ST640 strain) was resistant to cefalexin (cephalosporins) and chloramphenicol (phenicols); and WU664 (an ST538 strain) was resistant to ampicillin and cephalosporin antibiotics (cefpodoxime, cefovecin, ceftiofur, cefazolin, cefalexin, and cefixime), as well as β-lactam /beta-lactamase inhibitors including amoxicillin/clavulanic acid and piperacillin/tazobactam.

Collectively, the data confirms that W-UPEC isolates surviving chlorination appear to represent clinically-important UPEC pathotypes of public health concern, and include the emerging antibiotic resistant ESBL strains associated with the O25b-ST131 lineage and other public health strains of global public health significance (i.e., ST95, ST127, ST640).

3.6 Characterization of W-UPEC in full scale wastewater effluents

The screening protocols developed for identification and characterization of clinically-important UPEC from chlorinated sewage samples was subsequently used to identify and isolate UPEC from full scale WWTPs from the City of Calgary (**Supplementary Figure S1**). A microbial library of presumptive *E. coli* isolates from partially-treated (e.g., secondary-treated) and UV-disinfected wastewater effluents was generated, and represented the 'survivor population' of *E. coli* from 3 different WWTPs. These samples were collected from different days throughout the year. This collection of *E. coli* isolates was negatively selected against

501 naturalized wastewater strains (i.e., those possessing the *uspC-IS30-flhDC* marker). Of the 261
502 *E. coli* isolates randomly selected for analysis (from 1212 isolates collected), only 13 isolates
503 (5%) possessed the *uspC-IS30-flhDC* marker and were therefore considered as naturalized
504 wastewater *E. coli*. The remaining 248 *E. coli* isolates were subsequently tested for the presence
505 of the *uidA* gene, five UPEC-related virulence genes (*papC*, *sfa/foc*, *fyuA* , *chuA*, and *iroN*) as
506 well as the ST131, O25b-ST131, *stx*1 and *stx*2 PCR markers.

507 One-hundred and fifty-five isolates possessed one or more of the UPEC-related virulence
508 genes. Nineteen isolates possessed 3 or more virulence genes. Of these 19 isolates, however,
509 only 1 was ST131 positive. Fifteen additional ST131 positive isolates were also identified and
510 which possessed at least one other UPEC-related virulence gene. The 19 *E. coli* isolates
511 containing more than 3 virulence genes and the 15 ST-131 positive isolates were further
512 characterized genetically (n=34).

513 Rep-PCR was used to look at clonal relatedness between the whole genome characterized W-
514 UPEC isolates from chlorinated sewage (i.e., isolates from Sample Set 1) and isolates screened
515 from the effluents of full-scale wastewater treatment plants (Sample Set 2). In total, 16 unique
516 rep-PCR fingerprints were obtained through this comparison. **Table 6** lists all rep-PCR patterns
517 represented by more than two isolates. Pattern III was the most abundant clonal group, and
518 contained the five ST131 W-UPEC strains isolated from chlorinated sewage (Sample Set 1), as
519 well as seven ST131 strains isolated from full scale wastewater-treated effluents (Sample Set 2).
520 Interestingly, all *E. coli* strains belonging to rep-PCR Pattern III (i.e., those collected from
521 chlorinated sewage or full scale treated wastewater) were also PCR positive for the O25b-ST131
522 sub-lineage marker (**Figure 3**). The remaining nine ST131 strains obtained from full-scale
523 treated wastewater effluents formed the second most abundant clonal group (Pattern IV). Rep-

524 PCR Patterns III and IV encompassed all ST131 strains (**Figure 3**). Interestingly, six of the
525 ST131 strains (2F5, 2F6, 4B8, 1F2A, 5A5, and 3C4) were all isolated from the final UV-
526 disinfected effluents from two different WWTPs, and on different sampling days, suggesting that
527 this clonal population of ST131W-UPEC appear to routinely survive the wastewater treatment
528 train among different WWTPs throughout the year. Similar to our findings observed in Sample
529 Set 1 (i.e., chlorinated sewage), no *stx1* or *stx2* positive *E. coli* isolates were observed in treated
530 effluents from the WWTPs.

531 In total, seven different rep-PCR patterns were represented by at least two W-UPEC strains
532 from either the chlorinated sewage samples (Sample Set 1) or from full scale effluents (Sample
533 Set 2). Strain 3B7, isolated from a secondary-treated effluent at full scale from a WWTP, was
534 clonally-related to W-UPEC strains WU1025 and WU1157 (ST127) isolated from chlorinated
535 sewage samples. Strains 2B8, 3H3, 2E3, and 4F6, were isolated from secondary-treated
536 wastewater effluents, and were clonally-related to the W-UPEC strain WU965 isolated from
537 chlorinated sewage samples (unknown ST). Strains 3B9 and 4G1 isolated from secondary-treated
538 wastewater effluents from two different WWTPs in Calgary were clonally-related to the W-
539 UPEC strains WU1274 (ST95) and WU1151 (ST95) from chlorinated sewage. In total, 14 out of
540 34 strains isolated from full-scale WWTP effluents were clonally similar to the whole genome
541 characterized W-UPEC strains found in chlorinated sewage samples, suggesting that W-UPEC
542 found in full-scale treated wastewater effluents (Sample Set 2) also represented clinically-
543 important strains.

544 To further validate the clinical importance of these isolates found in treated wastewater
545 effluents (Sample Set 2), antibiotic resistance testing was performed (**Table 5**). Four of these W-
546 UPEC strains were sensitive to all antibiotics, while the other 30 strains showed clinical

547 resistance to one or more antibiotics. Twenty-six of the 34 isolates (76%) were resistant to more
548 than 3 antibiotics. Remarkably, one isolate, 4C1, collected from a secondary-treated wastewater
549 effluent sample, represented an ST131 strain that was clinically-resistant to 19 different
550 antibiotics. Of the sixteen ST131 isolates obtained from wastewater effluents, 14 (88%) were
551 resistant to more than 4 antibiotics. W-UPEC isolates 2F5 and 2F6 represented two O25b-
552 ST131 isolates that: a) were ESBLs; b) had identical clinical antibiotic resistance patterns (15
553 different antibiotics); c) possessed resistance to the same 12 antibiotics that O25b-ST131 strains
554 from chlorinated samples were resistant to; d) had identical rep-PCR patterns to each other, and
555 for which the rep-PCR patterns were virtually identical to the O25b-ST131 isolates obtained
556 from chlorinated sewage; and e) originated from two different WWTPs in Calgary and for which
557 the point of origin was different than the O25b-ST131 strains isolated from chlorinated samples
558 coming from the third WWTP in Calgary. The data suggest that antibiotic resistant clonal O25b-
559 ST131 strains of UPEC can be routinely found in finished treated effluents of all of Calgary's
560 WWTPs.

561 In total, 7/35 isolates collected from treated-wastewater samples were considered as ESBLs
562 by phenotypic resistance testing. Five of these belonged to the ST131 lineage (4B8, 2F5, 2F6,
563 4C1, 5B5) and two to non-ST131 lineages (4C8, 1E1A).

564

565 **4. Discussion**

566 Collectively, the data presented in this study provide compelling evidence that clinically -
567 important pathotypes of *E. coli* comprise a major proportion of the *E. coli* population surviving
568 wastewater treatment – a finding similar to that reported in other research studies (Adefisoye and
569 Okoh 2016; Anastasi et al., 2010; Boczek et al., 2007; Ebomah et al., 2018). In two separate

570 studies by Anastasi and colleagues (2010, 2013), certain strains of *E. coli* were shown to have
571 enhanced survivability to wastewater treatment processes, including chlorination and UV
572 irradiation, and that isolates surviving treatment were dominated by UPEC pathotypes. Likewise,
573 Adefisoye and Okoh (2016) and Ebomah *et al*. (2018) observed that >40% of *E. coli* isolates
574 from treated wastewater effluents, and from a beach impacted by wastewater, were deemed to be
575 ExPEC. Calhau *et al*. (2015), observed a dominance of UPEC-related strains in hospital-treated
576 wastewaters, including ST131 ESBL strains. In the present study 94/377 (25%) isolates from
577 chlorinated sewage possessed one or more of the UPEC-related virulence genes. Similar to the
578 findings described by Calhau *et al*. (2015) we also identified a predominance of ST131 ESBL
579 strains in both chlorinated sewage and in treated wastewater effluents, and for which these
580 isolates were obtained from several wastewater treatment plants across Alberta, Canada.
581 Of 24 wastewater UPEC isolates subjected to whole genome sequencing in the present study,
582 11 sequence types were represented, and dominated by ST131, followed by ST95, ST127, and
583 ST640. Other STs included ST6331, ST349, ST80, ST625, ST538, ST357, and ST117. The
584 majority of these sequence types have also been observed in clinical UTI cases (15, 16, 56). In a
585 U.S. study by Yamaji *et al*. (2018) the most common STs among 233 clinical UPEC strains were
586 found to be ST95, ST127, ST73, ST69, and ST131. In a British study, the most common
587 sequence types were ST73, ST131, ST69, ST95, ST10, and ST127 (Gibreel et al., 2012b). The
588 relative abundance and diversity of UPEC sequence types observed in treated wastewater
589 effluents and chlorinated sewage compared to other *E. coli* pathotypes (i.e., STEC), suggests that
590 UPEC may be more adapted to survival and persistence in environmental water matrices
591 compared to other pathotypes. Anastasi and colleagues (2013) noted an absence of STEC
592 representation in a library of *E. coli* isolates collected from treated wastewater effluents. Franz *et*

593 *al*. (2015) also observed a lack of representation of STEC in surface and wastewater libraries of
594 *E. coli*, but found that gastrointestinal pathotypes included mostly enteroaggregative *E. coli*
595 (EAEC). Similar to that reported by Anastasi *et al*. (2013), we did not find any *stx* positive *E.*
596 *coli* isolates in our chlorinated sewage or wastewater effluent libraries.

597 Although previous research supports the idea that UPEC may be particularly adapted to
598 survive wastewater treatment, a major challenge rests with the fact that no single virulence gene,
599 or designated set, can be used to conclusively identify this genetically heterogeneous group of
600 pathogenic *E. coli* (Barber et al., 2016; Schreiber et al., 2017). In fact, ExPEC appear to have
601 exploited an array of molecular strategies and redundant cellular mechanisms to attach and
602 colonize to host cell surfaces, and have acquired diverse toxin production systems and invasion
603 strategies that allow it to diversify and exploit multiple niches for survival and growth (i.e., gut,
604 urethra, bladder, kidney, and blood) (Barber et al., 2016; Nielubowicz and Mobley, 2010).
605 Indeed, many of the so-called virulence genes described from UPEC are also found in
606 commensal strains of *E. coli* (Cyoia et al., 2015; Qin et al., 2013), and for which these genes
607 facilitate adhesion and colonization of the gut –i.e., a general prerequisite for survival in the gut
608 for both commensals and pathogens. Consequently, virulence gene profiling among clinical
609 UPEC strains has been shown to be highly variable (Kanamaru et al., 2003; Mao et al., 2012;
610 Marrs et al., 2005; Marrs et al., 2002; Spurbeck et al., 2012; Vigil et al., 2011; Yun et al., 2014).
611 From a human clinical perspective, UTI symptomology in clinically-ill patients provide the
612 necessary and accessory evidence for labelling a clinical isolate as a UPEC strain. Likewise,
613 bloodborne infections caused by systemic ExPEC, result in a range of clinical symptoms that
614 support culture confirmation of *E. coli* isolated from the blood. Virulence gene profiling (and
615 sequence typing) in these cultured isolates simply confirms the pathogenic nature of the isolates.

Unfortunately, the lack of a clearly defined set of virulence genes makes it difficult to predict and assess the true pathogenic potential of presumptive strains found in environmental samples. This has led to some uncertainty regarding the public health risks posed when these presumptive UPEC isolates are observed in treated wastewater samples. Another challenge rests with the fact that animal UPEC often have similar virulence profiles as human clinical UPEC, suggesting that animal reservoirs may also contribute to the occurrence of presumptive UPEC in environmental samples but for which the true zoonotic potential of these animal strains still remains somewhat unclear (Singer, 2015).

As such, this study provides very important evidence validating the genomic relatedness between ExPEC strains in treated sewage/wastewater and strains causing clinical disease in humans. Furthermore, it supports the emerging hypothesis that these pathotypes may be adapted to survive water treatment. Our evidence for clinical relevance is based on a number of comparative genomic approaches, including SNP core genome analysis, pairwise whole genome sequence comparisons, virulence gene composition, antibiotic-resistance gene composition and phenotypic antibiotic resistance profiling. Across the core genome (~0.4 MB analyzed), W-UPEC varied from their most closely related C-UPEC counterpart by only a few SNPs (in one case, only 1 SNP differed). Whole genome pairwise sequence-analysis demonstrated that most W-UPEC strains had >96% similarity to one or more C-UPEC strains found in NCBI. Of all the W-UPEC strains examined, WU1214 had the highest similarity to its C-UPEC counterpart (U341), with 99.49% similarity at the whole genome level. In most cases, virulence gene profiling demonstrated a similar complement of virulence genes between a particular W-UPEC strain and their most closely related C-UPEC counterparts found in NCBI databases. For example, based on whole genome sequencing all O25b-ST131 W-UPEC isolates collected from

chlorinated sewage possessed the same set of 80 virulence genes that their most closely related C-UPEC counterparts possessed. These W-UPEC strains also contained 60 different antibiotic resistance genes, and were found to be phenotypically-resistant to 16 different antibiotics at clinical minimum inhibitory concentration breakpoints. Genetically and phenotypically, the O25b-ST131 isolates from chlorinated sewage were classified as ESBLs, and were genetically similar to the O25b-ST131 strains found in the treated effluents of three different WWTPs in Calgary, based on rep-PCR and virulence gene profiling. The observation that these O25b-ST131 strains were found in multiple treatment plants and at different times supports the idea that these strains may be evolutionarily adapted to survive wastewater treatment compared to other *E. coli* pathotypes.

Several other sequence types of clinical importance were also identified and characterized in the chlorinated sewage population, including those associated with ST95 and ST127 – both considered as globally emerging pathogens of significant concern (O'Hara et al., 2019). Clonally-related isolates from multiple sequence types were found in both treated wastewater effluents as well as chlorinated sewage. As further evidence of enhanced survival of UPEC during wastewater treatment, a single sample of chlorinated sewage from a WWTP in Calgary, Alberta, contained 13 different UPEC isolates comprised of 6 different sequence types (ST131, ST80, ST625, ST127, ST95 and an unknown ST), and for which 11 of the isolates from 4 sequence types (ST131, ST80, ST127, and ST95) displayed >96% genome similarity to clinical UPEC. Collectively, our data demonstrate that the UPEC observed in treated wastewater effluents are very likely to be pathogenic to humans. Although our screening approach to detect and isolate clinically-relevant UPEC from sewage/wastewater was extremely effective, the process likely targets select strains, and therefore underestimates overall representation of the

diversity of clinical UPEC strains that might be present. Based on our data, we estimate that >25% of the surviving *E. coli* population found in treated wastewater effluents represent potentially pathogenic UPEC, a number that is in alignment with other studies (Anastasi et al., 2010; Anastasi et al., 2013; Adefisoye and Okoh 2016; Ebomah et al., 2018; Calhau et al., 2015)

Our data is also in alignment with Paulshus *et al*. (2019), who recently used whole genome sequencing to characterize ESBLs from sewage pump stations in suburban communities in Norway. Forty-five repetitive samplings were conducted over 15 months of this study generating a library of 3123 *E. coli* isolates. Up to 10% of *E. coli* isolates observed in this study were considered ESBLs, of which 44% of these were characterized as ST131 and ST648 strains. The fact that these strains were present throughout the sampling period and in such abundance, led the authors to speculate that ExPEC may represent: a) a resident population within wastewater systems, or b) that these ExPEC strains circulate abundantly in the community and for which infections may be underreported (Paulshus et al., 2019). In a survey of wastewater treatment plants in the U.S., Hoelle *et al*. (2019) also observed a relative abundance of ESBL strains associated with ST131, O25b-ST131, and ST648 lineages, findings that are in alignment with ours. Our O25b-ST131 W-UPEC isolates were confirmed as ESBLs based on phenotypic resistance profiling and were resistant to penicillins, β-lactam/β-lactamase inhibitor combinations, and all eight cephalosporins (cefazolin, cefixime, ceftriaxone, cefalexin, cefpodoxime, cefovecin, ceftazidime, and ceftiofur). The strains obtained from chlorinated sewage also possessed the $bla_{\text{CTX-M-15}}$, $bla_{\text{OXA-1}}$, and *ampC* genes and for which the $bla_{\text{CTX-M-15}}$ and $bla_{\text{OXA-1}}$ genes are known to confer resistance to β-lactam/β-lactamase inhibitor combination antibiotics (Sugumar et al., 2014). In addition to being ESBLs, our O25b-ST131 strains were also resistant to aminoglycosides and fluoroquinolones, both of which are also commonly used

antimicrobials to treat UTI (Ghafourian et al., 2015; Goodlet et al., 2019). The emergence of highly multidrug resistant UPEC greatly complicates clinical treatment and sometimes results in treatment failure of UTI or even death (Can et al., 2015; Owens et al., 2011).

The apparent ability of UPEC to survive wastewater has also been noted in other *E. coli* strains (Anastasi et al., 2010; Anastasi et al., 2013), including naturalized wastewater strains of *E. coli* (Zhi et al., 2017; Zhi et al., 2016; Zhi et al., 2019). Naturalized wastewater strains of *E. coli* have been shown to be genetically distinct from the cryptic and the human/animal enteric clades of *E. coli* based on comparative whole genome sequencing and phylogrouping (Zhi et al., 2019). They can be distinguished from enteric and cryptic clades through possession of the *uspC-IS30-flhDC* marker (Zhi et al., 2016; Zhi et al., 2019). Naturalized wastewater strains possess an over-abundance of stress-adaptive genes compared to enteric clades (Zhi et al., 2019) (believed to be acquired through horizontal gene transfer), and have been shown to be resistant to chlorine (Zhi et al., 2017) and heat (Wang et al., 2020). Interestingly, their extreme resistance to heat, is conferred by the locus of heat resistance (LHR), and for which we have demonstrated that LHR cross protects these strains against chlorine (Wang et al., 2020). The dominance of stress-related genes in naturalized wastewater strains of *E. coli*, as compared to enteric clades, suggests the evolution of survival strategies aimed at overcoming the harsh conditions of a wastewater treatment plant, and for which these strains may be specifically adapted to live and survive in this environment. Based on the thousands of *E. coli* genomes sequenced to date, these naturalized strains have only ever been found in sewage and wastewater (Zhi et al., 2019). It is believed that these naturalized wastewater strains of *E. coli* are non-pathogenic in nature, having abandoned life in the gut of an animal for life in sewage/ wastewater. Remarkably, these naturalized wastewater *E. coli* strains possess upwards of 36 different UPEC-related virulence

genes, including : a) the ferrienterobactin multi-enzyme sythetase complex and transport system; b) UPEC fimbrial genes; c) the curli virulence gene (*csgB*); d) the outer membrane protein gene known as *omp*A; and e) the flagellar biosynthesis protein encoded by *flhA* (Zhi et al., 2019). Some naturalized strains also possessed the yersiniabactin siderophore biosynthetic pathway and receptor/transport system found in many clinical UPEC, as well as the virulence genes *fyuA* and *irp*1/2 (Zhi et al., 2019). It is interesting to note that the virulence genes highlighted above have been identified as being important in the survival of clinical UPEC in the nutrient-deprived and osmotically/hydrologically-stressed environment of the urinary tract [i.e., low solutes and repetitive flushing] (Mann et al., 2017). For example, human UPEC strains have a predominance of iron-sequestering virulence genes (e.g., ferribactin, yersinabactin, enterobactin-related genes and transport pathways) to enhance their survival in iron-deficient urine (Terlizzi et al., 2017). An important finding noted by Zhi et al., (2019) was that many of these UPEC-related virulence genes found could also be found in the environmental cryptic clades (Zhi et al., 2019). Thus, the existence of these so-called virulence genes in naturalized wastewater *E. coli* strains, environmental cryptic clades of *E. coli*, and in clinical UPEC, suggests that the genes may actually provide an adaptive evolutionary advantage for survival in extraintestinal niches including non-host environmental niches such as water, sewage and wastewater.

Our data, and that of others, also raises some important questions about the public health risks associated with UPEC/ExPEC infections acquired through waterborne routes of transmission. Greater than 10 million physician visits occur each year in the U.S. due to urinary tract infections and for which the vast majority of these infections are caused by UPEC (Flores-Mireles et al., 2015; Foxman 2010; Pitout and DeVinney 2017; Russell et al., 2018; Xie et al., 2006). As with many clinical infections, the number of reported cases likely underestimates the

731 true scope of the health issue. Recently, it has been demonstrated that swimming in natural
732 water bodies has been epidemiologically-linked to increased prevalence of urinary tract
733 infections (UTIs), largely caused by UPEC (Soraas et al. 2013). Urinary pathogenic *E. coli* have
734 also been identified as extremely important emerging foodborne pathogens (Liu et al., 2018;
735 Markland et al., 2015), and although foodborne UPEC are typically attributed to animal
736 reservoirs (Todd, 1997; Singer, 2015), it is important to note that wastewater effluent discharges
737 can comprise a significant proportion of water available for agricultural production (animal and
738 crop production). It is estimated that 50% of drinking water treatment plants in the U.S. are
739 impacted by wastewater effluents, and in some cases wastewater effluents account for >50%
740 (and as high as 90%) of river flow volumes in the U.S. (Rice and Westerhoff, 2015; Rice and
741 Westerhoff, 2017). The U.S. EPA estimates that as many as 40,000 sewer overflows occur each
742 year, and up to 500,000 km of coastlines, rivers and streams do not currently meet ambient
743 microbial water quality guidelines for recreation as a result of human and animal waste
744 contamination (US EPA, 2007a; US EPA, 2007b). For these reasons there is growing concern
745 that water could represent an important transmission nexus for UPEC and other ExPEC strains.
746 The plausibility of waterborne disease transmission (i.e., ingestion) associated with
747 urinary and septicemic *E. coli* is supported by recent research that suggests that ExPEC strains
748 typically establish infection by first *asymptomatically* colonizing the gastrointestinal tract of an
749 animal host, subsequently invading the gut lining and disseminating to the urinary tract (e.g.,
750 UPEC) or bloodstream (e.g., NMEC) (Bower et al., 2005; Russell et al., 2017). Therefore,
751 ingestion of wastewater-contaminated food (i.e., crop irrigation), drinking water (i.e., untreated
752 or inadequately treated), or recreational water (i.e., ingestion) offer possible transmission
753 pathways for ExPEC infections from water. Unfortunately, UTIs or systemic septicemia have not

754 been considered as clinical endpoints for most epidemiological studies examining water as a
755 vehicle of transmission, focusing primarily on gastrointestinal illness. For example, the most
756 recent recreational water quality standards promulgated by the U.S.EPA in 2012 did not include
757 UTIs (or systemic septicemia) in their epidemiological benchmarks, but rather, gastrointestinal
758 illness as the clinical outcome (US EPA, 2012). As mentioned previously, UTIs have been
759 epidemiologically-linked to recreational water exposures (Soraas et al. 2013), and ESBL *E.coli*
760 have been routinely found in surface water samples and irrigation water samples (Franz et al.,
761 2015; Gekenidis et al., 2018; Gomi et al., 2017; Njage and Buys, 2015; Nuesch-Inderbinen et al.,
762 2015). In a study by Tanner *et al*. (2019), ESBLs were found in 6.4% of drinking water samples
763 that failed bacteriological water quality parameters in the U.S. (i.e., total coliforms), for which
764 several of the ESBLs identified were *E. coli*, and which led the authors to conclude that drinking
765 water may be an underestimated vehicle for transmission of ExPEC into the community.
766 Arguably, gastrointestinal illness may still be a more sensitive measure of the human health
767 outcomes associated with exposure to contaminated water, largely due to the diversity of
768 etiological agents that can cause gastroenteritis (viruses, bacteria, protozoa). However, given the
769 emerging global distribution and dramatic increase in the prevalence of antibiotic resistant UTIs
770 (i.e., ESBLs) and septicemia caused by ExPEC over the last decade, we would encourage future
771 research aimed at addressing the contribution that waterborne transmission may play in this
772 important and emerging public health issue. This call for research is further galvanized if we
773 consider that pathogens may not only be emerging resistance to antibiotics, but also to our most
774 important barrier for infectious disease control in society– water treatment and sanitation.
775

776 **5. Conclusion**

- Clinically-relevant UPEC strains readily survive municipal wastewater treatment processes.
- UPEC strains found in treated wastewater include ST131 ESBL-producing *E. coli* and represent strains of global public health concern.
- The resistance of UPEC to wastewater treatment implies that a potentially significant public health risk may exist for downstream drinking water systems impacted by wastewater discharge, as well as for crops irrigated with water contaminated by treated wastewater, or recreational sites impacted by municipal wastewater effluents.

Acknowledgements

This work was funded by: a) an Alberta Innovates grant to NN and NA under the Water Innovation Program (WIP); b) a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant to NN; c) a Canadian Institutes of Health Research Grant (CIHR) to NA and NN under the Healthy Environments Signature Initiative; and d) research funding from the City of Calgary to NN and NA.

Competing interests

The authors of this paper solemnly declare that there are no competing interests associated with the work presented in this manuscript.

800 **References**

801 Adefisoye, M.A. and Okoh, A.I., 2016. Identification and antimicrobial resistance prevalence of
802 pathogenic *Escherichia coli* strains from treated wastewater effluents in Eastern Cape, South
803 Africa. Microbiologyopen 5 (1), 143-151.

804 American Public Health Association., American Water Works Association and Water
805 Environment Federation, 1999. *Standard methods for the examination of water and*
806 *wastewater*. Washington DC: American Public Health Association.

807 Anastasi, E.M., Matthews, B., Gundogdu, A., Vollmerhausen, T.L., Ramos, N.L., Stratton, H.,
808 Ahmed, W. and Katouli, M., 2010. Prevalence and persistence of *Escherichia coli* strains with
809 uropathogenic virulence characteristics in sewage treatment plants. Appl Environ Microbiol 76
810 (17), 5882-5886.

811 Anastasi, E.M., Wohlsen, T.D., Stratton, H.M. and Katouli, M., 2013. Survival of *Escherichia coli*
812 in two sewage treatment plants using UV irradiation and chlorination for disinfection. Water
813 Res 47 (17), 6670-6679.

814 Ballestero-Tellez, M., Docobo-Perez, F., Portillo-Calderon, I., Rodriguez-Martinez, J.M., Racero,
815 L., Ramos-Guelfo, M.S., Blazquez, J., Rodriguez-Bano, J. and Pascual, A., 2017. Molecular
816 insights into fosfomycin resistance in *Escherichia coli*. J Antimicrob Chemother 72 (5), 1303-
817 1309.

818 Barber, A.E., Norton, J.P., Wiles, T.J. and Mulvey, M.A., 2016. Strengths and limitations of model
819 systems for the study of urinary tract infections and related pathologies. Microbiol Mol Biol Rev
820 80 (2), 351-367.

821 Beatson, S.A., Ben Zakour, N.L., Totsika, M., Forde, B.M., Watts, R.E., Mabbett, A.N., Szubert,
822 J.M., Sarkar, S., Phan, M.D., Peters, K.M., Petty, N.K., Alikhan, N.F., Sullivan, M.J., Gawthorne,
823 J.A., Stanton-Cook, M., Nhu, N.T., Chong, T.M., Yin, W.F., Chan, K.G., Hancock, V., Ussery, D.W.,
824 Ulett, G.C. and Schembri, M.A., 2015. Molecular analysis of asymptomatic bacteriuria
825 *Escherichia coli* strain VR50 reveals adaptation to the urinary tract by gene acquisition. Infection
826 and Immunity 83 (5), 1749-1764.

827 Bertels, F., Silander, O.K., Pachkov, M., Rainey, P.B. and van Nimwegen, E., 2014. Automated
828 reconstruction of whole-genome phylogenies from short-sequence reads. Mol Biol Evol 31 (5),
829 1077-1088.

830 Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M.,
831 Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G.,
832 Alekseyev, M.A. and Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its
833 applications to single-cell sequencing. J. of Comput. Biol. 19 (5), 455-477.
834

Boczek, L.A., Rice, E.W., Johnston, B. and Johnson, J.R., 2007. Occurrence of antibiotic-resistant uropathogenic *Escherichia coli* clonal group A in wastewater effluents. Appl Environ Microbiol 73 (13), 4180-4184.

Bower, J.M., Eto, D.S. and Mulvey, M.A., 2005. Covert operations of uropathogenic *Escherichia coli* within the urinary tract. Traffic 6 (1), 18-31.

Bunny, K.L., Hall, R.M. and Stokes, H.W., 1995. New mobile gene cassettes containing an aminoglycoside resistance gene, *aacA7*, and a chloramphenicol resistance gene, *catB3*, in an integron in pBWH301. Antimicrob Agents Chemother 39 (3), 686-693.

Cabiscol, E., Tamarit, J. and Ros, J., 2000. Oxidative stress in bacteria and protein damage by reactive oxygen species. Int Microbiol 3 (1), 3-8.

Calhau, V., Mendes, C., Pena, A., Mendonca, N. and Da Silva, G.J., 2015. Virulence and plasmidic resistance determinants of *Escherichia coli* isolated from municipal and hospital wastewater treatment plants. Journal of Water and Health 13 (2), 311-318.

Can, F., Azap, O.K., Seref, C., Ispir, P., Arslan, H. and Ergonul, O., 2015. Emerging *Escherichia coli* O25b/ST131 clone predicts treatment failure in urinary tract infections. Clin Infect Dis 60 (4), 523-527.

Chen, L., Zheng, D., Liu, B., Yang, J. and Jin, Q., 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. Nucleic Acids Res 44 (D1), D694-697.

Clermont, O., Dhanji, H., Upton, M., Gibreel, T., Fox, A., Boyd, D., Mulvey, M.R., Nordmann, P., Ruppe, E., Sarthou, J.L., Frank, T., Vimont, S., Arlet, G., Branger, C., Woodford, N. and Denamur, E., 2009. Rapid detection of the O25b-ST131 clone of *Escherichia coli* encompassing the CTX-M-15-producing strains. J Antimicrob Chemother 64 (2), 274-277.

Coque, T.M., Novais, A., Carattoli, A., Poirel, L., Pitout, J., Peixe, L., Baquero, F., Canton, R. and Nordmann, P., 2008. Dissemination of clonally related *Escherichia coli* strains expressing extended-spectrum beta-lactamase CTX-M-15. Emerg Infect Dis 14 (2), 195-200.

Cutler, D. and Miller, G., 2005. The role of public health improvements in health advances: the twentieth-century United States. Demography 42 (1), 1-22.

Cyoia, P.S., Rodrigues, G.R., Nishio, E.K., Medeiros, L.P., Koga, V.L., Pereira, A.P., Vespero, E.C., Houle, S., Dozois, C.M., Nakazato, G. and Kobayashi, R.K., 2015. Presence of virulence genes and pathogenicity islands in extraintestinal pathogenic *Escherichia coli* isolates from Brazil. J Infect Dev Ctries 9 (10), 1068-1075.

Dolejska, M., Frolkova, P., Florek, M., Jamborova, I., Purgertova, M., Kutilova, I., Cizek, A., Guenther, S. and Literak, I., 2011. CTX-M-15-producing *Escherichia coli* clone B2-O25b-ST131 and *Klebsiella spp.* isolates in municipal wastewater treatment plant effluents. J Antimicrob Chemother 66 (12), 2784-2790.

Drazic, A., Tsoutsoulopoulos, A., Peschek, J., Gundlach, J., Krause, M., Bach, N.C., Gebendorfer, K.M. and Winter, J., 2013. Role of cysteines in the stability and DNA-binding activity of the hypochlorite-specific transcription factor HypT. PLoS One 8 (10), e75683.

Du, Z., Nandakumar, R., Nickerson, K.W. and Li, X., 2015. Proteomic adaptations to starvation prepare *Escherichia coli* for disinfection tolerance. Water Res 69, 110-119.

Ebomah, K.E., Adefisoye, M.A. and Okoh, A.I., 2018. Pathogenic *Escherichia coli* strains recovered from selected aquatic resources in the Eastern Cape, South Africa, and Its significance to public health. Int J Environ Res Public Health 15 (7), 1506.

Flores-Mireles, A.L., Walker, J.N., Caparon, M. and Hultgren, S.J., 2015. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. Nat Rev Microbiol 13 (5), 269-284.

Foxman, B., 2010. The epidemiology of urinary tract infection. Nat Rev Urol 7 (12), 653-660.

Frampton, E.W., Restaino, L. and Blaszko, N., 1988. Evaluation of the beta-Glucuronidase Substrate 5-Bromo-4-Chloro-3-Indolyl-beta-D-Glucuronide (X-GLUC) in a 24-Hour Direct Plating Method for *Escherichia coli*. J Food Prot 51 (5), 402-404.

Francisco, A.P., Vaz, C., Monteiro, P.T., Melo-Cristino, J., Ramirez, M. and Carrico, J.A., 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13, 87.

Franz, E., Veenman, C., van Hoek, A.H., de Roda Husman, A. and Blaak, H., 2015. Pathogenic *Escherichia coli* producing extended-spectrum beta-lactamases isolated from surface water and wastewater. Sci Rep 5, 14372.

Gekenidis, M.T., Qi, W., Hummerjohann, J., Zbinden, R., Walsh, F. and Drissner, D., 2018. Antibiotic-resistant indicator bacteria in irrigation water: High prevalence of extended-spectrum beta-lactamase (ESBL)-producing *Escherichia coli*. PLoS One 13 (11), e0207857.

Ghafourian, S., Sadeghifard, N., Soheili, S. and Sekawi, Z., 2015. Extended spectrum beta-lactamases: definition, classification and epidemiology. Curr Issues Mol Biol 17, 11-21.

Gibreel, T.M., Dodgson, A.R., Cheesbrough, J., Bolton, F.J., Fox, A.J. and Upton, M., 2012a. High metabolic potential may contribute to the success of ST131 uropathogenic *Escherichia coli*. J Clin Microbiol 50 (10), 3202-3207.

Gibreel, T.M., Dodgson, A.R., Cheesbrough, J., Fox, A.J., Bolton, F.J. and Upton, M., 2012b. Population structure, virulence potential and antibiotic susceptibility of uropathogenic *Escherichia coli* from Northwest England. J Antimicrob Chemother 67 (2), 346-356.

903 Gomi, R., Matsuda, T., Matsumura, Y., Yamamoto, M., Tanaka, M., Ichiyama, S. and Yoneda, M.,
904 2017. Whole-genome analysis of antimicrobial-resistant and extraintestinal pathogenic
905 *Escherichia coli* in river water. Appl Environ Microbiol 83 (5), e02703-16.

906 Goodlet, K.J., Benhalima, F.Z. and Nailor, M.D., 2019. A systematic review of single-dose
907 aminoglycoside therapy for urinary tract infection: is it time to resurrect an old strategy?
908 Antimicrob Agents Chemother 63(1), e02165-18.

909 Gray, M.J., Wholey, W.Y., Parker, B.W., Kim, M. and Jakob, U., 2013. NemR is a bleach-sensing
910 transcription factor. J Biol Chem 288 (19), 13789-13798.

911 Hillion, M. and Antelmann, H., 2015. Thiol-based redox switches in prokaryotes. Biol Chem 396
912 (5), 415-444.

913 Hoelle, J., Johnson, J.R., Johnston, B.D., Kinkle, B., Boczek, L., Ryu, H. and Hayes, S., 2019. Survey
914 of US wastewater for carbapenem-resistant Enterobacteriaceae. J Wat Health 17 (2), 219-226.

915 Jia, B., Raphenya, A.R., Alcock, B., Waglechner, N., Guo, P., Tsang, K.K., Lago, B.A., Dave, B.M.,
916 Pereira, S., Sharma, A.N., Doshi, S., Courtot, M., Lo, R., Williams, L.E., Frye, J.G., Elsayegh, T.,
917 Sardar, D., Westman, E.L., Pawlowski, A.C., Johnson, T.A., Brinkman, F.S., Wright, G.D. and
918 McArthur, A.G., 2017. CARD 2017: expansion and model-centric curation of the comprehensive
919 antibiotic resistance database. Nucleic Acids Res 45 (D1), D566-D573.

920 Joensen, K.G., Tetzschner, A.M., Iguchi, A., Aarestrup, F.M. and Scheutz, F., 2015. Rapid and
921 easy *in silico* serotyping of *Escherichia coli* isolates by use of whole-genome sequencing data. J
922 Clin Microbiol 53 (8), 2410-2426.

923 Johnning, A., Kristiansson, E., Fick, J., Weijdegard, B. and Larsson, D.G., 2015. Resistance
924 mutations in *gyrA* and *parC* are common in *Escherichia* communities of both fluoroquinolone-
925 polluted and uncontaminated aquatic environments. Front Microbiol 6, 1355.

926 Kanamaru, S., Kurazono, H., Ishitoya, S., Terai, A., Habuchi, T., Nakano, M., Ogawa, O. and
927 Yamamoto, S., 2003. Distribution and genetic association of putative uropathogenic virulence
928 factors iroN, iha, kpsMT, ompT and usp in *Escherichia coli* isolated from urinary tract infections
929 in Japan. J Urol 170 (6 Pt 1), 2490-2493.

930 Korvin, D., Graydon, C., McNeil, L. and Mroczek, M., 2014. Banding Profile of Rep-PCR
931 Experiments Differs With Varying Extension Times and Annealing Temperatures. J Exper Micro
932 Immunol 18, 146-149.

933 Landini, P., Egli, T., Wolf, J. and Lacour, S., 2014. sigmaS, a major player in the response to
934 environmental stresses in *Escherichia coli*: role, regulation and mechanisms of promoter
935 recognition. Environ Microbiol Rep 6 (1), 1-13.

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Ponten, T., Ussery, D.W., Aarestrup, F.M. and Lund, O., 2012. Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50 (4), 1355-1361.

Letunic, I. and Bork, P., 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47 (W1), W256-W259.

Liu, C.M., Stegger, M., Aziz, M., Johnson, T.J., Waits, K., Nordstrom, L., Gauld, L., Weaver, B., Rolland, D., Statham, S., Horwinski, J., Sariya, S., Davis, G.S., Sokurenko, E., Keim, P., Johnson, J.R. and Price, L.B., 2018. *Escherichia coli* ST131-H22 as a foodborne uropathogen. Mbio 9 (4), e00470-18.

Mann, R., Mediati, D.G., Duggin, I.G., Harry, E.J. and Bottomley, A.L., 2017. Metabolic adaptations of uropathogenic *E. coli* in the urinary tract. Front Cell Infect Microbiol 7, 241.

Mao, B.H., Chang, Y.F., Scaria, J., Chang, C.C., Chou, L.W., Tien, N., Wu, J.J., Tseng, C.C., Wang, M.C., Chang, C.C., Hsu, Y.M. and Teng, C.H., 2012. Identification of *Escherichia coli* genes associated with urinary tract infections. J Clin Microbiol 50 (2), 449-456.

Markland, S.M., LeStrange, K.J., Sharma, M. and Kniel, K.E., 2015. Old friends in new places: exploring the role of extraintestinal *E. coli* in intestinal disease and foodborne illness. Zoonoses Public Health 62 (7), 491-496.

Marrs, C.F., Zhang, L. and Foxman, B., 2005. *Escherichia coli* mediated urinary tract infections: are there distinct uropathogenic E. coli (UPEC) pathotypes? Fems Microbiology Letters 252 (2), 183-190.

Marrs, C.F., Zhang, L., Tallman, P., Manning, S.D., Somsel, P., Raz, P., Colodner, R., Jantunen, M.E., Siitonen, A., Saxen, H. and Foxman, B., 2002. Variations in 10 putative uropathogen virulence genes among urinary, faecal and peri-urethral *Escherichia coli*. J Med Microbiol 51 (2), 138-142.

Mathers, A.J., Peirano, G. and Pitout, J.D., 2015. *Escherichia coli* ST131: The quintessential example of an international multiresistant high-risk clone. Adv Appl Microbiol 90, 109-154.

Matsumura, Y., Pitout, J.D.D., Peirano, G., DeVinney, R., Noguchi, T., Yamamoto, M., Gomi, R., Matsuda, T., Nakano, S., Nagao, M., Tanaka, M. and Ichiyama, S., 2017. Rapid identification of different *Escherichia coli* sequence type 131 Clades. Antimicrob Agents Chemother 61 (8), e00179-17.

Melnyk, R.A., Youngblut, M.D., Clark, I.C., Carlson, H.K., Wetmore, K.M., Price, M.N., Iavarone, A.T., Deutschbauer, A.M., Arkin, A.P. and Coates, J.D., 2015. Novel mechanism for scavenging of hypochlorite involving a periplasmic methionine-rich peptide and methionine sulfoxide reductase. Mbio 6 (3), e00233-00215.

970 Nicolas-Chanoine, M.H., Bertrand, X. and Madec, J.Y., 2014. *Escherichia coli* ST131, an intriguing
971 clonal group. Clin Microbiol Rev 27 (3), 543-574.

972 Nielubowicz, G.R. and Mobley, H.L., 2010. Host-pathogen interactions in urinary tract infection.
973 Nat Rev Urol 7 (8), 430-441.

974 Njage, P.M. and Buys, E.M., 2015. Pathogenic and commensal *Escherichia coli* from irrigation
975 water show potential in transmission of extended spectrum and AmpC beta-lactamases
976 determinants to isolates from lettuce. Microb Biotechnol 8 (3), 462-473.

977 Nuesch-Inderbinen, M., Zurfluh, K., Peterhans, S., Hachler, H. and Stephan, R., 2015.
978 Assessment of the prevalence of extended-spectrum beta-lactamase-producing
979 *Enterobacteriaceae* in ready-to-eat salads, fresh-cut fruit, and sprouts from the Swiss market. J
980 Food Prot 78 (6), 1178-1181.

981 O'Hara, R.W., Jenks, P.J., Emery, M. and Upton, M., 2019. Rapid detection of extra-intestinal
982 pathogenic *Escherichia coli* multi-locus sequence type 127 using a specific PCR assay. J Med
983 Microbiol 68 (2), 188-196.

984 Ohkoshi, Y., Sato, T., Suzuki, Y., Yamamoto, S., Shiraishi, T., Ogasawara, N. and Yokota, S.I.,
985 2017. Mechanism of reduced susceptibility to fosfomycin in *Escherichia coli* clinical isolates.
986 Biomed Research International 2017, 5470241.

987 Owens, R.C., Jr., Johnson, J.R., Stogsdill, P., Yarmus, L., Lolans, K. and Quinn, J., 2011.
988 Community transmission in the United States of a CTX-M-15-producing sequence type ST131
989 *Escherichia coli* strain resulting in death. J Clin Microbiol 49 (9), 3406-3408.

990 Pagès, H., Aboyoun, P., Gentleman, R., and DebRoy, S., 2019. Biostrings: efficient manipulation
991 of biological strings. R package version 2.52.0.

992 Parker, B.W., Schwessinger, E.A., Jakob, U. and Gray, M.J., 2013. The RclR protein is a reactive
993 chlorine-specific transcription factor in *Escherichia coli*. J Biol Chem 288 (45), 32574-32584.

994 Paulshus, E., Thorell, K., Guzman-Otazo, J., Joffre, E., Colque-Navarro, P., Kuhn, I., Mollby, R.,
995 Sorum, H. and Sjoling, A., 2019. Repeated isolation of ESBL positive *Escherichia coli* ST648 and
996 ST131 from community wastewater - are sewage systems important sources of emerging clones
997 of antibiotic resistant bacteria? Antimicrob Agents Chemother 63 (9): e00823-19 .

998 Pitout, J.D. and DeVinney, R., 2017. *Escherichia coli* ST131: a multidrug-resistant clone primed
999 for global domination. F1000Res 6(F1000 Faculty Rev): 195.

1000 Qin, X., Hu, F., Wu, S., Ye, X., Zhu, D., Zhang, Y. and Wang, M., 2013. Comparison of adhesin
1001 genes and antimicrobial susceptibilities between uropathogenic and intestinal commensal
1002 *Escherichia coli* strains. PLoS One 8 (4), e61169.

Ramirez, M.S. and Tolmasky, M.E., 2010. Aminoglycoside modifying enzymes. Drug Resist Updat 13 (6), 151-171.

Rice, J. and Westerhoff, P., 2015. Spatial and temporal variation in de facto wastewater reuse in drinking water systems across the U.S.A. Environ Sci Technol 49 (2), 982-989.

Rice, J. and Westerhoff, P., 2017. High levels of endocrine pollutants in US streams during low flow due to insufficient wastewater dilution. Nature Geosci 10, 587.

Russell, C.W., Fleming, B.A., Jost, C.A., Tran, A., Stenquist, A.T., Wambaugh, M.A., Bronner, M.P. and Mulvey, M.A., 2018. Context-dependent requirements for *fimH* and other canonical virulence factors in gut colonization by extraintestinal pathogenic *Escherichia coli*. Infect Immun 86 (3): e00746-17.

Russell, C.W., Richards, A.C., Chang, A.S. and Mulvey, M.A., 2017. The rhomboid protease GlpG promotes the persistence of extraintestinal pathogenic *Escherichia coli* within the gut. Infect Immun 85 (6): e00866-16.

Schreiber, H.L.t., Conover, M.S., Chou, W.C., Hibbing, M.E., Manson, A.L., Dodson, K.W., Hannan, T.J., Roberts, P.L., Stapleton, A.E., Hooton, T.M., Livny, J., Earl, A.M. and Hultgren, S.J., 2017. Bacterial virulence phenotypes of *Escherichia coli* and host susceptibility determine risk for urinary tract infections. Sci Transl Med 9(382), eaaf1283.

Sims, G.E. and Kim, S.H., 2011. Whole-genome phylogeny of *Escherichia coli*/*Shigella* group by feature frequency profiles (FFPs). Pro Natl l Acad of Sci USA. 108 (20), 8329-8334.

Singer, R.S., 2015. Urinary tract infections attributed to diverse ExPEC strains in food animals: evidence and data gaps. Front Microbiol 6, 28.

Soraas, A., Sundsfjord, A., Sandven, I., Brunborg, C. and Jenum, P.A., 2013. Risk factors for community-acquired urinary tract infections caused by ESBL-producing *Enterobacteriaceae*--a case-control study in a low prevalence country. PLoS One 8 (7), e69581.

Spurbeck, R.R., Dinh, P.C., Jr., Walk, S.T., Stapleton, A.E., Hooton, T.M., Nolan, L.K., Kim, K.S., Johnson, J.R. and Mobley, H.L., 2012. *Escherichia coli* isolates that carry *vat, fyuA, chuA, and yfcV* efficiently colonize the urinary tract. Infect Immun 80 (12), 4115-4122.

Sugumar, M., Kumar, K.M., Manoharan, A., Anbarasu, A. and Ramaiah, S., 2014. Detection of OXA-1 beta-lactamase gene of *Klebsiella pneumoniae* from blood stream infections (BSI) by conventional PCR and *in-silico* analysis to understand the mechanism of OXA mediated resistance. PLoS One 9 (3), e91800.

Tanner, W.D., VanDerslice, J.A., Goel, R.K., Leecaster, M.K., Fisher, M.A., Olstadt, J., Gurley, C.M., Morris, A.G., Seely, K.A., Chapman, L., Korando, M., Shabazz, K.A., Stadsholt, A., VanDeVelde, J., Braun-Howland, E., Minihane, C., Higgins, P.J., Deras, M., Jaber, O., Jette, D. and

1037 Gundlapalli, A.V., 2019. Multi-state study of *Enterobacteriaceae* harboring extended-spectrum
1038 beta-lactamase and carbapenemase genes in U.S. drinking water. Sci Rep 9 (1), 3938.

1039 Taskin, B., Gozen, A.G. and Duran, M., 2011. Selective quantification of viable *Escherichia coli*
1040 bacteria in biosolids by quantitative PCR with propidium monoazide modification. Appl Environ
1041 Microbiol 77 (13), 4329-4335.

1042 Tenaillon, O., Skurnik, D., Picard, B. and Denamur, E., 2010. The population genetics of
1043 commensal *Escherichia coli*. Nat Rev Microbiol 8 (3), 207-217.

1044 Terlizzi, M.E., Gribaudo, G. and Maffei, M.E., 2017. UroPathogenic *Escherichia coli* (UPEC)
1045 infections: virulence factors, bladder responses, antibiotic, and non-antibiotic antimicrobial
1046 strategies. Front Microbiol 8, 1566.

1047 Todd, E.C., 1997. Epidemiology of foodborne diseases: a worldwide review. World Health Stat Q
1048 50 (1-2), 30-50.

1049 Touchon, M., Hoede, C., Tenaillon, O., Barbe, V., Baeriswyl, S., Bidet, P., Bingen, E., Bonacorsi,
1050 S., Bouchier, C., Bouvet, O., Calteau, A., Chiapello, H., Clermont, O., Cruveiller, S., Danchin, A.,
1051 Diard, M., Dossat, C., Karoui, M.E., Frapy, E., Garry, L., Ghigo, J.M., Gilles, A.M., Johnson, J., Le
1052 Bouguenec, C., Lescat, M., Mangenot, S., Martinez-Jehanne, V., Matic, I., Nassif, X., Oztas, S.,
1053 Petit, M.A., Pichon, C., Rouy, Z., Ruf, C.S., Schneider, D., Tourret, J., Vacherie, B., Vallenet, D.,
1054 Medigue, C., Rocha, E.P. and Denamur, E., 2009. Organised genome dynamics in the *Escherichia*
1055 *coli* species results in highly diverse adaptive paths. PLoS Genet 5 (1), e1000344.

1056 Toval, F., Kohler, C.D., Vogel, U., Wagenlehner, F., Mellmann, A., Fruth, A., Schmidt, M.A.,
1057 Karch, H., Bielaszewska, M. and Dobrindt, U., 2014. Characterization of *Escherichia coli* isolates
1058 from hospital inpatients or outpatients with urinary tract infection. J Clin Microbiol 52 (2), 407-
1059 418.

1060 Trastoy, R., Manso, T., Fernandez-Garcia, L., Blasco, L., Ambroa, A., Perez Del Molino, M.L., Bou,
1061 G., Garcia-Contreras, R., Wood, T.K. and Tomas, M., 2018. Mechanisms of bacterial tolerance
1062 and persistence in the gastrointestinal and respiratory environments. Clin Microbiol Rev 31 (4),
1063 e00023-18.

1064 United States Environmental Protection Agency, 2007a. Protocol for developing nutrient TMDLs.
1065 Washington, DC: United States Environmental Protection Agency.

1066 United States Environmental Protection Agency, 2007b. Protocol for developing pathogen
1067 TMDLs. Washington, DC: United States Environmental Protection Agency.

1068 United States Environmental Protection Agency, 2012. Recreational Water Quality Criteria.
1069 Washington, DC: United States Environmental Protection Agency.

United States Environmental Protection Agency, 2010. EPA Microbiological Alternate Test Procedure (ATP) Protocol for Drinking Water, Ambient Water, Wastewater and Sewage Sludge Monitoring Methods. Washington, DC: United States Environmental Protection Agency.

Vigil, P.D., Stapleton, A.E., Johnson, J.R., Hooton, T.M., Hodges, A.P., He, Y. and Mobley, H.L., 2011. Presence of putative repeat-in-toxin gene tosA in *Escherichia coli* predicts successful colonization of the urinary tract. Mbio 2 (3), e00066-00011.

Vimont, S., Boyd, A., Bleibtreu, A., Bens, M., Goujon, J.M., Garry, L., Clermont, O., Denamur, E., Arlet, G. and Vandewalle, A., 2012. The CTX-M-15-producing *Escherichia coli* clone O25b: H4-ST131 has high intestine colonization and urinary tract infection abilities. PLoS One 7 (9), e46547.

Wang, Z., Fang, Y., Zhi, S., Simpson, DJ., Gill, A., McMullen, L.M., Neumann, N. F., and Gänzle, M.G. 2020. The locus of heat resistance confers resistance to chlorine and other oxidizing chemicals in *Escherichia coli*. Appl Environ Microbiol 86(4). pii: e02123-19. doi: 10.1128/AEM.02123-19.

White, A.P., Sibley, K.A., Sibley, C.D., Wasmuth, J.D., Schaefer, R., Surette, M.G., Edge, T.A. and Neumann, N.F., 2011. Intergenic sequence comparison of *Escherichia coli* isolates reveals lifestyle adaptations but not host specificity. Appl Environ Microbiol 77 (21), 7620-7632.

Winter, J., Ilbert, M., Graf, P.C., Ozcelik, D. and Jakob, U., 2008. Bleach activates a redox-regulated chaperone by oxidative protein unfolding. Cell 135 (4), 691-701.

Xie, J., Foxman, B., Zhang, L. and Marrs, C.F., 2006. Molecular epidemiologic identification of *Escherichia coli* genes that are potentially involved in movement of the organism from the intestinal tract to the vagina and bladder. J Clin Microbiol 44 (7), 2434-2441.

Yamaji, R., Friedman, C.R., Rubin, J., Suh, J., Thys, E., McDermott, P., Hung-Fan, M. and Riley, L.W., 2018. A population-based surveillance study of shared genotypes of *Escherichia coli* isolates from retail meat and suspected cases of urinary tract infections. mSphere 3(4), e00179-18.

Yun, K.W., Kim, D.S., Kim, W. and Lim, I.S., 2015. Molecular typing of uropathogenic *Escherichia coli* isolated from Korean children with urinary tract infection. Korean J Pediatr 58 (1), 20-27.

Yun, K.W., Kim, H.Y., Park, H.K., Kim, W. and Lim, I.S., 2014. Virulence factors of uropathogenic *Escherichia coli* of urinary tract infections and asymptomatic bacteriuria in children. J Microbiol Immunol Infect 47(6), 455-461.

Zhang, Y. and Lin, K., 2012. A phylogenomic analysis of *Escherichia coli*/Shigella group: implications of genomic features associated with pathogenicity and ecological adaptation. BMC Evol Biol 12, 174.

1104 Zhi, S., Banting, G., Li, Q., Edge, T.A., Topp, E., Sokurenko, M., Scott, C., Braithwaite, S., Ruecker,
1105 N.J., Yasui, Y., McAllister, T., Chui, L. and Neumann, N.F., 2016. Evidence of naturalized stress-
1106 tolerant strains of *Escherichia coli* in municipal wastewater teatment plants. Appl Environ
1107 Microbiol 82 (18), 5505-5518.

1108 Zhi S., Banting G., Ruecker.N. and Neumann N., 2017. Stress resistance in naturalised waste
1109 water *E. coli* strains. J Environ Eng Sci 12 (2), 42-50.

1110 Zhi, S., Banting, G., Stothard, P., Ashbolt, N.J., Checkley, S., Meyer, K., Otto, S. and Neumann,
1111 N.F., 2019. Evidence for the evolution, clonal expansion and global dissemination of water
1112 treatment-resistant naturalized strains of *Escherichia coli* in wastewater. Water Res 156, 208-
1113 222.

1114

Table 1. List of PCR gene targets, primer sequences and associated references.

Target Gene	Forward Primer (5' – 3')	Reverse Primer (5' – 3')	Amplicon Size (bp)	Reference
papC	GTGGCAGTATGAGTAATGACCGTTA	ATATCCTTTCTGCAGGGATGCAATA	202	(White et al., 2011)
sfa-foc	CTCCGGAGAACTGGGTGCATCTTAC	CGGAGGAGTAATTACAAACCTGGCA	407	(White et al., 2011)
chuA	CTGAAACCATGACCGTTACG	TTGTAGTAACGCACTAAACC	652	(Spurbeck et al., 2012)
iroN	AAGTCAAAGCAGGGGTTGCCCG	GACGCCGACATTAAGACGCAG	665	(White et al., 2011)
fyuA	GTAAACAATCTTCCCGCTCGGCAT	TGACGATTAACGAACCGGAAGGGA	850	(Spurbeck et al., 2012)
uidA	CGCAAGGTGCACGGGAATA Probe: FAM-ACCCGACGCGTCCGATCACCT-NFQMGB	CAGGCACAGCACATCAAAGAGA	143	This study (Taskin et al., 2011)
uspC-IS30-flhDC	CGGGGAACAAATGAGAACAC	TGGAGAAACGACGCAATC	386	(Zhi et al., 2016)
ST131	AGCAACGATATTTGCCCATT	GGCGATAACAGTACGCCATT	580	(Matsumura et al., 2017)
O25b-ST131	TCCAGCAGGTGCTGGATCGT	GCGAAATTTTTCGCCGTACTGT	347	(Clermont et al., 2009)
rep-PCR	5'-GTGGTGGTGGTGGTG-3'			(Korvin et al., 2014)

Table 2. General genome characteristics of sequenced wastewater UPEC strains.

	Strain Designation							
	WU153	**WU664**	**WU965**	**WU1022**	**WU1025**	**WU1030**	**WU1033**	**WU1036**
Genome size (Mb)	5.3	4.8	5	4.9	5.1	5.3	5.1	5.3
GC (%)	50.34	50.58	50.52	50.83	50.44	50.62	50.52	50.62
Average coverage	430	373	549	526	340	374	424	507
Genes (coding)	4706	4384	4579	4588	4779	5025	4782	5045
tRNA number	94	71	72	69	73	74	77	74
	Strain Designation							
	WU1038	**WU1149**	**WU1151**	**WU1155**	**WU1157**	**WU1214**	**WU1265**	**WU1266**
Genome size (Mb)	5	4.9	5.2	5.3	4.9	5.1	5.3	5.3
GC (%)	50.57	50.42	50.68	50.63	50.55	50.34	50.62	50.62
Average coverage	496	440	459	569	445	550	351	343
Genes (coding)	4623	4556	4887	5033	4516	4687	5024	5032
tRNA number	75	73	77	72	75	74	75	73
	Strain Designation							
	WU1274	**WU1630**	**WU1635**	**WU1752**	**WU2356**	**WU3165**	**WU3598**	**WU3707**
Genome size (Mb)	5.2	4.9	4.8	5.2	5.2	5	5.2	5
GC (%)	50.63	50.65	50.69	50.61	49.78	50.73	50.5	50.67
Average coverage	456	404	468	389	386	423	375	332
Genes (coding)	4933	4555	4477	4818	5098	4698	4835	4633
tRNA number	76	76	74	74	62	79	76	78

Table 3. Distance matrix SNP analysis based on core genome (labelled 'SNP') and pairwise whole genome similarity (labelled 'Similarity') of W-UPEC and C-UPEC strains

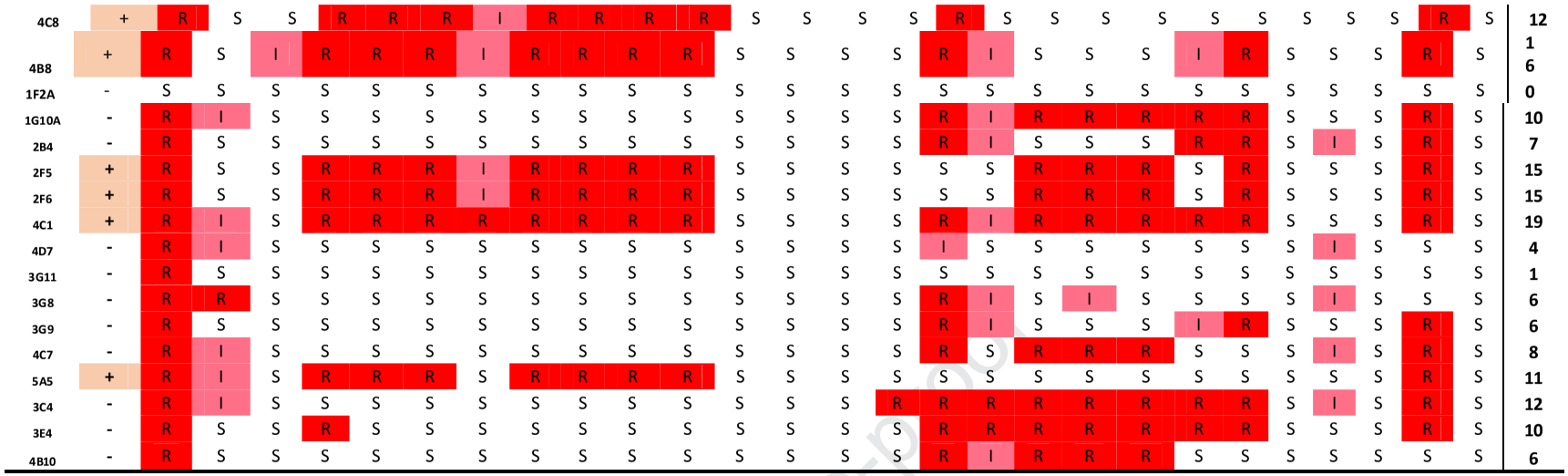
WU1030			WU1036			WU1155			WU1265			WU1266		
Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity
U308	3	96.66%	U308	2	96.05%	U308	3	96.58%	U308	2	96.64%	U308	2	96.66%
U309	7	96.72%	U309	6	96.11%	U309	7	96.64%	U309	6	96.71%	U309	6	96.73%
U310	8	96.73%	U310	7	96.12%	U310	8	96.65%	U310	7	96.71%	U310	7	96.73%
U244	9	97.23%	U244	8	96.63%	U244	9	97.15%	U244	8	97.22%	U244	8	97.24%
U272	20	96.70%	U272	19	96.10%	U272	20	96.62%	U272	19	96.69%	U272	19	96.69%
WU1157			**WU3707**			**WU1025**			**WU1151**			**WU1752**		
Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity
U60	4	98.54%	U139	3	98.93%	U64	5	97.50%	U23	18	97.39%	U215	16	96.50%
U203	6	98.30%	U238	4	96.76%	U339	7	96.75%	U294	21	97.46%	U288	18	96.05%
U70	17	96.04%	U48	6	98.18%	U26	9	97.09%	U295	22	96.83%	U291	19	98.79%
U44	21	96.18%	U162	14	97.60%	U41	9	97.36%	U292	25	97.07%	U192	20	97.00%
U321	22	96.15%	U289	17	98.56%	U223	11	96.77%	U32	25	96.27%	U293	23	97.20%
U334	23	97.36%	U8	46	96.37%	U38	11	97.26%	U67	26	96.09%			
U339	29	96.62%	U137	46	96.71%	U182	12	96.13%						
U64	29	96.09%	U204	47	96.82%	U44	29	96.41%						
U163	29	96.94%	**WU1274**			**WU965**			**WU1214**			**WU1033**		
U116	31	96.66%	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity	Clinical UPEC	SNP	Similarity
U26	31	96.90%	U294	12	96.65%	U173	13	98.58%	U341	7	99.49%	U290	25	98.03%
U41	31	96.29%	U295	13	96.91%	U154	14	97.53%	U57	9	97.08%			
U38	33	96.57%												
U115	33	97.01%												
U223	33	96.92%												
U182	34	96.19%												

Table 4. Serotype prediction of W-UPEC strains and their most closely related clinical- UPEC strains as determined by genome sequence analysis.

W-UPEC Strain	Genome-predicted Serotype of W-UPEC isolate	Most Closely Related C-UPEC strain(s)	Genome-predicted Serotype of C-UPEC isolate
WU1030	025:H4	**U244, U308, U309, U310**	O25:H4
WU1036	025:H4		
WU1155	025:H4		
WU1265	025:H4		
WU1266	025:H4		
WU1025	O6:H31	**U26, U339, U41, U64**	O6:H31
WU1033	01:H7	**U290**	O1:H7
WU1151	O18:H7	**U23, U292, U294, U295**	O18:H7
WU1214	O2:H7	**U341, U547**	O2:H7
WU1157	O6:H31	**U115, U203, U334, U60**	O6:H31
WU1274	O18:H7	**U23, U67, U294, U295**	O18:H7
WU1752	O1:H7	**U192, U291, U293**	O2:H7
		U215	O1:H7
WU3707	02:H4	**U139, U162, U289, U48**	O2:H4
WU965	O6:H1	**U173,U154**	O6:H1

Table 5. Phenotypic antibiotic resistance profiles (dark red = complete resistance [R]; light red = intermediate resistance [I]; and white = susceptible [S]) of W-UPEC strains isolated from chlorinate sewage (strains labelled with the prefix WU) and full scale WWTPs (all other strains) *.

Strain Name	Extended-spectrum β-lactamases	Penicillins	β-Lactam/β-lactamase inhibitor combinations		Cephalosporins								Carbapenems			Aminoglycosides			Quinolones and fluoroquinolones			Tetracyclines		Nitrofurantoins	Phenicols	Lipopeptides	Sulfonamides	Fosfomycins	No. of resistant antibiotics
	ESBL	AM	AMC	TZP	CN	CPD	CFO	CAZ	CFT	CZ	CFM	CRO	IPM	ETP	MEM	AN	GM	TM	CIP	ENR	MRB	DO	TE	FT	C	PB	SXT	FOS	
WU1157	-	S	S	S	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	1
WU3598	-	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	1
WU3165	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I	I	S	S	S	I	S	S	S	3
WU2356	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I	S	S	S	1
WU1030	+	R	I	S	R	R	R	I	R	R	R	R	S	S	S	S	R	R	R	R	R	S	S	S	S	S	S	S	16
WU1630	-	S	S	S	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I	S	S	S	2
WU664	-	R	R	R	R	R	R	S	I	R	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	9
WU1155	+	R	I	S	R	R	R	R	R	R	R	R	S	S	S	S	R	I	R	R	R	S	S	S	S	S	S	S	16
WU1036	+	R	I	S	R	R	R	R	R	R	R	R	S	S	S	S	R	I	R	R	R	S	S	S	S	S	S	S	16
WU1265	+	R	I	S	R	R	R	R	R	R	R	R	S	S	S	S	R	R	R	R	R	S	S	S	S	S	S	S	16
WU1266	+	R	I	S	R	R	R	R	R	R	R	R	S	S	S	S	R	R	R	R	R	S	S	S	S	S	S	S	16
2E3	-	R	R	I	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	R	S	R	S	R	S	7
3B1	-	R	R	I	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I	R	S	R	S	R	S	7
3B9	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I	S	S	S	1
4G1	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I	S	S	S	1
1G6	-	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	S	S	R	S	3
2B8	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I	S	S	S	S	1
4F6	-	R	R	S	S	S	S	S	I	S	S	S	S	S	S	S	S	S	S	S	S	R	R	S	R	S	R	S	7
4F9	-	R	I	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	R	S	R	S	R	S	6
4G9	-	R	I	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	R	S	R	S	R	S	6
1E1A	+	R	S	S	R	R	R	S	R	R	R	R	S	S	S	S	R	S	R	I	S	S	S	S	I	S	R	S	14
1G4	-	S	S	S	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	1
2E4	-	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	I	S	S	S	R	R	S	I	S	S	S	6
3H3	-	R	I	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	3
4C2	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	I	R	S	S	S	R	S	3



4C8	+	R	S	S	R	R	R	I	R	R	R	R	S	S	S	S	R	S	S	S	S	S	S	S	S	S	R	S	12
4B8	+	R	S	I	R	R	R	I	R	R	R	R	S	S	S	S	R	I	S	S	S	I	R	S	S	S	R	S	16
1F2A	-	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	0
1G10A	-	R	I	S	S	S	S	S	S	S	S	S	S	S	S	S	R	I	R	R	R	R	R	S	S	S	R	S	10
2B4	-	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	I	S	S	S	R	R	S	I	S	R	S	7
2F5	+	R	S	S	R	R	R	I	R	R	R	R	S	S	S	S	S	S	R	R	R	S	R	S	S	S	R	S	15
2F6	+	R	S	S	R	R	R	I	R	R	R	R	S	S	S	S	S	S	R	R	R	S	R	S	S	S	R	S	15
4C1	+	R	I	S	R	R	R	R	R	R	R	R	S	S	S	S	R	I	R	R	R	R	R	S	S	S	R	S	19
4D7	-	R	I	S	S	S	S	S	S	S	S	S	S	S	S	S	I	S	S	S	S	S	S	S	I	S	S	S	4
3G11	-	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	S	1
3G8	-	R	R	S	S	S	S	S	S	S	S	S	S	S	S	S	R	I	S	I	S	S	S	S	I	S	S	S	6
3G9	-	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	I	S	S	S	I	R	S	S	S	R	S	6
4C7	-	R	I	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	R	R	R	S	S	S	I	S	R	S	8
5A5	+	R	I	S	R	R	R	S	R	R	R	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	S	11
3C4	-	R	I	S	S	S	S	S	S	S	S	S	S	S	S	R	R	R	R	R	R	R	R	S	I	S	R	S	12
3E4	-	R	S	S	R	S	S	S	S	S	S	S	S	S	S	S	R	R	R	R	R	R	R	S	S	S	R	S	10
4B10	-	R	S	S	S	S	S	S	S	S	S	S	S	S	S	S	R	I	R	R	R	S	S	S	S	S	S	S	6

***Antibiotic acronyms:** ESBL (Extended spectrum beta-lactamase), AM (Ampicillin), AMC (Amoxicillin/Clavulanic Acid), TZP (Piperacillin/ Tazobactam), CN (Cefalexin), CPD (Cefpodoxime), CFO (Cefovecin), CAZ (Ceftazidime), CFT (Ceftiofur), CZ (Cefazolin), CFM (Cefixime), CRO (Ceftriaxone), IPM (Imipenem), ETP (Ertapenem), MEM (Meropenem), AN (Amikacin), GM (Gentamicin), TM (Tobramycin), CIP (Ciprofloxacin), ENR (Enrofloxacin), MRB (Marbofloxacin), DO (Doxycycline), TE (Tetracycline), FT (Nitrofurantoin), C (Chloramphenicol), PB (Polymyxin B), SXT (Trimethoprim/Sulfamethoxazole), FOS (Fosfomycin).

Table 6. Comparative rep-PCR analysis of the wastewater UPEC strains isolated from chlorinated sewage or wastewater treatment plant effluents (WWTP-Effluent), along with their associated sequence types based on PCR (ST131 and O25b-ST131) or whole genome sequence analysis (for isolates from chlorinated sewage).

Rep-PCR Pattern	W-UPEC Strain	Sequence Type	Source	O25b-ST131	Rep-PCR Pattern	W-UPEC Strain	Sequence Type	Source	O25b-ST131
Pattern I	1025	ST127	Chlorinated sewage	No	Pattern IV	1F2A	ST131	WWTP-Effluent	No
	1157	ST127	Chlorinated sewage	No		2B4	ST131	WWTP-Effluent	No
	3B7	ND	WWTP-Effluent	No		3C4	ST131	WWTP-Effluent	No
Pattern II	965	unknown	Chlorinated sewage	No		3E4	ST131	WWTP-Effluent	No
	2B8	ND*	WWTP-Effluent	No		3G8	ST131	WWTP-Effluent	No
	3H3	ND*	WWTP-Effluent	No		4B10	ST131	WWTP-Effluent	No
	2E3	ND*	WWTP-Effluent	No		4B8	ST131	WWTP-Effluent	No
	4F6	ND*	WWTP-Effluent	No		4C7	ST131	WWTP-Effluent	No
Pattern III	1030	ST131	Chlorinated sewage	Y		5A5	ST131	WWTP-Effluent	No
	1155	ST131	Chlorinated sewage	Y	Pattern V	3B9	ND*	WWTP-Effluent	No
	1036	ST131	Chlorinated sewage	Y		4G1	ND*	WWTP-Effluent	No
	1265	ST131	Chlorinated sewage	Y		1274	ST95	Chlorinated sewage	No
	1266	ST131	Chlorinated sewage	Y		1151	ST95	Chlorinated sewage	No
	2F6	ST131	WWTP-Effluent	Y		1752	unknown	Chlorinated sewage	No
	4C1	ST131	WWTP-Effluent	Y	Pattern VI	3B1	ND*	WWTP-Effluent	No
	1G10A	ST131	WWTP-Effluent	Y		2E4	ND*	WWTP-Effluent	No
	2F5	ST131	WWTP-Effluent	Y		4F9	ND*	WWTP-Effluent	No
	4D7	ST131	WWTP-Effluent	Y		4G9	ND*	WWTP-Effluent	No
	3G11	ST131	WWTP-Effluent	Y	Pattern VII	3707	ST95	Chlorinated sewage	No
	3G9	ST131	WWTP-Effluent	Y		1033	ST95	Chlorinated sewage	No

*ND- Not determined

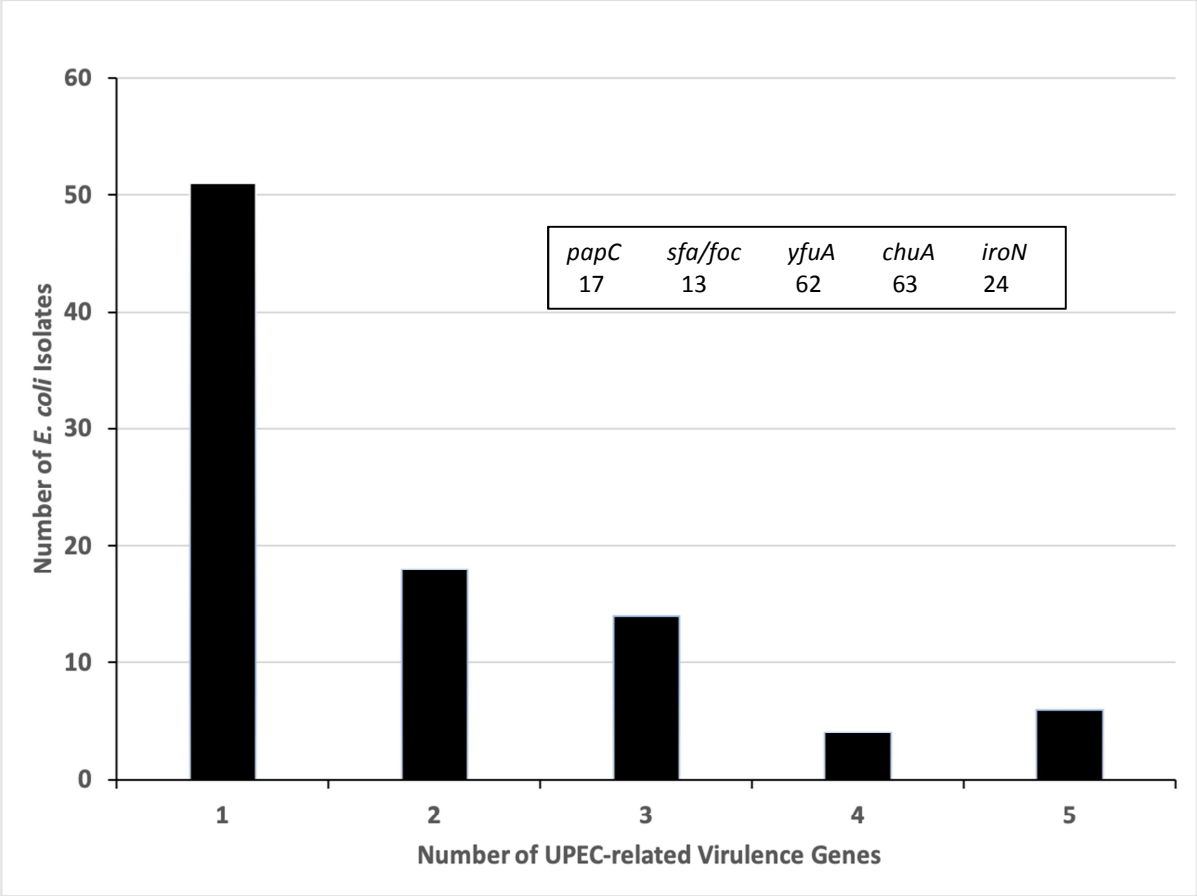


Figure 1. Number (non-cumulative) of chlorine-tolerant *E. coli* isolates possessing one or more UPEC-related virulence genes. Inset highlights the number of isolates in which the specific UPEC-related gene was observed (n=94).

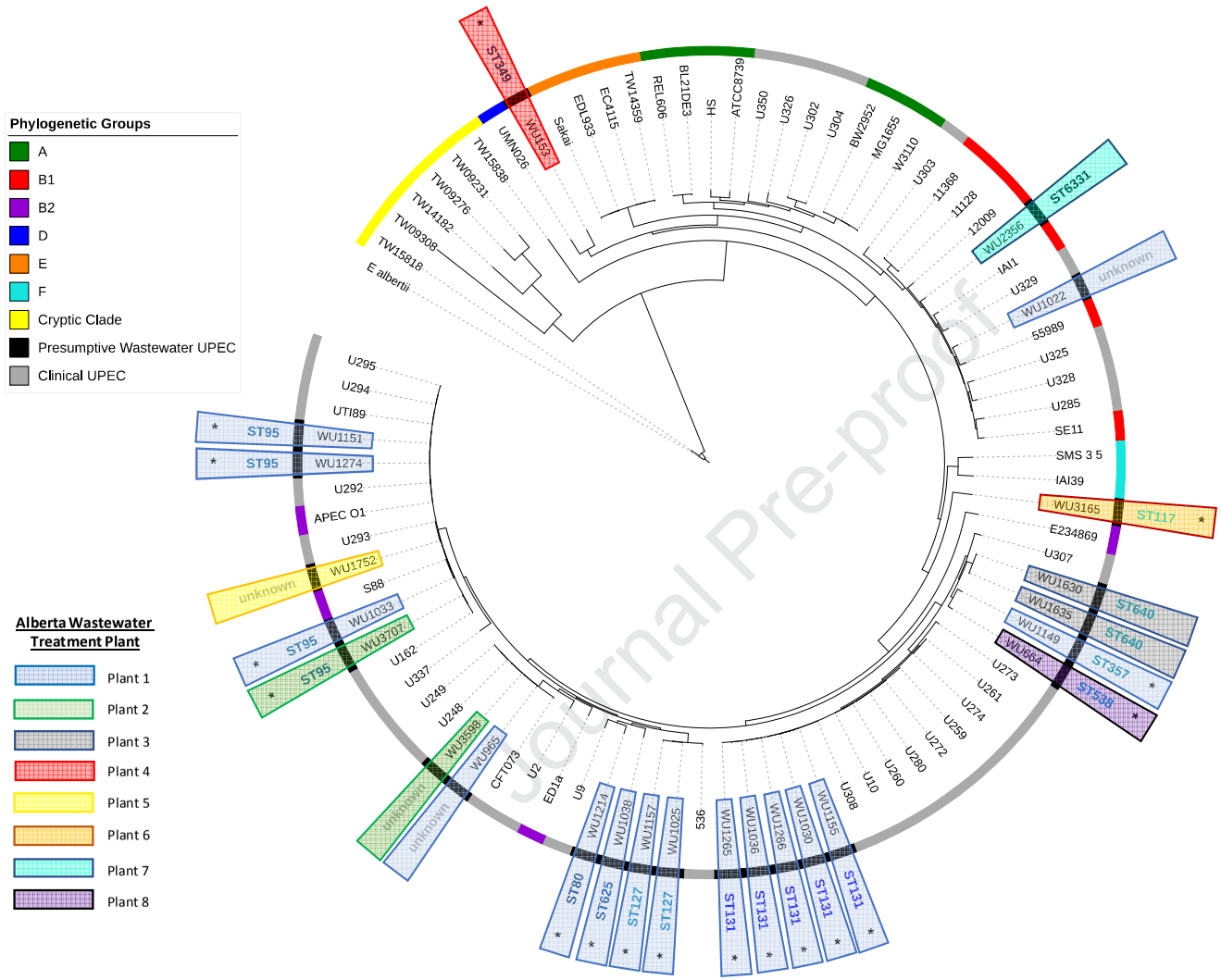


Figure 2. Whole genome maximum likelihood phylogenetic tree of wastewater UPEC strains. W-UPEC strains were isolated from chlorinated sewage (designated with the prefix 'WU' and colored black in the concentric circle), and genome sequences compared to clinical UPEC strains from NCBI (designated with the prefix 'U' and colored grey in the concentric circle), along with various reference phylogroup strains of *E. coli* (other colored lines of the circle as represented in the upper figure legend). Wastewater UPEC strains originated from eight different wastewater treatment plants, with each specific treatment plant represented by a colored polygon (as reflected in the lower figure legend). Within each of the polygons are the predicted sequence types of the W-UPEC isolates. *E. albertii* was used as the outgroup.

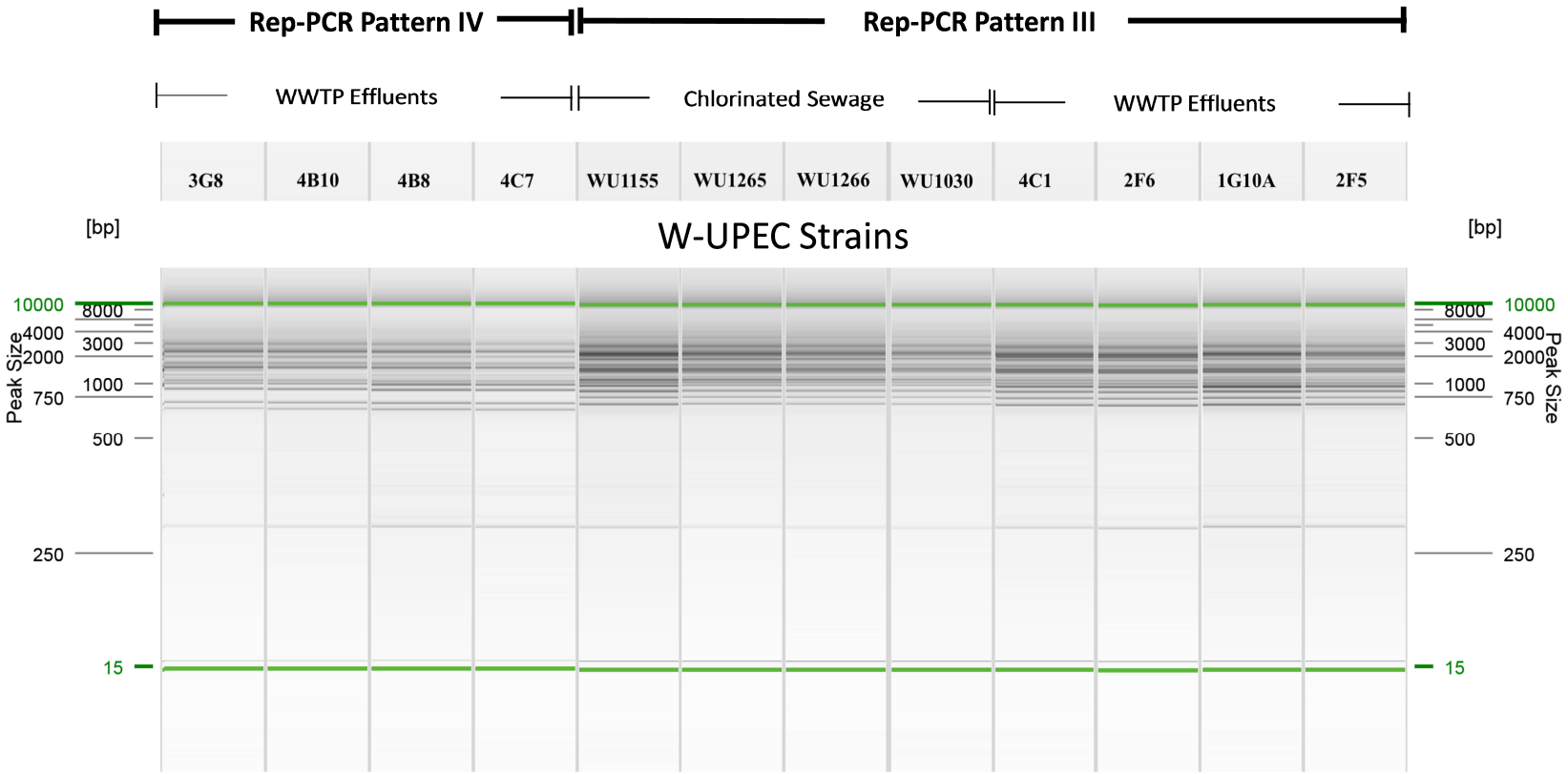


Figure 3. $(GTG)_5$ rep-PCR profiles of ST131 strains of W-UPEC isolated from chlorinated sewage or full-scale wastewater treatment plant effluents. Rep-PCR Pattern 3 represented all O25b-ST131 isolates, whereas Pattern IV was associated with non-O25b isolates of the ST131 lineage and isolated from full-scale WWTP effluents.

HIGHLIGHTS

- Urinary pathogenic *E. coli* (UPEC) appear to be resistant to wastewater treatment.
- Based on comparative genomics, the UPEC found in treated wastewater are clinically important.
- Pathogenic UPEC strains, such as ST-131 ESBL, can be frequently found in effluents.
- UPEC comprise a major portion of the *E. coli* discharged from wastewater treatment plants.
- Unlike shigatoxin producing *E. coli,* UPEC appear adapted to survive wastewater treatment.

Conflict of Interest Statement

The authors of this paper solemnly declare that there are no competing interests associated with the work presented in this manuscript.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: