Brief Report

Molecular and metabolic characteristics of wastewater associated Escherichia coli strains

Mahboobeh Behruznia and David M. Gordon
Division of Ecology and Evolution, Research School of Biology, The Australian National University, Canberra, ACT, 2601, Australia.

Summary
We previously characterized the genetic diversity of Escherichia coli strains isolated from septic tanks in the Canberra region, Australia. In this study, we used repetitive element palindromic (REP) PCR fingerprinting to identify dominant REP-types belonging to phylogroups A and B1 strains across septic tanks. Subsequently, 76 E. coli strains were selected for whole-genome sequencing and phenotype microarrays. Comparative genome analysis was performed to compare septic tank E. coli genomes with a collection of 433 E. coli isolates from different hosts and fresh water. Clonal complexes (CCs) 10 (n = 15) and 399 (n = 10) along with sequence type (ST) 401 (n = 9) were the common lineages in septic tanks. CC10 strains have been detected from animal hosts and freshwater, whereas CC399 and ST401 strains appeared to be associated with septic tanks as they were uncommon in isolates from other sources. Comparative genome analysis revealed that CC399 and ST401 were genetically distinct from other isolates and carried an abundance of niche-specific traits involved in environmental adaptation. These strains also showed distinct metabolic characteristics, such as the ability to utilize pectin, which may provide a fitness advantage under nutrient-limited conditions. The results of this study characterized the adaptive mechanisms allowing E. coli to persist in wastewater.

Introduction
Escherichia coli is a diverse species that mainly inhabits the lower intestinal tracts of warm-blooded animals and is released into the environment through faeces or wastewater (Jang et al., 2017). The presence of E. coli in drinking and recreational waters has long been considered as an indicator of recent faecal contamination (Leclerc et al., 2001). However, previous research has established that E. coli can proliferate in the environment outside a host and become adapted to various natural habitats (Ishii et al., 2006; Walk et al., 2007; Brennan et al., 2010; Byappanahalli et al., 2012), confounding its use as a faecal indicator (van Elsas et al., 2011).

Escherichia coli exhibits a strong clonal genetic structure, with the majority of strains belonging to one of the four main phylogenetic groups – A, B1, B2 and D (Clermont et al., 2013). Each phylogenetic group differs in their phenotypic and genotypic traits and has different ecological niches and life history characteristics (Tenaillon et al., 2010). For example, phylogroups B2 and D strains are specialists and restricted to endothermic vertebrates (Gordon and Cowling, 2003). Phylogroup B2 strains harbour several virulence-associated genes that may enhance colonization and persistence in the host (Nowrouzian et al., 2006; Diard et al., 2010). On the other hand, phylogroup A and B1 strains are known as generalists as they can be isolated from a wide range of host and non-host environments (Gordon and Cowling, 2003). Persistent and free-living E. coli strains usually belong to phylogroups A and B1 (Power et al., 2005; Tymensen et al., 2015; Saraceno et al., 2021). Members of these phylogroups have the highest diversity of genes associated with mobile genetic elements (MGEs), which can promote rapid adaptation to different ecological niches (Touchon et al., 2020).

Recent studies have indicated that E. coli can persist in wastewater and survive the treatment process in wastewater treatment plants (WWTPs) (Anastasi et al., 2013; Zhi et al., 2016; Zhi et al., 2019). However, most studies have focused on sampling antibiotic-resistant E. coli in WWTPs, which may result in the underrepresentation of specific non-resistant lineages (Raven et al., 2019; Lenart-Boroń et al., 2020). Little is known about the origin of persistent E. coli strains in wastewater.
wastewater and the adaptive mechanisms that can contribute to their survival.

We previously characterized the genetic diversity of \textit{E. coli} isolated from 100 septic tanks in the Canberra region, Australia (Behruznia et al., 2021). A septic tank represents a unique environment for investigating persistent \textit{E. coli} strains because it represents a closed system and the inputs to the septic tank are largely unambiguous. Additionally, previous studies have characterized the faecal populations of \textit{E. coli} in humans living in the Canberra region (Blyton et al., 2014; Gordon et al., 2015; Gordon et al., 2017). In the present study, we genotyped one representative of each unique REP-type belonging to phylogroups A and B1 using (CGG)$_4$ and ERIC PCR assays (Versalovic et al., 1991; Adamus-Bialek et al., 2009) to identify shared REP types across septic tanks. Subsequently, 76 \textit{E. coli} strains were selected for whole-genome sequencing, and their carbon source utilization patterns were determined. Comparative genome analysis was performed to compare septic tank \textit{E. coli} genomes with a collection of 433 \textit{E. coli} strains collected from the faeces of native Australian birds (\(n = 107\)) (Blyton et al., 2015), mammals (\(n = 59\)) (Gordon and Cowling, 2003), humans living in the Canberra region (\(n = 93\)) (Blyton et al., 2014; Gordon et al., 2015) and freshwater (\(n = 174\)) (Power et al., 2005; Blyton and Gordon, 2017).

**Results and discussion**

**Identification of shared REP types across septic tanks**

Overall, 27 REP types were detected in more than one tank: 12 were detected in two tanks each; 10 in three tanks; two in four tanks; and three were observed in five different tanks. One example of each REP type from each septic tank in which it was found was selected for whole-genome sequencing. In total, 76 \textit{E. coli} strains were subjected to genome sequencing. The REP types selected for sequencing represented 81.6% (40 out of 49) of phylogroup A and 58% (36 out of 62) of phylogroup B1 strains. The raw sequence read files have been deposited in NCBI and are associated with SRA data PRJNA695020.

**Sequence type diversity and distribution**

Among the 76 isolates sequenced, 38 different STs were identified (Supporting Information Table S1). The most frequent STs were ST10 (\(n = 13, 17.1\%\)), ST635 (\(n = 7, 9.2\%\)), in phylogroup A, and ST401 (\(n = 9, 12.0\%\)) in phylogroup B1. Each of the following STs was represented by three isolates (4%): ST155, ST191, ST206, ST399, ST409 and ST5295. STs were grouped into 13 clonal complexes (CCs) and 12 singletons (STs that did not share alleles at five out of seven loci with any other ST) (Feil et al., 2004). Major CCs were CC10 (\(n = 15, 19.7\%\)), CC399 (\(n = 10, 13.1\%\)), ST401 (\(n = 9, 12.0\%\)) without a CC, CC206 (\(n = 6, 8.0\%\)), and CC155 (\(n = 5, 6.6\%\)). The phylogroup membership of the strains was verified using the ClermontTyper v1.4.0 (http://clermonttyping.iarm-research.center) (Beghain et al., 2018). The \textit{in silico} phylogroup of ST401 strains did not match the \textit{in vitro} PCR results, and these strains were primarily classified as phylogroup A strains with the \textit{in vitro} PCR assay. The discrepancy between the results can be due to a single nucleotide polymorphism (SNP) or insertions and deletions affecting the binding region of the primer TspE4.C2 in ST401 strains (Beghain et al., 2018). ST401 strains were further referred based on their \textit{in silico} phylogroup in this study.

A total of 214 STs were found among the 433 phylogroups A and B1 isolates from birds, humans, mammals and freshwater samples, of which 21 STs were also detected in septic tanks (Supporting Information Table S2). Comparing the distribution of STs in septic tanks to those of the \textit{E. coli} collection indicated that CC10, CC155 and CC206 strains had been previously isolated from different host and freshwater samples. However, CC399 and ST401 strains were rare in isolates from other sources. ST401 was only found in isolates from freshwater (2/150 isolates, 1.3%), and CC399 was detected in isolates from humans (3/93 isolates, 3.6%). To explore this further, we investigated the distribution of CC399 and ST401 isolates available in the MLST database (http://enterobase.warwick.ac.uk). Members of these two clonal groups had a similar distribution in different host and non-host environments. Treated or untreated wastewater accounted for approximately 45% of CC399 and ST401 isolates, followed by raw meat products with about 30% of the total isolates (Table 1).

**Table 1.** Distribution by source of CC399 and ST401 isolates in the MLST database.

| CC      | Wastewater | Meat sample | Human host | Animal host | Freshwater |
|---------|------------|-------------|------------|-------------|------------|
| 399 \(n = 371\) (%) | 166 (44.7%) | 111 (29.9%) | 79 (21.3%) | 7 (1.9%) | 8 (2.2%) |
| 401 \(n = 100\) (%) | 45 (45%) | 37 (37.0%) | 12 (12%) | 2 (2%) | 4 (4%) |

STs were identified by the MLST University of Warwick scheme using MLST 2.0 (Larsen et al., 2012) available on the Centre for Genomic Epidemiology website.

© 2022 The Authors. \textit{Environmental Microbiology Reports} published by Society for Applied Microbiology and John Wiley & Sons Ltd., \textit{Environmental Microbiology Reports}, 14, 646–654
These isolates were less common in animal hosts and were rarely detected in freshwater (last accessed: January 2022).

**Genome comparison analysis**

The pan-genome analysis of the 509 A1 and B1 isolates revealed a total of 60,461 genes, of which 2259 were classified as core genome. A phylogenetic tree of all isolates based on SNPs in the core genome showed that isolates clustered by their phylogroups and septic tank strains were interspersed with other *E. coli* strains in the phylogenetic tree (Fig. 1). CC10 septic tank strains were clustered with both host and freshwater isolates. ST401 strains were virtually identical to two isolates recovered from freshwater samples (ST401 and ST3168). These isolates were collected during the East Coast Low 2015 storm event as part of a verification monitoring program (Hunter Water, 2021) and are likely to have originated from sewage overflow during the storm.

A principal component analysis (PCA) based on the variable genome of the 509 strains separated the collection into two main clusters (Fig. 2). Group 1 contained phylogroup A strains, and group 2 included phylogroup B1 strains. The PCA also showed that CC399 and ST401 strains grouped together and formed a distinct subcluster within groups 1 and 2 despite their different phylogroup membership. This result suggested that these two clonal groups might represent a genetically distinct population associated with septic tanks. Genes with the

![Fig. 1. A phylogenetic tree was constructed by the HARVEST suite based on SNPs in the core genome for 509 *E. coli* strains belonging to phylogroups A and B1 (Treangen et al., 2014). *Escherichia coli* strains were collected from a variety of animal hosts, freshwater and septic tanks. The inner ring indicates the source of the strains, and the outer ring shows their phylogroup membership. The position of CC399 and ST401 strains is marked in the tree. Visualization of the phylogenetic tree was performed using iTOL (Letunic and Bork, 2018).](image)
known function that were significantly overrepresented or underrepresented in the CC399-ST401 subcluster are listed in Supplement Tables S3A and S3B.

**Niche-specific genes in CC399 and ST401 strains**

The conditions in the wastewater are generally very harsh for *E. coli* and other enteric bacteria. Survival in the septic tanks requires overcoming temperature fluctuations (from 11°C in winter to 29°C in summer), hypo-osmotic stress (dilution in water) and limited nutrient availability. Wastewater contains a variety of hazardous substances and chemicals, such as detergents, antibiotics, pesticides and heavy metals (copper, zinc, nickel, lead and cadmium) (Palmquist and Hanaeus, 2005). Additionally, predation and competition with the indigenous microbial communities negatively influence *E. coli* persistence in aquatic environments (Wanjigi et al., 2016; Korajkic et al., 2019).

*Escherichia coli* variable genome is associated with MGEs such as transposable elements, plasmids and phages that are acquired independently through horizontal gene transfer (HGT) and play a vital role in niche adaptation and diversification of the species (Touchon et al., 2020). Previous research has demonstrated the role of MGEs in the evolution of virulence (Denamur et al., 2021) and antibiotic resistance (Partridge et al., 2018) in pathogenic and commensal *E. coli* strains. CC399 and ST401 strains possessed an abundance of adaptive genes for survival in wastewater, which may be classified into several groups:

i. Cation binding proteins and efflux transporters, involved in capturing and expelling heavy metal ions, oxidizing compounds and antimicrobial agents (Delmar et al., 2015; Blanco et al., 2016). These genes are often encoded in plasmids, and their overrepresentation in CC399 and ST401 strains may indicate selective pressure for acquiring and maintaining these traits to overcome the toxicity of metals and hazardous chemicals in wastewater.

ii. Heat shock response genes, responsible for preventing and repairing protein damage through interactions with the misfolded proteins. The presence of a heat resistance genomic island, known as the locus of heat resistance, has been shown to protect *E. coli* not only against heat but also against chlorine and other oxidizing chemicals (Wang et al., 2020). The heat

---

**Fig. 2.** Principal component analysis based on the variable genome of 509 *E. coli* strains labelled by phylogroup. Group 1 contained phylogroup A, and group 2 contained phylogroup B1. CC399 and ST401 strains formed a distinct subcluster within clusters 1 and 2. Genomes were annotated using Prokka (Seemann, 2014), and pan-genome analysis was conducted using Roary (Page et al., 2015). Statistical analyses were performed using R Studio Version 1.4.1717.

© 2022 The Authors. *Environmental Microbiology Reports* published by Society for Applied Microbiology and John Wiley & Sons Ltd., *Environmental Microbiology Reports*, 14, 646–654
resistance genes in CC399 and ST401 strains may promote environmental resistance to more than just heat stress and contribute to long-term survival in septic tanks.

iii. Conjugal elements, associated with niche adaptation by promoting HGT and integration of genomic islands (Shi et al., 2015). Habitat sharing can facilitate the acquisition of locally adaptive traits between CC399 and ST401 strains regardless of their phylogroups’ genetic background (Touchon et al., 2020).

iv. A toxin–antitoxin system, important in regulating nutritional starvation and intraspecific competition in environments with limited resources (Christensen-Dalsgaard et al., 2010). Toxin production is particularly advantageous in E. coli present in biofilms and the sludge layer of the septic tank, where close proximity occurs.

Antimicrobial resistance determinants of CC399 and ST401 strains

Virulence genes were determined using VirulenceFinder 2.0 (Joenssen et al., 2014), and antimicrobial resistance (AMR) genes were determined using ResFinder 4.1 (Zankari et al., 2017) tools available on the Centre for Genomic Epidemiology (CGE) website (www.genomicepidemiology.org). Overall, 41 virulence factors (VFs) were identified among the 76 septic tank strains, but most of these traits (n = 23) were detected in three strains only. Commonly observed VFs were glutamate decarboxylase (gad) and long polar fimbriae (lpfA), which were detected in 79% and 64.5% of the strains respectively. The average number of VFs varied with respect to the clonal group of the strains. CC10 had an average of four VFs, and CC399 and ST401 had an average of two VFs, including gad and lpfA. One ST10 strain was identified as a potential extraintestinal pathogen E. coli following Johnson et al. (2018). The isolate carried VFs that aid adhesion (afa/draBC) and iron acquisition system (iutA). ST297 in phylogroup B1 harboured 13 VFs, including the enterohemolysin gene (ehxA) and two subunits of the Shiga toxin 1 (stx1A and stx1B). This isolate was identified as an enterohemorrhagic E. coli strain based on its virulence profile (Robins-Browne et al., 2004). ST327 in phylogroup B1 was a putative atypical enteropathogenic (EPEC) strain as it harboured VFs associated with EPEC such as eae gene (Kaper et al., 2004). The genetic characteristics of the three E. coli strains indicated that they are likely to be human pathogens. A total of 13 different AMR determinants were identified. As the multi-drug transporter mdtA1 gene was present in all strains, this gene was not included in the analysis. Overall, 85.5% (n = 65) of the isolates had no AMR, 6.5% had one or two (n = 5) and 8.0% (n = 6) had three or more AMR determinants. Escherichia coli strains carrying more than two different classes of AMR genes were classified as multi-drug resistance (MDR). MDR strains belonged to ST10 (n = 4), ST450 (n = 1) and ST401 (n = 1). The AMR genes detected were against aminoglycosides (aadA2, aph(6)-Ib, aph(3’)-lb, aph(3’)-Ibc, and aac(3)-Ila), β-lactam (blaTEM-1A and blaTEM-1B), sulphonamides (sul1 and sul2), tetracycline (tetA and tetB) and trimethoprim (fdrA1 and fdrA12). The VFs and AMR determinants of the septic tank strains are provided in Supporting Information Table S2.

The tertiary treatment process in WWTPs incorporates UV or chlorine disinfection to eradicate remaining E. coli and other faecal coliforms in the wastewater and makes the treated water safe to recycle or release into the environment (Levine and Asano, 2004). High concentrations of antibiotics and antiseptic compounds in the WWTPs (Brecht et al., 2014) and disinfection processes such as chlorine treatment (Liu et al., 2018) might favour the development of antibiotic-resistant E. coli and provide them with an adaptive advantage. Zhi et al. (2019) characterized chlorinated resistant CC399 strains from various WWTPs in Canada. These strains were detected to harbour several AMR determinants that may have contributed to their chlorine resistance in wastewater. Constantinides et al. (2020) found that STs 635 and 401 were the dominant E. coli lineages recovered from hospital sink drains in the United Kingdom, and some of the ST635 strains carried the β-lactamase gene, blacTX-M-15. The AMR determinants identified among CC399 and ST401 strains in this study were mdtA and acrD multidrug efflux pumps, which are known to be important in the transport of a broad spectrum of chemically unrelated drugs and newly encountered toxic compounds (Rosenberg et al., 2000; Yardeni et al., 2018). One ST401 strain was detected with six acquired AMR genes, suggesting the possibility of AMR acquisition among these clonal groups.

Metabolic profiling

Biolog® GEN III microplates were used to explore the carbon source utilization pattern of the 76 E. coli strains. Metabolic activity in each well was assessed by an increase in OD reading at 600 nm. Nine carbon sources that were not used by at least 5% of the tested E. coli strains were excluded from the analysis. PCA based on the carbon source utilization patterns revealed clustering in relation to phylogroup (Supporting Information Fig. 1A). Phylogroup A strains were mainly concentrated in the upper quadrant of the PCA plot, while B1 isolates largely clustered in the lower quadrant. CC399 and ST401 strains were dispersed between phylogroups A and B1.
strains and concentrated in the left quadrant of the PCA (Supporting Information Fig. 1B).

An analysis of variance was conducted for each carbon source to analyse carbon source utilization patterns of the CC399-ST401 subcluster with phylogroup A and B1 strains. In total, 29 carbon sources were differentially utilized among groups (p < 0.05) (Supporting Information Table S4). Phylogroup A strains were more efficient in converting amino acids into microbial biomass. By contrast, phylogroup B1 strains had a higher growth yield on sucrose and the phenolic compound, hydroxyphenyl acetic acid. CC399 and ST401 strains had the highest growth yields on simple carbohydrates (glucose, mannose, maltose and trehalose) and pectin. Conversely, these strains reached the lowest growth yields on sugar acids (such as glucuronamide, galacturonic acid and glucuronic acid) and amino acids. Arabinol was exclusively metabolized by a high proportion of CC399 (70%) and ST401 (66%) strains, suggesting that this trait may be part of the E. coli variable genome. The ability to grow at 8% NaCl and pH 5 varied significantly, and CC399 and ST401 strains showed reduced yields at pH 5% and 8% NaCl relative to the other phylogroup A and B1 strains.

Carbon source availability is expected to be different in host and non-host environments (Savageau, 1983), and metabolic heterogeneity observed among E. coli strains might reflect adaptation to different ecological niches (Franz et al., 2008; Méric et al., 2013). The concentration of free amino acids is estimated to be 1000 times higher in the gut than in soil and sediment (Kemp and Mudrochova, 1973). It seems likely that amino acids are not ecologically significant carbon sources for CC399 and ST401 strains in wastewater. Likewise, a previous study by Méric et al. (2013) observed that plant-associated E. coli strains reached a significantly higher growth yield on sucrose and raffinose, the most abundant carbohydrates in plant tissues, but were less efficient at utilizing amino acids compared with human-associated E. coli strains. Higher growth yields of CC399 and ST401 strains on carbohydrates may be favoured under nutrient-limited conditions and retain a higher survival rate in wastewater.

CC399 and ST401 strains also reached higher growth yields on pectin. Pectin is a complex polysaccharide in plant cell walls, and plant-based food constitutes a major yield on pectin. Pectin is a complex polysaccharide in plant tissues, but was less efficient metabolized by a high proportion of CC399 (70%) and ST401 strains (Gordon et al., 2002; Zhi et al., 2019). The results of this study confirmed that E. coli CC399 and ST401 strains are adapted to survive in wastewater and identified unique genetic and metabolic characteristics that may contribute to their survival. The examination of the wastewater associated lineages in the future assessment of water quality can help determine the likely sources of E. coli in a water body.

Host colonization

The finding of this study raises the question of whether the adaptation to non-host environments can affect host colonization. The genetic and metabolic characteristics of CC399 and ST401 strains may indicate their limited ability to persist in the gut and explain their low frequency in animal hosts. Variable genome analysis demonstrated that CC399 and ST401 strains lack VFs and E. coli common pilus operon (ecpRABCDE), the most common fimbrial structure in commensal and pathogenic strains (Rendón et al., 2007). The presence of adhesins and fimbrae are correlated with the ability of E. coli strains to efficiently colonize the gut (Nowrouzian et al., 2006). The lack of VFs was also reported by Zhi et al. (2019), who suggested that virulence determinants are essentially host-adaptation genes and are not necessary for survival in wastewater. The ability to utilize sugar acids such as galacturonate, glucuronate and gluconate has been shown to play an important role in the colonization of the mammalian gut (Chang et al., 2004; Mandrand-Berthelot et al., 2004). Furthermore, the acid sensitivity of these strains may suggest that they may not be able to survive the low pH in the stomach. Therefore, CC399 and ST401 strains are likely to be transient host colonizers, which reside in the gut for a short period, usually days to weeks (Tullus et al., 1992; Nowrouzian et al., 2005). Consistent with the literature, this research found that the population structure of E. coli was different in the host and non-host environment, and selection plays a key role in determining the fate of E. coli strains (Gordon et al., 2002; Zhi et al., 2019). The results of this study confirmed that E. coli CC399 and ST401 strains are adapted to survive in wastewater and identified unique genetic and metabolic characteristics that may contribute to their survival. The examination of the wastewater associated lineages in the future assessment of water quality can help determine the likely sources of E. coli in a water body.

References

Adamus-Bialek, W., Wojtasik, A., Majchrzak, M., Soosnowski, M., and Parniewski, P. (2009) (CGG)4 based PCR as a novel tool for discrimination of uropathogenic Escherichia coli strains: comparison with enterobacterial repetitive intergenic consensus-PCR. J Clin Microbiol 47: 3937–3944.

Anastasi, E.M., Wohlsen, T.D., Stratton, H.M., and Katouli, M. (2013) Survival of Escherichia coli in two sewage treatment plants using UV irradiation and chlorination for disinfection. Water Res 47: 6670–6679.
Beghain, J., Bridier-Nahmisas, A., Le Nagard, H., Denamur, E., and Clermont, O. (2018) ClermontTyping: an easy-to-use and accurate method for Escherichia genus strain phyotyping. *Microbial Genomics* 4: e000192.

Behrouznia, M., and Gordon, D.M. (2017) Sex-dependent competitive dominance of phylogenetic group B2 Escherichia coli strains within human hosts. *Environ Microbiol Rep* 6: 605–610.

Berthe, T., Ratajczak, M., Clermont, O., Denamur, E., and Petit, F. (2013) Evidence for coexistence of distinct Escherichia coli populations in various aquatic environments and their survival in estuary water. *Appl Environ Microbiol* 79: 4684–4693.

Blanco, P., Hernando-Amado, S., Reales-Calderon, J.A., Corvina, F., Lira, F., Alcalde-Rico, M., et al. (2016) Bacterial multidrug efflux pumps: much more than antibiotic resistance determinants. *Microorganisms* 4: 14.

Blyton, M.D.J., Cornall, S.J., Kennedy, K., Colligon, P., and Gordon, D.M. (2014) Sex-dependent competitive dominance of phylogenetic group B2 Escherichia coli strains within human hosts. *Environ Microbiol Rep* 6: 605–610.

Blyton, M.D.J., and Gordon, D.M. (2017) Genetic attributes of E. coli isolates from chlorinated drinking water. *PLoS One* 12: e0169445.

Blyton, M.D.J., Pi, H., Vangchhia, B., Abraham, S., Trott, D. J., Johnson, J.R., and Gordon, D.M. (2015) Genetic structure and antimicrobial resistance of Escherichia coli and cryptic clades in birds with diverse human associations. *Appl Environ Microbiol* 81: 5123–5133.

Bréchet, C., Plantin, J., Sauget, M., Thouret, M., Talon, D., Cholley, P., et al. (2014) Wastewater treatment plants release large amounts of extended-spectrum β-lactamase-producing Escherichia coli into the environment. *Clin Infect Dis* 58: 1658–1665.

Brennan, F.P., Abram, F., Chinaila, F.A., Richards, K.G., and Flaherty, V. (2010) Characterization of environmentally persistent Escherichia coli isolates leached from an Irish soil. *Appl Environ Microbiol* 76: 2175–2180.

Byappanahalli, M.N., Yan, T., Hamilton, M.J., Ishii, S., Fujikura, R.S., Whitman, R.L., and Sadowsky, M.J. (2012) The population structure of Escherichia coli isolated from subtropical and temperate soils. *Sci Total Environ* 417–418: 276–279.

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., et al. (2004) Carbon nutrition of Escherichia coli in the mouse intestine. *Proc Natl Acad Sci U S A* 101: 7427–7432.

Christensen-Dalsgaard, M., Jørgensen, M.G., and Gerdes, K. (2010) Three new RelE-homologous mRNA interferases of Escherichia coli differentially induced by environmental stresses. *Mol Microbiol* 75: 333–348.

Clermont, O., Christensen, J.K., Denamur, E., and Gordon, D.M. (2013) The Clermont Escherichia coli phylotyping method revisited: improvement of specificity and detection of new phylo-groups. *Environ Microbiol Rep* 5: 58–65.

Constantinides, B., Chau, K.K., Quan, T.P., Rodger, G., Andersson, M.I., Jeffery, K., et al. (2020) Genomic surveillance of Escherichia coli and Klebsiella spp. in hospital sink drains and patients. *Microb Genom* 6: mgen000391.
phylogenetic background predict Escherichia coli extraintestinal virulence better than does ecological source. J Infect Dis 219: 121–132.
Kaper, J.B., Nataro, J.P., and Mobley, H.L.T. (2004) Pathogenic Escherichia coli. Nat Rev Microbiol 2: 123–140.
Kemp, A.L.W., and Mudrochova, A. (1973) The distribution and nature of amino acids and other nitrogen-containing compounds in Lake Ontario surface sediments. Geochim Cosmochim Acta 37(9): 2191–2206.
Kolde, R. (2012) Phatmap: Pretty Heatmaps. R Package Version 1.
Korajkic, A., Wanjugi, P., Brooks, L., Cao, Y., and Harwood, V.J. (2019) Persistence and decay of fecal microbiota in aquatic habitats. Microbiol Mol Biol Rev 83: e00005-00019.
Larsen, M.V., Cosentino, S., Rasmussen, S., Friss, C., Hasman, H., Marvig, R.L., et al. (2012) Multilocus sequence typing of total-genome-sequenced bacteria. J Clin Microbiol 50: 1355–1361.
Leclerc, H., Mossel, D.A.A., Edberg, S.C., and Struijk, C.B. (2001) Advances in the bacteriology of the coliform group: their suitability as markers of microbial water safety. Annu Rev Microbiol 55: 201–234.
Lenart-Boroń, A.M., Kulik, K., and Jelonkiewicz, E. (2020) Antimicrobial resistance and ESBL genes in E. coli isolated in proximity to a sewage treatment plant. J Environ Sci Health A 55: 1571–1580.
Letunic, I, Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucleic Acids Res 44(W1): W242–W245.
Levine, A.D., and Asano, T. (2004) Recovering sustainable water from wastewater. Environ Sci Technol 38: 201a–208a.
Liu, S.-S., Qu, H.-M., Yang, D., Hu, H., Liu, W.-L., Qiu, Z.-G., et al. (2018) Chlorine disinfection increases both intracellular and extracellular antibiotic resistance genes in a full-scale wastewater treatment plant. Water Res 136: 131–136.
Mantrand-Berthelot, M.A., Condemine, G., Hugouvieux-Cotte-Pattat, N., and Stewart, V. (2004) Catabolism of hexuronides, hexuronates, aldonates, and aldarates. EcoSal Plus 1: 1–22.
Méric, G., Kemsley, E.K., Falush, D., Sagger, E.J., and Lucchini, S. (2013) Phylogenetic distribution of traits associated with plant colonization in Escherichia coli. Environ Microbiol 15: 487–501.
Nowrouzian, F.L., Adlerberth, I., and Wold, A.E. (2006) Enhanced persistence in the colonic microbiota of Escherichia coli strains belonging to phylogenetic group B2: role of virulence factors and adherence to colonic cells. Microbes Infect 8: 834–840.
Nowrouzian, F.L., Wold, A.E., and Adlerberth, I. (2005) Escherichia coli strains belonging to phylogenetic group B2 have superior capacity to persist in the intestinal microflora of infants. J Infect Dis 191: 1078–1083.
Page, A.J., Cummins, C.A., Hunt, M., Wong, V.K., Reuter, S., Holden, M.T.G., et al. (2015). Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31(22): 3691–3693.
Palmquist, H., and Hanaeus, J. (2005) Hazardous substances in separately collected grey- and blackwater from ordinary Swedish households. Sci Total Environ 348: 151–163.
Partridge, S.R., Kwong, S.M., Firth, N., and Jensen, S.O. (2018) Mobile genetic elements associated with antimicrobial resistance. Clin Microbiol Rev 31: e00088-00017.
Power, M.L., Littlefield-Wyer, J., Gordon, D.M., Veal, D.A., and Slade, M.B. (2005) Phenotypic and genotypic characterization of encapsulated Escherichia coli isolated from blooms in two Australian lakes. Environ Microbiol 7: 631–640.
Raven, K.E., Ludden, C., Goulouris, T., Blane, B., Naydenova, P., Brown, N.M., et al. (2019) Genomic surveillance of Escherichia coli in municipal wastewater treatment plants as an indicator of clinically relevant pathogens and their resistance genes. Microbial Genomics 5: e00267.
Rendón, M.A., Saldaña, Z., Erdem, A.L., Monteiro-Neto, V., Vázquez, A., Kaper, J.B., et al. (2007) Commensal and pathogenic Escherichia coli use a common pilus adherence factor for epithelial cell colonization. Proc Natl Acad Sci U S A 104: 10637–10642.
Robins-Browne, R.M., Bordun, A.-M., Tauschek, M., Bennett-Wood, V.R., Russell, J., Oppedissano, F., et al. (2004) Escherichia coli and community-acquired gastroenteritis, Melbourne, Australia. Emerg Infect Dis 10: 1797–1805.
Rosenberg, E.Y., Ma, D., and Nikaido, H. (2000) AcrD of Escherichia coli is an aminoglycoside efflux pump. J Bacteriol 182: 1754–1756.
Saraceno, M., Gómez Lugo, S., Ortiz, N., Gómez, B.M., Sabio y García, C.A., Frankel, N., and Graziano, M. (2021) Unraveling the ecological processes modulating the population structure of Escherichia coli in a highly polluted urban stream network. Sci Rep 11: 14679.
Savageau, M.A. (1983) Escherichia coli habitats, cell types, and molecular mechanisms of gene control. Am Nat 122: 732–744.
Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. Bioinformatics 30(14): 2068–2069.
Shi, Q., Strauss, M.R., Caron, J.J., Wang, H., Chung, Y.S., Guarné, A., and Peters, J.E. (2015) Conformational toggling controls target site choice for the heteromeric transposase element Tn7. Nucleic Acids Res 43: 10734–10745.
Tenaillon, O., Skurnik, D., Picard, B., and Denamur, E. (2010) The population genetics of commensal Escherichia coli. Nat Rev Microbiol 8: 207–217.
Toucheon, M., Perrin, A., de Sousa, J.A.M., Vangchhia, B., Burn, S. & O’Brien, C.L. et al. (2020) Phylogenetic background and habitat drive the genetic diversification of Escherichia coli bioRxiv: 2020.2002.2012.945709.
Treangen, T.J., Ondov, B.D., Koren, S., and Phillippy, A.M. (2014) The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15(11): 1–15.
Tullius, K., Kühn, I., Orskov, I., Orskov, F., and Möllby, R. (1992) The importance of P and type 1 fimbriae for the persistence of Escherichia coli in the human gut. Epidemiol Infect 108: 415–421.

© 2022 The Authors. Environmental Microbiology Reports published by Society for Applied Microbiology and John Wiley & Sons Ltd., Environmental Microbiology Reports, 14, 646–654.
Tymensen, L.D., Pyrdok, F., Coles, D., Koning, W., McAllister, T.A., Jokinen, C.C., et al. (2015) Comparative accessory gene fingerprinting of surface water Escherichia coli reveals genetically diverse naturalized population. *J Appl Microbiol* **119**: 263–277.

Vallenet, D., Calteau, A., Dubois, M., Amours, P., Bazin, A., Beuvin, M., et al. (2019) MicroScope: an integrated platform for the annotation and exploration of microbial gene functions through genomic, pangenomic and metabolic comparative analysis. *Nucleic Acids Res* **48**: D579–D589.

van Elsas, J.D., Semenov, A.V., Costa, R., and Trevors, J.T. (2011) Survival of Escherichia coli in the environment: fundamental and public health aspects. *ISME J* **5**: 173–183.

Versalovic, J., Koeuth, T., and Lupski, J.R. (1991) Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Res* **19**: 6823–6831.

Walk, S.T., Alm, E.W., Calhoun, L.M., Mladonicky, J.M., and Whittam, T.S. (2007) Genetic diversity and population structure of Escherichia coli isolated from freshwater beaches. *Environ Microbiol* **9**: 2274–2288.

Wang, Z., Fang, Y., Zhi, S., Simpson, D.J., Gill, A., McMullen, L.M., et al. (2020) The locus of heat resistance confers resistance to chlorine and other oxidizing chemicals in Escherichia coli. *Appl Environ Microbiol* **86**: e02123-02119.

Wanjigi, P., Fox, G.A., and Harwood, V.J. (2016) The interplay between predation, competition, and nutrient levels influences the survival of Escherichia coli in aquatic environments. *Microb Ecol* **72**: 526–537.

Yardeni, E.H., Zomot, E., and Bibi, E. (2018) The fascinating but mysterious mechanistic aspects of multidrug transport by MdfA from Escherichia coli. *Res Microbiol* **169**: 455–460.

Zankari, E., Allesøe, R., Joensen, K.G., Cavaco, L.M., Lund, O., and Aarestrup, F.M. (2017) PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *J Antimicrob Chemother* **72**: 2764–2768.

Zhi, S., Banting, G., Li, Q., Edge, T.A., Topp, E., Sokurenko, M., et al. (2016) Evidence of naturalized stress-tolerant strains of Escherichia coli in municipal wastewater treatment plants. *Appl Environ Microbiol* **82**: 5505–5518.

Zhi, S., Banting, G., Stoatham, P., Ashbolt, N.J., Checkley, S., Meyer, K., et al. (2019) 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.

**Supporting Information**

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

**Fig. S1.** A-B. Principal Components Analysis based on carbon sources utilization patterns of 76 septic tank *E. coli* strains discriminated strains according to their phylogroup (A). CC399 and ST401 strains were dispersed between phylogroups A and B1 strains and were concentrated in the left quadrant of the PCA (B). The PCA plots were created using JMP 16.2.0.

**Fig. S2.** The metabolic differences among the CC10, CC399 and ST401 strains from septic tanks were plotted as a heatmap using R package pheatmap (Kolde2012). Red colour indicates the presence of a complete metabolic pathway, while blue colour represents the absence of a pathway. The metabolic pathways present in the sequenced strains were recovered using the tools available in the MicroScope platform (Vallenet et al., 2019).

**Table S1.** List of STs, virulence factors and antibiotic resistance genes identified in septic tank strains.

**Table S2.** Details of all *E. coli* isolates used for comparative genome analysis.

**Table S3.** A and B. Genes overrepresented and underrepresented among CC399 and ST401 strains compared to other *E. coli* strains. Pearson’s chi-square test of homogeneity was used to determine whether there is a significant difference between the gene presence and absence data.

**Table S4.** List of carbon sources that were differentially utilized between *E. coli* phylogroup A, B1 and CC399-ST401 strains. OD readings are normalized against the negative control well. Only significant differences are shown (p < 0.05).