Texas Bacterial Source Tracking Program (FY20-FY21)

Texas Water Resources Institute TR-542
May 2022

Texas Water Resources Institute
make every drop count

Lucas Gregory, Brian Hux, Terry J. Gentry,
Anna Gitter, Carlos Monserrat, Kristina D. Mena
Texas Water Resources Institute

Texas Bacterial Source Tracking Program (FY20-FY21)

Prepared by:

Lucas Gregory[1], Brian Hux[2], Terry J. Gentry[2], Anna Gitter[3], Carlos Monserrat[3], Kristina D. Mena[3]

[1] Texas Water Resources Institute
[2] Texas A&M Agrilife Research – Department of Soil and Crop Sciences
[3] University of Texas Health Science Center at Houston School of Public Health – El Paso Campus

Texas Water Resources Institute TR-542

May 2022

State Nonpoint Source Grant Program
Texas State Soil and Water Conservation Board Project 20-52

Prepared for:
Texas State Soil and Water Conservation Board

Cover photo:
Leon River. Photo by June Wolfe.

Table of Contents

List of Tables 3
List of Figures 4
List of Acronyms 6
Executive Summary 7
Introduction 9
Expansion of the Texas *E. coli* BST Library 10
BST Library Refinement 12
Leon River 15
Texas *E. coli* BST Library Use 16
  Leon River Watershed 16
  Leon River at Coryell CR 183 18
  Leon River at Comanche CR 382 21
  Walnut Creek at FM 1476 23
  Pecan Creek at SH 22 25
Leon River Watershed Temporal BST Comparison 26
  Leon River at Coryell CR 183 28
  Leon River at Comanche CR 382 29
  Walnut Creek at FM 1476 30
  Pecan Creek at SH 22 31
Evaluating Site-Specific Health Risks with Quantitative Microbial Risk Assessment 32
  Leon River at Coryell CR 183 33
  Leon River at Comanche CR 382 34
  Walnut Creek at FM 1476 35
  Pecan Creek at SH 22 35
Other Watersheds 37
Evaluation of the Texas *E. coli* BST Library 38
Defining Host Class 38
Developing a Feral Hog Host Class 39
Evaluation of Library-Independent PCR Markers 41
Education and Outreach 42
Literature Cited 42

List of Tables

Table 1. Number of samples collected per source, the number of samples testing positive for *E. coli* , screened, validated, archived and added to the Texas *E. coli* BST Library. ...11

Table 2. Texas *E. coli* BST Library (ver. 04-22, cross-library validation) composition and rates of correct classification (RCCs) by Jackknife analysis of ERIC-RP composite data sets using an 80% similarity cutoff and 3- and 7-way splits. 13

Table 3. Monthly counts of *E. coli* isolates from water samples for Leon River at Coryell CR 183 (TCEQ Station ID 11929) between September 2020 through August 2021. 19

Table 4. Monthly counts of *E. coli* isolates from water samples for Leon River at Comanche CR 382 (TCEQ Station ID 11933) between September 2020 through August 2021. 21

Table 5. Monthly counts of *E. coli* isolates from water samples for Walnut Creek at FM 1476 (TCEQ Station ID 17379) between September 2020 through August 2021. 23

Table 6. Monthly counts of *E. coli* isolates from water samples for Pecan Creek at SH..22 (TCEQ Station ID 17547) between September 2020 through August 2021. 25

Table 7. Reference pathogens of public health concern for each fecal source. 33

Table 8. Texas *E. coli* BST Library (ver. 4-way: 05-22, cross-library validation) composition and rates of correct classification (RCCs) by Jackknife analysis of ERIC-RP composite data sets using an 80% similarity cutoff and 4 and 8-way splits..... 40

Table S 1. QMRA parameters and assumptions. 46

Table S 2. Prevalence of infection and infectious potential of pathogens for non-human fecal sources. 47

Table S 3. Dose response models utilized to estimate the risk of infection and illness... 47

List of Figures

Figure 1. Texas *E. coli* BST Library (ver. 04-22) composition by 3-way split of source classes (1,942 isolates from 1,775 different fecal source samples). 14

Figure 2. Texas *E. coli* BST Library (ver. 04-22) composition by 7-way split of source classes (1,942 isolates from 1,775 different fecal source samples). 14

Figure 3. Sampling locations in the Leon River watershed evaluated for BST analysis ...17

Figure 4. Source classification of *E. coli* isolates (combined n=406) from Leon River watershed using a 3-way split. 18

Figure 5. Source classification of *E. coli* isolates (combined n=406) from Leon River watershed using a 7-way split. 18

Figure 6. Source classification of *E. coli* isolates (combined n=100) from Leon River at Coryell CR 183 (TCEQ Station ID 11929) using a 3-way split. 20

Figure 7. Source classification of *E. coli* isolates (combined n=100) from Leon River at Coryell CR 183 (TCEQ Station ID 11929) using a 7-way split. 20

Figure 8. Source classification of *E. coli* isolates (combined n=102) from Leon River at Comanche CR 382 (TCEQ Station ID 11933) using a 3-way split. 22

Figure 9. Source classification of *E. coli* isolates (combined n=102) from Leon River at Comanche CR 382 (TCEQ Station ID 11933) using a 7-way split. 22

Figure 10. Source classification of *E. coli* isolates (combined n=103) from Walnut Creek at FM 1476 (TCEQ Station ID 17379) using a 3-way split. 24

Figure 11. Source classification of *E. coli* isolates (combined n=103) from Walnut Creek at FM 1476 (TCEQ Station ID 17379) using a 7-way split. 24

Figure 12. Source classification of *E. coli* isolates (combined n=101) from Walnut Creek at SH 22 (TCEQ Station ID 17547) using a 3-way split. 26

Figure 13. Source classification of *E. coli* isolates (combined n=101) from Walnut Creek at SH 22 (TCEQ Station ID 17547) using a 7-way split. 26

Figure 14. Comparison chart between Leon River watershed *E. coli* identification overall percentage results using a 7-way split of source class from 2011–2012 (n=566; 131 samples) vs. 2020–2021 (n=406; 48 samples). 28

Figure 15. Comparison chart between Leon River watershed *E. coli* identification for TCEQ Station ID 11929 percentage results using a 3-way split of source class from 2011–2012 (n=60; 12 samples) vs. 2020–2021 (n=100; 12 samples). 29

Figure 16. Comparison chart between Leon River watershed *E. coli* identification for TCEQ Station ID 11933 percentage results using a 3-way split of source class from 2011–2012 (n=50; 10 samples) vs. 2020–2021 (n=102; 12 samples). 30

Figure 17. Comparison chart between Leon River watershed *E. coli* identification for TCEQ Station ID 17379 percentage results using a 3-way split of source class from 2011–2012 (n=40; 8 samples) vs. 2020–2021 (n=103; 12 samples). 31

Figure 18. Comparison chart between Leon River watershed *E. coli* identification for TCEQ Station ID 17547 percentage results using a 3-way split of source class from 2011–2012 (n=20; 4 samples) vs. 2020–2021 (n=101; 12 samples). 32

Figure 19. Reference pathogen health risks at station 11929. 34

Figure 20. Reference pathogen health risks at station 11933.....34

Figure 21. Reference pathogen health risks at station 17379. 35

Figure 22. Reference pathogen health risks at station 17547.....36

Figure 23. Source-specific and overall health risks for each site. Each station is represented by a specific color: 17379-orange, 11929-purple, 11933-blue, 17547-green. 37

List of Acronyms

ARCC	Average rate of correct classification
BST	Bacterial source tracking
cfu	Colony forming units
DNA	Deoxyribonucleic acid
E. coli	*Escherichia coli*
EPA	Environmental Protection Agency
ERIC-PCR	Enterobacterial repetitive intergenic consensus sequence
ERIC-RP	Composite DNA banding pattern from ERIC and RP
NA-MUG	Nutrient agar with 4-methylumbelliferyl-β-D-glucuronide
NRI	Texas A&M Natural Resources Institute
MPN	Most probable number
mTEC	Membrane Thermotolerant *Escherichia coli*
OSSF	On-site sewage facility
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
RARCC	Random average rate of correct classification based on library composition
RCC	Rate of correct classification
RP	RiboPrinting
rRNA	Ribosomal ribonucleic acid
RWQC	Recreational Water Quality Criteria
SARA	San Antonio River Authority
SCSC	Texas A&M AgriLife Research, Department of Soil and Crop Sciences
TCEQ	Texas Commission on Environmental Quality
TP	Temple
TMDL	Total maximum daily load
TSSWCB	Texas State Soil and Water Conservation Board
TWRI	Texas Water Resources Institute
QMRA	Quantitative microbial risk assessment
UTH SPH	University of Texas Health Science Center (UTHealth) at Houston School of Public Health El Paso Campus, Environmental Microbiology Laboratory
WPP	Watershed protection plan
WWTF	Wastewater treatment facility

Executive Summary

The 2020 Texas Integrated Report of Surface Water Quality identified 338 waterbodies as being impaired due to excessive bacteria in Texas. To identify bacterial sources and help address these impairments, Texas established the Bacterial Source Tracking (BST) Program in 2006. To support the maintenance, expansion, and use of the Texas BST Library and other BST tools, the Texas Water Resources Institute (TWRI), University of Texas Health Science Center at Houston School of Public Health El Paso Campus, Environmental Microbiology Laboratory (UTH SPH), and the Texas A&M AgriLife Research, Department of Soil and Crop Sciences (SCSC) collaborated with the Texas State Soil and Water Conservation Board (TSSWCB) in fiscal years 2020 and 2021 to:

(1) Continue personnel support and operation and maintenance of analytical infrastructure at public BST laboratories.

(2) Continue delivery of information and materials that give an overview of BST activities in Texas to date and describe the use, capabilities, and applicability of BST and the services provided by the State-supported analytical labs to local, state, and national stakeholder audiences.

(3) Expand the Texas *Escherichia coli* (*E. coli*) BST Library through known source sample collection in the Leon River watersheds.

(4) Support BST efforts in the Leon River watersheds.

(5) Evaluate and refine the Texas *E. coli* BST Library by assessing geographic and temporal stability, composition, average rates of correct classification, diversity of source isolates of the updated library, and working to develop/refine source-specific bacteria markers for library independent BST.

(6) Provide statistical characterization of the Texas *E. coli* BST Library and integration of BST results and QMRA to evaluate the human health significance.

(7) Provide outreach regarding BST.

Major findings from the project include:

- The Texas *E. coli* BST Library was expanded and refined, with the current version now containing 1,942 isolates from 1,775 known source fecal samples retrieved from 4,351 individual known source samples in over 20 watersheds. An additional 38 isolates from the Leon River watershed were added to the BST Library.
- BST analysis in the Leon River watershed indicate that wildlife (non-avian and avian) are the leading contributors of *E. coli* in the watershed, followed by domestic animals and humans.
- Analysis of the Texas *E. coli* BST Library and qPCR markers identified:
  - the need for continued evaluation of geographic impacts on source identification as the statewide library continues to expand and;

- potential application of new human-specific, dog-specific, and seagull-specific qPCR markers for future BST projects in Texas.

- A quantitative microbial risk assessment was modelled from the data collected from the four sites studied in the Leon River watershed showed relatively low human health risks for recreational activities.

- BST analyses in the Leon River watershed provided the opportunity to conduct a temporal assessment of *E. coli* isolates and fecal pollution contribution in a watershed that was previously evaluated in 2011-2012 and has been supported by a watershed protection plan.
  - There was a slight decrease in the overall percentage of source pollution from wildlife and human contributions with domestic animals and unidentified sources slightly increasing, however this current report sampled at four sites with normal rainfall while the previous Leon River report in 2011-2012 sampled at 15 sites under extreme drought conditions.

- Outreach of the BST Program resulted in:
  - AgriLife SCSC gave a presentation on Bacterial Source Tracking (BST) on Tributaries of Trinity and Galveston Bays for Water and Sediment Quality Subcommittee of Galveston Bay Council, online meeting, June 10, 2020.
  - AgriLife SCSC gave a presentation on Bacterial Source Tracking (BST) on Tributaries of Trinity and Galveston Bays for Galveston Bay Council Quarterly Meeting, online meeting, July 15, 2020.
  - AgriLife SCSC gave a presentation on Use of Bacterial Source Tracking to Characterize Texas Watersheds at North Central Texas Council of Governments meeting, online meeting, January 7, 2021.
  - AgriLife SCSC gave a presentation on April 1, 2022, entitled “Use of Bacterial Source Tracking for Characterization of Watersheds” at a webinar hosted by the North Central Texas Council of Governments.
  - Discussed potential BST projects with multiple other cities and river authorities.

Introduction

Bacteria are the number one cause of water quality impairments in Texas. Bacterial Source Tracking (BST) is a valuable tool for identifying human and animal sources of fecal pollution to support development of watershed plans, total maximum daily loads (TMDLs), and other strategies for addressing these impairments. Comprehensive BST has been completed by UTSPH EP and AgriLife SCSC in numerous watersheds throughout Texas with support provided by the TSSWCB. As a result of these joint efforts over the last decade, the Texas *E. coli* BST Library (ver. 04-22) currently contains 1,942 *E. coli* isolates obtained from 1,775 different domestic sewage, wildlife, livestock, and pet fecal samples. Despite its expansiveness, continued development and refinement of the library to include additional known source isolates from additional Texas watersheds and different animal hosts are needed to further increase its utility. Looking to the future, library independent BST holds much promise. It is already being used to support BST analyses in Texas. However, to improve its ability to address the needs in Texas, further work is needed to develop and evaluate new markers. To further strengthen the statistical integrity of current BST work, different statistical methods need to be evaluated to calculate confidence intervals and provide a range of certainty/uncertainty with current library-dependent BST work. Evaluating the temporal integrity of the BST library in a watershed with previous BST analysis is necessary as well. Quantitative microbial risk assessment (QMRA) is a valuable tool that can integrate BST results and improve risk estimations for specific water bodies. Efforts to delineate QMRA outputs to inform policy and best practices can increase the utility of BST work. Finally, continued outreach and technology transfer is needed to expand awareness and understanding of BST, foster dialogue and collaboration, and bring water resource managers up to speed on advances in BST technologies, methodologies, applications, and results.

According to the *2020 Texas Integrated Report*, there are 338 impairments due to excessive bacteria. One key to effectively abating these impairments is the identification and assessment of fecal pollution sources. Proper evaluation of these sources is needed to target best management practices and develop bacterial TMDLs or watershed protection plans (WPPs). This information may also be useful to properly assess risk in contact recreation, as many waterborne pathogens causing human illnesses do not colonize nonhuman hosts. Use of genetic and biochemical tests which allow identification of the original host species is referred to as bacterial source tracking. The premise behind BST is that genetic and phenotypic tests can identify bacterial strains that are host specific so that the original host species and source of fecal contamination can be identified. While there has been some controversy concerning host specificity and survival of *E. coli* in the environment (Gordon, Bauer et al. 2002), this indicator organism has the advantage of being correlated with the presence of fecal contamination and being used for human health risk assessments. Thus, *E. coli* BST has direct regulatory significance and standardized culturing techniques for water samples available, such as EPA Method 1603 (USEPA 2005).

BST is a valuable tool for identifying sources of fecal pollution. Comprehensive BST has been completed by UTSPH EP and AgriLife SCSC for the following watersheds: (1) Lake Waco and Belton Lake, (2) San Antonio area, (3) Lake Granbury, (4) Buck Creek, (5)

Leon and Lampasas Rivers, (6) Little Brazos River tributaries, (7) Big Cypress Creek, (8) Leona River, (9) Attoyac Bayou, (10) Galveston Bay, (11) Big Elm Creek, (12) Mission River, and (13) Aransas River. A Texas *E. coli* BST Library has been developed based on known source isolates from these and other (i.e. Upper Trinity River and Upper Oyster Creek) watersheds. The Texas *E. coli* BST Library (ver. 03-20) currently contains 1,912 *E. coli* isolates obtained from 1,653 different domestic sewage, wildlife, livestock, and pet fecal samples. While this represents a significant step towards development of a statewide *E. coli* BST library, there remains a need for continued expansion of the library to include additional known source isolates from different Texas watersheds and different animal hosts. As the library is expanded, this will allow continued evaluation of the library for geographical stability and the diversity of source specific isolates to identify specific needs for future expansion and refinement of the library. Further, use of the Texas *E. coli* BST Library provides for significant cost and time savings for the identification of nonpoint source pollution in the development of TMDLs and WPPs. Lastly, the state of BST science, methodologies, application and confidence continues to evolve. Continued outreach and technology transfer is needed to foster dialogue and collaboration and bring water resource managers up to speed on advances in BST technologies, methodologies, applications, and results.

Expansion of the Texas *E. coli* BST Library

The Texas *E. coli* BST Library is a key component of the Texas BST Program, successfully identifying sources of *E. coli* in more than a dozen watersheds across Texas over the past decade. The Texas *E. coli* BST Library is dynamic, with new isolates being added with each successive BST project. In an effort to expand the Texas *E. coli* BST Library and to continue to support BST analyses in the Leon River watersheds, the project aimed to collect approximately 50 known source fecal samples, from which 75 *E. coli* isolates were fingerprinted for potential addition to the library. Over the course of the project, multiple trips were made to gather known source samples. Specific arrangements were made to meet with landowners to collect both livestock and wildlife samples. Human wastewater treatment facility (WWTF) samples were collected from both the inlets and outlets of functioning WWTFs in the watersheds. On-site sewage facility (OSSF) samples were collected from septic pump trucks operating in the watershed areas. Lastly, roadkill was also utilized as a source of wildlife samples when opportunities presented themselves. BST analyses in the Leon River watershed also provided the opportunity to conduct a temporal assessment of *E. coli* isolates and fecal pollution contribution in a watershed that was previously evaluated in 2011-2012 and was supported by WPPs.

Known-source sampling in the Leon River watershed resulted in a total of 50 unique samples being collected between September 2020 and August 2021. Samples collected were held on ice until being transported to AgriLife SCSC for processing within 96 hours of collection. Table 1 describes the number of samples collected per source, the number of samples testing positive for *E. coli*, screened, validated, archived, and added to the Texas *E. coli* BST library.

Table 1. Number of samples collected per source, the number of samples testing positive for *E. coli*, screened, validated, archived, and added to the Texas *E. coli* BST Library.

Source	Samples Collected	Samples (+) for *E.coli*	Isolates archived	Isolates screened by ERIC	Isolates RP in local library	Self-validated (isolate/sample)	TXSV 04-22 (isolate/sample)
Human	**11**	**10**	**102**	**102**	**32**	**24/9**	**23/9**
Sewage	8	7	58	58	22	18/6	17/6
Septic	3	3	44	44	10	6/3	6/3
Cattle	**6**	**6**	**19**	**19**	**7**	**2/2**	**2/2**
Other non-avian livestock	**10**	**9**	**28**	**28**	**13**	**5/4**	**3/3**
Goat	4	4	13	13	5	1/1	1/1
Horse	3	2	6	6	3	1/1	0/0
Sheep	3	3	9	9	5	3/2	2/2
Pets	**5**	**4**	**14**	**14**	**7**	**2/2**	**2/2**
Cat	1	1	4	4	2	1/1	0/0
Dog	4	3	10	10	5	1/1	2/2
Avian Wildlife	**0**	**0**	**0**	**0**	**0**	**0/0**	**0/0**
Non-Avian Wildlife	**18**	**18**	**51**	**51**	**16**	**8/7**	**8/7**
Bobcat	2	2	6	6	4	2/2	2/2
Coyote	2	2	6	6	3	1/1	1/1
Deer	1	1	3	3	1	0/0	0/0
Hog, Feral	11	11	31	31	5	2/2	2/2
Fox	1	1	2	2	1	1/1	1/1
Raccoon	1	1	3	3	2	2/1	2/1
Total	**50**	**47**	**214**	**214**	**75**	**41/24**	**38/23**

Of the 50 fecal known-source samples processed, 47 had culturable *E. coli* as determined using EPA Method 1603 (modified mTEC and NA-MUG positive). A total of 214 isolates from these samples were collected and archived. All 214 *E. coli* isolates were screened with ERIC-PCR, and 75 known-source isolates from 47 unique known-source fecal samples were DNA fingerprinted using RiboPrinting (ERIC-RP). After screening the 214 *E. coli* isolates for clonality, 84 unique *E. coli* isolates were identified. AgriLife SCSC analyzed 75 of these 84 unique *E. coli* isolates with RP. The chosen 75 *E. coli* isolates from 47 known-source samples were fingerprinted by ERIC-RP.

The 75 known-source isolates from the Leon River local library were screened using the traditional self-validation step (a stringent seven-way split of source classes and an 80% similarity cutoff) resulting in 41 self-validated isolates from 24 samples. These 41 self-validated isolates from the local library were combined with the similarly screened isolates from all previous watershed studies in order to perform serial Jackknife analyses to create the Texas *E. coli* BST library ver. 04-22, which contains 1,942 isolates,

including 38 from the Leon River watershed. The composition, average rates of correct classification, and diversity of this new version of the library are detailed in Table 2.

During the project period, the Texas *E. coli* BST Library was used to identify fecal pollution source contributions in the Leon River watershed as part of this project and other watersheds as part of multiple projects funded by the Texas Commission on Environmental Quality (TCEQ), San Antonio River Authority (SARA), and other cities and organizations.

BST Library Refinement

AgriLife SCSC evaluated the geographical and temporal stability, composition, average rates of correct classification (accuracy), and diversity of source specific isolates, while continuing to further develop and refine the Texas *E. coli* BST library with new known-source isolates. To increase its accuracy and utility, the updated Texas *E. coli* BST Library with pooled self-validated local watershed libraries as described in Table 2 (3,839 isolates) was refined through cross-validation. To attempt to remove cosmopolitan (non-specific*) E. coli* source isolates, repetitive Jackknife analyses of the combined self-validated libraries were performed to remove isolates that cross-identified between human, domestic animals, and wildlife with the goal of 100% average rate of correct classification (ARCC) using a 3-way split of source classes. In the first round of serial Jackknife analysis, 1,851 isolates were removed leaving 1,988 isolates. Four additional rounds of Jackknife analysis were performed, resulting in 1,942 isolates with a 100% ARCC using a three-way split of source classes and a 93% ARCC using a seven-way split. A total of 19% of the isolates were singletons (i.e., unique fingerprints; Table 2). The Texas *E. coli* BST Library ver. 04-22 contains 1,942 isolates obtained from 1,775 individual fecal samples. Library composition is based on three- and seven-way source class splits (Figures 1 and 2 respectively).

Table 2. Texas *E. coli* BST Library (ver. 04-22, cross-library validation) composition and rates of correct classification (RCCs) by Jackknife analysis of ERIC-RP composite data sets using an 80% similarity cutoff and three- and seven-way splits.

Source Class	Number of Isolates	Number of Samples	Library Composition and Expected Random Rate of Correct Classification	Calculated Rate of Correct Classification (RCC)	RCC to Random Ratio***	Left Unidentified (unique patterns)
HUMAN	**447**	**408**	**23%**	**100**	**4.3**	**23%**
DOMESTIC ANIMALS	**562**	**530**	**29%**	**100**	**3.4**	**20%**
Pets	91	82	5%	83	16.6	43%
Cattle	247	230	13%	94	7.2	11%
Avian Livestock	98	95	5%	86	17.2	28%
Other Non-Avian Livestock	126	123	6%	88	14.6	14%
WILDLIFE	**933**	**837**	**48%**	100	**2.1**	**17%**
Avian Wildlife	273	259	14%	79	5.6	19%
Non-Avian Wildlife	660	578	34%	92	2.7	16%
%Overall	**1942**	**1775**		**ARCC** ** **= 3-way 100% 7-way 93%**		**19%**

*RARCC, expected random average rate of correct classification based on library composition

**ARCC = average rate of correct classification: the proportion of all identification attempts which were correctly identified to source class for the entire library, which is similar to the mean of the RCCs for all source classes when the number of isolates in each source class is similar

***An RCC/Random Ratio greater than 1.0 indicates that the rate of correct classification is better than random. For example, the rate of correct classification for human is 4.5-fold greater than random chance based on library composition.

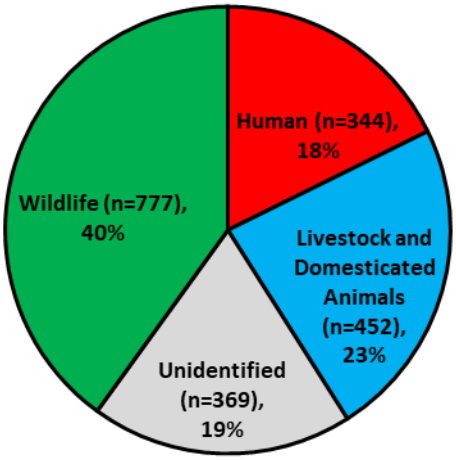

Figure 1. Texas *E. coli* BST Library (ver. 04-22) composition by three-way split of source classes (1,942 isolates from 1,775 different fecal source samples).

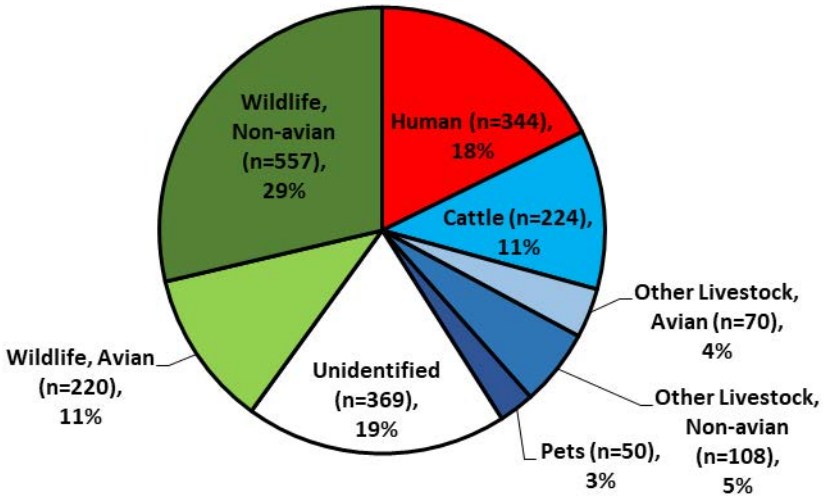

Figure 2. Texas *E. coli* BST Library (ver. 04-22) composition by seven-way split of source classes (1,942 isolates from 1,775 different fecal source samples).

Leon River

An additional 75 isolates from the Leon River watershed known-source isolates were added to generate the Texas *E. coli* BST Library ver. 04-22.

Of the 50 total known source fecal samples collected by the Texas A&M Natural Resources Institute (NRI) from the watershed, AgriLife SCSC successfully isolated *E. coli* from 47 individual samples. From these samples, 214 isolates (at least three isolates per known source sample) were screened using ERIC-RP and included in the local watershed library. AgriLife SCSC did subsequent library evaluation and used Jackknife analysis of the ERIC-RP to identify isolates that correctly classified using a 7-way split of source classes (i.e., human, pets, cattle, other non-avian livestock, avian livestock, avian wildlife, and non-avian wildlife). Isolates with unique fingerprints (left unidentified using an 80% similarity cutoff) were also included to create the local self-validated library. In total, 41 isolates were self-validated in the local library.

The 41 local self-validated source isolates from the watershed were then added to the current library of Texas *E. coli* BST self-validated source isolates from previous watershed projects across Texas. A series of Jackknife analyses were run on the combined libraries, removing all isolates that cross-identified between human, domestic animals, and wildlife. After each removal, the Jackknife was run again with the goal of 100% ARCC using a three-way split of source classes. After four iterations of cross-watershed validation, the resulting Texas *E. coli* BST Library (ver. 04-22) contained 1,942 isolates from 1,775 samples, resulting in a 100% ARCC with a three-way split of source classes and a 93% ARCC using the seven-way split of source classes. A total of 19% of the isolates were identified as singletons (unique fingerprints left unidentified using an 80% similarity cutoff) and were kept in the library in order to reflect the diversity of patterns potentially seen in unknown water samples. After cross-watershed validation, 38 isolates (51% of the local library samples) were included in the Texas *E. coli* BST Library (ver. 04-22). The 38 isolates were comprised of individual fecal samples from sewage (17), septic (6), cattle (2), goat (1), sheep (2), dog (2), bobcat (2), coyote (1), feral hog (2), fox (1), and raccoon (2). The 75 isolates were included with the new Leon River isolates when evaluating and generating the newest version of the library as described in the previous section (Table 2).

Texas *E. coli* BST Library Use

Leon River Watershed

NRI collected water samples from the Leon River watershed at four different sites from September 2020 through August 2021 (Figure 3). Samples were delivered to AgriLife SCSC for processing using EPA Method 1603. For twelve monthly samples, up to eight isolates from each sample were tested for each of the four sites, confirmed as *E. coli* (modified mTEC and NA-MUG positive), and archived. Collectively, the geometric mean for the *E. coli* colony counts for all four sites combined was 301 cfu/100mL. All isolates were fingerprinted using ERIC-PCR and RP. A total of 406 *E. coli* isolates were fingerprinted using ERIC-RP and compared against Texas *E. coli* BST Library v. 04-22 for source determination. Overall results for the Leon River watershed isolates are shown in Figures 4 and 5.

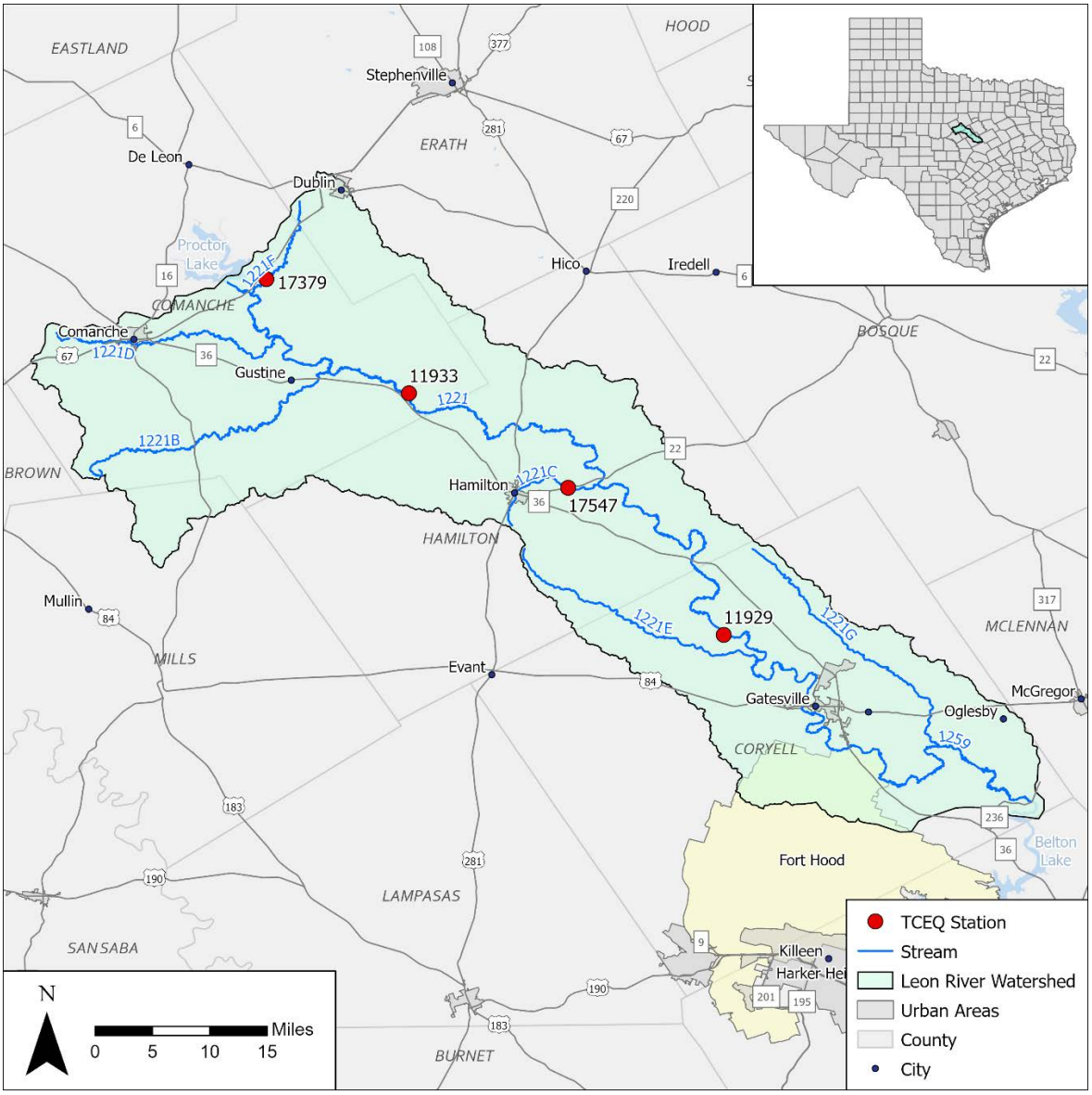

Figure 3. Sampling locations in the Leon River watershed evaluated for BST analysis

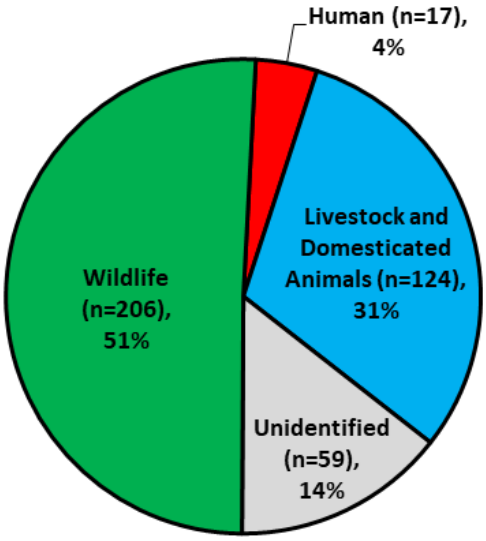

Figure 4. Source classification of *E. coli* isolates (combined n=406) from Leon River watershed using a three-way split.

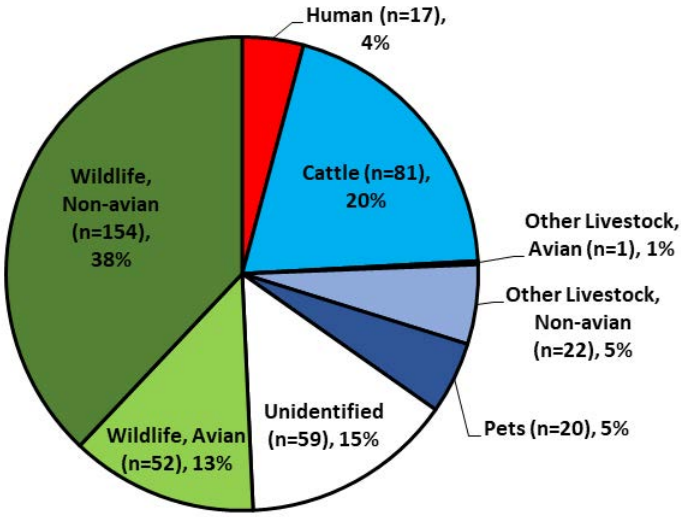

Figure 5. Source classification of *E. coli* isolates (combined n=406) from Leon River watershed using a seven-way split.

Leon River at Coryell CR 183

NRI collected water samples monthly from the Leon River at Coryell CR 183 (TCEQ Station ID 11929) from September 2020 through August 2021. *E. coli* colony counts ranged from 154 to 3,700 cfu/100 mL, and the geometric mean was 500 cfu/100 mL (Table 3). All isolates were fingerprinted using ERIC-PCR and RP. A total of 100 *E. coli* isolates were fingerprinted using ERIC-RP and compared against Texas *E. coli* BST Library v. 04-22 for source determination. Overall results for these isolates are shown in Figures 6 and 7.

Using a three-way split, 38% of the isolates were classified as originating from wildlife, 44% from livestock and domesticated animals, and 6% from humans. Using the more detailed seven-way split, 32% of the isolates were classified as originating from cattle, 30% from non-avian wildlife, 8% from avian wildlife, 7% for other non-avian livestock, 6% from humans, 4% from pets, and 1% for other avian livestock. The source could not be identified for 12% of the isolates.

Table 3. Monthly counts of *E. coli* isolates from water samples for Leon River at Coryell CR 183 (TCEQ Station ID 11929) between September 2020 through August 2021.

Sampling Months	*E. coli* (cfu/100mL)
September '20	660
October '20	268
November '20	159
December '20	500
January '21	160
February '21	670
March '21	990
April '21	3700
May '21	690
June '21	154
July '21	590
August '21	560
Geometric Mean	500

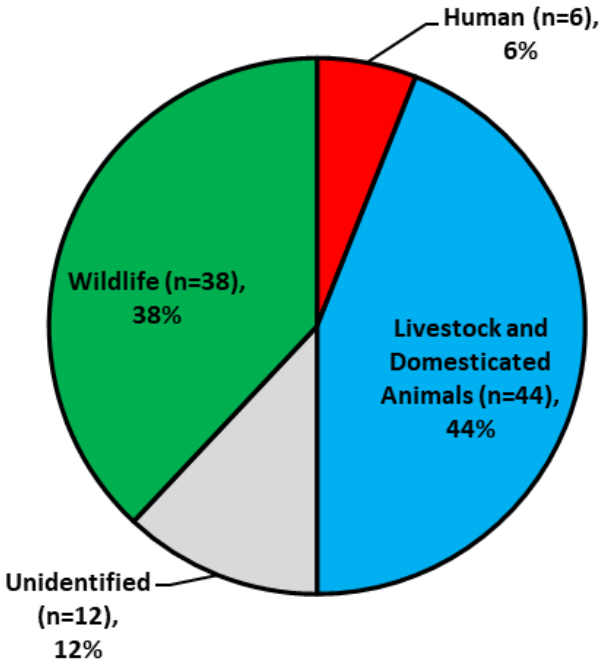

Figure 6. Source classification of *E. coli* isolates (combined n=100) from Leon River at Coryell CR 183 (TCEQ Station ID 11929) using a three-way split.

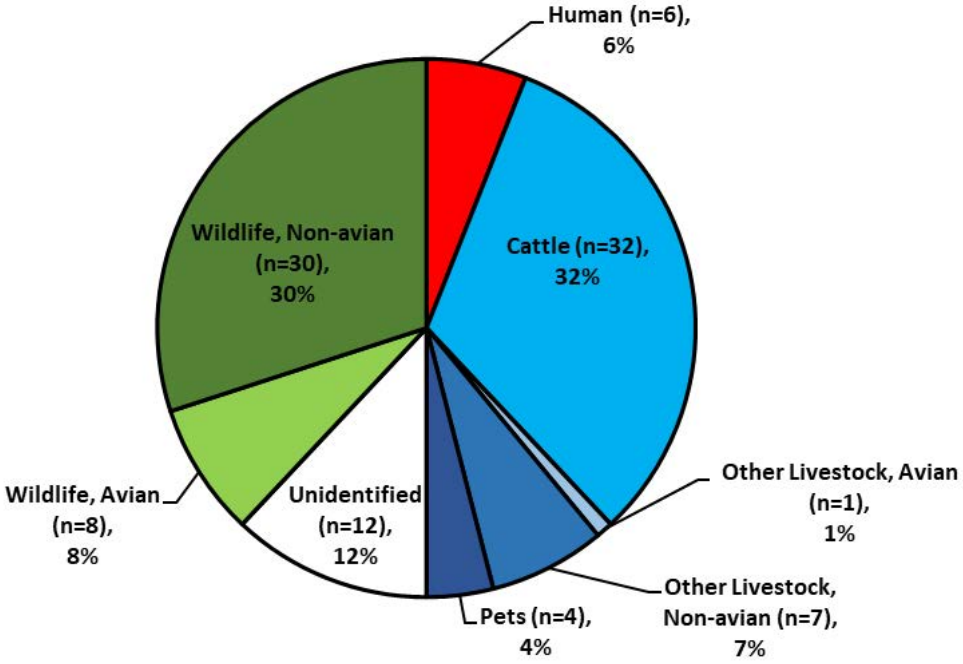

Figure 7. Source classification of *E. coli* isolates (combined n=100) from Leon River at Coryell CR 183 (TCEQ Station ID 11929) using a seven-way split.

Leon River at Comanche CR 382

NRI collected water samples monthly from the Leon River at Comanche CR 382 (TCEQ Station ID 11933) from September 2020 through August 2021. *E. coli* colony counts ranged from 154 to 2,100 cfu/100 mL, and the geometric mean was 352 cfu/100 mL (Table 4). All isolates were fingerprinted using ERIC-PCR and RP. A total of 102 *E. coli* isolates were fingerprinted using ERIC-RP and compared against Texas *E. coli* BST Library v. 04-22 for source determination. Overall results for the Leon River at Comanche CR 382 isolates are shown in Figures 8 and 9.

Using a three-way split, 50% of the isolates were classified as originating from wildlife, 25% from livestock and domesticated animals, and 6% from humans. Using the more detailed seven-way split, 41% of the isolates were classified as originating from non-avian wildlife, 17% from cattle, 9% from avian wildlife, 6% from humans, 5% from pets, and 4% for other non-avian livestock. The source could not be identified for 18% of the isolates.

Table 4. Monthly counts of *E. coli* isolates from water samples for Leon River at Comanche CR 382 (TCEQ Station ID 11933) between September 2020 through August 2021.

Sampling Months	*E. coli* (cfu/100mL)
September '20	481
October '20	209
November '20	174
December '20	131
January '21	271
February '21	272
March '21	820
April '21	2100
May '21	1900
June '21	129
July '21	154
August '21	220
Geometric Mean	352

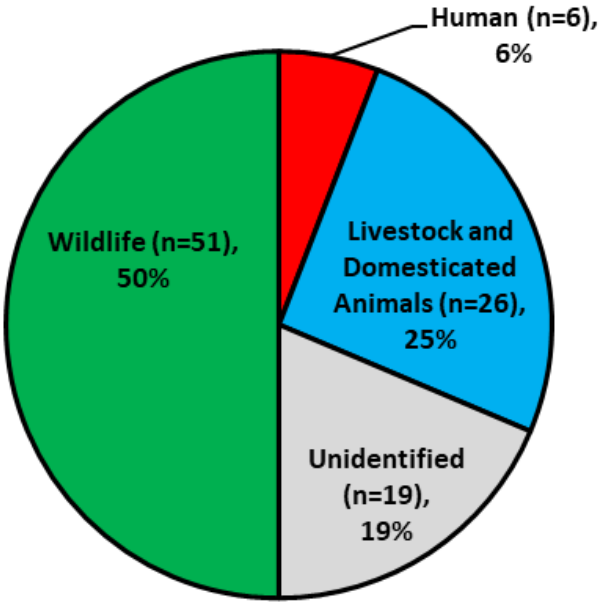

Figure 8. Source classification of *E. coli* isolates (combined n=102) from Leon River at Comanche CR 382 (TCEQ Station ID 11933) using a three-way split.

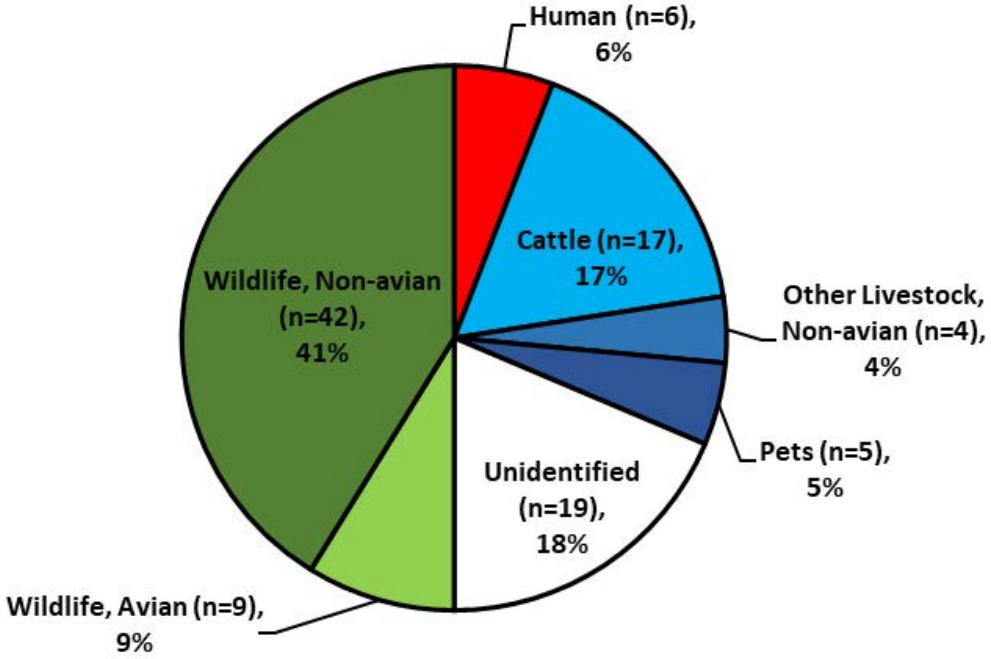

Figure 9. Source classification of *E. coli* isolates (combined n=102) from Leon River at Comanche CR 382 (TCEQ Station ID 11933) using a seven-way split.

Walnut Creek at FM 1476

NRI collected water samples monthly from Walnut Creek at FM 1476 (TCEQ Station ID 17379) from September 2020 through August 2021. *E. coli* colony counts ranged from 56 to 4,900 cfu/100 mL, and the geometric mean was 276 cfu/100 mL (Table 5). All isolates were fingerprinted using ERIC-PCR and RP. A total of 103 *E. coli* isolates were fingerprinted using ERIC-RP and compared against Texas *E. coli* BST Library v. 04-22 for source determination. Overall results for the Walnut Creek at FM 1476 isolates are shown in Figures 10 and 11.

Using a three-way split, 48% of the isolates were classified as originating from wildlife, 31% from livestock and domesticated animals, and 5% from humans. Using the more detailed seven-way split, 34% of the isolates were classified as originating from non-avian wildlife, 16% from cattle, 14% from avian wildlife, 8% for other non-avian livestock, 7% from pets, and 5% from humans. The source could not be identified for 16% of the isolates.

Table 5. Monthly counts of *E. coli* isolates from water samples for Walnut Creek at FM 1476 (TCEQ Station ID 17379) between September 2020 through August 2021.

Sampling Months	*E. coli* (cfu/100mL)
September '20	283
October '20	73
November '20	56
December '20	180
January '21	370
February '21	580
March '21	227
April '21	152
May '21	4900
June '21	670
July '21	226
August '21	139
Geometric Mean	276

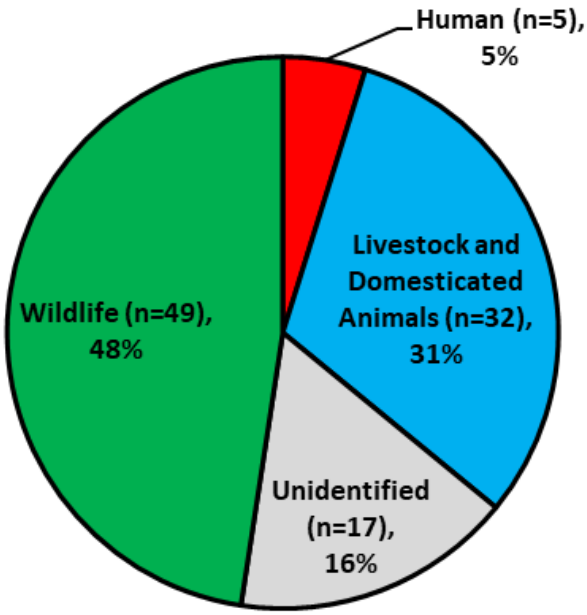

Figure 10. Source classification of *E. coli* isolates (combined n=103) from Walnut Creek at FM 1476 (TCEQ Station ID 17379) using a three-way split.

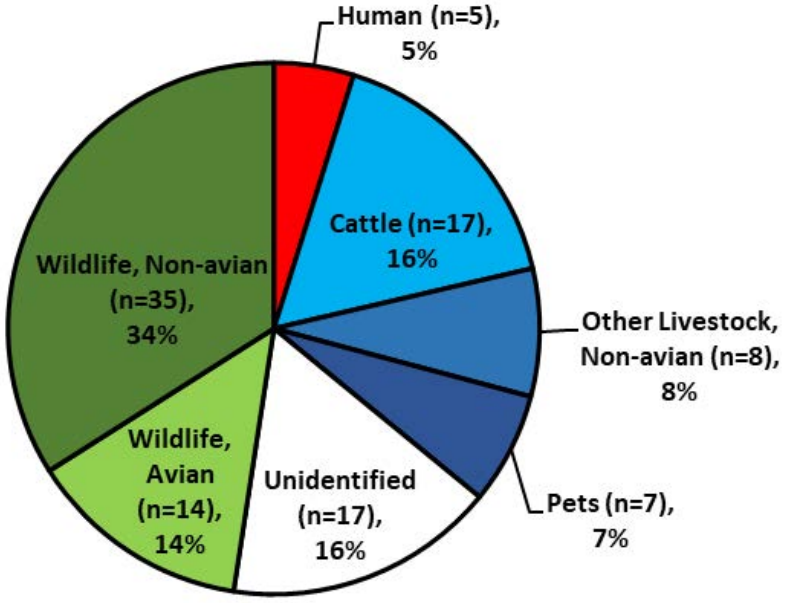

Figure 11. Source classification of *E. coli* isolates (combined n=103) from Walnut Creek at FM 1476 (TCEQ Station ID 17379) using a seven-way split.

Pecan Creek at SH 22

NRI collected water samples monthly from Pecan Creek at SH 22 (TCEQ Station ID 17547) from September 2020 through August 2021. *E. coli* colony counts ranged from 18 to 7,000 cfu/100 mL, and the geometric mean was 168 cfu/100 mL (Table 6). All isolates were fingerprinted using ERIC-PCR and RP. A total of 101 *E. coli* isolates were fingerprinted using ERIC-RP and compared against Texas *E. coli* BST Library v. 04-22 for source determination. Overall results for the Pecan Creek at SH 22 isolates are shown in Figures 12 and 13.

Using a three-way split, 67% of the isolates were classified as originating from wildlife, 22% from livestock and domesticated animals, and 0% from humans. Using the more detailed seven-way split, 46% of the isolates were classified as originating from non-avian wildlife, 21% from avian wildlife, 15% from cattle, 4% from pets, and 3% for other non-avian livestock. The source could not be identified for 11% of the isolates.

Table 6. Monthly counts of *E. coli* isolates from water samples for Pecan Creek at SH 22 (TCEQ Station ID 17547) between September 2020 through August 2021.

Sampling Months	*E. coli* (cfu/100mL)
September '20	227
October '20	62
November '20	18
December '20	64
January '21	36
February '21	158
March '21	730
April '21	260
May '21	7000
June '21	1900
July '21	26
August '21	38
Geometric Mean	168

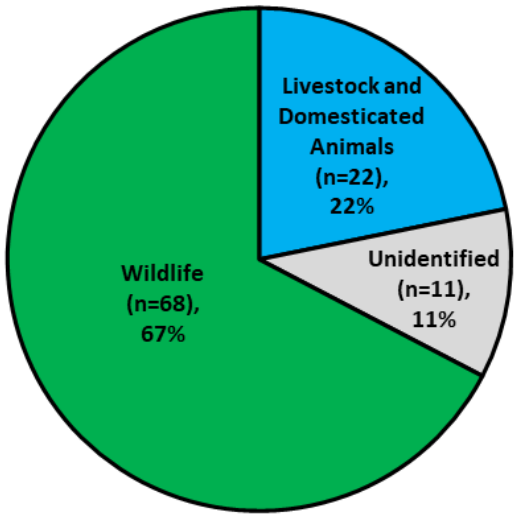

Figure 12. Source classification of *E. coli* isolates (combined n=101) from Walnut Creek at SH 22 (TCEQ Station ID 17547) using a three-way split.

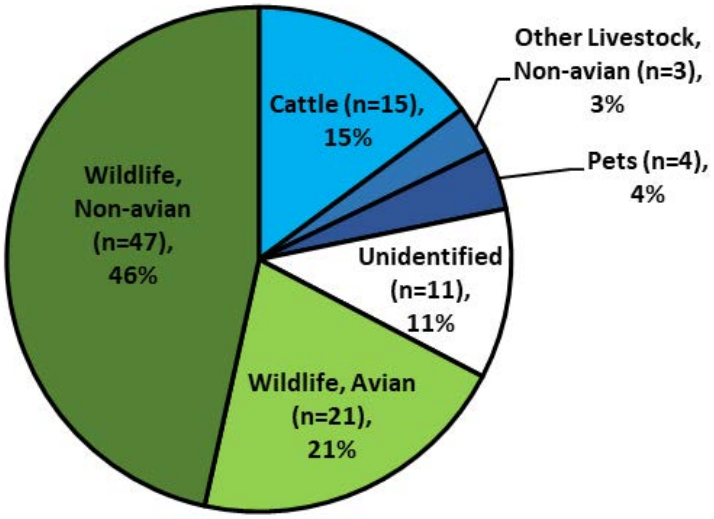

Figure 13. Source classification of *E. coli* isolates (combined n=101) from Walnut Creek at SH 22 (TCEQ Station ID 17547) using a seven-way split.

Leon River Watershed Temporal BST Comparison

Between February 2011 and January 2012, UTSPH-EP received 116 water samples from flowing water from 15 sampling stations within the Leon River watershed as part of TSSWCB Project 10-51. The average number of water samples collected among the sites was seven samples per site ranging from 2 to 12. The water samples were processed using EPA Method 1603. Following enumeration, cultures were shipped to El Paso, TX for genetic typing by UTHealth Houston School of Public Health. Collectively, the

geometric mean of data from all 15 sites was 60.6 cfu/100 mL. Up to 5 isolates per sample, were analyzed with ERIC-PCR and RP. A total of 566 isolates from 114 water samples were fingerprinted using ERIC-RP and identified using the Texas *E. coli* BST Library v. 10-12 for source determination. The number of isolates identified varied between 20 to 60 isolates per site.

Between September 2020 and August 2021, AgriLife SCSC received 12 water samples of flowing water from four sampling station from the Leon River watershed. Samples were processed using EPA Method 1603. Up to 8 isolates from each sample were isolated from all four sites each month, and the minimum number of isolates identified was 100 isolates per site. Collectively, the geometric mean of colony count data from all four sites ranged from 168 cfu/100mL to 500 cfu/100mL. The 2011-2012 water samples collected at the same four sites in comparison had a geometric mean of colony count data ranging from 16 cfu/100mL to 163 cfu/100 mL. All isolates were fingerprinted using ERIC-PCR and RP. A total of 406 *E. coli* isolates were fingerprinted using ERIC-RP and compared against Texas *E. coli* BST Library v. 04-22 for source determination. BST results for all 566 watershed isolates identified from the 2011-2012 Leon River watershed iteration were compared with the overall identification results for all 406 watershed isolates analyzed from the 2020–2021 Leon River watershed iteration in Figure 14.

Wildlife was the major contributor of *E. coli* contamination in the Leon River watershed for both iterations, however there was a slight percentage decrease of 44% to 38% (non-avian) and 16% to 13% (avian) in 2020–2021. Regarding the human-derived *E. coli* isolates, there was a slight decrease from 8% to 4% in 2020–2021. Also, other avian livestock decreased from 2% to 1% in 2020–2021. All other source classes, including unidentified, slightly increased in 2020–2021 (Figure 14).

The 2011-2012 Leon River project documented water quality and quantity conditions observed in the Leon River watershed under exceptional drought conditions that meteorologists characterized as the worst 1-year drought documented in Texas since record keeping began in 1895. Normal average annual rainfall for the area is approximately 30 inches. However. precipitation between February 1 and September 30, 2011 averaged 4.33 inches, and precipitation between October 1, 2011 and January 24, 2012 averaged 7.34 inches. These extreme drought conditions might account for the drastically lower *E. coli* geometric means when compared to the 2020–2021 Leon River watershed report.

Following the initial BST study in the Leon River watershed, several rounds of implementation funding were dedicated to repairing and replacing failing OSSFs. These efforts were focused in Hamilton and Coryell counties. Centralized wastewater treatment facilities in the cities of Comanche and Hamilton and another system operated by the Upper Leon River Municipal Water District were all upgraded to improve their treatment processes. These actions all have direct impacts on instream water quality and BST results suggest that they have reduced contributions of human derived *E. coli* into the Leon River watershed.

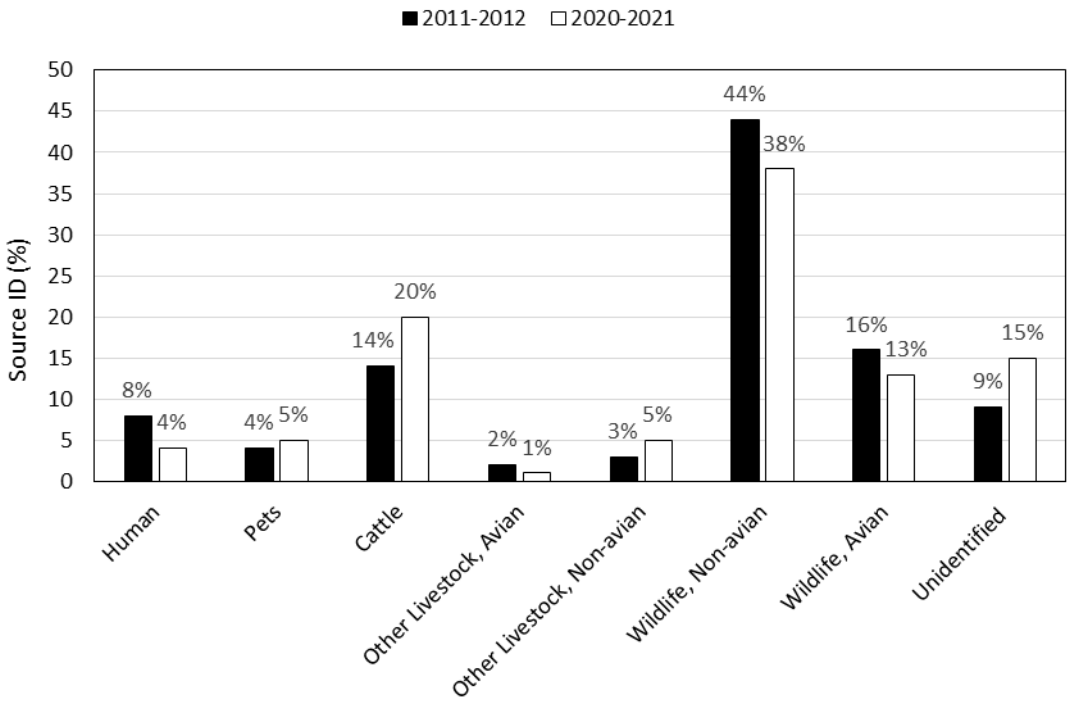

Figure 14. Comparison chart between Leon River watershed *E. coli* identification overall percentage results using a seven-way split of source class from 2011-2012 (n=566; 131 samples) vs. 2020–2021 (n=406; 48 samples).

Leon River at Coryell CR 183

During 2011–2012, TCEQ Station ID 11929 produced 12 samples and 60 isolates were identified with an overall geometric mean of 76 cfu/100 mL. Of those 60 *E. coli* isolates, 5 were human, 15 were livestock and domesticated animals, 36 were wildlife, and 4 were left unidentified. During the 2020–2021 Leon River watershed iteration, 12 samples were received, and 100 isolates were identified with an overall geometric mean of 500 cfu/100 mL. Of those 100 *E. coli* isolates, 6 were human, 44 were livestock and domesticated animals, 38 were wildlife, and 12 were left unidentified. A percentage comparison chart of site 11929 is shown in Figure 15. The human and wildlife source class contributions have decreased while the livestock and domestic animals and the unidentified source class contributions have increased.

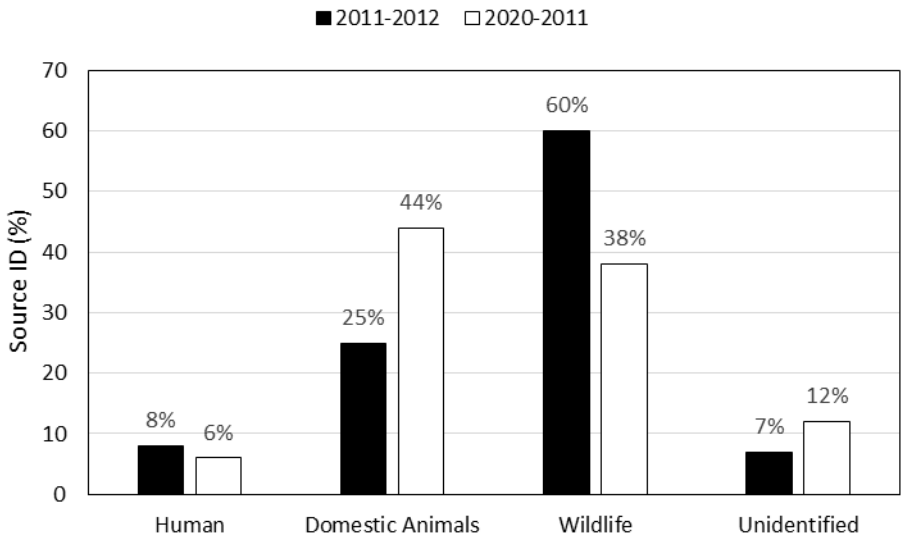

Figure 15. Comparison chart between Leon River watershed *E. coli* identification for TCEQ Station ID 11929 percentage results using a three-way split of source class from 2011–2012 (n=60; 12 samples) vs. 2020–2021 (n=100; 12 samples).

Leon River at Comanche CR 382

During 2011–2012, TCEQ Station ID 11933 produced 10 samples and 50 isolates were identified with an overall geometric mean of 118 cfu/100 mL. Of those 50 *E. coli* isolates, 3 were human, 10 were livestock and domesticated animals, 34 were wildlife, and 3 were left unidentified. During the 2020–2021 Leon River watershed iteration, 12 samples were received, and 102 isolates were identified with an overall geometric mean of 352 cfu/100 mL. Of those 102 *E. coli* isolates, 6 were human, 26 were livestock and domesticated animals, 51 were wildlife, and 19 were left unidentified. A percentage comparison chart of site 11933 is shown in Figure 16. The wildlife source class contributions have decreased while the livestock and domestic animals and the unidentified source class contributions have increased. The human source class contributions remained the same.

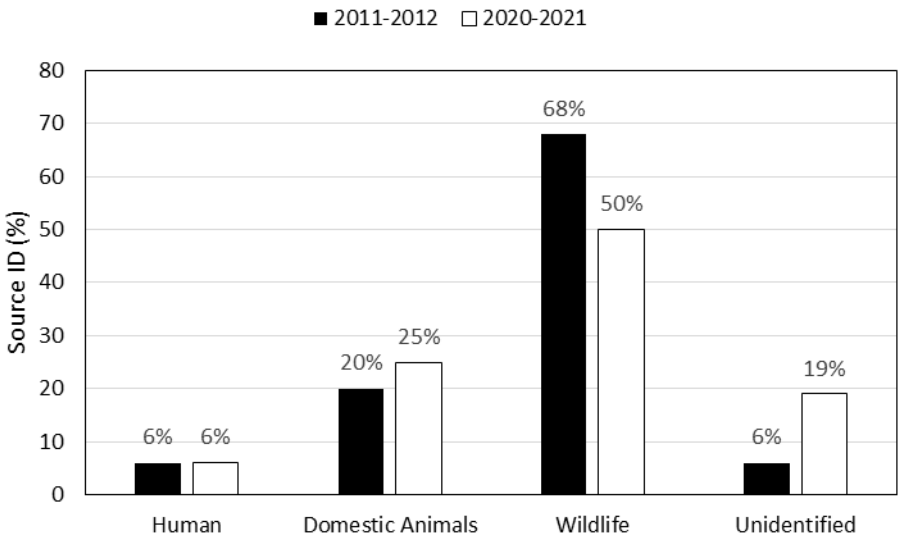

Figure 16. Comparison chart between Leon River watershed *E. coli* identification for TCEQ Station ID 11933 percentage results using a three-way split of source class from 2011–2012 (n=50; 10 samples) vs. 2020-2021 (n=102; 12 samples).

Walnut Creek at FM 1476

During 2011–2012, TCEQ Station ID 17379 produced eight samples and 40 isolates were identified with an overall geometric mean of 163 cfu/100 mL. Of those 40 *E. coli* isolates, 3 were human, 10 were livestock and domesticated animals, 26 were wildlife, and 1 was left unidentified. During the 2020–2021 Leon River watershed iteration, 12 samples were received, and 103 isolates were identified with an overall geometric mean of 276 cfu/100 mL. Of those 103 *E. coli* isolates, 5 were human, 32 were livestock and domesticated animals, 49 were wildlife, and 17 were left unidentified. A percentage comparison chart of site 17379 is shown in Figure 17. The human and wildlife source class contributions have decreased while the livestock and domestic animals and the unidentified source class contributions have increased.

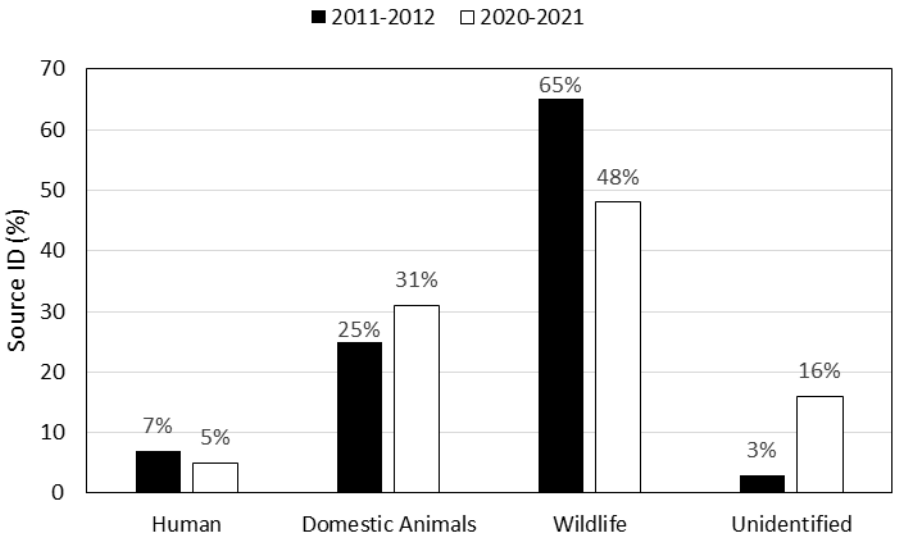

Figure 17. Comparison chart between Leon River watershed *E. coli* identification for TCEQ Station ID 17379 percentage results using a three-way split of source class from 2011-2012 (n=40; 8 samples) vs. 2020-2021 (n=103; 12 samples).

Pecan Creek at SH 22

During 2011-2012, TCEQ Station ID 17547 received four samples and 20 isolates were identified with an overall geometric mean of 16 cfu/100 mL. Of those 20 *E. coli* isolates, 2 were human, 4 were livestock and domesticated animals, 12 were wildlife, and 2 were left unidentified. During the 2020-2021 Leon River watershed iteration, 12 samples were received, and 101 isolates were identified with an overall geometric mean of 168 cfu/100 mL. Of those 101 *E. coli* isolates, 0 were human, 22 were livestock and domesticated animals, 68 were wildlife, and 11 were left unidentified. A percentage comparison chart of site 17547 is shown in Figure 18. The livestock and domestic animals, wildlife, and the unidentified source class contributions have increased, while the human source class contributions reduced to zero.

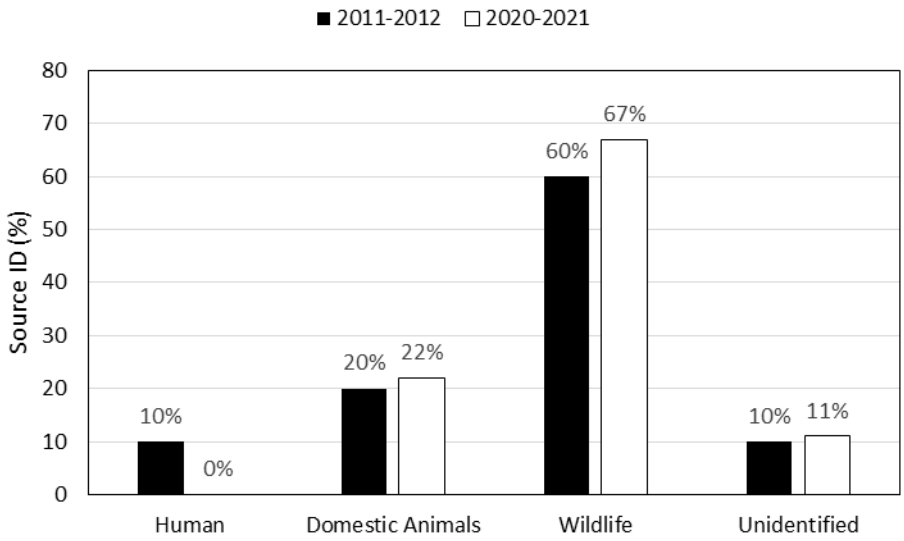

Figure 18. Comparison chart between Leon River watershed *E. coli* identification for TCEQ Station ID 17547 percentage results using a three-way split of source class from 2011-2012 (n=20; 4 samples) vs. 2020-2021 (n=101; 12 samples).

Evaluating Site-Specific Health Risks with Quantitative Microbial Risk Assessment

The United States Recreational Water Quality Criteria (RWQC) of 2012 provided guidance regarding the development of alternative site-specific water quality standards using the quantitative microbial risk assessment (QMRA) framework (EPA, 2012). This mathematical framework consists of four phases - hazard identification, exposure assessment, dose-response, and risk characterization - to estimate potential human health risks associated with exposure to specific microbial contaminants (Haas, Rose, and Gerba, 2014). Results of BST, coupled with the site-specific *E. coli* concentrations measured in environmental samples, can be utilized in a QMRA to estimate human health risks that may be present due to waterborne pathogens in recreational waters. Estimated health risks developed from these site-specific conditions can be compared to the U. S. EPA risk threshold of 36 cases of a gastrointestinal illness per 1,000 recreators, or 0.036 (EPA, 2012). This approach of incorporating environmental BST and microbiological data into a QMRA to inform of health risks and provide guidance to water quality managers is currently being reviewed and implemented by public agencies in California, Canada, and more broadly in the U.S. (SCCWRP, 2022; Health Canada, 2021; EPA, 2012). This framework was applied to the four Leon River sites sampled during this project to evaluate not only human health risks, but the utility of using QMRA to supplement BST analyses in the state.

The three-way split source-classification and geometric average of *E. coli* isolates at TCEQ stations 11929, 11933, 17379, and 17547 were incorporated into a QMRA framework to estimate human health risks for recreational activities at the four sites.

The unidentified category of each source-classification was divided to match the representation of the other three fecal sources. All sites except for 17379 are listed to meet primary contact 1 recreational standard (126 cfu/100 mL), while site 17379 is required to meet the water quality standards for secondary contact recreation two (1,030 cfu/100 mL). Exposure scenarios that included ingestion of water while swimming were conducted for sites 11929, 11933, and 17547, while site 17379 was evaluated for incidental ingestion of water while fishing. A potential exposure dose for a variety of pathogens was developed based on the concentration of *E. coli* isolates and three-way split source-classification. The three source categories were represented by a primary fecal polluter to narrow down the pathogens of greatest public health concern. The human source is represented by primary sewage (e.g. from a failing septic system), the domestic animal source is represented by cattle, and wildlife is represented by deer. Since not all *E. coli* isolates that are measured in a water body are pathogenic, reference pathogen doses were estimated, which take into consideration the prevalence of the pathogen in the host source and infectivity of the pathogen to humans (EPA, 2010; Soller et al., 2010; Soller et al., 2014). Reference pathogens for each fecal source used in the QMRA, that have been identified to pose a human health risk, are listed in Table 7.

Table 7. Reference pathogens of public health concern for each fecal source.

Fecal Source	Reference Pathogens of Concern
Human (sewage)	Norovirus, *Salmonella*, *Cryptosporidium*, *Giardia*, *Campylobacter*, *E. coli* O157:H7
Domestic Animals (Cattle)	*Salmonella*, *Cryptosporidium*, *Giardia*, *Campylobacter*, *E. coli* O157:H7
Wildlife (Deer)	*Cryptosporidium* and *Giardia*

Leon River at Coryell CR 183

Station 11929 had a geometric average of 500 cfu/100 mL with a recategorized three-way source split of human (6.8%), domestic animals (50%), and wildlife (43.2%). The human health risks associated with each reference pathogen at the site are depicted in Figure 19. Reference pathogen health risks at station 11929. The health risks associated with each fecal source is identified as blue-human, orange-domestic animals, and green-wildlife. The red dashed line indicates the 0.036 U.S. EPA risk threshold.

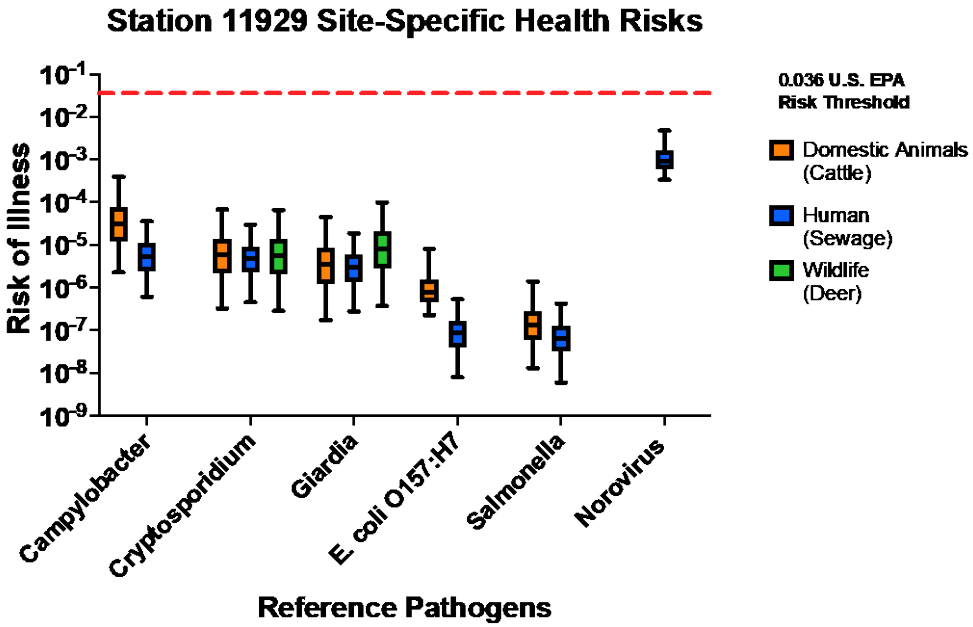

Figure 19. Reference pathogen health risks at station 11929.

Leon River at Comanche CR 382

Station 11933 had a geometric average of 352 cfu/100 mL with a recategorized three-way split of human (7.0%), wildlife (62.0%) and domestic animals (31%). The human health risks associated with each reference pathogen at the site are depicted in Figure 20.

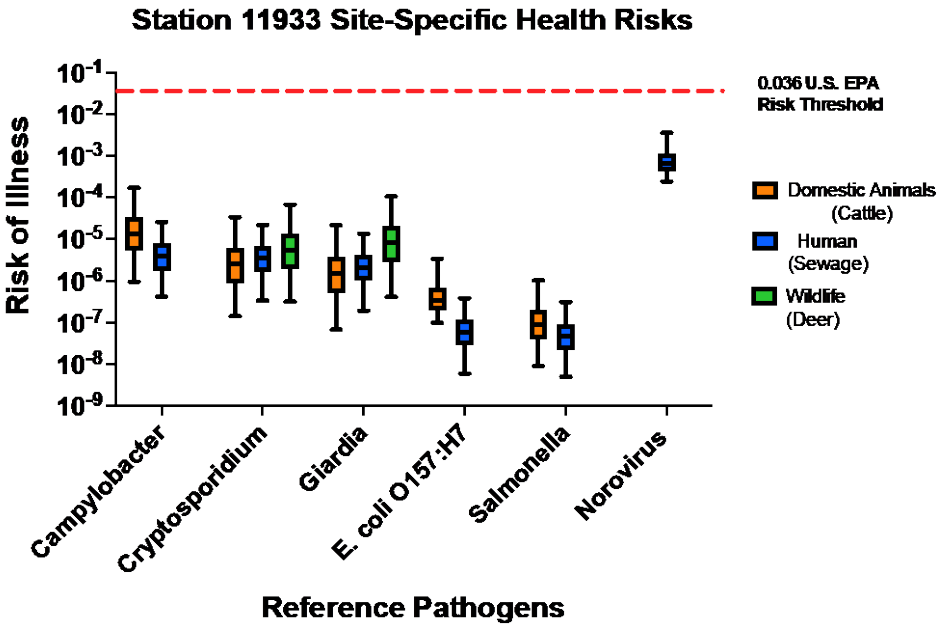

Figure 20. Reference pathogen health risks at station 11933.

Walnut Creek at FM 1476

Station 17379 had a geometric average of 276 cfu/100mL but is only required to meet secondary contact 2 recreational water quality standards. The recategorized three-way split included human (6.0%), wildlife (57.0%), and domestic animals (37.0%). The health risks associated with each reference pathogen at the site are depicted in Figure 21.

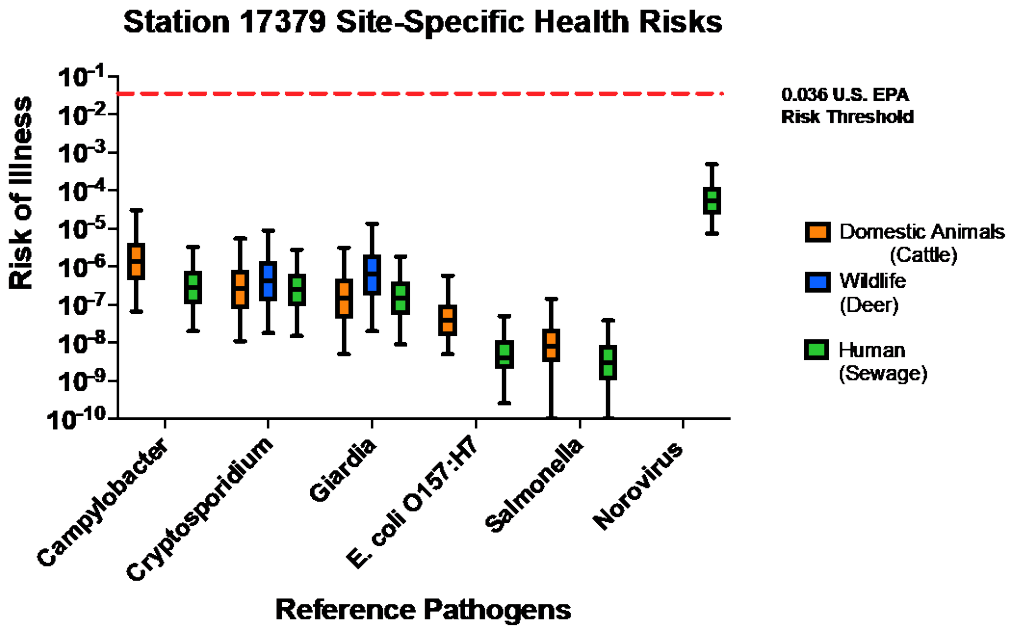

Figure 21. Reference pathogen health risks at station 17379.

Pecan Creek at SH 22

Lastly, station 17547 had a geometric average of 168 cfu/100mL, and a recategorized three-way split that only included wildlife (75.6%) and domestic animals (24.4%). Human fecal isolates were not detected at this site and were therefore not included in the risk estimate. The human health risks associated with wildlife and domestic animals are depicted in Figure 22.

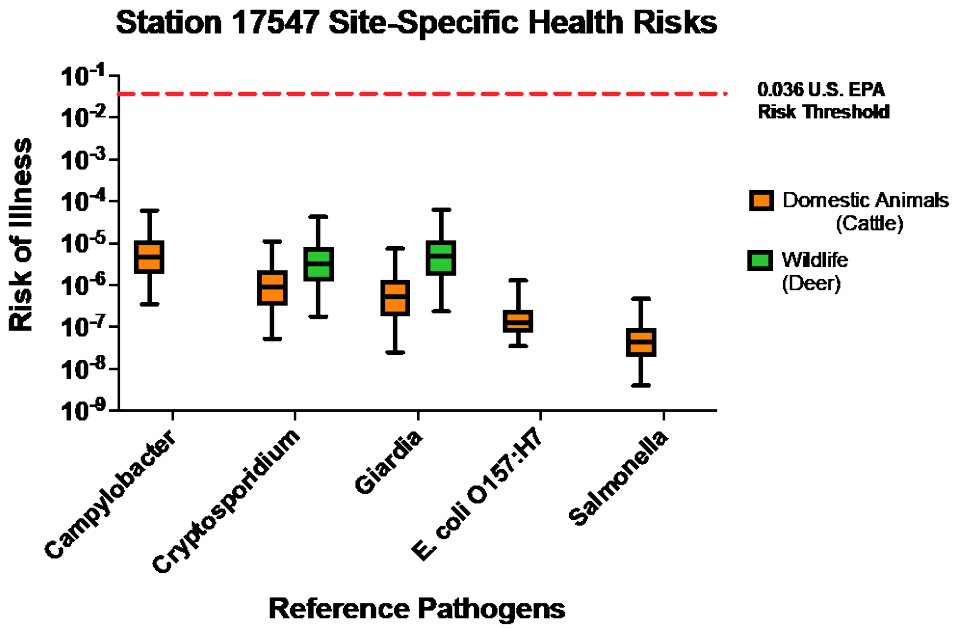

Figure 22. Reference pathogen health risks at station 17547.

Reference pathogens from both human (sewage) and the domestic animal (cattle) sources contribute towards a greater health risk than the pathogens from wildlife (deer). For the human source, norovirus likely contributes the greatest risk for human health, followed by *Campylobacte*r from domestic animals. The 95th percentile health risk for a gastrointestinal illness, amongst all reference pathogens at all four sites, did not exceed the U.S. EPA recreational risk threshold of 0.036. The overall health risk associated with each fecal source at all four sites was also estimated and shown in Figure 23.

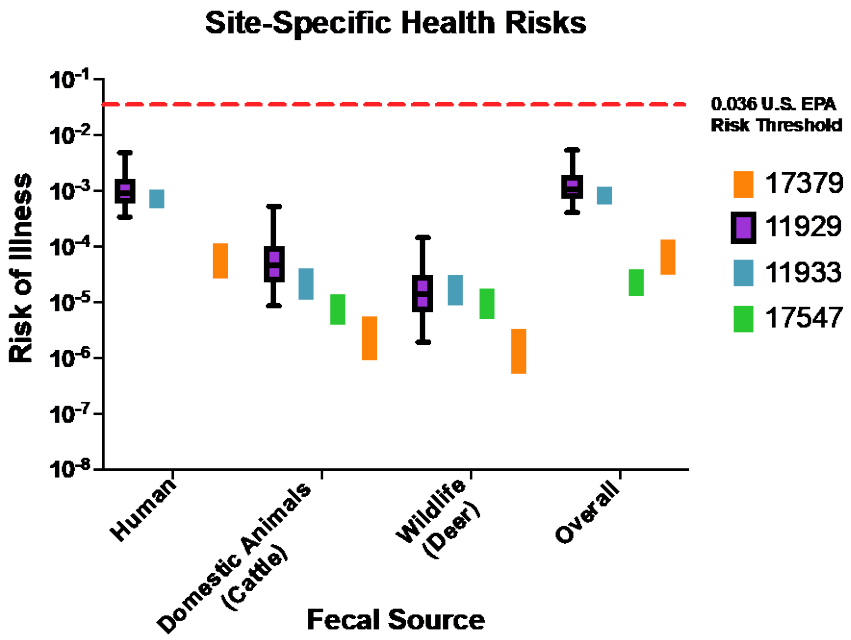

Figure 23. Source-specific and overall health risks for each site. Each station is represented by a specific color: 17379-orange, 11929-purple, 11933-blue, 17547-green.

When considering the different fecal sources impacting each site, none of the estimated health risks exceeded the U.S. EPA risk threshold of 0.036. While three of the four sites did not meet water quality standards for *E. coli*, health risks were not found to be elevated, which is likely due to the significant proportion of *E. coli* originating from wildlife. These findings (of estimated human health risks associated with fecal indicator bacteria or pathogens from wildlife being lower than that of human and cattle sources) parallel similar QMRA studies conducted in both fresh and marine waters (Schoen and Ashbolt, 2010; Soller et al., 2010, 2014; Boehm and Soller, 2020). Health risks from fecal pathogens do exist, especially for pathogens that originate from humans and domestic animals (i.e. cattle). However, it is important to not only consider if microbiological water quality standards are being exceeded, but also which fecal sources are contributing to these exceedances. Targeting bacteria pollution from human and domestic animal sources may not always result in the greatest reduction of the overall *E. coli* contribution to a waterbody, but it will be the most protective for human health. This application of QMRA with BST provides valuable information regarding relative risks from pathogen sources that can aid watershed managers in decision making that can better target protecting public health.

Other Watersheds

AgriLife SCSC continued BST analysis in support of the San Antonio River Authority's (SARA's) watershed characterization efforts. A total of 104 *E. coli* isolates from 10 water samples collected in 2021-2022 were isolated, verified as *E. coli*, fingerprinted by ERIC-RP, and compared against Texas *E. coli* BST Library ver. 03-20 for source identification.

Additionally, AgriLife SCSC provided two watersheds in Central Texas with BST analysis. A total of 415 *E. coli* isolates from 51 water samples collected in 2020 were isolated, verified as *E. coli*, fingerprinted by ERIC-RP, and compared against Texas *E. coli* BST Library ver. 03-20 for source identification.

Evaluation of the Texas *E. coli* BST Library

The Texas *E. coli* BST Library was first developed as a way to counter one of the biggest drawbacks of library-dependent source tracking methods—the need to collect large numbers of known source samples for every watershed study. An identification library should reflect the large host inter- and intra-species variation in *E. coli*, so that the DNA fingerprints of *E. coli* isolates found in the water can be matched and their sources identified. We theorized that there would be enough geographical and temporal stability in the host specificity of *E. coli* populations to allow the DNA fingerprints from known source isolates from different watershed studies to be pooled together. Developing a statewide BST library using *E. coli* isolates from local watershed libraries allows for time and cost savings.

Defining Host Class

BST is based on the premise that different strains of *E. coli* have adapted to different gut environments to become host specific. There are many caveats to consider when dealing with *E. coli* populations. There are different strains of *E. coli* in a single individual while different strains of the bacteria will exist in different individuals. Similar strains may be present in similar environments. Strains of *E. coli* present may be dependent upon animal species; gut type; diet; environment; and interactions with other individuals/species. The questions of temporal and geographical variations are especially relevant to a pooled library, such as the Texas *E. coli* BST Library, since it is built over time from different watersheds. While library-independent markers attempt to identify a limited specific animal species, our library-dependent approach can work with these realities of *E. coli* populations while still giving practical results by using host classes. The most supportable division of known sources is into three host classes: human, domestic animals, and wildlife. These embrace the adaptations to a shared environment, allow the use of a wide variety of wildlife, and do not penalize cross-identification seen between livestock. The division of water isolates into human, domestic animals and wildlife is also practical for making decisions about best management practices. There is also more statistical strength when small numbers of isolates are divided into fewer categories. To ensure a variety of host classes are included in collections for the library, domestic animals and wildlife are further divided into subsets. Below are the three- and seven-way split categories that were used for categorizing *E. coli* isolates and which we have most frequently used for characterizing watersheds:

Three-way split

1. Domesticated animals and livestock (livestock and pets)
2. Wildlife (including feral hogs)
3. Humans

Seven-way split

1. Cattle
2. Other livestock, non-avian (non-avian livestock other than cattle; sheep, etc.)
3. Other livestock, avian (chickens, etc.)
4. Pets (dogs, cats)
5. Avian wildlife (ducks, geese, sparrows, etc.)
6. Non-avian wildlife (deer, feral hogs, coyotes, etc.)
7. Humans

For any *E. coli* isolate that could not be matched to a group in the Texas *E. coli* BST Library, its source category was designated as being "unidentified."

Developing a Feral Hog Host Class

To expand the utility of the Texas *E. coli* BST Library, a four-way split of host classes was explored. This creates a new category specifically for feral hogs that is separate from the wildlife pool of known *E. coli* isolates. Instead of refining the current version of the Texas *E. coli* BST Library, a new four-way split version was established starting with the total 3,839 of all known source *E. coli* isolates that have ever been archived and DNA fingerprinted from all previous watersheds, of which 268 *E. coli* isolates originated from feral hogs.

Repetitive Jackknife analyses of the combined self-validated libraries was performed to remove isolates that cross-identified between human, domestic animals, feral hogs, and wildlife to attempt to remove cosmopolitan (non-specific*) E. coli* source isolates. The goal of this analysis was a 100% average rate of correct classification (ARCC) using a four-way split of source classes. In the first round of serial Jackknife analysis, 1,557 isolates were removed leaving 2,282 isolates. Three additional rounds of Jackknife analysis were performed, resulting in 2,116 isolates with a 100% ARCC using a four-way split of source classes and an 80% ARCC using an eight-way split. A total of 19% of the isolates were singletons (i.e., unique fingerprints; Table 8). The utilization of this four-way library variation with the separate feral hog category looks promising due to the 100% ARCC at the initial source class split. Also, a larger pool of isolates remain after the serial Jackknife analysis when compared to the current three-way Texas *E. coli* BST library (ver. 04-22). However, future work needs to be done by analyzing a select number of isolates against each library and comparing the results before a determination can be made.

Table 8. Texas *E. coli* BST Library (ver. 4-way: 05-22, cross-library validation) composition and rates of correct classification (RCCs) by Jackknife analysis of ERIC-RP composite data sets using an 80% similarity cutoff and four and eight-way splits

Source Class	Number of Isolates	Number of Samples	Library Composition and Expected Random Rate of Correct Classification	Calculated Rate of Correct Classification (RCC)	RCC to Random Ratio***	Left Unidentified (unique patterns)
HUMAN	**430**	**358**	**20%**	**100**	**5.0**	**24%**
DOMESTIC ANIMALS	**847**	**769**	**40%**	**100**	**2.5**	**12%**
Pets	132	119	6%	83	13.8	27%
Cattle	347	316	16%	73	4.6	7%
Avian Livestock	145	124	7%	56	8.0	17%
Other Non-Avian Livestock	223	210	11%	58	5.3	8%
FERAL HOGS	**77**	**72**	**4%**	**100**	**25**	**23%**
WILDLIFE	**762**	**669**	**36%**	**100**	**2.8**	**17%**
Avian Wildlife	292	263	14%	69	4.9	17%
Non-Avian Wildlife	470	406	26%	84	3.2	17%
Overall	**2,116**	**1,868**		**ARCC** ** **= 4-way 100% 8-way 80%**		**19%**

*RARCC, expected random average rate of correct classification based on library composition

**ARCC = average rate of correct classification: the proportion of all identification attempts which were correctly identified to source class for the entire library, which is similar to the mean of the RCCs for all source classes when the number of isolates in each source class is similar

***An RCC/Random Ratio greater than 1.0 indicates that the rate of correct classification is better than random. For example, the rate of correct classification for human is 4.5-fold greater than random chance based on library composition.

Evaluation of Library-Independent PCR Markers

In an effort to expand the BST toolbox for future projects, additional library-independent markers and platforms were evaluated by AgriLife SCSC. As detailed below, the dog specific marker DogBact, the gull specific marker LeeSeaGull, and the Bacteroidales human-associated marker HF183/BacR287 was used for qPCR-based analysis of water samples and preliminary tests.

Studies have shown that fecal pollution is associated with a decrease in the resilience and diversity of marine coastal systems. Galveston, Texas has been identified as hotspot for bacterial pollution by the Texas Beach Watch in long-term analysis monitoring fecal pollution along the Texas Coast. In collaboration with the Texas General Land Office the project "Integrative assessment of Bacterial Pollution" (contract number: 21-060-025-D274), samples that are identified as high (>104MPN/100mL) are sent to Texas A&M, College Station, and evaluated using qPCR. Data will be used to pinpoint stations that exhibit a history of bacterial pollution as well as confirming the presence of human, dog, or gull waste through the testing of host specific markers.

AgriLife SCSC evaluated qPCR for the dog marker, DogBact, and the gull marker, LeeSeaGull. The protocols followed Dick et al. (2005) for DogBact and Lee et al. (2013) for LeeSeaGull. Both were evaluated using synthetic standards of known copy number, synthesized by Integrated DNA Technologies (IDTDNA, Coralville, Iowa, USA) and constructed from a portion of the 16S rRNA gene of *Bacteroides* for DogBact and *Catellicoccus* for LeeSeaGull. The DogBact marker was then optimized with the standard curve R^2 value of 0.996 and LeeSeaGull was optimized at a R^2 of 0.991 (R^2 values should be ≥ 0.980). Note that R^2 is a measure of the variability in the qPCR assay calibration curve.

Lastly, in order to characterize human sources of fecal pollution in recreational waters, an EPA Method 1696-based approach was by evaluated by AgriLife SCSC using standard reference for the Bacteroidales human-associated qPCR marker HF183/BacR287 (Ref), internal amplification control (IAC), method blank (MB), and three water samples from Burton Creek in College Station, TX (unknown 1, 2, 3). Two components of the EPA protocol were adjusted: (1) a synthetic fragment from a similar DNA sequence was used for the standard reference and the internal amplification control sequence instead of plasmid DNA, and (2) for multiplexing, two fluorescence probes, FAM and SUNN, were used instead of the two fluorescence probes, FAM and VIC from the EPA method. The SUNN probe was used instead as an equivalent alternative, and it was more cost-effective through Integrated DNA Technologies (IDTDNA, Coralville, Iowa, USA). The marker was optimized with the standard curve R^2 value of 0.999 (R^2 values should be ≥ 0.980).

These markers have only been optimized at this point and need additional testing to validate their accuracy. The next step is to evaluate unknown isolates with each marker then verify their correct classification with other BST methods.

Education and Outreach

Education and outreach activity during the project included a combination of presentations, website updates and discussions about BST at various events. A total of four presentations were given regarding the use, application and results of BST and specific BST projects around the state. Two of these were regarding BST results for Galveston and Trinity Bays and were presented to the Galveston Bay Council. The other two presentations were to the North Central Texas Council of Governments regarding the use and application of BST in the greater Dallas-Fort Worth metroplex.

The website was routinely updated throughout the course of the project and led to various phone calls regarding the use of BST in other parts of the state. During the project period, a total of 662 site visits occurred from 601 unique users. These website visits and presentations led to additional conversations with individuals from municipalities, river authorities and water districts regarding the use, costs, and ability to perform BST in other parts of the state.

Literature Cited

Atwill ER, Pereira MD, Alonso LH, Elmi C, Epperson WB, Smith R, Riggs W, Carpenter LV, Dargatz DA, Hoar B. 2006. Environmental load of *Cryptosporidium parvum* oocysts from cattle manure in feedlots from the central and western United States. Journal of environmental quality;35(1):200-6.

Berry ED, Woodbury BL, Nienaber JA, Eigenberg RA, Thurston JA, Wells JE. 2007. Incidence and persistence of zoonotic bacterial and protozoan pathogens in a beef cattle feedlot runoff control–vegetative treatment system. Journal of Environmental Quality; 36(6):1873-82.

Boehm AB, Graham KE, Jennings WC. 2018. Can we swim yet? Systematic review, meta-analysis, and risk assessment of aging sewage in surface waters. Environmental Science & Technology; 52(17):9634-45.

Boehm AB and Soller JA. 2020. Refined ambient water quality thresholds for human-associated fecal indicator HF183 for recreational waters with and without co-occurring gull fecal contamination. Microbial Risk Analysis; 16:100139

Crockett CS. 2007. The Role of Wastewater Treatment in Protecting Water Supplies Against Emerging Pathogens. Water Environment Research; 79(3):221–32.

Dick, L.K., Bernhard, A.E., Brodeur, T.J., Domingo, J.S., Simpson J.M., Walters, S.P., Field, K.G. 2005. Host distributions of uncultivated fecal Bacteroidales bacteria reveal genetic markers for fecal source identification. *Applied Environmental Microbiology*. 71(6), 3184-3191.

Dorevitch S, Panthi S, Huang Y, Li H, Michalek AM, Pratap P, Wroblewski M, Liu L, Scheff PA, Li A. 2011. Water ingestion during water recreation. Water research; 45(5):2020-8.

Dufour AP, Behymer TD, Cantu R, Magnuson M, Wymer LJ. 2017. Ingestion of swimming pool water by recreational swimmers. Journal of water and health;15(3):429-37.

Eftim SE, Hong T, Soller J, Boehm A, Warren I, Ichida A, et al. 2017. Occurrence of norovirus in raw sewage – A systematic literature review and meta-analysis. Water Research; 111:366–74.

Eisenberg JN, Seto EYW, Olivieri AW, Spear RC. 1996. Quantifying Water Pathogen Risk in an Epidemiological Framework. Risk Analysis;16(4):549–63.

EPA. 2006. National primary drinking water regulations: Long Term 2 Enhanced Surface Water Treatment Rule (LT2ESWTR). Available from: https://pubmed.ncbi.nlm.nih.gov/11800007/

EPA. 2010. Quantitative Microbial Risk Assessment to Estimate Illness in Freshwater Impacted by Agricultural Animal Sources of Fecal Contamination. Office of Water 822-R-10-005. Retrieved from: https://www.epa.gov/sites/default/files/2015-11/documents/quantitiave-microbial-risk-fecal.pdf

EPA. 2012. 2012 Recreational Water Quality Criteria. Office of Water 820-F-12-058. Retrieved from: https://www.epa.gov/sites/default/files/2015-10/documents/rwqc2012.pdf

EPA. 2019. *Method 1696: Characterization Of Human Fecal Pollution In Water By Taqman Quantitative Polymerase Chain Reaction (qPCR) Assay*. NEPIS, pp. 1-56.

Fayer R, Trout JM, Graczyk TK, Lewis EJ. 2000. Prevalence of *Cryptosporidium*, *Giardia* and *Eimeria* infections in post-weaned and adult cattle on three Maryland farms. Veterinary parasitology;93(2):103-12.

Fossler CP, Wells SJ, Kaneene JB, Ruegg PL, Warnick LD, Eberly LE, Godden SM, Halbert LW, Campbell AM, Bolin CA, Zwald AG. 2005. Cattle and environmental sample-level factors associated with the presence of Salmonella in a multi-state study of conventional and organic dairy farms. Preventive veterinary medicine;67(1):39-53.

Garcia-Aljaro C, Bonjoch X, Blanch AR. 2005. Combined use of an immunomagnetic separation method and immunoblotting for the enumeration and isolation of Escherichia coli O157 in wastewaters. Journal of Applied Microbiology. 98(3):589–97.

García-Presedo I, Pedraza-Díaz S, González-Warleta M, Mezo M, Gómez-Bautista M, Ortega-Mora LM, Castro-Hermida JA. 2013. The first report of Cryptosporidium bovis, C. ryanae and Giardia duodenalis sub-assemblage A-II in roe deer (Capreolus capreolus) in Spain. Veterinary parasitology;197(3-4):658-64.

Gordon, D. M., Bauer, S., & Johnson, J. R. 2002. The genetic structure of Escherichia coli populations in primary and secondary habitats. *Microbiology (Reading, England)*, *148*(Pt 5), 1513–1522. https://doi.org/10.1099/00221287-148-5-1513

Health Canada. 2021. https://www.canada.ca/en/health-canada/programs/consultation-guidelines-recreational-water-quality-understanding-managing-risks-recreational-waters/document.html

Haas, C. N.; Rose, J. B.; Gerba, C. P. 1999 Quantitative microbial risk assessment; John Wiley & Sons: New York

Haas, C. N.; Rose, J. B.; Gerba, C. P. 1999 Quantitative microbial risk assessment; John Wiley & Sons: New York

Haas CN, Rose JB, Gerba CP. 1999. Quantitative Microbial Risk Assessment; John Wiley & Sons: New York.

Haas CN, Rose JB, Gerba CP. 2014. Quantitative Microbial Risk Assessment. Second Edition. Quantitative Microbial Risk Assessment: Second Edition. Wiley Blackwell.

Harwood VJ, Levine AD, Scott TM, Chivukula V, Lukasik J, Farrah SR, et al. 2005. Validity of the Indicator Organism Paradigm for Pathogen Reduction in Reclaimed Water and Public Health Protection. Applied and Environmental Microbiology; 71(6):3163–70.

Heitman TL, Frederick LM, Viste JR, Guselle NJ, Morgan UM, Thompson, R.C. and Olson, M.E. 2002. Prevalence of Giardia and Cryptosporidium and characterization of Cryptosporidium spp. isolated from wildlife, human, and agricultural sources in the North Saskatchewan River Basin in Alberta, Canada. Canadian journal of microbiology; 48(6): 530-541.

Hoar BR, Atwill ER, Elmi C, Farver TB. 2001. An examination of risk factors associated with beef cattle shedding pathogens of potential zoonotic concern. Epidemiology & Infection;127(1):147-55.

Hutchison ML, Walters LD, Avery SM, Synge BA, Moore A. 2004. Levels of zoonotic agents in British livestock manures. Letters in Applied Microbiology; 39(2):207-14.

Kitajima M, Haramoto E, Iker BC, Gerba CP. 2014. Occurrence of Cryptosporidium, Giardia, and Cyclospora in influent and effluent water at wastewater treatment plants in Arizona. Science of the Total Environment;484(1):129–36.

Koivunen J, Siitonen A, Heinonen-Tanski H. 2003. Elimination of enteric bacteria in biological-chemical wastewater treatment and tertiary filtration units. Water Research; 37(3):690–8.

Lee, C., Marion, M., Lee, J. 2013. Development and Application of a quantitative PCR assay targeting *Catellicoccus marimammalium* for assessing gull-specific fecal contamination at Lake Erie beaches. *Science of the Total Environment*, 454-455, 1-8.

Lemarchand K, Lebaron P. 2003. Occurrence of *Salmonella* spp. and *Cryptosporidium* spp. in a French coastal watershed: relationship with fecal indicators. FEMS Microbiology Letters; 218(1):203–9.

McGinnis SM, Burch T, Murphy HM. 2022. Assessing the risk of acute gastrointestinal illness (AGI) acquired through recreational exposure to combined sewer overflow-impacted waters in Philadelphia: A quantitative microbial risk assessment. Microbial Risk Analysis; 20:100189.

Medema GJ, Teunis PFM, Havelaar AH, Haas CN. 1996. Assessment of the dose-response relationship of *Campylobacter jejuni*. International Journal of Food Microbiology;30(1–2):101–11.

Messner MJ, Berger P, Nappier SP. 2014. Fractional Poisson-A Simple Dose-Response Model for Human Norovirus. Risk Analysis;34(10):1820–9.

Nasser AM. 2016. Removal of Cryptosporidium by wastewater treatment processes: A review. Journal of Water and Health. IWA Publishing; 14: 1–13.

Padia, R., Karthikeyan, R., Mukhtar, S. and Parker, I. 2012. Occurrence and fate of *E. coli* from various non-point sources in a subtropical watershed. *J. Nat. Environ. Sci*; *3*(1): 9-18.

Parker, I.D., Lopez, R.R., Padia, R., Gallagher, M., Karthikeyan, R., Cathey, J.C., Silvy, N.J. and Davis, D.S. 2013. Role of free-ranging mammals in the deposition of Escherichia coli into a Texas floodplain. *Wildlife Research*; *40*(7): 570-577.

Paziewska A, Bednarska M, Nieweglowski H, Karbowiak G, Bajer A. 2007. Distribution of Cryptosporidium and Giardia spp. in selected species of protected and game mammals from North-Eastern Poland. Annals of Agricultural and Environmental Medicine;14(2).

Rose JB, Gerba CP. 1991. Use of risk assessment for development of microbial standards. In: Water Science and Technology. IWA Publishing; 29–34.

Rose JB, Farrah SR, Harwood VJ, Levine AD, Lukasik J, Menendez P, et al. 2004. Reduction of pathogens, indicator bacteria, and alternative indicators by wastewater treatment and reclamation processes. Water Environment Research Foundation.

SCCWRP. 2022. https://www.sccwrp.org/about/research-areas/microbial-water-quality/microbial-source-tracking/

Schoen ME and Ashbolt NJ. 2010. Assessing Pathogen Risk to Swimmers at Non-Sewage Impacted Recreational beaches. Environmental Science & Technology; 44(7): 2286-2291.

Schoen ME, Ashbolt NJ, Jahne MA, Garland J. 2017. Risk-based enteric pathogen reduction targets for non-potable and direct potable use of roof runoff, stormwater, and greywater. Microbial Risk Analysis; 5:32–43.

Soller, J. A., Schoen, M. E., Bartrand, T., Ravenscroft, J. E., & Ashbolt, N. J. 2010. Estimated human health risks from exposure to recreational waters impacted by human and non-human sources of faecal contamination. *Water research*, *44*(16), 4674-4691.

Soller, J. A., Schoen, M. E., Varghese, A., Ichida, A. M., Boehm, A. B., Eftim, S., ... & Ravenscroft, J. E. 2014. Human health risk implications of multiple sources of faecal indicator bacteria in a recreational waterbody. *Water research*, *66*, 254-264.

Stampi S, Varoli O, Zanetti F, de Luca G. 1993. Arcobacter cryaerophilus and thermophilic campylobacters in a sewage treatment plant in Italy: Two secondary treatments compared. Epidemiology and Infection;110(3):633–9.

Sturdee AP, Bodley-Tickell AT, Archer A, Chalmers RM. 2003. Long-term study of *Cryptosporidium* prevalence on a lowland farm in the United Kingdom. Veterinary parasitology;116(2):97-113.

Teunis PFM, Nagelkerke NJD, Haas CN. 1999. Dose Response Models for Infectious Gastroenteritis. Risk Analysis; 19(6):1251–60.

Teunis PFM, Ogden ID, Strachan NJC. 2008. Hierarchical dose response of *E. coli* O157:H7 from human outbreaks incorporating heterogeneity in exposure. Epidemiology and Infection;136(6):761–70.

Van Abel N, Schoen ME, Kissel JC, Meschke JS. 2017. Comparison of risk predicted by multiple norovirus dose–response models and implications for quantitative microbial risk assessment. Risk Analysis; 37(2):245-64.

Wade SE, Mohammed HO, Schaaf SL. 2000. Prevalence of *Giardia spp*., *Cryptosporidium parvum* and *Cryptosporidium muris* (*C. andersoni*) in 109 dairy herds in five counties of southeastern New York. Veterinary Parasitology;93(1):1-1.

Wesley IV, Wells SJ, Harmon KM, Green A, Schroeder-Tucker L, Glover M, Siddique I. 2000. Fecal shedding of Campylobacter and Arcobacter spp. in dairy cattle. Applied and Environmental Microbiology; 66(5):1994-2000.

Yang J, Schneider OD, Jjemba PK, Lechevallier MW. 2015. Microbial Risk Modeling for Main Breaks. Journal - American Water Works Association; 107(2): E97–108.

Table S 1. QMRA parameters and assumptions.

Parameter	Value	Units	Source
Geometric mean of *E. coli* at monitoring site	11929: 500	cfu/100 mL	Field data
	11933: 352		
	17379: 276		
	17547: 168		
***E. coli* in Human Sewage**	6.7, 8.0[a]	CFU/L	Rose et al., 2004
***E. coli* in Cattle Waste**	2.525, 7.24[a]	CFU/g	Padia et al., 2012; USEPA 2010
***E. coli* in Deer**	4.146, 7.748[a]	CFU/g	Parker et al., 2013
Human (Sewage)			
***E. coli* O157:H7**	-1, 3.3[a]	cfu/L	Garcia-Aljaro et al., 2005
Salmonella	0.5, 5[a]	cfu/L	Koivunen et al., 2003; Lemarchand and Lebaron, 2003
Cryptosporidium	-0.52, 3.7[a]	oocysts/L	Harwood et al., 2005; Crockett, 2007; Yang et al., 2015; Nasser 2016; Schoen et al., 2017
Giardia	0.51, 4.2[a]	cysts/L	Harwood et al., 2005; Kitajima et al., 2014
Campylobacter	2.9, 4.6[a]	MPN/L	Stampi et al., 1993
Norovirus	4.7, 1.5[a]	gc/L	Eftim et al., 2017
Wildlife (Deer)			
Giardia	0.04, 3.07[a]	cysts/g	Heitman et al., 2002; Paziewska et al., 2007
Cryptosporidium	0.26, 2.35[a]	oocysts/g	Garcia-Presedo et al., 2013; Paziewska et al., 2007
Domestic Animals (Cattle)			
***E. coli* O157:H7**	3.08, 1.49[b]	organisms/g	USEPA 2010
Campylobacter	1.8, 4.5[a]	organisms/g	USEPA 2010
Salmonella	2.6, 4.6[a]	organisms/g	USEPA 2010
Cryptosporidium	-0.3, 3.2[a]	organisms/g	USEPA 2010
Giardia	0.2, 3.5[a]	organisms/g	USEPA 2010
Ingestion Volume			
Swimming	1.15, 0.55[c,d]	ml/hour	Boehm et al. (2018); Dufour et al. (2017); McGinnis et al., 2022
Fishing	1.28, 1.72[b,e]	ml/hour	Dorevitch et al. (2011); McGinnis et al., 2022

[a]log10-uniform distribution; [b]lognormal distribution;[c]log10-normal distribution; [d]assumed duration of 30 minutes; [e]assumed duration of 1 hour

Table S 2. Prevalence of infection and infectious potential of pathogens for non-human fecal sources.

Fecal Source	Reference Pathogen	Prevalence of Infection in Source	Infectious Potential (midpoint of range)	Source
Domestic Animals (Cattle)	*E. coli* O157:H7	9.7-28%	83.5[a]	Berry et al., 2007; Soller et al., 2010
	Campylobacter	5-38%	83.5[a]	Hoar et al., 2001; Wesley et al., 2000; Soller et al., 2010
	Salmonella	5-18%	49.5[b]	Hutchinson et al., 2004; Fossler et al., 2005; Soller et al., 2010
	Cryptosporidium	0.6-23%	83.5[a]	Atwill et al., 2006; Sturdee et al., 2003; Soller et al., 2010
	Giardia	0.2-37%	83.5[a]	Fayer et al., 2000; Wade et al., 2000; Soller et al., 2010
Wildlife (Deer)	*Giardia*	0.15-21.2%	16.5[c]	Heitman et al., 2002; Paziewska et al., 2007
	Cryptosporidium	0.15-14.4%	16.5[c]	Garcia-Presedo et al., 2013; Paziewska et al., 2007

[a]infectious potential range for high (67-100%); [b]infectious potential range for medium (33-66%); [c]infectious potential range (0-33%)

Table S 3. Dose response models utilized to estimate the risk of infection and illness.

Pathogen	Probability of Infection	Probability of Illness|Infection	References
Salmonella	$1-(1+dose/2884)^{-0.3126}$	0.17-0.4	Haas et al., 1999; Teunis et al., 1999
Campylobacter	$1-(1+(dose/7.59))^{-0.145}$	0.1-0.6	Medema et al., 1996
E. coli O157:H7	$1-(1+(dose/48.8))^{-0.248}$	0.2-0.6	Teunis et al., 2008
Cryptosporidium	$1-exp(-0.09*dose)$	0.3-0.7	U.S. EPA, 2006
Giardia	$1-exp(-0.01982*dose)$	0.2-0.7	Rose and Gerba, 1991; Eisenberg et al., 1996
Norovirus	$0.72*(1-exp(-dose/1))$	0.3-0.8	Messner et al., 2014; Van Abel et al., 2017