

Pathogenicity Potential of the Bacteria in Selected Locations of Rapid Creek

Murray, Kelsey E.¹, Erickson, Molly², Schaefer, Morgan², Kunza, Lisa. A.^{1,2,3},
DeVeaux, Linda. C.^{1,2}

South Dakota School of Mines and Technology

¹Biomedical Engineering Program

²Department of Chemistry and Applied Biological Sciences

³Program of Atmospheric and Environmental Sciences

1.1 Introduction

Bacterial levels, particularly fecal coliforms such as *E. coli*, are standard water quality indicators of fecal contamination. The EPA estimates 25% of national groundwater systems exceed the Total Coliform Rule, suggesting widespread dissemination of enteric organisms into the environment [1, 2]. South Dakota surface water is not immune to fecal contamination, and according to the 2016 South Dakota Integrated Report for Surface Water Quality Assessment, sections of Rapid Creek are affected by unacceptably high levels of indicator bacteria, particularly *E. coli* [3].

Certain *E. coli* strains are a normal inhabitant of the human intestinal tract, and when ingested do not cause disease, while other strains are highly pathogenic to humans, causing severe and potentially fatal illnesses. Pathogenic strains are considered emerging zoonotic infectious agents, and have been linked to outbreaks of severe diarrhea and hemolytic uremic syndrome, an often lethal condition in children. Infection by pathogenic variants, predominantly Shiga-toxigenic *E. coli* serotype O157:H7, accounts for approximately 73,000 annual cases of reported illnesses in the U.S. alone [4]. Although disease-causing in humans, many pathogenic *E. coli* strains are common inhabitants of the guts of both domestic and wild animals, particularly ruminants, where they appear to cause no symptoms. Both municipal drinking water and water used recreationally are known to harbor Shiga toxin-producing organisms, and transmission of Shiga-toxigenic *E. coli* (STEC) to humans often occurs through fecal loading of local water sources by domestic and wild animals, storm water drainage, septic tank leakage, and wastewater and municipal sewage discharge [5].

The difference between harmless and dangerous bacteria lies in the genetic make-up of the bacteria. Pathogenic organisms have acquired genes that allow for production of Shiga toxins and other factors which allow successful invasion of human hosts. Bacteria have many mechanisms, collectively known as horizontal gene transfer, by which they can share their genes to other members of the microbial community, regardless of species. The more potentially harmful genes a single bacterium acquires, the more likely the bacterium will be able to cause disease in humans. Routine water quality testing provides a snapshot of fecal coliform abundance; however, such sampling does not account for the potential human health risk due to bacterial presence.

In order to better inform the public about water safety, West Dakota Water Development District funded the current project to assess the pathogenicity potential of bacteria in Rapid Creek. Using a novel water quality monitoring metric developed by our research group at South Dakota School of Mines and Technology, we tested for the presence of approximately 20 known virulence genes in surface water samples and bacteria isolates, allowing for an extensive characterization of the pathogenicity potential of the resident bacteria.

2.1 Materials and Methods

2.1.1 Sampling Scheme. The sampling effort for this project was a collaborative effort with H2E and USGS, with Emily Martinez, Galen Hoogestraat, and Jaime Haueter performing sample collection. Samples were collected from four sites along Rapid Creek: Dark Canyon (WQM 69), Canyon Lake (USGS 6413220), East St. Patrick Street (WQM 92), and the below the wastewater treatment plant (USGS 06421500) [Refer to figure 4.1]. Samples were collected monthly on the following dates: November 21 and December 13, 2016, and January 10, February 15, March 13, April 10, May 8, June 12, July 10, August 7, September 20, and October 11, 2017.

2.1.2 Sample Collecting and Processing. Surface water was collected in sterile polypropylene bottles and were transported on ice to the laboratory, where they were processed within 24 hours of collection. Samples were agitated prior to filtration to resuspend settled solids and homogenize the sample. A subsample of 100 mL of each sample was filtered through 0.45 μ m mixed cellulose ester membranes using a vacuum manifold. Total DNA from filter paper was extracted from surface water using PowerWater DNA Extraction Kit (MoBio Laboratories), and plasmid DNA was extracted using GeneJet Plasmid Kit (Thermo Scientific) per the respective manufacturer protocols. Bacterial isolate DNA was prepared with the GNOME DNA Isolation Kit (MP Biomedicals). All DNA samples were stored at -20 °C.

Environmental bacteria were collected by membrane filtration and were cultured on MacConkey agar to select for gram-negative, enteric bacilli and differentiate colonies based on lactose fermentation. Isolated colonies were replica plated onto Sorbitol MacConkey Agar, and Sorbitol MacConkey Agar supplemented with potassium tellurite. DNA was extracted as described above.

2.1.3 Pathogenicity Profiling (Path-STREAM). DNA extracts were used as template in polymerase chain reactions (PCR) using oligonucleotide primer pairs (Fisher Scientific) specific for *E. coli* pathogenicity genes associated with the virulence profiling tool designed by our research group. Presence of a particular gene was confirmed if the appropriately sized DNA fragment was visible when amplified DNA samples were subjected to agarose gel electrophoresis and UV transillumination upon ethidium bromide staining.

3.1 Results

3.1.1 Path-STREAM Analysis of Surface Water. In order to assess the temporal and spatial virulence profiles of the bacterial communities within Rapid Creek, Path-STREAM was implemented in surface water samples from various sites over the course of 12 months. Genes detected in the highest percentage of Rapid Creek surface water samples were *stx*₂ (55%), *eae* (22%), *exhA* (20%) and *espP* (17%), and genes infrequently detected were *eaeA* (3%) and *einV* (2.5%) [Figure 1]. The *stx*₁ gene was not detected in the total bacterial population of Rapid Creek. The *stx*₂ gene was detected at least once during every month of sampling, whereas *eaeA* was only detected in June 2017, *einV* was only detected in February 2017, and *espP* was frequently detected between February and April 2017. The highest number of gene detections (28 total) was in February 2017, and the lowest number of gene detections was in December 2016 (4 total). Genes detected from Dark Canyon, Canyon Lake, and East St. Patrick Street were relatively uniform, with 43, 44, and 44 respective gene detections. Below the wastewater treatment plant, however, there were 69 total gene detections over the course of the project.

There may be an unanticipated inverse relationship between water temperature and gene detection. The total number of genes detected from November 2016 to April 2017, when the average water temperature was 3.5°C, was 108, compared to 79 gene detections between May and October 2017, when the average water temperature was 14.9°C. As expected, an inverse relationship between water temperature and dissolved oxygen (DO) was observed. However, there was no relationship between DO and gene detection.

3.1.2 Path-STREAM Analysis of Bacterial Isolates from Surface Water. Path-STREAM analysis of surface water demonstrated the presence of multiple virulence genes within Rapid Creek. In order to determine whether a single organism carried multiple pathogenicity genes, or if the various virulence genes were carried individually, surface water bacteria were isolated and characterized for their respective pathogenicity gene profiles. Bacterial isolates were obtained from the July 2107 samples by culturing surface water filtrates on MacConkey agar, which selects for gram-negative and enteric bacilli, and differentiates based on lactose fermentation. A freeze-thaw lysis method was used to obtain template for subsequent pathogenicity gene and 16S PCR reactions. In total, 255 isolates were characterized: 22 from Dark Canyon, 61 from Canyon Lake, 86 from East St. Patrick Street, and 86 from below the wastewater treatment plant. The varying numbers of isolates obtained from each site were due to differential bacterial loads from each respective site.

Of the 255 bacterial isolates, 31% did not carry any Path-STREAM virulence genes, 32% carried only one virulence gene, 30.5% carried two virulence genes, 4% carried 3 virulence genes and 1.5% carried 4 virulence genes [Figure 2]. Interestingly, *stx*₁ was detected in isolates, but was noticeably absent from the community analysis. In this case, *stx*₁ was presumably below the limit of detection of the assay, an obstacle that can be overcome by further development of the assay, possibly in the form of a microarray. In isolates carrying a single virulence gene, *stx*₁, *stx*₂, *eaeA*, *einV*, or *espP* were detected, with the highest percentage of isolates carrying the *eaeA* (13%) or *stx*₂ gene (7%). In isolates carrying two virulence genes, the *stx*₂/*eaeA* combination was the most

prevalent (13% of isolates). Combinations of three or four virulence genes were detected within a single organism in a small fraction of the total isolates. While isolates carrying 3 or 4 virulence genes in tandem were relatively rare, they were more likely to be associated with Canyon Lake or the wastewater treatment plant [Figure 3].

When analyzing bacterial isolates by site, Dark Canyon had 22 total gene detections, Canyon Lake had 106, East St. Patrick St. had 83, and below the wastewater treatment plant had 120 gene detections. *Stx*₂ was the most detected gene with 109 total detections, followed by *eaeA* with 107, *espP* with 51, *einV* with 39, and *stx*₁ with 25 total gene detections. While *stx*₁ was not detected in the total bacterial community of Rapid Creek, the gene was present in a small fraction of isolates. The *eae* and *exhA* genes were not detected in any bacterial isolates, but they were present in the total bacterial population of Rapid Creek.

3.1.3 16S rRNA Sequencing of Bacterial Isolates from Surface Water. From the 255 bacterial isolates that were characterized for their respective pathogenicity profiles, 16 isolates were chosen for 16S rRNA sequencing for species identification: 2 isolates were obtained from Dark Canon, 5 from Canyon Lake, 4 from E. St. Patrick St., and 5 from the wastewater treatment plant. *Escherichia coli*, *Aeromonas veronii*, *Hafnia alvei*, organisms from the *Enterobacter cloacae* complex, and *Panotaea agglomerans* were identified as species that carry the virulence genes detected within Rapid Creek [Table 1]. All species of bacteria identified as carrying *E. coli* virulence genes are highly related, and all have been previously reported as *stx* carriers [Figure 3] [6-11]. Interestingly, certain gene combinations were carried by multiple species of bacteria, which may be indicative of high levels of horizontal gene transfer among the resident bacteria. For example, *stx*₂/*eaeA*/*einV* was present in 6 total organisms from 4 different species: *Hafnia alvei*, *Enterobacter cloacae* complex, *Escherichia coli*, and *Aeromonas veronii*. Overall, most of the virulence gene-carrying organisms were identified as *E. coli*, as 7 of the 16 isolates were from this species. Certain organisms were associated with specific sites; *Hafnia alvei*, for instance, was identified only in Canyon Lake, and *Aeromonas veronii* was solely associated with the wastewater treatment plant. *E. coli*, however, was identified at all sites along Rapid Creek.

4.1 Discussion

E. coli is a genetically, phenotypically, and metabolically diverse and abundant species of bacteria. The primary niche for *E. coli* is the mammalian gut, thus when found in the environment, *E. coli* are often indicators of fecal contamination. Surface water is particularly susceptible to fecal inputs, and can therefore harbor potentially pathogenic organisms, largely attributed to loading from domestic and wild animals, stormwater runoff, septic tank leakage, and wastewater and municipal discharge [12]. Current water quality testing relies on measuring fecal coliforms, or, more specifically, *E. coli*. However, since bacteria have varying levels of harmful attributes, identification of a water sample as carrying *E. coli* provides limited information about the potential danger of consumption or recreational use of the water.

When bacteria adapt to new environments, they are able to acquire and exchange genes that confer a selective advantage. But, often the genes that allow survival in a secondary habitat are also genes that contribute to pathogenicity in humans [13]. Horizontal gene transfer is a source of bacterial evolution that allows for genes to be passed easily between bacteria, even among distantly related or unrelated species, and the pathogenic capacity of a bacterium is directly related to the acquisition and expression of virulence genes. Shiga toxins are critical pathogenicity factors that contribute to the disease-causing potential of *E. coli*. In addition to Shiga toxin, other *E. coli* pathogenicity factors including adhesins, toxins, invasins, cytotoxic necrotizing factors, protein secretion systems, and iron uptake systems may be shared between individuals in a given bacterial population, especially in environments where bacterial levels are high [14]. Therefore, existing harmless *E. coli* may acquire genes carried by other organisms, transforming them into dangerous human pathogens. In surface water, where there are diverse populations of bacterial communities originating from various sources, the likelihood of acquisition of novel combinations of virulence attributes within a single organism is high.

Because of the availability of DNA from neighboring cells within an environment, the entire genetic complement must be considered. Historically, environments are fairly isolated, with minimal mixing; however, human activity has affected this in a number of ways. Both urbanization and agricultural practices have introduced new sources of fecal contamination to waterways that were not previously impacted, and animal-human interactions result in spreading of their biota to the environment. Under certain stresses, including pollution of waterways, wastewater treatment processes, and changes in nutrients within a waterbody, the rate of lateral gene transfer increases, often resulting in exchange of traits that allow for pathogenicity in humans.

4.1.1 Indicator Bacteria and Pathogenicity Profiling. *E. coli* and/or coliform levels are commonly used as indicators of fecal contamination. Portions of Rapid Creek are currently listed as impaired by SD DENR. This is of particular concern during high recreational use. In the current work, pathogenicity gene presence was not related to the abundance of *E. coli* or total fecal coliforms. It has been established that neither presence nor abundance of virulence factors is directly linked to fecal indicator bacteria levels [19-21]. For example, it is possible for surface water to have an *E. coli* load that is below the EPA's threshold of impairment, while prevalence of virulence factors is high [16, 19, 22]. Conversely, it is possible for impaired sites to completely lack *E. coli* virulence genes [16]. Thus, the tool developed by our lab group improves upon the current water quality metric.

4.1.2 Pathogenicity Genes. A wide and diverse array of microorganisms are capable of producing Shiga toxin, including STEC, a human pathogen. Such bacteria possess at least one gene for Shiga toxin, an essential virulence factor that contributes to the disease-causing potential. The Shiga toxin genes (*stx*), which comprise two antigenically distinct variants (*stx1* and *stx2*), have not only been found in the coliform bacteria *E. coli*, *Enterobacter*, and *Citrobacter*, and the aquatic bacterium *Aeromonas*, but in the soil bacterium *Acinetobacter*, the food pathogens *Campylobacter* and *Shigella*, and even an insect endosymbiont, *Hamiltonella* [23, 24]. Bacteria, particularly *E. coli*, are capable of encoding and expressing more than one Shiga toxin variant. In addition to Shiga toxin, other virulence factors including adhesins, toxins, invasins, cytotoxic necrotizing factors, protein secretion systems, and iron uptake systems may be shared between individuals in a given bacterial population [25]. The *eae* virulence factors (*eae*, *eaeA*, *eaeY2*), encode proteins

responsible for the attaching and effacing lesions in the gut mucosa of infected patients, and *ehxA*, a protein that damages red blood cells, genes are often transmitted in *E. coli* along with Shiga toxins. Presence of *einV* is indicative of enteroinvasive *E. coli* (EIEC), which can cause a Shigellosis-like disease with watery diarrhea and dysentery. EIEC's mechanism of pathogenesis involves epithelial invasion of the colon leading to inflammation and ulceration of the mucosal lining. The *espP* gene encodes a serine protease which has homologs in *Neisseria gonorrhoeae*, *Haemophilus influenzae*, and *E. coli* O157:H7. Serine proteases contribute to mucosal hemorrhage in hemorrhagic colitis and hemolytic uremic syndrome. These virulence factors in pathogenic *E. coli* originated from horizontal gene transfer events, and such gene sharing is observed in environments where bacterial levels are high.

4.1.3 Potential Health Risk. Virulence gene profiling of the bacterial community within surface water provides a foundation for risk assessment associated with contacting fecally contaminated waterways [27]. The presence of *E. coli* pathogenicity genes that are carried by various organisms within Rapid Creek indicates significant horizontal gene transfer among the resident bacteria, and their presence may indicate a potential risk to human health. Almost all water samples containing a Shiga toxin gene also had *eaeA*, which is required for full expression of virulence within EHEC strains [16]. Even though there were fewer gene detections during months where recreational-use is presumed to be higher, it should be noted there is a potential risk of infection as long as such pathogenicity factors are present within the system, even when the total *E. coli* numbers are not high [28].

With the elevated levels of total *E. coli* in portions of Rapid Creek, there is a potential to create new combinations of these genes through horizontal gene transfer, and thus, new human pathogens. Such evolution has been documented recently in Germany, where the acquisition of the *stx2* gene by an *E. coli* subtype containing the *eaeA* gene caused an outbreak in which more than 50 people died from ingestion of the pathogen [29]. The *stx1*, *stx2*, and *eaeA* genes occur concurrently in many strains of pathogenic *E. coli*, and are all found in *E. coli* O157:H7. Since this gene combination was present within several different species within Rapid Creek, there is potential for this gene combination to be transferred to *E. coli* or other opportunistic pathogens. Presence of pathogenicity factors in Rapid Creek indicates a potential risk to human health. Future studies should focus on the magnitude of pathogenicity gene presence within the waterways.

Acknowledgements | We are grateful for West Dakota Water Development District for funding this project and the joint work done with H2E and USGS. This work was completed in part with support from South Dakota School of Mines and Technology.

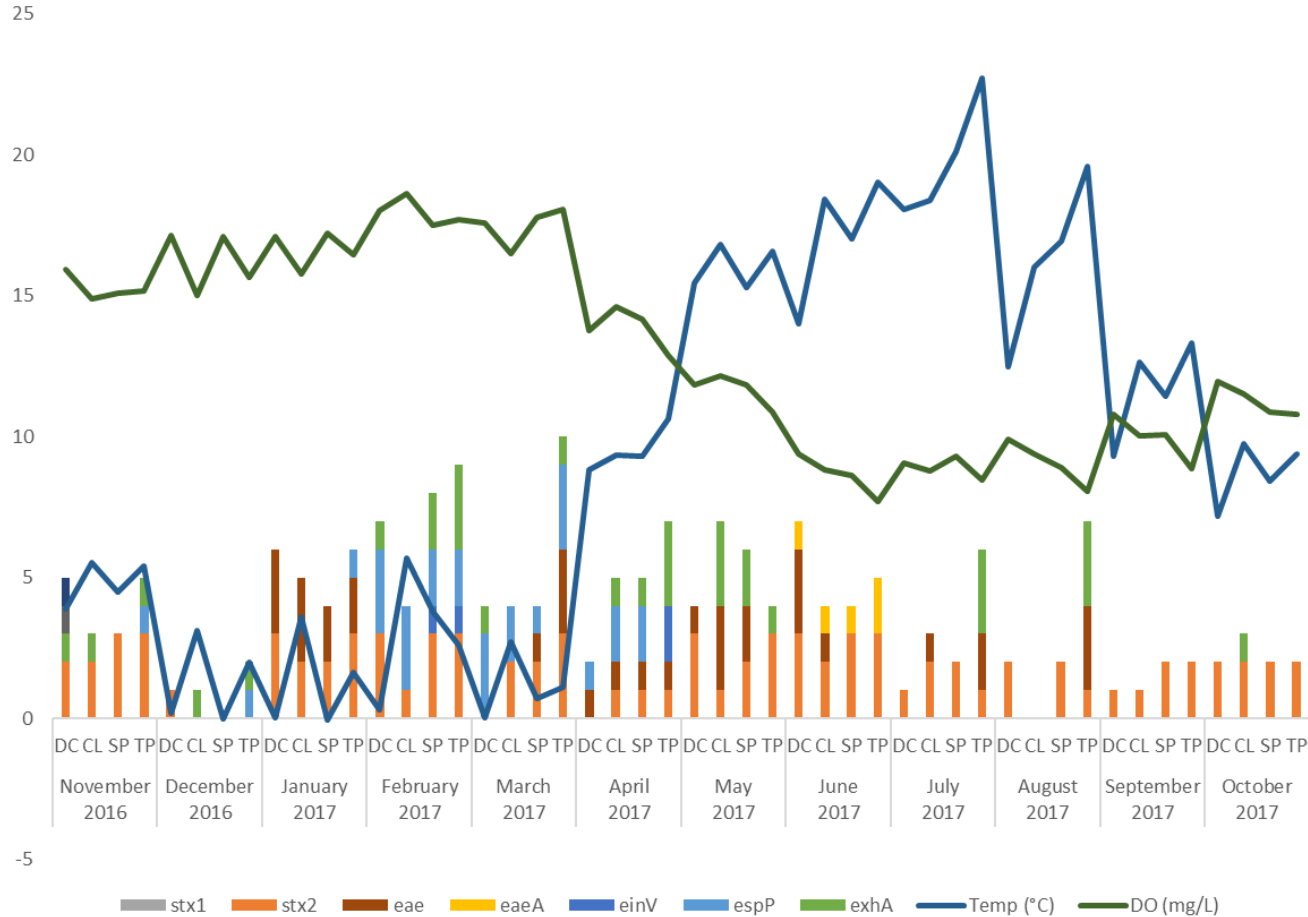


Figure 1 | Path-STREAM analysis of Rapid Creek. Genes detected using Path-STREAM included *stx₂*, *eae*, *eaeA*, *einV*, *espP* and *ehxA*. The *stx₁* gene was not detected in any samples. The magnitude of the bar represents the number of replicate samples positive for each respective gene. Sites are abbreviated as follows: DC: Dark Canyon, CL: Canyon Lake, SP: East St. Patrick Street, TP: below the wastewater treatment plant.

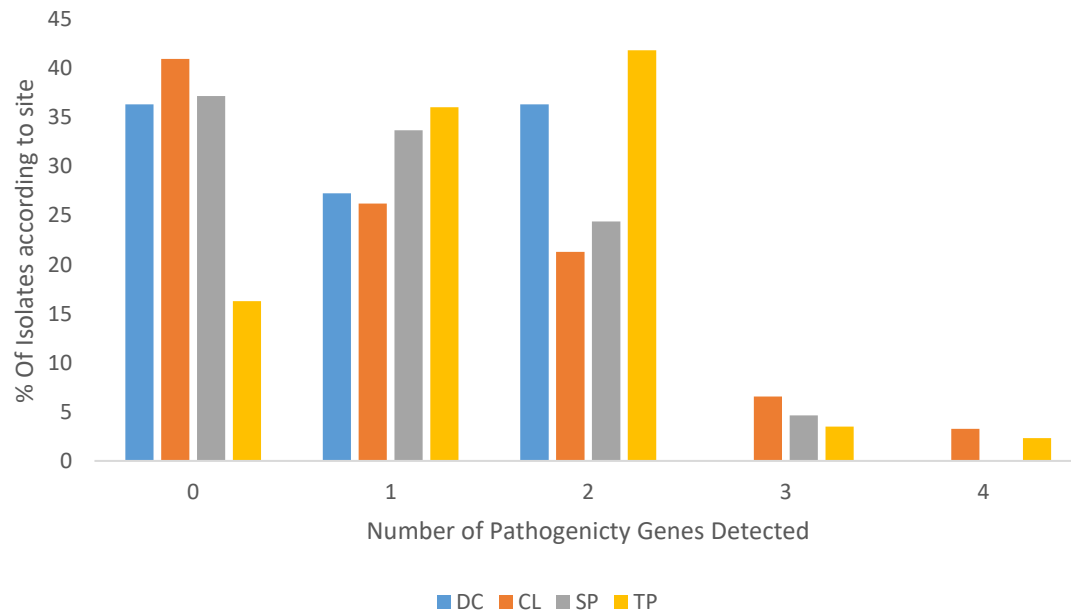


Figure 2 | Number of pathogenicity genes detected in Rapid Creek isolates, expressed as a percentage of isolates from each individual site. Sites are abbreviated as follows: DC: Dark Canyon, CL: Canyon Lake, SP: East St. Patrick Street, TP: below the wastewater treatment plant.

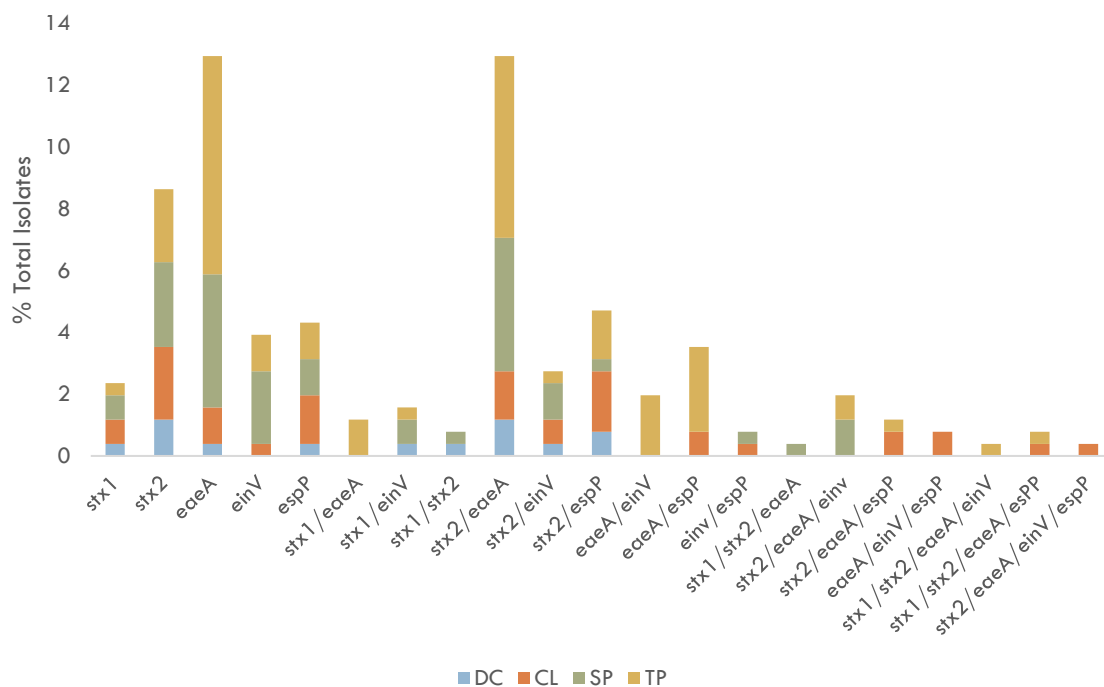


Figure 3 | Percent of bacterial isolates carrying specific virulence gene combinations. Sites are abbreviated as follows: DC: Dark Canyon, CL: Canyon Lake, SP: East St. Patrick Street, TP: below the wastewater treatment plant.

Table 1 | Species identification of bacterial isolates carrying various combinations of virulence genes.

Gene Combination	Site	Species	% Identity
<i>stx1/einV</i>	DC	<i>Escherichia coli</i>	96
<i>stx1/stx2/eaeA</i>	SP	<i>Escherichia coli</i>	88
<i>stx1/stx2/eaeA/einV</i>	TP	<i>Escherichia coli</i>	98
<i>stx1/stx2/eaeA/espP</i>	CL	<i>Hafnia alvei</i>	97
	TP	<i>Aeromonas veronii</i>	95
<i>stx2/eaeA</i>	DC	<i>Pantoea agglomerans</i>	97
<i>stx2/eaeA/einV</i>	CL	<i>Hafnia alvei</i>	97
	SP	<i>Enterobacter cloacae</i> complex	95
	SP	<i>Escherichia coli</i>	97
	SP	<i>Escherichia coli</i>	97
	TP	<i>Aeromonas veronii</i>	98
	TP	<i>Escherichia coli</i>	97
<i>stx2/eaeA/espP</i>	CL	<i>Hafnia alvei</i>	96
	CL	<i>Hafnia alvei</i>	96
	TP	<i>Aeromonas veronii</i>	99
	CL	<i>Escherichia coli</i>	96

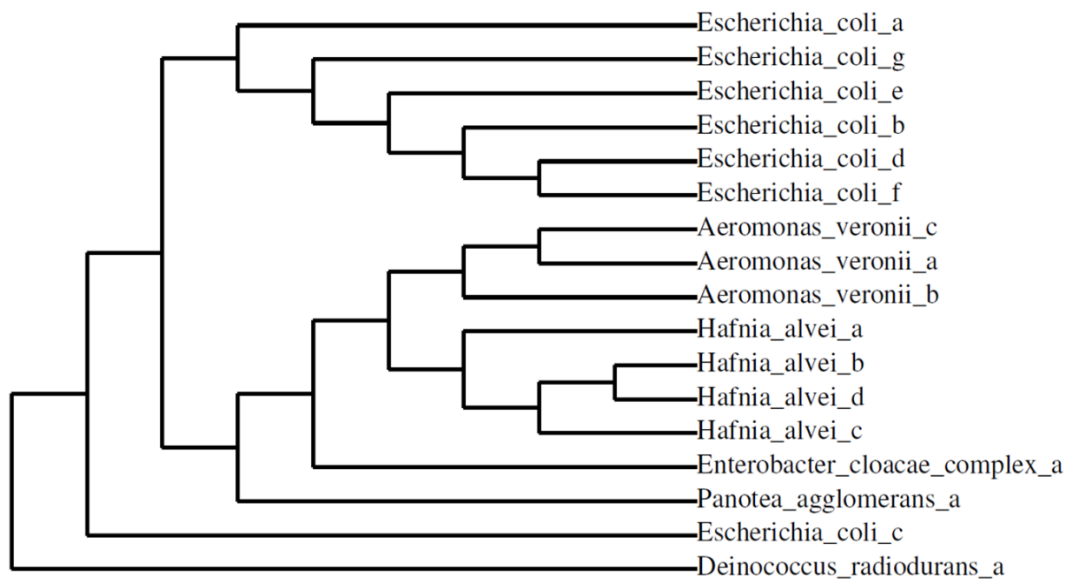


Figure 3 | Phylogenetic tree of isolates sequenced from Rapid Creek [30-36].

5.1 Literature Cited

1. Daniel SL, Hung KF, Janezic KJ, Ferry B, Hendricks EW, Janiga BA, Johnson T, Murphy S, Roberts ME, Scott SM: **Phenotypic and genotypic characterization of Escherichia coli isolated from untreated surface waters.** *The open microbiology journal* 2013, **7**:9-19.
2. Macler BA, Merkle JC: **Current knowledge on groundwater microbial pathogens and their control.** *Hydrogeology Journal* 2000, **8**(1):29-40.
3. Pirner SM: **The 2016 South Dakota Integrated Report for Surface Water Quality Assessment.** In.: South Dakota Department of Environment and Natural Resources; 2016.
4. CDC: **National Enteric Disease Surveillance: Shiga toxin-producing Escherichia coli (STEC) Annual Report, 2012.** National Center for Emerging and Zoonotic Infectious Diseases; Division of Foodborne, Waterborne, and Environmental Diseases Centers for Disease Control 2012.
5. Ishii S, Sadowsky MJ: **Escherichia coli in the environment: implications for water quality and human health.** *Microbes and Environments* 2008, **23**(2):101-108.
6. Probert WS, McQuaid C, Schrader K: **Isolation and identification of an Enterobacter cloacae strain producing a novel subtype of Shiga toxin type 1.** *Journal of clinical microbiology* 2014, **52**(7):2346-2351.
7. Mezzatesta ML, Gona F, Stefani S: **Enterobacter cloacae complex: clinical impact and emerging antibiotic resistance.** *Future microbiology* 2012, **7**(7):887-902.
8. Parker JL, Shaw JG: **Aeromonas spp. clinical microbiology and disease.** *Journal of Infection* 2011, **62**(2):109-118.
9. Albert MJ, Alam K, Islam M, Montanaro J, Rahaman A, Haider K, Hossain MA, Kibriya A, Tzipori S: **Hafnia alvei, a probable cause of diarrhea in humans.** *Infection and immunity* 1991, **59**(4):1507-1513.
10. Figueras MJ, Aldea MJ, Fernández N, Aspíroz C, Alperi A, Guarro J: **Aeromonas hemolytic uremic syndrome. A case and a review of the literature.** *Diagnostic microbiology and infectious disease* 2007, **58**(2):231-234.
11. Alperi A, Figueras M: **Human isolates of Aeromonas possess Shiga toxin genes (stx1 and stx2) highly similar to the most virulent gene variants of Escherichia coli.** *Clinical Microbiology and Infection* 2010, **16**(10):1563-1567.
12. Agarwal M, Tomar RS, Jyoti A: **Detection of water-borne pathogenic bacteria: where molecular methods rule.** *Int J of Multi Curr Researc* 2014, **2**:351-358.
13. Brinkmeyer R, Amon RMW, Schwarz JR, Saxton T, Roberts D, Harrison S, Ellis N, Fox J, DiGuardi K, Hochman M et al: **Distribution and persistence of Escherichia coli and Enterococci in stream bed and bank sediments from two urban streams in Houston, TX.** *Science of The Total Environment* 2015, **502**:650-658.
14. Gerrish RS, Lee JE, Reed J, Williams J, Farrell LD, Spiegel KM, Sheridan PP, Shields MS: **PCR versus hybridization for detecting virulence genes of enterohemorrhagic Escherichia coli.** *Emerg Infect Dis* 2007, **13**(8):1253-1255.
15. Ahmed W, Sawant S, Huygens F, Goonetilleke A, Gardner T: **Prevalence and occurrence of zoonotic bacterial pathogens in surface waters determined by quantitative PCR.** *water research* 2009, **43**(19):4918-4928.
16. Masters N, Wiegand A, Ahmed W, Katouli M: **Escherichia coli virulence genes profile of surface waters as an indicator of water quality.** *water research* 2011, **45**(19):6321-6333.
17. Ram S, Vajpayee P, Shanker R: **Prevalence of multi-antimicrobial-agent resistant, shiga toxin and enterotoxin producing Escherichia coli in surface waters of river Ganga.** *Environmental science & technology* 2007, **41**(21):7383-7388.
18. Poma V, Mamani N, Iñiguez V: **Impact of urban contamination of the La Paz River basin on thermotolerant coliform density and occurrence of multiple antibiotic resistant enteric pathogens in river water, irrigated soil and fresh vegetables.** *SpringerPlus* 2016, **5**(1):499.
19. Smith CJ, Olszewski AM, Mauro SA: **Correlation of Shiga toxin gene frequency with commonly used microbial indicators of recreational water quality.** *Applied and environmental microbiology* 2009, **75**(2):316-321.
20. Bonetta S, Pignata C, Lorenzi E, De Ceglia M, Meucci L, Bonetta S, Gilli G, Carraro E: **Detection of pathogenic Campylobacter, E. coli O157: H7 and Salmonella spp. in wastewater by PCR assay.** *Environmental Science and Pollution Research* 2016:1-8.

21. Haack SK, Duris JW, Fogarty LR, Kolpin DW, Focazio MJ, Furlong ET, Meyer MT: **Comparing wastewater chemicals, indicator bacteria concentrations, and bacterial pathogen genes as fecal pollution indicators.** *Journal of environmental quality* 2009, **38**(1):248-258.
22. Shelton DR, Karns JS, Coppock C, Patel J, Sharma M, Pachepsky YA: **Relationship between eae and stx virulence genes and Escherichia coli in an agricultural watershed: implications for irrigation water standards and leafy green commodities.** *Journal of food protection* 2011, **74**(1):18-23.
23. Mauro SA, Koudelka GB: **Shiga toxin: expression, distribution, and its role in the environment.** *Toxins* 2011, **3**(6):608-625.
24. Degnan PH, Yu Y, Sisneros N, Wing RA, Moran NA: **Hamiltonella defensa, genome evolution of protective bacterial endosymbiont from pathogenic ancestors.** *Proceedings of the National Academy of Sciences* 2009, **106**(22):9063-9068.
25. Pass M, Odedra R, Batt R: **Multiplex PCRs for identification of Escherichia coli virulence genes.** *Journal of clinical microbiology* 2000, **38**(5):2001-2004.
26. Tozzoli R, Caprioli A, Morabito S: **Detection of toxB, a plasmid virulence gene of Escherichia coli O157, in enterohemorrhagic and enteropathogenic E. coli.** *Journal of clinical microbiology* 2005, **43**(8):4052-4056.
27. Bugarel M, Martin A, Fach P, Beutin L: **Virulence gene profiling of enterohemorrhagic (EHEC) and enteropathogenic (EPEC) Escherichia coli strains: a basis for molecular risk assessment of typical and atypical EPEC strains.** *BMC microbiology* 2011, **11**(1):1.
28. Sidhu JP, Ahmed W, Hodggers L, Toze S: **Occurrence of virulence genes associated with diarrheagenic pathotypes in Escherichia coli isolates from surface water.** *Applied and environmental microbiology* 2013, **79**(1):328-335.
29. Brzuszkiewicz E, Thürmer A, Schuldes J, Leimbach A, Liesegang H, Meyer F-D, Boelter J, Petersen H, Gottschalk G, Daniel R: **Genome sequence analyses of two isolates from the recent Escherichia coli outbreak in Germany reveal the emergence of a new pathotype: Entero-Aggregative-Haemorrhagic Escherichia coli (EAHEC).** *Archives of microbiology* 2011, **193**(12):883-891.
30. Dereeper A, Audic S, Claverie J-M, Blanc G: **BLAST-EXPLORER helps you building datasets for phylogenetic analysis.** *BMC evolutionary biology* 2010, **10**(1):8.
31. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M: **Phylogeny. fr: robust phylogenetic analysis for the non-specialist.** *Nucleic acids research* 2008, **36**(suppl_2):W465-W469.
32. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic acids research* 2004, **32**(5):1792-1797.
33. Castresana J: **Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis.** *Molecular biology and evolution* 2000, **17**(4):540-552.
34. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood.** *Systematic biology* 2003, **52**(5):696-704.
35. Anisimova M, Gascuel O: **Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative.** *Systematic biology* 2006, **55**(4):539-552.
36. Chevenet F, Brun C, Bañuls A-L, Jacq B, Christen R: **TreeDyn: towards dynamic graphics and annotations for analyses of trees.** *BMC bioinformatics* 2006, **7**(1):439.