

ARTICLE OPEN



Differential survival of potentially pathogenic, septicemia- and meningitis-causing *E. coli* across the wastewater treatment train

Daniel Yu^{1,2}, Kanghee Ryu^{1,2}, Simon J. G. Otto^{1,2,3,4}, Paul Stothard⁵, Graham Banting¹, Norma Ruecker⁶, Norman F. Neumann^{1,2} and Shuai Zhi^{7,8}✉

A growing body of evidence indicates that extraintestinal pathogenic *E. coli* (ExPEC) readily survive wastewater treatment, raising concerns about the public health risks associated with exposure to wastewater-contaminated environments. In this study, *E. coli* isolates recovered from chlorinated sewage or treated wastewater effluents in Canada were screened for ExPEC virulence markers. Eighty-six isolates were identified as presumptive ExPEC, clustering within major pandemic lineages including ST131, ST95, and ST73 according to multilocus sequence typing analyses. Across the whole, core, and accessory genome, 37 isolates were extremely similar to clinical bloodborne *E. coli* (BBEC) and neonatal meningitic *E. coli* (NMEC) strains, suggesting that these wastewater isolates may exhibit a similar phenotypically related pathogenic potential. Interestingly, ExPEC strains also shared accessory gene content with naturalized wastewater strains, suggesting a common genetic capacity for surviving water treatment. Collectively, these findings suggest that *E. coli* strains that may cause septicemia and meningitis are surviving wastewater treatment and may be transmissible through wastewater effluents.

npj Clean Water (2022)5:33; <https://doi.org/10.1038/s41545-022-00177-y>

INTRODUCTION

Whether for drinking, agricultural, or recreational purposes, access to safe, clean water is of paramount importance to public health. With the incredible burden that waterborne pathogens place on human health, the risks associated with exposure to contaminated water sources need to be properly evaluated. For instance, an estimated 1.6 million deaths and 105 million disability-adjusted life years lost were associated with water quality or sanitation issues in 2016, of which roughly half were directly attributable to waterborne diarrheal disease¹. While the public health benefits of access to clean water are most apparent with drinking water^{2,3}, effective wastewater sanitation also represents an important public health intervention. Indeed, across 39 nations, inadequate wastewater treatment was found to correlate with increased disease mortality, irrespective of national income, development, and overall sanitation⁴. As such, the treatment and sanitation of all urban water supplies represent one of the most important preventative strategies against infectious diseases.

The ability to reliably eliminate pathogens from the water supply is key for infectious disease prevention; however, what would happen if pathogens could evolve resistance against water treatment? Microbes are incredibly adaptable, and they have a remarkable ability to tolerate normally lethal stressors. For instance, the model bacterium *Escherichia coli* possesses a myriad of adaptive stress response mechanisms^{5,6} and stress-specific transcription factors^{5,7–11}, many of which have been found to confer resistance against water treatment-associated stressors including chlorination¹², osmotic stress, and oxidant- and UV-mediated oxidative stress⁵. Beyond these stress-specific response

systems, cross-resistance may also be conferred by other stress resistance strategies in *E. coli*, such as through the production of heat shock proteins¹³ and a recently characterized genetic island known as the locus of heat resistance (LHR) which, alongside heat resistance, also appear to provide resistance against chlorination and oxidative stress^{14,15}. As such, microbes may already possess the genetic capacity to survive water treatment.

Reflecting this, growing evidence suggests that water treatment may be selected for the evolution of resistance to water disinfection processes. Indeed, certain subpopulations of *E. coli* have been found to be particularly resistant to water treatment, including naturalized wastewater strains that appear to have adapted to survive in wastewater treatment plants¹⁶. Interestingly, these strains appear to be enriched across wastewater treatment¹⁷, and demonstrate enhanced resistance to chlorine, oxidants, and heat, as well as an increased capacity for biofilm formation^{15,18}. Beyond these naturalized strains, however, concerning evidence suggests that extraintestinal pathogenic *E. coli* (ExPEC), including strains causing urinary tract infections, also appear to survive wastewater treatment. Indeed, *E. coli* isolates harboring typical uropathogenic *E. coli* (UPEC) virulence genes and pathogenicity islands have been recovered from finished wastewater effluents¹⁹, and it has been estimated that anywhere from 41.7% to 59.5% of *E. coli* isolates surviving wastewater treatment represent potential UPEC according to virulence gene screening^{20–22}.

More recently, comparative genomic analyses performed by Zhi and colleagues similarly identified several *E. coli* strains isolated from chlorinated sewage and treated effluents as clinically-

¹School of Public Health, University of Alberta, Edmonton, AB, Canada. ²Antimicrobial Resistance—One Health Consortium, Calgary, AB, Canada. ³Lead, HEAT-AMR Research Group, School of Public Health, University of Alberta, Edmonton, AB, Canada. ⁴Thematic Area Lead, Healthy Environments, Centre for Health Communities, School of Public Health, University of Alberta, Edmonton, AB, Canada. ⁵Department of Agriculture, Food and Nutritional Sciences, University of Alberta, Edmonton, AB, Canada. ⁶City of Calgary, Water Quality Services, Calgary, AB, Canada. ⁷School of Medicine, Ningbo University, Ningbo, China. ⁸The Affiliated Hospital of the Medical School, Ningbo University, Ningbo, China. ✉email: zhishuai@nbu.edu.cn

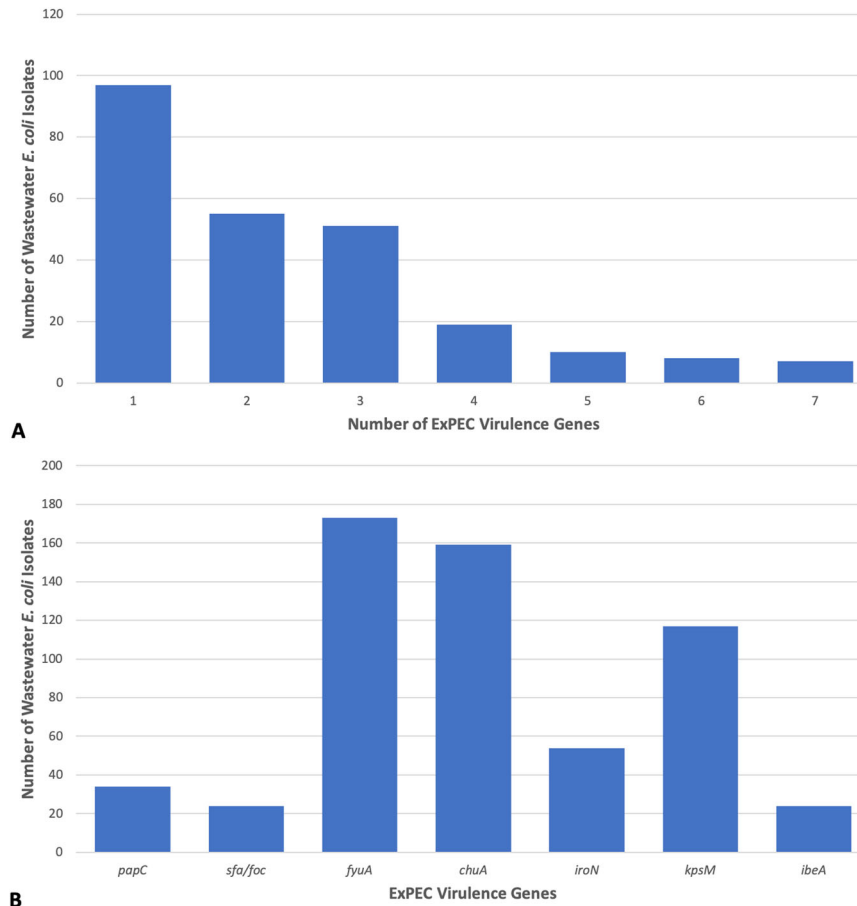


Fig. 1 ExPEC virulence marker screening panel results. Number **A** of ExPEC virulence genes harbored by each wastewater *E. coli* isolate and **B** their frequency across all wastewater *E. coli* isolates as determined by genetic screening with ExPEC PCR marker panel and targeted gene identification in the genome sequences of the wastewater *E. coli* isolates.

relevant UPEC²³. *E. coli* strains recovered from wastewater matrices clustered in clinically relevant sequence types, such as ST131 and ST95, and were found to be almost identical to clinical UPEC strains across the whole and core genome, as they shared 96.00–99.49% whole genome similarity and differed by as few as 2 single nucleotide polymorphisms (SNPs) in a ~0.4 Mb core genome backbone compared to clinical UPEC strains. Many of the wastewater isolates also possessed the exact complement of virulence and antibiotic resistance genes harbored by their closest clinical UPEC match. Concerningly, five wastewater isolates belonged to the emerging pandemic O25b-ST131 clonal group and were characterized as extended-spectrum beta-lactamase (ESBL) producing strains. Collectively, the finding that ExPEC appear to dominate the population of *E. coli* surviving wastewater treatment has led to the suggestion that either ‘higher-than-expected’ levels of ExPEC infection exist in the community, or that ExPEC strains naturally occur in wastewater²⁴.

Importantly, the evolution towards water treatment resistance may extend to other ExPEC pathotypes, including the bloodborne *E. coli* (BBEC) and neonatal meningitic *E. coli* (NMEC). For instance, *E. coli* strains isolated from treated wastewater have been shown to cluster within sequence types such as ST95 and ST131^{19,23,25}, lineages that have recently been associated with NMEC²⁶ and BBEC^{26,27} outbreaks, respectively. Furthermore, while Adefisoye and Okoh characterized the majority of *E. coli* isolates recovered from wastewater effluent as potential UPEC, 14.8% were identified as potential NMEC²². This points to the possibility that, similar to the findings described for UPEC, the NMEC and BBEC strains may also have evolved resistance against water treatment.

Herein, we demonstrate through a comprehensive comparative whole-genome analysis, that NMEC and BBEC strains are common constituents in full-scale treated wastewater effluents and chlorinated sewage—raising the worrying prospect that diseases such as urinary tract infections, septicemia, and meningitis could be water transmissible.

RESULTS

Identification of presumptive ExPEC from chlorinated sewage and final wastewater effluent

Among all 637 wastewater *E. coli* isolates collected in this study, 247 possessed at least one screened ExPEC-associated virulence gene, of which 7 harbored all seven (Fig. 1a). Of the seven screened ExPEC virulence genes, *fyuA* and *chuA* were the most prevalent, while *sfa/foc* and *ibeA* were the least common (Fig. 1b). All isolates were also screened for the major ExPEC pandemic lineage ST131 marker, of which 22 were positive. According to the screening criteria used, 86 isolates were identified as presumptive wastewater ExPEC (W-ExPEC) and were selected for genome sequencing. The genome characteristics of the sequenced isolates are summarized in Supplementary Table 1.

Core genome similarity between wastewater and clinical ExPEC (C-ExPEC) strains

Based on an upper maximum difference of 250 core genome SNPs across a ~417 kbp backbone, 37 W-ExPEC strains were found to share close core genome similarity with anywhere from 1 to 48

clinical BBEC and NMEC strains (Supplementary Table 2). Some W-ExPEC strains in particular displayed a remarkable level of similarity to a clinical counterpart at the core genome level, including the wastewater strains 2F5 and 2F6, which both varied from BBEC_156 by only 5 SNPs (Table 1); and wastewater strain 3G9, which could be distinguished from BBEC_211 by only 3 SNPs (Table 1), and wastewater strain 1G6, which differed from the clinical strain NMEC_9 by only 7 SNPs (Table 2). Some wastewater strains were even found to exhibit close core genome similarity with multiple clinical strains, such as strains 1F2A and 2B4, which differed from BBEC_268 by 6 and 4 SNPs, respectively, and BBEC_267 by 5 SNPs each. Interestingly, several wastewater strains previously identified to be genetically similar to clinical UPEC strains²³ also shared high core genome similarity with a clinical NMEC or BBEC strain in this analysis. Of note were WU1030, WU1036, WU1155, WU1265, and WU1266, as each differed from BBEC_265 and BBEC_158 by 6 or fewer core genome SNPs; WU1274, which differed from BBEC_162 by 8 SNPs; and WU1157, which could be distinguished from BBEC_71 by only 7 SNPs (Table 1).

Pairwise whole-genome similarity between W-ExPEC and C-ExPEC strains

Identified W-ExPECs demonstrating high core genome similarity with clinical BBEC and NMEC strains ($n = 37$) were selected for pairwise whole-genome comparisons against the local genome repository of 320 clinical NMEC and BBEC isolates. Additional comparisons were also performed for each W-ExPEC strain against 46 representative intestinal pathogenic *E. coli* and naturalized wastewater *E. coli* strains as further evidence that any significant similarities observed between W-ExPEC and C-ExPEC strains were reflective of a shared extraintestinal pathogenic phenotype. All pairwise comparisons were evaluated against an upper median whole genome similarity value of 96.03%, which was previously found to be reflective of a shared pathogenic phenotype²³.

Interestingly, all 37 W-ExPEC strains sharing close core genome similarity with a clinical counterpart also exhibited >96.03% whole genome similarity with at least 1 clinical BBEC (Table 1) or NMEC (Table 2). Whole-genome similarity values ranged from 96.04% to 99.74%, with some W-ExPEC strains exhibiting high similarity with up to 48 C-ExPEC counterparts (Supplementary Table 2). Of note was the wastewater strain 1G6, which exhibited 99.72% whole-genome similarity with NMEC_9; and wastewater strains 3B9 and 4G1, which shared 99.62% and 99.58% whole genome similarity, respectively, with BBEC_38 (Table 1), and 99.59% and 99.55% whole genome similarity, respectively, with NMEC_4 (Table 2). Interestingly, wastewater strains previously identified to share close genetic similarity to clinical UPEC strains²³ were found to exhibit greater similarity to clinical BBEC strains in this analysis. For instance, the wastewater strains WU1030, WU1036, WU1155, WU1265, and WU1266 all shared >99.10% whole genome similarity with the clinical strain BBEC_265. Similarly, WU1030, WU1155, WU1265, and WU1266 were also all found to exhibit >99.40% whole genome similarity with BBEC_158 (Table 1).

In contrast, W-ExPEC isolates exhibited significantly lower similarity with intestinal pathogenic and naturalized wastewater *E. coli* strains, with similarity values ranging from 65.9% to 95.3%, with a mean of 84.9% (Supplementary Table 3). Indeed, none of the 86 presumptive W-ExPEC isolates shared >96.03% whole genome similarity with any intestinal pathogenic or naturalized wastewater *E. coli* strains.

Core genome phylogenetic and sequence typing of W-ExPEC and C-ExPEC strains

To understand the evolutionary relationship between the W-ExPEC strains and their clinical counterparts, a maximum-likelihood core genome phylogenetic analysis was performed with the 37

W-ExPEC isolates, 38 of the closest clinical ExPEC strains at the whole and core genome levels, and *E. coli* strains from various known phylogroups^{28–30}. Of the 37 W-ExPEC strains, 34 were distributed amongst clinical NMEC and BBEC strains throughout phylogroup B2, a major phylogroup known to harbor ExPEC pathotypes^{31,32}, whereas the remaining 3 clustered with their closest clinical matches in phylogroup A (Fig. 2). Although phylogroup A is classically considered to be non-pathogenic^{33,34} and negatively associated with the UPEC pathotype³⁵, studies have reported clinical ExPEC strains belonging to this phylogroup³⁶. Reflecting this, the latter three W-ExPEC isolates grouped separately from commensal *E. coli* strains and into two separate sub-clusters alongside clinical BBEC strains within phylogroup A (Fig. 2). Interestingly, although phylogroup D has been positively associated with ExPEC pathotypes other than UPEC³⁵, none of the ExPEC strains clustered within this phylogroup.

W-ExPEC and C-ExPEC strains were found to sub-structure extensively by sequence type within the phylogenetic tree (Fig. 2). Although 2 of the 37 W-ExPEC strains were designated as unknown sequence types (ST), the remaining 35 were distributed across 8 STs, including those of clinical importance. The major ExPEC lineage-associated sequence type ST131³⁷ was the most represented amongst the W-ExPEC strains, with 18 of 37 isolates clustering within this sequence type. Confirming previous analyses, these 18 W-ExPEC isolates were also confirmed to be positive for the ST131 marker by the PCR screening panel. The NMEC-associated lineage ST95²⁷ was also well represented, identified in 8 wastewater isolates. Of the remaining W-ExPEC isolates, 2 each were designated as ST73, ST10, and ST127, while 1 isolate each was designated as ST357, ST538, and ST44. The 38 C-ExPEC strains were distributed in a similar manner across the 8 STs represented, as each W-ExPEC strain that was assigned a sequence type belonged to the same lineage as their closest clinical match.

Pan-genome similarity and accessory genome phylogenetics of W-ExPEC and C-ExPEC strains

Across the core and whole-genome, W-ExPEC and C-ExPEC strains appear to be highly similar; however, measures of genetic similarity alone may not completely reflect whether the W-ExPEC strains share a similar pathogenic phenotype with their clinical counterparts. In contrast to the core genome, which encodes for essential housekeeping genes, the accessory genome includes genes linked to adaptation, virulence, and antibiotic resistance, which are likely reflective of the predominant lifestyle of a given strain³⁸. As such, a pan-genomic analysis was performed to determine whether the W-ExPEC strains shared a similar accessory gene profile, and thus a similar pathogenic potential, with their clinical counterparts.

A pan-genome was calculated for all 86 presumptive W-ExPEC strains, 38 of the closest clinical NMEC/BBEC matches, 9 reference UPEC strains, 5 naturalized wastewater strains^{17,39}, and 4 laboratory reference strains (Supplementary Table 4). The pan-genome consisted of 26,865 genes, including 2133 core genes, indicating a high level of pan-genomic diversity across the strains analyzed. Reflecting this, the core genes comprised only 8% of the pan-genome, whereas 77% consisted of genes that were shared by fewer than 15% of the strains analyzed (Supplementary Fig. 1). As the W-ExPEC strains possessed an average of 4729 genes, the core genes comprised roughly 45% of the W-ExPEC genome. Although this percentage is higher than typical estimates for *E. coli* core genomes⁴⁰, the majority of the strains analyzed either shared a similar ecological niche (i.e., treated wastewater matrices) or a presumed similar extraintestinal pathogenic potential (i.e., wastewater and clinical ExPEC).

To assess the similarity of wastewater and clinical ExPEC strains at the accessory genome level, all isolates were clustered based on

Table 1. Pairwise whole genome similarity and core genome SNP distances of W-ExPEC strains and the two closest clinical BBEC counterparts.

Strain name (sequence type)	Closest clinical BBEC	Core genome SNP differences ^c	Whole-genome similarity (%)	Strain name (sequence type)	Closest clinical BBEC	Core genome SNP differences ^c	Whole-genome similarity (%)
1F2A (ST131)	BBEC_268 (ST131)	6	97.80	1G6 (ST95) ^{a,b}	BBEC_41 (ST95)	30	97.38
	BBEC_267 (ST131)	5	97.37				
1G10A (ST131) ^a	BBEC_259 (ST131)	12	97.50	3B9 (ST95) ^{a,b}	BBEC_38 (ST95)	18	99.62
	BBEC_210 (ST131)	9	97.43		BBEC_35 (ST95)	24	98.48
2B4 (ST131)	BBEC_268 (ST131)	4	98.82	4G1 (ST95) ^{a,b}	BBEC_38 (ST95)	20	99.58
	BBEC_267 (ST131)	5	98.46		BBEC_35 (ST95)	26	98.44
2F5 (ST131) ^a	BBEC_156 (ST131)	5	98.82	4G9 (ST95) ^b	BBEC_41 (ST95)	15	97.00
2F6 (ST131) ^a	BBEC_54 (ST131)	7	98.54				
3E4 (ST131) ^a	BBEC_267 (ST131)	47	98.03	WU1033 (ST95) ^b	BBEC_41 (ST95)	23	96.51
	BBEC_280 (ST131)	51	97.48				
3G8 (ST131)	BBEC_267 (ST131)	6	97.41	WU1151 (ST95) ^b	BBEC_38 (ST95)	33	97.56
	BBEC_268 (ST131)	7	97.29		BBEC_162 (ST95)	21	97.49
3G9 (ST131) ^a	BBEC_211 (ST131)	3	98.80	WU1274 (ST95) ^b	BBEC_38 (ST95)	20	97.07
					BBEC_162 (ST95)	8	96.08
3G11 (ST131)	BBEC_211 (ST131)	9	96.89	WU1752 (unnamed) ^b	BBEC_273 (ST95)	29	96.80
					BBEC_283 (ST95)	22	96.47
4C1 (ST131) ^a	BBEC_202 (ST131)	8	98.96	WU3707 (ST95) ^b	BBEC_281 (ST95)	40	96.08
	BBEC_54 (ST131)	5	98.35				
4C7 (ST131) ^a	BBEC_267 (ST131)	49	98.71	3H3 (ST73)	BBEC_171 (ST73)	38	96.45
	BBEC_280 (ST131)	53	98.70		BBEC_151 (ST73)	34	96.35
4D7 (ST131) ^a	BBEC_202 (ST131)	5	98.22	4F9 (ST73)	BBEC_171 (ST73)	47	96.98
	BBEC_9 (ST131)	13	97.54		BBEC_151 (ST73)	43	96.72
5A5 (ST131)	BBEC_280 (ST131)	12	98.04	WU965 (ST73)	BBEC_5 (ST73)	15	97.80
	BBEC_267 (ST131)	8	96.85				
WU1030 (ST131)	BBEC_265 (ST131)	5	99.74	1F4 (ST10)	BBEC_285 (ST8453)	239	96.85
	BBEC_158 (ST131)	6	99.55				
WU1036 (ST131)	BBEC_265 (ST131)	4	99.12	2F12 (ST10)	BBEC_29 (ST10)	12	96.60
	BBEC_158 (ST131)	5	98.93		BBEC_225 (ST10)	11	96.54
WU1155 (ST131)	BBEC_265 (ST131)	5	99.66	WU1025 (ST127)	BBEC_69 (ST127)	10	97.57
	BBEC_158 (ST131)	6	99.47		BBEC_71 (ST127)	23	96.57
WU1265 (ST131)	BBEC_265 (ST131)	4	99.72	WU1157 (ST127)	BBEC_71 (ST127)	7	98.51
	BBEC_158 (ST131)	5	99.53				
WU1266 (ST131)	BBEC_265 (ST131)	5	99.74	WU1149 (ST357)	BBEC_141 (ST357)	11	96.29
	BBEC_158 (ST131)	6	99.55		BBEC_131 (ST357)	14	96.10
2H7 (ST44)	BBEC_88 (ST44)	18	97.49	WU664 (ST538)	BBEC_28 (ST538)	34	98.90

^aWastewater strains with additional clinical BBEC matches than described in the table, based on a lower whole genome similarity threshold of 96.03% for pathotype similarity.

^bWastewater strains with clinical matches to both BBEC and NMEC strains.

^cCore genome size is ~417 kbp.

Table 2. Pairwise whole genome similarity and core genome SNP distance matrix of W-ExPEC strains and their closest clinical NMEC counterparts.

Strain name (sequence type)	Closest clinical BBEC	Core genome SNP difference ^c	Whole-genome similarity (%)
1G6 (ST95) ^{a,b}	NMEC_9 (ST95)	7	99.72
	NMEC_24 (ST95)	36	98.51
3B9 (ST95) ^b	NMEC_4 (ST95)	13	99.59
4G1 (ST95) ^b	NMEC_4 (ST95)	15	99.55
4G9 (ST95) ^{a,b}	NMEC_28 (ST95)	24	97.37
	NMEC_10 (ST95)	38	97.35
WU1033 (ST95) ^{a,b}	NMEC_10 (ST95)	28	98.20
	NMEC_24 (ST95)	29	97.31
WU1151 (ST95) ^b	NMEC_4 (ST95)	28	97.18
WU1274 (ST95) ^b	NMEC_4 (ST95)	15	97.01
WU1752 (ST95) ^b	NMEC_12 (ST2619)	20	97.75
	NMEC_26 (ST95)	16	97.06

^aWastewater strains with additional clinical NMEC matches than described in the table, based on a lower whole genome similarity threshold of 96.03% for pathotype similarity.

^bWastewater strains with clinical matches to both BBEC and NMEC strains.

^cCore genome size is ~417 kbp.

the binary presence and absence of accessory genes within the estimated pan-genome. According to the generated clustering tree, the isolates were grouped into three main clusters (Fig. 3a). The most basal cluster (Cluster 1) included *E. coli* isolates that did not share any significant similarities with any clinical ExPEC, wastewater ExPEC, laboratory reference or naturalized wastewater *E. coli* strains. As this cluster included strains harboring comparatively large numbers of virulence genes (Fig. 3b), Cluster 1 may represent enteric *E. coli* strains, which would be expected to be found in wastewater and treated sewage. In contrast, the naturalized wastewater *E. coli* (all ST635 strains), wastewater ExPEC, and clinical ExPEC strains were distributed throughout Clusters 2 and 3. Cluster 2 isolates were distributed in two main subclusters designated 2a and 2b (Fig. 3a), with 2 of the 86 sequenced wastewater *E. coli* isolates (2F11 and 1H6) grouping with previously characterized naturalized wastewater *E. coli* strains including WW223, WW10, and ABWA45^{17,39}. With no clinical ExPEC matches in the local repository, 2F11 and 1H6 may represent additional naturalized wastewater *E. coli* strains. Subcluster 2b consisted of a small group of W-ExPEC strains (4H1, 2C8, 2F12, 2H7, and 1F4) that grouped with clinical ExPEC strains (BBEC_29, BBEC_88, BBEC_285 and the UPEC strain *E. coli* 219) representing relatively minor ExPEC lineages including ST10 and ST44 (Fig. 3a). Although some wastewater strains (4E10 and 4G8) present in this cluster did not have a direct clinical match, this could be due to the under-representation of other minor ExPEC lineages in the local repository. The last and largest cluster included most of the wastewater *E. coli* isolates and their closest clinical NMEC and BBEC matches representing the major ExPEC lineages, including ST131, ST95, ST73, and ST127, as well as other important ExPEC lineages such as ST538 and ST357²⁶. While most W-ExPEC strains clustered closely with their closest clinical BBEC or NMEC counterpart in this analysis, some strains grouped closer to a reference UPEC strain, including WU664 (W-ExPEC) with U059 (C-ExPEC) and 4F9 (W-ExPEC) with CFT073 (C-ExPEC), suggesting that some of these wastewater isolates are more likely to exhibit a uropathogenic phenotype²³. Interestingly, even W-ExPEC strains in Cluster 3 without a clinical match in this analysis were found to group within clusters dominated by clinical ExPEC strains (Fig. 3a).

Virulence and antibiotic resistance gene composition of W-ExPEC and C-ExPEC strains

To better clarify the pathogenic potential of W-ExPEC strains, the composition of virulence genes (VG) and antibiotic resistance genes (ARG) were compared between the wastewater strains and their clinical counterparts. Virulence gene composition was roughly bimodally distributed across the strains analyzed (Fig. 3b), with subcluster 2a (i.e., naturalized wastewater strains) harboring the fewest virulence genes compared to the flanking strains in Clusters 1 and 3. Interestingly, the comparatively low abundance of virulence genes in this subcluster supports previous hypotheses that these naturalized wastewater strains are non-pathogenic and have become endogenous to wastewater treatment plants¹⁷. In contrast, each W-ExPEC strain with at least one close clinical ExPEC counterpart in this analysis possessed an extensive VG repertoire, ranging from 195 to 278 VGs (Supplementary Table 5). W-ExPEC strains from sequence types ST95, ST127, and ST73 consistently harbored the highest number of VGs, though there did not appear to be any discernable pattern of annotated VGs that could consistently discriminate a particular ExPEC ST lineage (Supplementary Table 5). Importantly, none of the 37 W-ExPEC strains possessed VGs characteristic of the major *E. coli* intestinal pathotypes, including the enteropathogenic *E. coli* (EPEC)-associated *eaeA* gene⁴¹, the Shiga-toxin producing *E. coli* (STEC)-associated *stx1* and *stx2* genes⁴², and the enterohaemorrhagic *E. coli* (EHEC) O157:H7-associated *rfbE* gene⁴³, suggesting that the W-ExPEC strains, if pathogenic, specifically possess extraintestinal pathogenic potential.

On a pairwise basis, the wastewater strains generally harbored a very similar repertoire of virulence genes as their clinical counterparts (Supplementary Table 5). For instance, the ST131 W-ExPEC strain 3G11 and clinical strain BBEC_211 shared 214 VGs, only differing by one VG uniquely harbored in each strain. Indeed, only 3G11 possessed *cah*, which encodes a calcium-binding and heat-extractable autotransporter protein associated with biofilm formation and colonization⁴⁴, while BBEC_211 uniquely harbored the *int* gene, encoding an integrase. Similar observations were made for various ST95 W-ExPEC and C-ExPEC strains. For example, while the wastewater strains 3B9 and 4G1 shared an astounding 276 virulence genes with the neonatal meningitic *E. coli* strain NMEC_4, the wastewater strains additionally possessed *neuE*, which is thought to encode for a K1 polysialic acid capsule biosynthesis protein, whereas NMEC_4 uniquely harbored *aaiW*, an uncharacterized protein.

In terms of antibiotic resistance genes, W-ExPEC strains possessed anywhere from 43 to 56 ARGs, though generally, their clinical counterparts harbored more. In particular, the ST131 W-ExPEC strains possessed the highest number of ARGs, followed by ST44 and ST73 strains (Supplementary Table 6). Interestingly, several ARGs of major clinical importance were found in the W-ExPEC strains. For instance, aminoglycoside modification enzymes (AME), representing the most common and important resistance mechanism against aminoglycoside antibiotics^{45,46}, were well-represented amongst the wastewater strains. Indeed, 7 W-ExPEC ST131 strains isolated from treated wastewater effluents (2B4, 1G10A, 3E4, 4C7, 3G8, 3G9, 4C1) harbored the *AAC(3)-Ild* gene, whereas five ST131 strains isolated from chlorinated sewage (WU1030, WU1036, WU1155, WU1265, WU1266) harbored both *AAC(3)-Ile* and *AAC(6')-Ib-cr*. Other W-ExPEC strains across the ST131 (2F5, 2F6, 2B4, 1G10A, 3E4), ST95 (1G6), ST73 (3H3, 4F9), and ST10 (2F12) lineages were also found to harbor *APH(3'')-Ib* and *APH(6)-Id*. Beyond AME genes, several W-ExPEC strains also possessed various beta-lactamase genes, including the ST131 W-ExPEC strains WU1030, WU1036, WU1155, WU1265, and WU1266, which all possessed *ampC*, *ampH*, *bla_{CTX-M-15}* and *bla_{OXA-1}*; W-ExPEC strain 4C1, which possessed *ampC*, *ampH*, *bla_{CTX-M-15}* and *bla_{TEM-181}*; and W-ExPEC

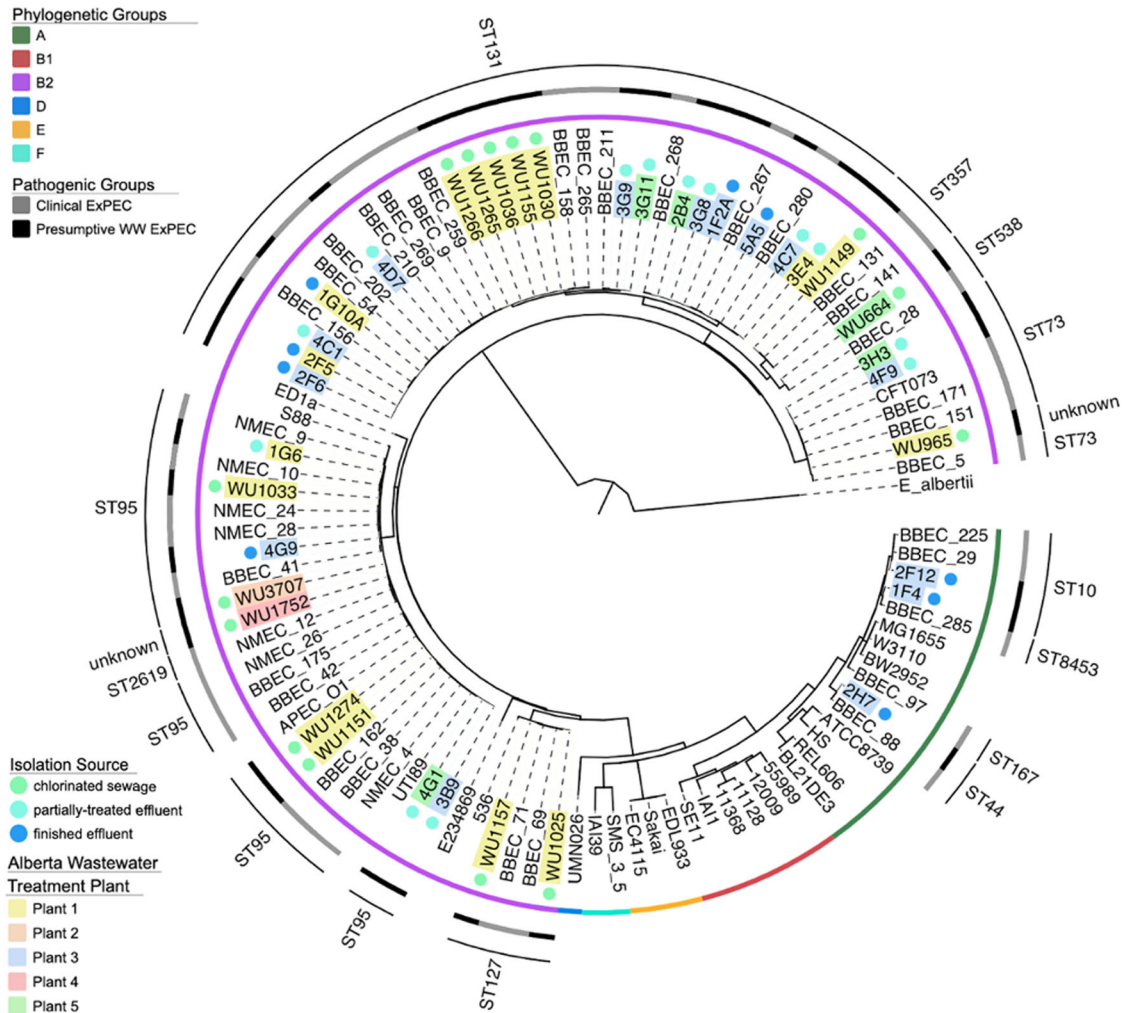


Fig. 2 Core genome maximum likelihood phylogenetic tree of wastewater ExPEC strains, their closest clinical counterparts and *E. coli* strains of known phylogroups. Presumptive wastewater ExPEC strains were collected from samples of chlorinated sewage, partially-treated wastewater effluent and finished wastewater effluent (indicated by the colored circles) from five wastewater treatment plants across Alberta, Canada (highlighted according to the lower legend). The core genome sequence variation of wastewater ExPEC strains (colored black in the outer circle) was compared to their closest clinical counterparts at the whole and core genome levels (colored gray in the outer circle) and various *E. coli* strains of known phylogroups (inner colored circle according to the upper legend). The main sequence type lineages of the wastewater and clinical ExPEC strains are indicated in the outermost circle. The tree is rooted with *E. albertii* as the outgroup.

strain 5A5, which harbored *ampC*, *ampH*, *bla*_{CTX-M-14} and *bla*_{OXA-1}. As these strains possessed various combinations of beta-lactamases, they may represent ESBL-producing *E. coli*, supporting previous analyses²³. Additional ARGs of clinical concern represented amongst the W-ExPEC strains include the sulfonamide-resistance genes *sul1*, *sul2* and *sul3*⁴⁷, and the tetracycline-efflux genes *tet(A)* and *tet(B)*⁴⁸.

On a pairwise basis, while some W-ExPEC strains differed from their corresponding C-ExPEC strains by as many as 11 antibiotic resistance genes, others shared incredibly similar ARG profiles. For instance, the ST131 W-ExPEC strains 2F5 and 2F6 differed from BBEC_156 by only 1 ARG, *bla*_{CTX-M-9} (Supplementary Table 6), whereas the ST95 W-ExPEC strains 3B9, 4G1, WU1151 and WU1274, shared the same repertoire of ARGs as the clinical strain BBEC_38 aside from *bla*_{TEM-181} (Supplementary Table 6). Remarkably, some wastewater strains shared an identical ARG composition with their clinical counterparts, including the W-ExPEC isolate 3E4 with BBEC_267; W-ExPEC strain 3G9 with BBEC_211; and W-ExPEC strains 3B9 and 4G1 with NMEC_4 (Supplementary Table 6).

DISCUSSION

The remarkable similarity between W-ExPEC and C-ExPEC strains across the core, whole, and accessory genome suggests that many strains surviving wastewater treatment may be highly pathogenic with septicemic and meningitic potential. Indeed, growing evidence indicates that, across the wastewater treatment train, certain pathogens appear to differentially survive the disinfection processes designed to eliminate them. Reflecting this, *E. coli* strains harboring UPEC-associated virulence genes and pathogenicity islands have been recovered from treated wastewater^{19–23}, making up anywhere from 40%²² to 60% of the surviving *E. coli* population following wastewater treatment^{20,21}. Concerningly, some strains isolated from treated wastewater matrices have even been found to belong to clinically relevant ExPEC lineages, including the pandemic O25b-ST131 clonal group²³. While this collectively indicates that UPEC strains have specifically evolved resistance to water treatment, growing evidence suggests that this observation may extend to other ExPEC pathotypes. Indeed, virulence gene screening has previously identified potential ExPEC strains in treated effluents⁴⁹, and it has been estimated that up to ~15% of *E. coli* strains

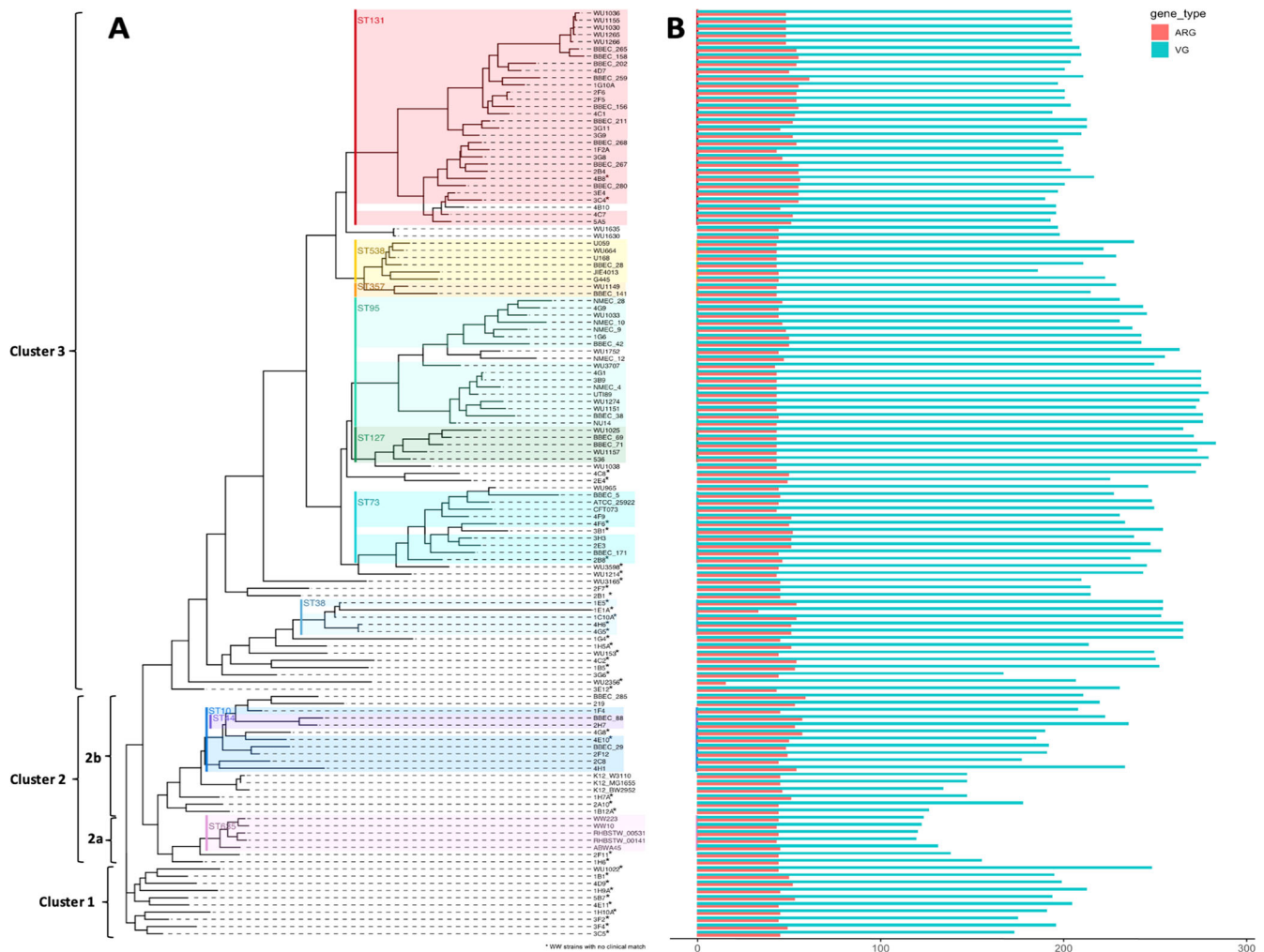


Fig. 3 Accessory genome clustering and virulence and antibiotic resistance gene screening of wastewater and clinical ExPEC strains. Binary accessory gene presence-absence clustering tree **A** of all original presumptive wastewater ExPEC strains ($n = 86$) and including those closely matching clinical ExPEC strains ($n = 37$ of the 86), as well as clinical ExPEC strains, naturalized wastewater *E. coli* strains, and laboratory reference strains. Included in this analysis are the number of antibiotic resistance genes (ARG) and virulence genes (VG) present in each isolate (**B**). The major sequence types represented are indicated in highlighted boxes in the clustering tree. Any original presumptive wastewater ExPEC strains that did not exhibit high genetic similarity with a clinical ExPEC strain at the whole and core genome levels are indicated in the clustering tree with an asterisk (*).

surviving wastewater treatment represent potential NMEC based on the presence of the *ibeA* gene²².

Unfortunately, virulence gene screening approaches are limited for characterizing potential ExPEC in wastewater since no single VG (or set of VGs) clearly differentiates the ExPEC pathotypes⁵⁰. This is further complicated by the identification of ExPEC virulence genes within commensal strains of *E. coli*^{51,52}, suggesting that these genes are generally required for adhesion and survival by all *E. coli* within the gut. Despite this, several recent comparative whole-genome studies^{23,24}, including the present study, provide very compelling evidence to suggest that ExPEC are particularly adept at surviving wastewater treatment, including strains known to cause urinary tract infections, septicemia and meningitis in humans. Concerningly, these surviving strains also appear to carry an abundance of ARGs that may confer resistance to antibiotics of concern for human medicine, and may include ESBL-producing strains^{23,53,54}.

In this analysis, of the 637 *E. coli* isolates collected from chlorinated sewage and treated wastewater, 86 presumptive ExPEC isolates were identified according to a virulence gene screening panel. Multilocus sequence typing clustered the W-ExPEC isolates into several major ExPEC-associated lineages including ST131, ST95,

and ST73, a finding observed in other studies^{19,23,25,55}. Several other ExPEC-associated sequence types of growing concern were also represented among the isolates, including ST10, ST44, ST127, ST357, and ST538²⁶. As several sequence types were represented in the surviving *E. coli* populations recovered from chlorinated sewage and full-scale treated effluents, water treatment resistance may have independently evolved in multiple ExPEC lineages. Through additional comparative genomic approaches, several wastewater isolates were found to be virtually identical to a clinical BBEC or NMEC strain across the core, whole, and accessory genome, suggesting *E. coli* strains surviving wastewater treatment may possess the capacity to cause extraintestinal disease. For instance, compared to the septicemic strain BBEC_156, wastewater strains 2F5 and 2F6 shared 98.82% whole genome similarity and differed by only 5 SNPs across a ~417 kbp core genome backbone. These wastewater strains also shared 203 VGs with BBEC_156 and harbored an ARG repertoire that differed only by the *bla*_{CTX-M-9} gene. Similarly, compared to BBEC_211, 3G9 shared 98.80% whole genome similarity, differed by only 3 SNPs across the core genome, and shared 212 VGs and an identical ARG repertoire. A remarkably high degree of genetic similarity was also observed between select W-ExPEC and clinical NMEC strains. Indeed, compared to NMEC_4,

wastewater strains 3B9 and 4G1 both shared over 99.5% whole genome similarity, differed by only 13 and 15 core genome SNPs respectively, while sharing 276 VGs and an identical ARG profile.

The observation that wastewater ExPEC strains demonstrate resistance to wastewater treatment, similar to the recently characterized naturalized wastewater *E. coli*, warrants further study. Naturalized wastewater *E. coli* strains appear to differentially survive the wastewater treatment process¹⁶, exhibit resistance to chlorine, heat, and oxidants^{15,18}, and possess an extensive repertoire of stress resistance genes compared to enteric strains¹⁷. Compared to ExPEC strains, naturalized wastewater strains possess far fewer virulence genes (Fig. 3); however, these strains do appear to share virulence genes in common with their clinical counterparts²³, suggesting that these 'virulence' factors may play an important role in survival in non-host environments. Reflecting this, the shared virulence genes between UPEC and naturalized wastewater strains include iron-transport proteins and siderophores that may facilitate the uptake of iron at dilute concentrations typically found in urine and wastewater²³. Considering that naturalized wastewater *E. coli* are also resistant to water treatment, the shared accessory genes between ExPEC and naturalized wastewater strains may explain the differential survival of ExPEC during water treatment^{19–23}. Indeed, the finding that naturalized wastewater *E. coli* and ExPEC strains share a common resistance to wastewater treatment and appear to exhibit some degree of similarity at the accessory genome raises questions concerning their relationship. Interestingly, previous whole genome phylogenetic analyses demonstrated that naturalized wastewater *E. coli* appear to have evolved from phylogroup A¹⁷, and it is notable that some ExPEC strains clustered within this phylogroup in this analysis (Fig. 2). While the exact evolutionary relationship between ExPEC and naturalized wastewater *E. coli* strains is outside the scope of this analysis, this finding does suggest that these distinct *E. coli* lineages may share a common evolutionary origin, offering an interesting direction for future study.

Regardless of the exact nature of the genetic relationship between naturalized wastewater *E. coli* and ExPEC strains, the underpinnings of their evolutionary emergence are critical to understanding from a public health viewpoint. A pre-requisite for the co-selection of water treatment resistance and pathogenesis in ExPEC strains requires that some level of sustained waterborne ExPEC transmission must occur, particularly from wastewater-contaminated environments. Presently, urinary tract infections (UTIs) account for more than 10 million physician visits each year in the US, of which the vast majority are caused by UPEC^{56–58}. As with many clinical infections, the number of reported cases likely underestimates the true prevalence of the health issue. Consequently, UTI rates in the community are likely higher than expected, which both: (a) exerts selective pressure on the pathogenesis (i.e., virulence genes) of ExPEC pathotypes in the population; and (b) provides a constant influx of ExPEC into wastewater treatment plants. Once within the wastewater treatment plant, microbial disinfection processes likely drive the selection of treatment resistance in ExPEC; however, and most importantly, the ability to maintain both phenotypes of water treatment resistance and pathogenesis requires that ExPEC surviving the wastewater treatment processes must cycle back into the human population. Importantly, recreational exposure to contaminated natural water bodies (i.e., swimming) has already been epidemiologically linked to an increased prevalence of UTIs, largely caused by UPEC⁵⁹. Considering that ESBL *E. coli* isolated from clinical, wastewater and recreational water samples from the same geographical location have been shown to be of the same clonal lineage⁶⁰, wastewater-contaminated recreational water may be an important route of ExPEC transmission.

Concerningly, the impact of wastewater contamination on ExPEC transmission may reach far beyond recreational water

sources. It has been estimated that 50% of drinking water treatment plants in the US are impacted by wastewater effluents and that wastewater discharge may account for 50% of the volume of water flowing in US rivers during the year, which can reach as high as 90% under low streamflow conditions^{61,62}. This is important to consider as these rivers may be used for drinking, irrigation, and recreational purposes, providing further routes of transmission for pathogens back into the population. Indeed, ESBL-producing Enterobacteriaceae, including *E. coli*, have been found in 6.4% of drinking water samples that failed bacteriological drinking water quality parameters in the US, suggesting that drinking water may be an underestimated vehicle for transmission of ExPEC in the community⁶³. Similarly, Paulshus and colleagues demonstrated that 10% of *E. coli* isolates recovered from community wastewater pumping stations represented ESBL-producing strains, of which 44% clustered within sequence types ST131 and ST648, suggesting that wastewater systems may be contributing a considerable amount of antibiotic-resistant ExPEC disseminating throughout the community²⁴. Despite this, most research efforts have classically focused on linking ExPEC transmission to the food supply. Indeed, *E. coli* isolates harboring ExPEC virulence genes^{64,65}, as well as those phylogenetically linked to extraintestinal disease in the clinic⁶⁶, have been isolated from food products including retail meat, eggs, milk, and produce—implicating food as an important source of ExPEC transmission. The present analysis, however, adds to a growing body of research that indicates that water may represent an important, yet overlooked, source of exposure for these pathogens^{67–69}. As far as we are aware, no epidemiological studies have examined the relationship between exposure to or consumption of contaminated water with meningitis or septicemia caused by ExPEC. In light of these findings, we would strongly encourage such studies to be undertaken.

Interestingly, several wastewater strains that were previously found to share high genetic similarity with clinical UPEC strains²³ also exhibited a high degree of similarity with a clinical BBEC or NMEC counterpart, emphasizing the importance of including a wide representation of relevant ExPEC isolates for comparative genomic purposes. For instance, while wastewater isolates WU1030, WU1036, WU1155, WU1265, and WU1266 each exhibited >96.03% similarity to a clinical UPEC²³, the degree of whole-genome similarity between these isolates and the septicemic strain BBEC_265 exceeded 99% (Table 1). Conversely, some wastewater isolates representing potential BBEC or NMEC strains in this study shared greater similarity with clinical UPEC counterparts in previous analyses²³. As these findings demonstrate, comparative genomic studies between wastewater and clinical ExPEC strains require a comprehensive pathotype database—and for this reason, we also believe that this study likely underestimated the occurrence of ExPEC strains present in wastewater effluents and chlorinated sewage.

In the present study, presumptive ExPEC strains were initially identified through virulence gene screening, which included sequence type markers for the ST131 lineage. While the use of virulence genes to characterize ExPEC strains is in line with other studies^{19–22,49,54,65}, no single gene marker clearly defines this pathotype, indicating our initial screen likely missed many potential ExPEC strains present in wastewater. Furthermore, although our PCR screen was optimized for strains in the ST131 pandemic lineage³⁷, and specifically the prominent O25b-ST131 clonal group⁶⁶, several other emerging ExPEC lineages have since been identified, including ST95, ST73, ST10, and ST127²⁶. Targeting ST131 in our assay, although not to the exclusion of non-ST131 strains, likely led to some bias in the selection process and the subsequent under-representation of other ExPEC sequence types in the final repository. In addition, although we created a relatively large local genome database of clinical BBEC and NMEC strains for comparative genomic purposes ($n = 320$),

there are limitations regarding just how many sequences can be reasonably compared. Thus, although several wastewater strains did not share >96.03% whole genome similarity with a clinical BBEC or NMEC strain in our study, these strains may still represent potential ExPEC. Indeed, several isolates that lacked a clinical match in this study still clustered within ExPEC-relevant sequence types, including ST131, ST73, and ST10 (Supplementary Table 2), and grouped within clinical ExPEC-dominated clusters (Fig. 3). Consequently, we believe that including more clinical strains into our repository would have led to the identification of a greater number of W-ExPEC strains, supporting the idea that ExPEC comprises a significant proportion of the viable *E. coli* found in treated wastewater effluents.

The prospect that pathogenic *E. coli* may have evolved resistance against water treatment is concerning, particularly considering that water treatment represents the single greatest intervention for control of infectious diseases in modern society. This is further compounded by the observation that many of these strains are also resistant to antibiotics. Importantly, this finding raises the question—if pathogenic *E. coli* are evolving resistance to water treatment, could this be occurring with other pathogenic microbes? The wastewater microbiome represents an incredibly diverse microbial community⁷⁰ upon which the same selection pressures can act. Indeed, several studies suggest a common paradigm for the evolution of water treatment resistance in various extraintestinal pathogenic taxa, including in *Legionella*⁷⁰, *Acinetobacter*⁷¹, and *Mycobacterium*⁷². Alongside these pathogenic genera, others including *Roseomonas*, *Pseudomonas*, *Aeromonas*, *Yersinia*, and *Escherichia*, have been found to be particularly abundant across all stages of wastewater treatment, even following disinfection⁷³. With the growing burden of various extraintestinal diseases worldwide, future research is needed to evaluate the risks associated with the potential waterborne transmission of ExPEC and other pathogens that might be evolving resistance to water treatment.

In the present study, wastewater *E. coli* strains recovered from chlorinated sewage and finished effluents were found to exhibit an extremely high degree of genetic similarity across the whole, core, and accessory genome to clinical BBEC and NMEC strains. As select wastewater strains were found to be virtually identical to a clinical BBEC or NMEC counterpart, the evidence suggests that clinically relevant BBEC and NMEC strains may readily survive municipal wastewater treatment processes. Taken together with previous findings suggesting that clinically relevant UPEC strains comprise a significant proportion of the surviving wastewater *E. coli* population following water treatment, ExPEC pathotypes overall may have evolved a differential capacity to resist water treatment disinfection. Considering the massive microbial diversity present in wastewater, these findings raise the prospect that other extraintestinal pathogenic microbes may be following a similar evolutionary path towards water treatment resistance. These findings prompt further research into environmental persistence and prevalence of these pathogens following wastewater treatment, and the public health risks associated with exposure to downstream impacted water environments—raising the concerning possibility that extraintestinal diseases such as urinary tract infections, septicemia, and meningitis could be transmissible through waterborne routes of exposure.

METHODS

Recovery of *E. coli* isolates from chlorinated sewage and treated effluents

To reflect the variability that exists in the disinfection processes utilized across wastewater treatment programs, both chlorine-stressed and water treatment-resistant *E. coli* isolates were collected for this analysis. Chlorine-stressed isolates were obtained from chlorinated sewage according to the United States Environmental Protection Agency's (U.S. EPA) Alternate Test

Procedure⁷⁴. Briefly, raw sewage samples obtained from 10 different treatment plants across Alberta, Canada, were treated with 3% sodium hypochlorite with a free chlorine residual of 0.3–0.5 ppm. Chlorine dose and contact time reduced the culturable *E. coli* population by ~4 log₁₀, as estimated by the Colilert QuantiTray[®] system (IDEXX Laboratories, Inc.), and verified in parallel samples. Following chlorine treatment, residual reactivity was neutralized with a 10% solution of sodium thiosulfate. The chlorine-treated wastewater samples were then inoculated into either ColiTag[®] or lauryl trypticase broth/BCG media as outlined by Method 9221.F in *Standard Methods for the Examination of Water and Wastewater*⁷⁵. Positive ColiTag[®] and LB/BCG cultures were then plated onto X-Gluc agar plates and incubated at 44.5 °C for 24 h to selectively grow *E. coli*. Blue colonies were picked and streaked onto non-selective blood agar plates and incubated at 35 °C for 24 h. All presumptive *E. coli* isolates were biochemically confirmed as *E. coli* using a Vitek[®] 2 Automated Bacterial Identification System (BioMerieux, St. Laurent, Quebec, Canada) according to the manufacturer's instructions. The resulting library of isolates represented the chlorine-resistant population of *E. coli*.

Wastewater treatment-resistant *E. coli* isolates were obtained from partially treated (i.e., secondary-treated) and finished effluents derived from three full-scale municipal wastewater treatment plants (WWTPs) in Calgary, Alberta. The wastewater treatment processes utilized in these treatment plants consist of grit removal, primary clarification, activated sludge, secondary clarification and UV disinfection at low or medium pressure doses of 25–30 mJ/cm² at peak flow, with maximal treatment capacities ranging from 140 to 1020 ML a day. To obtain the wastewater treatment-resistant isolates, samples of wastewater effluent were processed by standard membrane filtration (100 mL) onto X-Gluc agar plates. Blue colonies representing putative *E. coli* were picked, inoculated into a 96-well plate containing 100 µL of 1X Luria-Bertani (LB) broth and incubated overnight without shaking at 37 °C. A total of 1212 isolates were collected from wastewater effluent matrices, of which a random collection of 261 isolates were confirmed as *E. coli* using a Vitek[®] 2 Automated Bacterial Identification System and selected for further analysis.

Analysis of ExPEC-related virulence genes and molecular markers

All selected *E. coli* isolates were grown in TSB overnight at 37 °C and their genomic DNA (gDNA) extracted from the TSB culture using DNeasy Blood & Tissue kits (QIAGEN, Toronto, Canada) according to the manufacturer's instructions. A total of 376 presumed chlorine-tolerant isolates and 261 water treatment-resistant isolates were assayed against the *uspC-IS30-flhDC* marker by PCR to eliminate naturalized wastewater *E. coli* isolates from the library. The remaining isolates in the library were subsequently screened against the *uidA* gene to further confirm isolates as *E. coli*, and all confirmed isolates were then screened against a panel of ExPEC-associated virulence genes and molecular markers, as previously described²³. All PCR panel gene targets and primers used in this study are provided in Supplementary Table 7. Purified genomic DNA was quantified after extraction using the Qubit fluorimeter (Thermo Fisher Scientific Inc.), and all PCR reactions were performed on an ABI 2720 thermocycler (Applied Biosystems). The PCR reactions for the *uidA*⁷⁶ and *uspC-IS30-flhDC*¹⁶ markers were carried out according to previously described protocols. For the other molecular markers, the reaction mixtures consisted of 20–40 ng of gDNA template, 12.5 µL of 1X GoTaq Hotstart Mastermix (Promega, Madison, WI), and 500 nM of each primer. Cycling conditions varied according to the specific molecular marker. For *papC*, *sfa-foc*, *iroN* and *ibeA*, cycling conditions were as follows: 95 °C for 2 min, followed by 33 cycles of 30 s at 95 °C, 30 s at 63 °C and 45 s at 72 °C, followed by a 7 min incubation at 72 °C. For *fyuA* and *chuA*: 95 °C for 2 min, followed by 33 cycles of 30 s at 95 °C, 30 s at 63 °C and 1 min at 72 °C, followed by a 7 min incubation at 72 °C. For *kpsM*: 95 °C for 2 min, followed by 35 cycles of 20 s at 95 °C, 20 s at 62 °C and 45 s at 72 °C, followed by a 7 min incubation at 72 °C. For the ST131 marker: 95 °C for 2 min, followed by 35 cycles of 20 s at 95 °C, 20 s at 57 °C and 45 s at 72 °C, followed by a 7 min incubation at 72 °C. For the O25b–ST131 marker: 95 °C for 4 min, followed by 30 cycles of 5 s at 94 °C and 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). PCR screening results for the ExPEC-associated virulence genes were then confirmed through whole-genome sequencing for each wastewater *E. coli* isolate (see below).

Whole-genome sequencing and assembly of presumptive wastewater ExPEC strains

Chlorine-tolerant and wastewater treatment-resistant *E. coli* isolate harboring at least 3 ExPEC virulence genes as well as any isolates positive for the NMEC-associated *ibeA* gene or ExPEC-associated ST131 genetic marker (regardless of the presence of other virulence genes) were selected for whole-genome sequencing as presumptive wastewater ExPEC (W-ExPEC) strains. Genomic DNA from the presumptive W-ExPEC isolates ($n = 86$) was sent to Genome Quebec (Montreal, Canada) for sequencing using an Illumina HiSeq X platform (Illumina) with paired-end 150 nucleotide reads. Trimmomatic Version 0.39⁷⁷ was used to trim the low-quality reads with the following parameters: SLIDINGWINDOW = 4:15, LEADING = 3, TRAILING = 3, MINLEN = 36. De novo genomic assembly was then performed using SPAdes Version 3.11.1⁷⁸ with the ‘—careful’ and ‘-k 21,33,55,77’ options. Any SPAdes-assembled contigs shorter than 1000 bp were excluded from downstream analyses.

Core genome SNP analysis of W-ExPEC and C-ExPEC isolates

To evaluate the likelihood of our presumptive W-ExPEC isolates being pathogenic, whole genomes were compared against a genomic library of clinical ExPEC (C-ExPEC) strains associated with either blood-borne bacteremia (BBEC) or meningitis (NMEC). First, 320 clinical NMEC and BBEC isolates were downloaded from NCBI databases to construct a local repository of clinical ExPECs (C-ExPECs) (Supplementary Table 4). Core genome SNP analyses were performed for each of the W-ExPEC ($n = 86$) genomes and against the local repository of 320 C-ExPEC strains using REALPHY v1.13⁷⁹. Whole-genome assemblies of all W-ExPEC and C-ExPEC isolates ($n = 406$) were collected into a local folder to be used as input for REALPHY. One of the *E. coli* assemblies (strain 3G6) in the input REALPHY folder was randomly selected to be the reference sequence, against which the genome sequences of all other isolates were mapped to produce a core genome alignment for SNP assessment. MEGA-X v10.1.0⁸⁰ was then used to determine the number of pairwise SNPs differences amongst all 406 strains in this study.

Pairwise whole-genome comparisons of W-ExPEC and C-ExPEC isolates

W-ExPEC strains that were identified as sharing a high core genome similarity with at least one clinical BBEC or NMEC counterpart—based on having fewer than 250 differing SNPs in a ~417 kbp core genome backbone—were selected for further comparative genetic analyses ($n = 37$). In this case, whole-genome approaches were used to evaluate the overall genetic relatedness between W-ExPEC strains and their closest clinical counterpart.

First, pairwise whole-genome comparisons were performed between the 37 selected W-ExPEC strains and the local repository of 320 C-ExPEC strains using REALPHY v1.13⁷⁹ with default parameters. The whole-genome similarity for each pairwise comparison was estimated by REALPHY v1.13 based on the percentage of each C-ExPEC genome that mapped onto each W-ExPEC genome as a reference. A previously established upper median value for whole-genome sequence similarity between highly diverse STEC O157:H7 *E. coli* isolates²³ was set as a threshold for further evaluating the genetic similarity between the W-ExPECs and C-ExPECs. This threshold of 96.03% served as the lower limit for identifying W-ExPEC strains that shared a high degree of genetic similarity, and thus a presumed similar pathogenic potential, with clinical ExPEC strains.

Second, additional pairwise whole-genome comparisons were performed for each identified W-ExPEC ($n = 37$) against a local repository of 46 representative intestinal pathogenic *E. coli* and naturalized wastewater *E. coli* strains downloaded from NCBI (Supplementary Table 4). The W-ExPEC strains were compared to intestinal pathogenic strains (i.e., enterohaemorrhagic *E. coli* [EHEC], enteropathogenic *E. coli* [EPEC], enterotoxigenic *E. coli* [ETEC], etc.) to rule out the possibility that any significant similarities observed between the wastewater and clinical ExPEC strains could be explained by a shared propensity for general pathogenicity and not true extraintestinal pathogenicity, especially since certain intestinal pathotypes such as ETEC have been recovered from wastewater matrices after primary and secondary treatment⁸¹. Naturalized wastewater strains were also included in the analysis to assess their degree of similarity with W-ExPEC strains, controlling for the genetic similarity that might be driven through sharing a similar ecological niche (i.e., sewage/wastewater).

Core genome phylogenetic analysis, phylogrouping, and multilocus sequence typing of W-ExPEC and C-ExPEC isolates

W-ExPEC strains that shared $\geq 96.03\%$ similarity at the whole genome level with at least one C-ExPEC strain were selected for further analysis. A maximum-likelihood phylogenetic tree was generated using RAXML 8.2.12⁸² based on the core genome alignments produced by REALPHY v1.13 between W-ExPEC strains ($n = 37$) and their closest clinical counterparts ($n = 38$), as well as *E. coli* strains of known phylogroups^{28–30}.

The phylogroups of the W-ExPEC and C-ExPEC strains were predicted with the ClermonTyping method, using the ClermonTyper (version 21.03) webserver⁸³. Multilocus sequence typing (MLST) analyses were also performed on all presumptive W-ExPEC strains and clinical ExPEC strains using mlst 2.19.0 (<https://github.com/tseemann/mlst>) with the *Escherichia coli* #1 scheme. All information pertaining to the bacterial strains included in the phylogenetic analysis can be found in Supplementary Table 4. The phylogenetic tree was then visualized and annotated using the R packages ggplot2 version 3.3.3⁸⁴, ape version 5.4.1⁸⁵ and ggtree version 2.2.4⁸⁶.

Pan-genomic analysis, accessory genome clustering tree, and whole genome screening of virulence and antibiotic resistance genes

A pan-genomic analysis of all presumptive W-ExPEC strains ($n = 86$), clinical NMEC and BBEC strains representing the closest clinical counterparts, reference UPEC strains, naturalized wastewater strains, and laboratory reference strains was performed using Roary 3.13.0⁸⁷. Reference UPEC strains and naturalized wastewater strains were included in the analysis, as previous studies have demonstrated that these subpopulations can be isolated from chlorinated sewage and finished wastewater effluents^{16,17,23}. Laboratory reference *E. coli* strains (i.e., *E. coli* K12 MG1655, *E. coli* K12 W3110, *E. coli* K12 BW2592, and *E. coli* ATCC 25922) were included in the analysis to represent a non-pathogenic reference subgroup. All original presumptive W-ExPEC isolates identified were included, including any presumptive W-ExPEC strains without clinical NMEC and BBEC matches as they may still represent strains with extraintestinal pathogenic potential that lacked a suitable clinical counterpart in the repository used for this analysis. An accessory genome phylogenetic analysis was then conducted based on the binary presence and absence of accessory genes within the calculated pan-genome, as determined by Roary 3.13.0⁸⁷. The accessory genome phylogenetic tree was then visualized and annotated using the R packages ggplot2 version 3.3.3⁸⁴, ape version 5.4.1⁸⁵, and ggtree version 2.2.4⁸⁶.

The genomes of the selected W-ExPEC strains and their closest clinical counterparts were also screened for the presence of virulence genes and antibiotic resistance genes for comparison using ABRicate version 1.0.1 (<https://github.com/tseemann/abricate>), with a minimum query coverage of 90% and a minimum percent identity of $\geq 80\%$. For virulence gene identification, W-ExPEC and C-ExPEC genomes were screened against the Virulence Factor Database (VFDB; updated January 12, 2021)⁸⁸ and *Escherichia coli* Virulence Factor Database (Ecoli_vf; updated January 12, 2021) (https://github.com/phac-nml/ecoli_vf). For antibiotic resistance gene identification, *E. coli* genomes were screened against the Comprehensive Antibiotic Resistance Database (CARD; updated January 12, 2021)⁸⁹. The number of virulence and antibiotic resistance genes were then visualized and appended to the accessory genome tree.

DATA AVAILABILITY

All data needed to evaluate the conclusion in this paper are presented. The NCBI accession number of 62 *E. coli* genomes sequenced by this study include
 JAJLMO000000000, JAJLMQ000000000, JAJLMP000000000, JAJLMR000000000,
 JAJLMS000000000, JAJLMT000000000, JAJLMU000000000, JAJLMV000000000,
 JAJLMY000000000, JAJLMW000000000, JAJLMX000000000, JAJLND000000000,
 JAJLMZ000000000, JAJLNA000000000, JAJLNB000000000, JAJLNC000000000,
 JAJLNE000000000, JAJLNF000000000, JAJLNG000000000, JAJLNH000000000,
 JAJLNI000000000, JAJLNI000000000, JAJLNI000000000, JAJLNI000000000,
 JAJLNP000000000, JAJLNL000000000, JAJLNM000000000, JAJLNN000000000,
 JAJLNQ000000000, JAJLNR000000000, JAJLNS000000000, JAJLNT000000000,
 JAJLNU000000000, JAJLNW000000000, JAJLNV000000000, JAJLNX000000000,
 JAJLYO000000000, JAJLOC000000000, JAJLNZ000000000, JAJLOA000000000,
 JAJLOB000000000, JAJLOD000000000, JAJLOF000000000, JAJLOE000000000,
 JAJLOG000000000, JAJLOH000000000, JAJLOI000000000, JAJLOJ000000000,
 JAJLOK000000000, JAJLOL000000000, JAJLOM000000000, JAJLON000000000,
 JAJLOO00000000, JAJLOP000000000, JAJLOQ000000000, JAJLOR000000000,

JAJLOS000000000, JAJLOT000000000, JAJLOU000000000, JAJLOV000000000, JAJLOW000000000, JAJLOX000000000.

CODE AVAILABILITY

All codes are available from the corresponding author upon reasonable request.

Received: 15 February 2022; Accepted: 22 June 2022;

Published online: 20 July 2022

REFERENCES

- Prüss-Ustün, A. et al. Burden of disease from inadequate water, sanitation and hygiene for selected adverse health outcomes: an updated analysis with a focus on low- and middle-income countries. *Int. J. Hyg. Environ. Health* **222**, 765–777 (2019).
- Hunter, P. R., MacDonald, A. M. & Carter, R. C. Water supply and health. *PLoS Med.* **7**, 1–9 (2010).
- Cutler, D. & Miller, G. The role of public health improvements in health advances: the twentieth-century United States. *Demography* **42**, 1–22 (2005).
- Naik, K. S. & Stenstrom, M. K. Evidence of the influence of wastewater treatment on improved public health. *Water Sci. Technol.* **66**, 644–652 (2012).
- Trastoy, R. et al. Mechanisms of bacterial tolerance and persistence in the gastrointestinal and respiratory environments. *Clin. Microbiol. Rev.* **31**, 1–46 (2018).
- Landini, P., Egli, T., Wolf, J. & Lacour, S. sigmaS, a major player in the response to environmental stresses in *Escherichia coli*: Role, regulation and mechanisms of promoter recognition. *Environ. Microbiol. Rep.* **6**, 1–13 (2014).
- Gebendorfer, K. M. et al. Identification of a hypochlorite-specific transcription factor from *Escherichia coli*. *J. Biol. Chem.* **287**, 6892–6903 (2012).
- Gray, M. J., Wholey, W. Y., Parker, B. W., Kim, M. & Jakob, U. NemR is a bleach-sensing transcription factor. *J. Biol. Chem.* **288**, 13789–13798 (2013).
- Parker, B. W., Schwessinger, E. A., Jakob, U. & Gray, M. J. The RclR protein is a reactive chlorine-specific transcription factor in *Escherichia coli*. *J. Biol. Chem.* **288**, 32574–32584 (2013).
- Drazic, A. et al. Role of cysteines in the stability and DNA-binding activity of the hypochlorite-specific transcription factor HypT. *PLoS ONE* **8**, (2013).
- Imlay, J. A. Cellular defenses against superoxide and hydrogen peroxide. *Annu. Rev. Biochem.* **77**, 755–776 (2008).
- Gray, M. J., Wholey, W. Y. & Jakob, U. Bacterial responses to reactive chlorine species. *Annu. Rev. Microbiol.* **67**, 141–160 (2013).
- Winter, J., Ilbert, M., Graf, P. C. F., Özcelik, D. & Jakob, U. Bleach activates a redox-regulated chaperone by oxidative protein unfolding. *Cell* **135**, 691–701 (2008).
- Mercer, R., Nguyen, O., Ou, Q., McMullen, L. & Gänzle, M. G. Functional analysis of genes comprising the locus of heat resistance in *Escherichia coli*. *Appl. Environ. Microbiol.* **83**, 1–13 (2017).
- Wang, Z. et al. The locus of heat resistance confers resistance to chlorine and other oxidizing chemicals in *Escherichia coli*. *Appl. Environ. Microbiol.* **86**, 1–16 (2020).
- Zhi, S. et al. Evidence of naturalized stress-tolerant strains of *Escherichia coli* in municipal wastewater treatment plants. *Appl. Environ. Microbiol.* **82**, 5505–5518 (2016).
- Zhi, S. et al. Evidence for the evolution, clonal expansion and global dissemination of water treatment-resistant naturalized strains of *Escherichia coli* in wastewater. *Water Res.* **156**, 208–222 (2019).
- Zhi, S., Banting, G. S., Ruecker, N. J. & Neumann, N. F. Stress resistance in naturalised waste water *E. coli* strains. *J. Environ. Eng. Sci.* **12**, 42–50 (2017).
- Calhau, V., Mendes, C., Pena, A., Mendonça, N., & Da Silva, G. J. Virulence and plasmidic resistance determinants of *Escherichia coli* isolated from municipal and hospital wastewater treatment plants. *J. Water Health* **13**, 311–318 (2015).
- Anastasi, E. M. et al. Prevalence and persistence of *Escherichia coli* strains with uropathogenic virulence characteristics in sewage treatment plants. *Appl. Environ. Microbiol.* **76**, 5882–5886 (2010).
- Anastasi, E. M., Wohlsen, T. D., Stratton, H. M. & Katouli, M. Survival of *Escherichia coli* in two sewage treatment plants using UV irradiation and chlorination for disinfection. *Water Res.* **47**, 6670–6679 (2013).
- Adefisoye, M. A. & Okoh, A. I. Identification and antimicrobial resistance prevalence of pathogenic *Escherichia coli* strains from treated wastewater effluents in Eastern Cape, South Africa. *MicrobiologyOpen* **5**, 143–151 (2016).
- Zhi, S. et al. Characterization of water treatment-resistant and multidrug-resistant urinary pathogenic *Escherichia coli* in treated wastewater. *Water Res.* **182**, 115827 (2020).
- Paulshus, E. et al. Repeated isolation of extended-spectrum-lactamase-positive *Escherichia coli* sequence types 648 and 131 from community wastewater indicates that sewage systems are important sources of emerging clones of antibiotic-resistant bacteria. *Antimicrob. Agents Chemother.* **63**, 1–12 (2019).
- Raven, K. E. et al. Genomic surveillance of *Escherichia coli* in municipal wastewater treatment plants as an indicator of clinically relevant pathogens and their resistance genes. *Microb. Genomics* **5**, e000267 (2019).
- Riley, L. W. Pandemic lineages of extraintestinal pathogenic *Escherichia coli*. *Clin. Microbiol. Infect.* **20**, 380–390 (2014).
- Wu, Y. H. et al. The role of Sequence Type (ST) 131 in adult community-onset non-ESBL-producing *Escherichia coli* bacteraemia. *BMC Infect. Dis.* **14**, 1–7 (2014).
- Tenaillon, O., Skurnik, D., Picard, B. & Denamur, E. The population genetics of commensal *Escherichia coli*. *Nat. Rev. Microbiol.* **8**, 207–217 (2010).
- Touchon, M. et al. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genet.* **5**, e1000344 (2009).
- Zhang, Y. & Lin, K. A phylogenomic analysis of *Escherichia coli/Shigella* group: Implications of genomic features associated with pathogenicity and ecological adaptation. *BMC Evol. Biol.* **12**, 174 (2012).
- Johnson, J. R. & Russo, T. A. Extraintestinal pathogenic *Escherichia coli*: The other bad *E. coli*. *J. Lab. Clin. Med.* **139**, 155–162 (2002).
- Picard, B. et al. The link between phylogeny and virulence in *Escherichia coli* extraintestinal infection? *Infect. Immun.* **67**, 546–553 (1999).
- Escobar-Páramo, P. et al. Identification of forces shaping the commensal *Escherichia coli* genetic structure by comparing animal and human isolates. *Environ. Microbiol.* **8**, 1975–1984 (2006).
- Li, B. et al. Phylogenetic groups and pathogenicity island markers in fecal *Escherichia coli* isolates from asymptomatic humans in china. *Appl. Environ. Microbiol.* **76**, 6698–6700 (2010).
- Hutton, T. A. et al. Phylogroup and virulence gene association with clinical characteristics of *Escherichia coli* urinary tract infections from dogs and cats. *J. Vet. Diagnostic Investig.* **30**, 64–70 (2018).
- Micenková, L., Bosák, J., Vrba, M., Ševčíková, A. & Šmajš, D. Human extraintestinal pathogenic *Escherichia coli* strains differ in prevalence of virulence factors, phylogroups, and bacteriocin determinants. *BMC Microbiol.* **16**, 1–8 (2016).
- Nicolas-Chanoine, M. H., Bertrand, X. & Madec, J. Y. *Escherichia coli* ST131, an intriguing clonal group. *Clin. Microbiol. Rev.* **27**, 543–574 (2014).
- Mira, A., Martín-Cuadrado, A. B., D'Auria, G. & Rodríguez-Valera, F. The bacterial pan-genome: a new paradigm in microbiology. *Int. Microbiol.* **13**, 45–57 (2010).
- Zurfluh, K. et al. Complete genome sequence of *Escherichia coli* ABWA45, an *rmtB* encoding wastewater isolate. *Genome Announc.* **5**, 45–46 (2017).
- Lukjancenko, O., Wassenaar, T. M. & Ussery, D. W. Comparison of 61 sequenced *Escherichia coli* genomes. *Microb. Ecol.* **60**, 708–720 (2010).
- Donnenberg, M. S. et al. Role of the *eaeA* gene in experimental enteropathogenic *Escherichia coli* infection. *J. Clin. Investig.* **92**, 1412–1417 (1993).
- Franz, E., Klerks, M. M., De Vos, O. J., Termorshuizen, A. J. & Van Bruggen, A. H. C. Prevalence of Shiga toxin-producing *Escherichia coli* *stx1*, *stx2*, *eaeA*, and *rfbE* genes and survival of *E. coli* O157:H7 in manure from organic and low-input conventional dairy farms. *Appl. Environ. Microbiol.* **73**, 2180–2190 (2007).
- Bilge, S. S., Vary, J. C., Dowell, S. F. & Tarr, P. I. Role of the *Escherichia coli* O157:H7 O side chain in adherence and analysis of an *rfb* locus. *Infect. Immun.* **64**, 4795–4801 (1996).
- Torres, A. G. et al. Characterization of Cah, a calcium-binding and heat-extractable autotransporter protein of enterohaemorrhagic *Escherichia coli*. *Mol. Microbiol.* **45**, 951–966 (2002).
- Magnet, S., Smith, T. A., Zheng, R., Nordmann, P. & Blanchard, J. S. Aminoglycoside resistance resulting from tight drug binding to an altered aminoglycoside acetyltransferase. *Antimicrob. Agents Chemother.* **47**, 1577–1583 (2003).
- Ramirez, M. S. & Tolmasky, M. E. Aminoglycoside modifying enzymes. *Drug Resist. Updat.* **13**, 151–171 (2010).
- Sköld, O. Sulfonamide resistance: Mechanisms and trends. *Drug Resist. Updat.* **3**, 155–160 (2000).
- Roberts, M. C. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. *FEMS Microbiol. Rev.* **19**, 1–24 (1996).
- Mahfouz, N. et al. High genomic diversity of multi-drug resistant wastewater *Escherichia coli*. *Sci. Rep.* **8**, 1–12 (2018).
- Sarowska, J. et al. Virulence factors, prevalence and potential transmission of extraintestinal pathogenic *Escherichia coli* isolated from different sources: recent reports. *Gut Pathog.* **11**, 1–16 (2019).
- Bok, E. et al. Comparison of commensal *Escherichia coli* isolates from adults and young children in Lubuskie province, Poland: virulence potential, phylogeny and antimicrobial resistance. *Int. J. Environ. Res. Public Health* **15**, 1–19 (2018).
- Qin, X. et al. Comparison of adhesin genes and antimicrobial susceptibilities between uropathogenic and intestinal commensal *Escherichia coli* strains. *PLoS ONE* **8**, 1–7 (2013).
- Bréchet, C. et al. Wastewater treatment plants release large amounts of extended-spectrum β -lactamase-producing *Escherichia coli* into the environment. *Clin. Infect. Dis.* **58**, 1658–1665 (2014).

54. Diallo, A. A. et al. Persistence and prevalence of pathogenic and extended-spectrum beta-lactamase-producing *Escherichia coli* in municipal wastewater treatment plant receiving slaughterhouse wastewater. *Water Res.* **47**, 4719–4729 (2013).
55. Finn, T. J. et al. A comprehensive account of *Escherichia coli* sequence Type 131 in wastewater reveals an abundance of fluoroquinolone-resistant Clade A strains. *Appl. Environ. Microbiol.* **86**, 1–11 (2020).
56. Flores-Mireles, A. L., Walker, J. N., Caparon, M. & Hultgren, S. J. Urinary tract infections: epidemiology, mechanisms of infection and treatment options. *Nat. Rev. Microbiol.* **13**, 269–284 (2015).
57. Foxman, B. The epidemiology of urinary tract infection. *Nat. Rev. Urol.* **7**, 653–660 (2010).
58. Pitout, J. D. D. & DeVinney, R. *Escherichia coli* ST131: a multidrug-resistant clone primed for global domination. *F1000Research* **6**, 1–7 (2017).
59. Søråas, A., Sundsfjord, A., Sandven, I., Brunborg, C. & Jenum, P. A. 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**, 1–7 (2013).
60. Jørgensen, S. B. et al. A comparison of extended spectrum β -lactamase producing *Escherichia coli* from clinical, recreational water and wastewater samples associated in time and location. *PLoS ONE* **12**, 1–15 (2017).
61. Rice, J. & Westerhoff, P. Spatial and temporal variation in de facto wastewater reuse in drinking water systems across the U.S.A. *Environ. Sci. Technol.* **49**, 982–989 (2015).
62. Rice, J. & Westerhoff, P. High levels of endocrine pollutants in US streams during low flow due to insufficient wastewater dilution. *Nat. Geosci.* **10**, 587–591 (2017).
63. Tanner, W. D. et al. Multi-state study of Enterobacteriaceae harboring extended-spectrum beta-lactamase and carbapenemase genes in U.S. drinking water. *Sci. Rep.* **9**, 1–8 (2019).
64. Griffin, P. M., Manges, A. R. & Johnson, J. R. Food-borne origins of *Escherichia coli* causing extraintestinal infections. *Clin. Infect. Dis.* **55**, 712–719 (2012).
65. Wasiński, B. Extra-intestinal pathogenic *Escherichia coli*—threat connected with food-borne infections. *Ann. Agric. Environ. Med.* **26**, 532–537 (2019).
66. Liu, C. M. et al. *Escherichia coli* ST131-H22 as a foodborne uropathogen. *mBio* **9**, 1–11 (2018).
67. Graham, J. P., Amato, H. K., Mendizabal-Cabrera, R., Alvarez, D. & Ramay, B. M. Waterborne urinary tract infections: have we overlooked an important source of exposure?. *Am. J. Trop. Med. Hyg.* **105**, 12–17 (2021).
68. Boczek, L. A., Rice, E. W., Johnston, B. & Johnson, J. R. Occurrence of antibiotic-resistant uropathogenic *Escherichia coli* clonal group A in wastewater effluents. *Appl. Environ. Microbiol.* **73**, 4180–4184 (2007).
69. Rogers, B. A., Sidjabat, H. E. & Paterson, D. L. *Escherichia coli* O25b-ST131: a pandemic, multiresistant, community-associated strain. *J. Antimicrob. Chemother.* **66**, 1–14 (2011).
70. Numberger, D. et al. Characterization of bacterial communities in wastewater with enhanced taxonomic resolution by full-length 16S rRNA sequencing. *Sci. Rep.* **9**, 1–14 (2019).
71. Kelly, J. J. et al. Wastewater treatment alters microbial colonization of microplastics. *PLoS ONE* **16**, 1–19 (2021).
72. Cai, L. & Zhang, T. Detecting human bacterial pathogens in wastewater treatment plants by a high-throughput shotgun sequencing technique. *Environ. Sci. Technol.* **47**, 5433–5441 (2013).
73. Osunmakinde, C. O., Selvarajan, R., Mamba, B. B. & Msagati, T. A. M. Profiling bacterial diversity and potential pathogens in wastewater treatment plants using high-throughput sequencing analysis. *Microorganisms* **7**, 1–18 (2019).
74. U.S. Environmental Protection Agency. *EPA Microbiological Alternate Test Procedure (ATP) Protocol for Drinking Water, Ambient Water, Wastewater, and Sewage Sludge Monitoring Methods*. (U.S. Environmental Protection Agency, 2010).
75. American Public Health Association, American Water Works Association and Water Environment Federation. *Standard Methods for the Examination of Water and Wastewater* (Health Laboratory Science, 1999).
76. Taskin, B., Gozen, A. G. & Duran, M. Selective quantification of viable *Escherichia coli* bacteria in biosolids by quantitative PCR with propidium monoazide modification. *Appl. Environ. Microbiol.* **77**, 4329–4335 (2011).
77. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
78. Bankevich, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comput. Biol.* **19**, 455–477 (2012).
79. Bertels, F., Silander, O. K., Pachkov, M., Rainey, P. B. & Van Nimwegen, E. Automated reconstruction of whole-genome phylogenies from short-sequence reads. *Mol. Biol. Evol.* **31**, 1077–1088 (2014).
80. Kumar, S., Stecher, G., Li, M., Nkya, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
81. Omar, K. B. & Barnard, T. G. The occurrence of pathogenic *Escherichia coli* in South African wastewater treatment plants as detected by multiplex PCR. *Water SA* **36**, 172–176 (2010).
82. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
83. Beghain, J., Bridier-Nahmias, A., Nagard, H., Le, Denamur, E. & Clermont, O. ClermonTyping: an easy-to-use and accurate in silico method for *Escherichia coli* strain phylotyping. *Microb. Genomics* **4**, 1–8 (2018).
84. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. 2nd edn, (Springer, 2016).
85. Paradis, E. & Schliep, K. Ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526–528 (2019).
86. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T. Y. Ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol. Evol.* **8**, 28–36 (2017).
87. Page, A. J. et al. Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics* **31**, 3691–3693 (2015).
88. Chen, L., Zheng, D., Liu, B., Yang, J. & Jin, Q. VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, D694–D697 (2016).
89. Jia, B. et al. CARD 2017: expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **45**, D566–D573 (2017).

ACKNOWLEDGEMENTS

Funding was provided by Alberta Ministry of Jobs, Economy, and Innovation (Major Innovation Fund) for the Antimicrobial Resistance—One Health Consortium [S.J.G.O. and N.F.N.], the Canadian Institutes for Health Research (CIHR) [N.F.N.], the Natural Sciences and Engineering Research Council (NSERC) [N.F.N.], Alberta Innovates [N.F.N., P.S., and S.J.G.O.], National Natural Sciences Foundation of China (grant number: 82073514 [S.Z.]), and Ningbo Natural Sciences Foundation (grant number: 202003N4114 [S.Z.]).

AUTHOR CONTRIBUTIONS

N.F.N. and S.Z. conceptualized and planned the experiments. S.Z., D.Y., G.B. and K.R. did the experimental work and performed data analysis. D.Y., S.Z. and N.F.N. drafted the initial version of the manuscript. N.F.N., S.Z., D.Y., K.R., S.J.G.O., P.S., G.B., and N.R. revised the manuscript. All authors contributed to and approved the final version of the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41545-022-00177-y>.

Correspondence and requests for materials should be addressed to Shuai Zhi.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022