PRELIMINARY ANALYSIS OF *E. COLI* GENETIC DIVERSITY IN THE GUADALUPE RIVER: INITIAL APPROACH TO FECAL POLLUTION TRACKING

Samantha R. Hetherington[1], Teresa Le Sage Clements[2] and Dmitri Sobolev[1*]
[1]School of Arts and Sciences, University of Houston-Victoria, Victoria, TX 77901
[2]School of Education, Health Professions and Human Development, University of Houston-Victoria, Victoria, TX 77901
**Corresponding author; Email: sobolevd@uhv.edu*

Abstract.–Four sites in the Guadalupe River were sampled for coliform bacteria from the winter 2016 through spring 2017 to assess the microbiological pollution situation in the river and to identify pollution sources. Numbers of putative coliforms detected ranged from below detection limits to 260 CFUs per 100 mL. Over the time period sampled, coliform levels exhibited strong seasonality, probably linked to the amount of rainfall prior to the sampling date. Molecular analysis of confirmed *E. coli* isolates demonstrated that the spike in coliform numbers 03 December 2016 was dominated by only two unique fingerprints, suggesting the possibility of a watershed source for fecal contamination in the urban area. The diversity of *E. coli* fingerprints was much greater in the rural portion of the river, suggesting a variety of coliform sources, probably including wildlife. Further studies are needed to identify and manage sources of fecal pollution.

Keywords: water quality; bacterial source tracking; BOX A1R

The Guadalupe River is a part of the Guadalupe-San Antonio river system, which flows across the Texas coastal plain from Kerr County in the Hill Country to the San Antonio Bay on the Gulf of Mexico. The river is extensively used for both contact (such as swimming) and noncontact (fishing, boating) recreation. Throughout its course, the river is influenced by runoff from fields, cattle feedlots, and urban areas. Detection and monitoring of fecal pollution and identification of sources and patterns are essential for maintaining public health. Although studies measuring fecal pollution in the region have been conducted previously (Mott & Lehman 2005), this is the first attempt to apply molecular bacterial source tracking to the Guadalupe River.

OPEN ACCESS DOI 10.32011/txjsci_73_1_Article1
Received: 29 February 2020 | Accepted: 3 December 2020 | Published Online: 2 January 2021

Fecal pollution in various water bodies is inferred based on the presence of indicator organisms, such as *Escherichia coli*. This organism serves as a reliable predictor of waterborne illness (Wade et al. 2003), and water quality standards have been developed based on the numbers of indicator organisms detected (TCEQ 2007). Numbers exceeding the established criteria indicate a probability of human enteric pathogens being present (Simpson et al. 2002). Unfortunately, microbiological detection of indicator organisms alone fails to distinguish anthropogenic and agricultural sources of pollution from natural ones, or to assess whether different samples receive pollution from the same or different sources. The situation is complicated by the fact that the indicator organisms can persist in aquatic sediments and return to the water column when the sediments are disturbed (Ishii et al. 2007). Additionally, the amount of fecal pollution is expected to vary greatly depending on the rainfall, creating highly irregular pulses of nutrients and bacteria.

Molecular approaches can potentially distinguish among different sources of indicator organisms (e.g., human vs. livestock or wildlife) as well as to compare sites to determine if similar sources are responsible for fecal pollution in those locations. Genome-based methods may detect the organisms of interest directly in the water samples via amplification of organism specific markers (Vadde et al., 2019). Alternatively, indicator organisms can be recovered in a culture as a part of culture-based enumeration effort. Upon such recovery, a number of approaches can be used to generate a fingerprint pattern that can be matched with similar patterns from the same study or from a database. Those approaches include enzymatic digestion of the whole genome, followed by pulse-field gel electrophoresis (PFGE) (Barrett et al. 1994). Likewise, detecting the length of a restriction fragment containing the 16SrRNA gene, called ribotyping, allows to reliably distinguish between human- and animal-sourced isolates (Parveen et al. 1999). BOX-PCR takes advantage of a repetitive BOX motif throughout the genome to produce a molecular fingerprint. Although this technique is not without concerns, including

poor reproducibility (Yang & Yen 2012), it appears to be adequate for strain differentiation (Olive & Bean 1999). In this study, we attempted to evaluate the number of coliform cells present in the Guadalupe River water, link those numbers to the events in the watershed, and assess similarities of the fecal pollution sources at different sites based on the coliform communities.

MATERIALS & METHODS

River water samples were collected monthly in triplicate from existing structures at four locations: Seguin, Starcke Park riverwalk (29.5512910541º, -97.9718573009º); Gonzales, kayak launch upstream of Hwy 183 bridge (29.4841068321º, -97.4477686774º); Cuero, Hwy 71 kayak launch (presently defunct; 29.090537294º, -97.3298502305º); Victoria, Riverside Park boat ramp (28.8246949753º, -97.0154657525º; all coordinates were retrieved from Bing Maps by use of QGIS 2.8.6 software) along the Guadalupe River system on the Texas coastal plain. Sterile plasticware was used for collection immediately below the water surface. Samples were stored on ice and returned to the laboratory within three hrs.

For coliform enumeration and recovery, 10 to 100 ml of each of the sample triplicate were filtered through sterile nitrocellulose filters (pore size 0.22 μm). Filters were placed bacteria side up on the surface of EMB Levine medium (Levine 1918) and incubated at 40°C for 48 hrs. Characteristic golden sheen colonies were counted and numbers of coliform colony-forming units (CFU) were calculated on a per 100 mL basis.

For molecular analysis, colonies identified as putative coliforms were picked with a sterile toothpick, streaked on a fresh EMB medium plate and incubated for 24 hrs at 40°C. Subsamples of the streaks exhibiting the same golden sheen were collected with a toothpick, placed into 100 μL of 1x TAE buffer in a 96 well plate and incubated for 20 mins at 96°C to lyse the cells. Lysates were used at 1% as

templates for PCR amplification of *E. coli* specific β-glucuronidase gene fragment (ca. 500 bp) with primers uidA1318F and uidA1698R (Bower et al. 2005). The PCR mixture contained 10 µL 10X PerfectTaq buffer, 20 µL 5P solution, 0.2 µL each of 200 mM forward and reverse primer solutions (Operon Technologies), 2 µL of 20 mM dNTP solution (Fisher Scientific), 2.5 units PerfectTaq polymerase, and PCR-grade water to a total of 100 µL. All PCR chemicals, unless otherwise noted, were supplied as a part of the PerfectTaq kit (5 Prime, Inc.). Twenty five cycles with 30 secs of melting at 96°C, 30 secs of annealing at 60°C and 90 secs extension at 72°C were used for amplification; cycles were preceded by initial melting at 96°C and followed by final extension at 72°C; both for for 5 mins. Use of PerfectTaq with 5P solution obviated the need to individually select the annealing temperature for each primer set. Thermotolerant isolates positive for *E. coli*-specific β-glucuronidase gene fragment were presumed to be *E. coli* and retained for further analysis.

Isolates producing positive amplification of a fragment of expected length with primers uidA1318F and uidA1698R (Bower et al. 2005) were subjected to additional PCR amplification as described above, except that BOX-A1R primer (Kon et al. 2007) was used to produce a variety of fragment lengths (200-3000 bp expected). BOX-A1R fragments were separated on 1% agarose gel in 0.5x TAE buffer with 1 µg/mL ethidium bromide to produce a fingerprint, and standard DNA ladders were used as a reference. Gels were photographed and images of individual gels scaled in the direction of electrophoresis (no more than 10%) using GIMP image processing software to match DNA ladder markers across all gels. Individual bands within the fingerprints were identified and marked visually. Once a band was identified in any of the samples, it was numbered and presence or absence of a band in the same position was recorded in the remaining samples. Presence of a band was scored as one for a particular position and absence scored as zero, producing a binary string describing the band pattern.

Distances between the individual isolates were calculated as Jaccard distances (J') for each pair of isolates as:

$$J'_{1,2} = 1-[C/(2N-C)]$$

where C is the number of common bands between isolates 1 and 2 (i.e., the number of sites where both isolates scored band presence), and N is the total number of different bands in 1 and 2 (i.e., the number of sites where one of the two samples scored one and the other scored zero) (Tibayrenc et al. 1993). In this setting, Jaccard distances can range from 0 (all bands are common between the two sites) to 1 (no common bands observed).

Jaccard distances for each pair of isolates were represented as a distance matrix, and a distance-matrix based tree generation function of MEGA5 software was used to visualize relationships among individual isolates using UPGMA algorithm implemented in MEGA5 (Tamura et al. 2011). Rainfall data were obtained from publicly available sources (U.S. Climate Data 2020) for the location closest to the sampling site (Victoria, Cuero, and Gonzales; New Braunfels rainfall data were used as a proxy for Seguin), and the coliform CFU counts were plotted against the total precipitation in the week prior to sample collection date.

Results and Discussion

Over the course of this study, total coliform CFU counts varied from below detection limits to occasionally as high as 260 (± 20, CI 95) CFUs/100mL (Fig. 1), somewhat above the limits established for recreational waters (200 coliform CFU per 100 mL of water for contact recreation; TCEQ 2007), although our approach did not permit a direct comparison to regulatory standards. Coliform levels were consistently higher across all sites on the December 2016 sampling date, with an additional peak in early April of 2017 (Fig. 1). At all sites the coliform CFU counts were consistently low (near zero) when

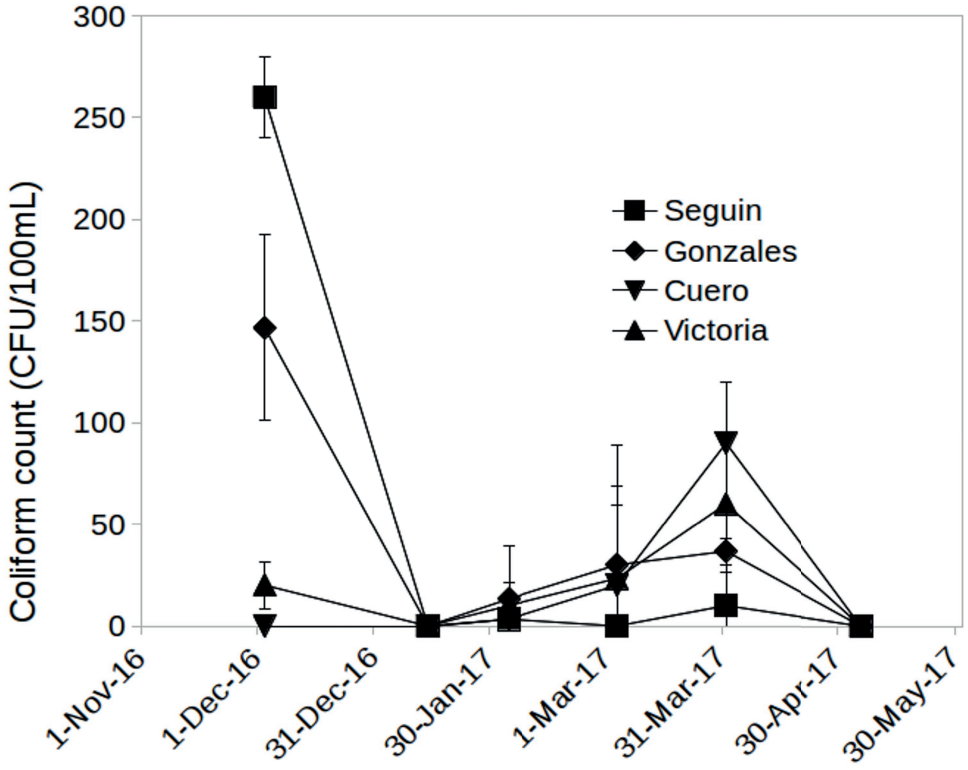

Figure 1. Total coliform counts in the Guadalupe River. Error bars, 95% confidence interval; not shown if less than the size of the symbol.

cumulative rainfall was about 0.3 inch (7.6 mm) or below in the week prior to sampling. On sampling dates with cumulative rainfall in the week prior exceeding 0.3 inch (7.6 mm), a positive relationship between rainfall and coliform CFU count was observed (Fig. 2), suggesting land runoff as a possible coliform source. The highest levels of coliform indicative of fecal pollution, were found at the urbanized Seguin site, suggesting that urban locales are the main pollution sources. On the other hand, the Seguin location was the only impoundment sampled. As coliform bacteria can settle out and persists in sediments (Ishii et al. 2007), disturbance of sediments during major rain events can release coliforms back into the water column. This effect is unlikely in the remaining sites were shallow waters and strong currents were observed, as those conditions would limit the retention of bacteria in the sediment.

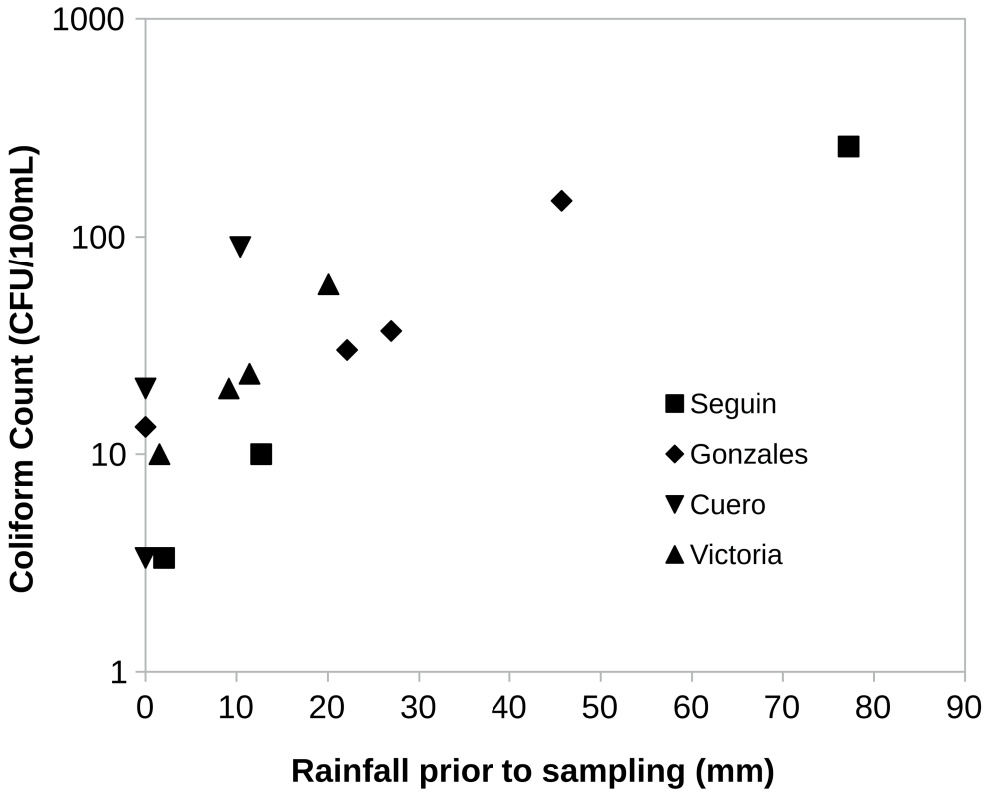

Figure 2. Relationship between rainfall over the week immediately preceding sample collection and coliform count.

Not all colonies identified as thermotolerant coliforms based on the EMB growth were confirmed as *E. coli* by PCR amplification of *E. coli*-specific β-glucuronidase gene; the colonies not confirmed were excluded from further analysis. Molecular fingerprints of *E. coli* bacteria demonstrated appreciable differences across sampling sites and times (Fig. 3). Isolates of *E. coli* recovered from the Seguin site during a coliform spike on 03 December 2016 (Fig. 3, series S031216; solid outline boxes) were primarily represented by two strains, as evidenced by BOX-A1R pattern typing. This suggests a watershed source for those organisms, rather than resuspension of sediment-preserved bacteria, which usually appear to be more diverse (Kon et al. 2007). Our study was not library-based and did not seek to identify

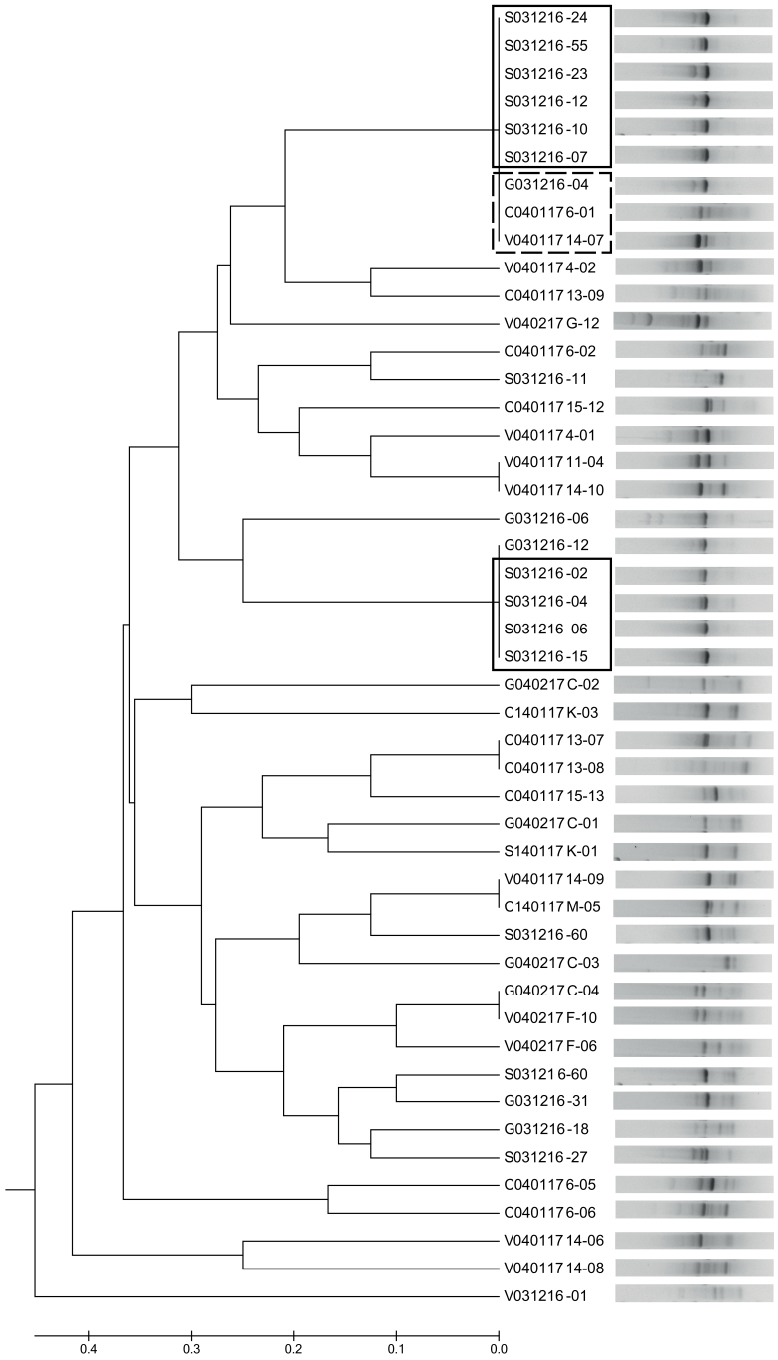

Figure 2. Grouping of *E. coli* isolates based on BOX-A1R fingerprint patterns. See text for explanation of highlighted parts of the figure. Scale = Jaccard distance. Image assembled from multiple individual gels scaled to match DNA size standards. First letter of the sample designation indicates site ("S" for Seguin, etc.), six digits that follow are the collection date in ddmmyy format, remaining characters are the individual isolate number.

specific sources; further study is needed to link the bacteria to potential sources.

Isolates identical to the ones recovered from other locations in small numbers (Fig. 3, dashed box), suggest that the same source may operate, to a smaller degree, in the rural areas as well. Nevertheless, no clear relationship among the *E. coli* strains recovered from various sites has emerged, which together with the great diversity of strains recovered, suggests each site is affected by different *E. coli* sources.

CONCLUSIONS

Based on our findings, the Guadalupe River currently does not experience a significant fecal pollution problem outside urban areas, and even in those areas the issue appears to be transient. We hypothesize that the presence and diversity of the coliforms in the dynamic river system is influenced primarily by intensity and stage of the rain event in the watershed; however, further studies of the processes influencing river water quality are needed. Identification of types of *E. coli* in the Guadalupe River is important for understanding the sources and routes for fecal contamination of the river water. This understanding is essential for making informed decisions in water quality management. We intend to use the experience gained in this preliminary study to specifically link organisms recovered from the river with host species responsible for fecal pollution. This may require using database-dependent approaches.

ACKNOWLEDGMENTS

This work was supported in part by NOAA B-WET grant number NA16NMF4630049 to TLC and DS. High school teachers participating in the educational portion of this project collected and processed water samples.

LITERATURE CITED

Barrett, T. J., H. Lior, J. H. Green, R. Khakhria, J. G. Wells, B. P. Bell, K. D. Greene, J. Lewis & P. M. Griffin. 1994. Laboratory investigation of a multistate food-borne outbreak of *Escherichia coli* O157:H7 by using pulsed field gel electrophoresis and phage typing. J. Clin. Microbiol. 32:3013–3017.

Bower P. A., C. O. Scopel, E. T. Jensen, M. M. Depas & S. L. McLellan. 2005. Detection of genetic markers of fecal indicator bacteria in Lake Michigan and determination of their relationship to *Escherichia coli* densities using standard microbiological methods. Appl. Environ. Microbiol. 71:8305–8313.

Ishii, S., D. L. Hansen, R. E. Hicks & M. J. Sadowski. 2007. Beach sand and sediments are temporal sinks and sources of *Escherichia coli* in Lake Superior. Environ. Sci. Technol. 41:2203–2209.

Kon, T., S. C. Weir, E. T. Howell, H. Lee & J. T. Trevors. 2007. Genetic relatedness of *Escherichia coli* isolates in interstitial water from a Lake Huron (Canada) beach. Appl. Environ. Microbiol. 73:1961–1967.

Levine, M. 1918. Differentiation of *B. coli* and *B. aerogenes* on a simplified eosin-methylene blue agar. J. Infect. Dis. 23:43–47.

Mott, J. B. & R. L. Lehman. 2005. Bacteria source tracking in Copano Bay. Phase II final report. Texas General Land Office, Austin. xv + 171 pp.

Olive D. M. & P. Bean. 1999. Principles and applications of methods for DNA-based typing of microbial organisms. J. Clin. Microbiol. 37:1661–1669.

Parveen, S., K. M. Portier, K. Robinson, L. Edminston & M. L. Tamplin. 1999. Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Appl. Environ. Microbiol. 65:3142–3147.

Simpson, J. M., J. W. Santo Domingo & D. J. Reasoner. 2002. Microbial source tracking: state of the science. Env. Sci. Technol. 36:5279–5288.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei & S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using likelihood, distance, and parsimony methods. Mol. Biol. Evol. 28:2731–2739.

TCEQ (Texas Commission on Environmental Quality). 2007. Texas water quality standards: criteria for recreation. TCEQ Staff Draft, 31 August 2007 https://www.tceq.texas.gov/assets/public/waterquality/tmdl/82lakehouston/82-bactstand-oct2007.pdf (Accessed 12 February 2020).

Tibayrenc, M., K. Neubauer, C. Barnabé, F. Guerrini, D. Skarecky & F. J. Ayala. 1993. Genetic characterization of six parasitic protozoa: parity between random-primer DNA typing and multilocus enzyme electrophoresis. Proc. Nat. Acad. Sci. 90:1335–1339.

Vadde, K. K., A. J. McCarthy, R. Rong & R. Secar. 2019. Quantification of microbial source tracking and pathogenic bacterial markers in water and sediments of Tiaoxi River (Taihu Watershed). Front. Microbiol. 10:699. doi: 10.3389/fmicb.2019.00699

Wade, T. J., N. Pai, J. N. S. Eisenberg & J. M. Colford. 2003. Do U.S. Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. Environ. Health Perspect. 111:1102–1109.

Yang, A. & C. Yen. 2012. PCR optimization of BOX-A1R PCR for microbial source tracking of *Escherichia coli* in waterways. J. Exp. Microbiol. Immunol. 16:85–89.

U.S. Climate Data. 2020. http://www.usclimatedata.com/climate/texas/united-states/3213. (Accessed 12 February 2020).