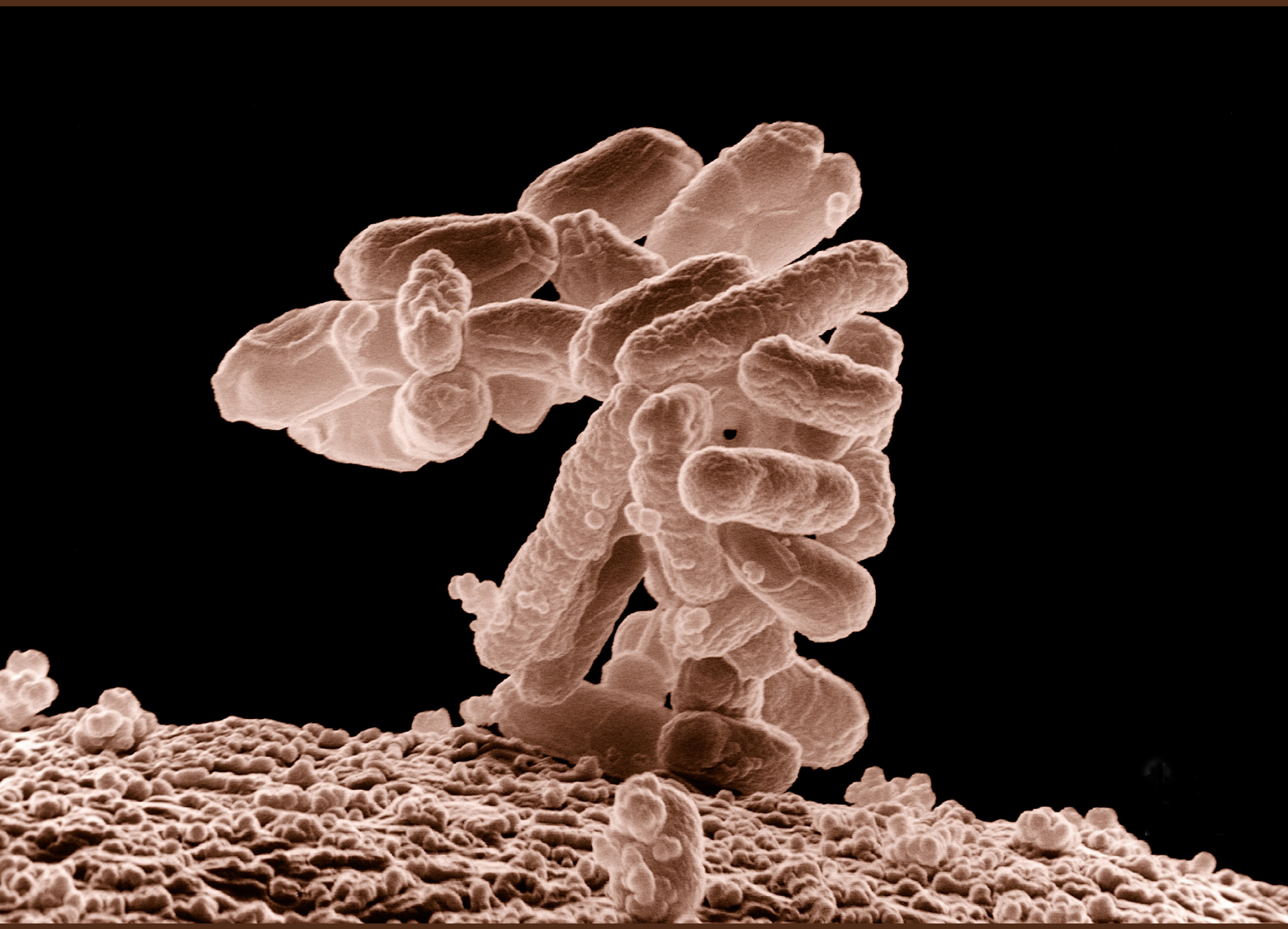
Texas BST Program Refinement, Expansion and Use – FY15

Texas Water Resources Institute TR-496
September 2016



Kevin Wagner, Texas Water Resources Institute

Terry J. Gentry and Maitreyee Mukherjee,Texas A&M University

George D. Di Giovanni, Elizabeth A. Casarez and Joy A. Truesdale,
University of Texas Health Science Center at Houston
School of Public Health, El Paso Campus

Texas Water Resources Institute
make every drop count

Texas' Bacterial Source Tracking Program for Fiscal Year 2015

STATE NONPOINT SOURCE GRANT PROGRAM

TSSWCB PROJECT 15-52

Prepared for:

TEXAS STATE SOIL AND WATER CONSERVATION BOARD

Prepared by:

KEVIN WAGNER

TEXAS WATER RESOURCES INSTITUTE

TERRY J. GENTRY

MAITREYEE MUKHERJEE

TEXAS A&M UNIVERSITY

GEORGE D. DI GIOVANNI

ELIZABETH A. CASAREZ

JOY A. TRUESDALE

UNIVERSITY OF TEXAS HEALTH SCIENCE CENTER AT HOUSTON SCHOOL OF PUBLIC HEALTH, EL PASO CAMPUS

SEPTEMBER 2016

TEXAS WATER RESOURCES INSTITUTE TR-496

Table of Contents

List of Figures

List of Tables

List of Acronyms

ARCC:	Average rate of correct classification
ATCC:	American Type Culture Collection
BMP:	Best management practice
bp:	Base pair
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 PCR
MDS:	Multi-dimensional scaling
MPN:	Most probable number
mTEC:	Membrane Thermotolerant *Escherichia coli*
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
SCSC:	Texas A&M University Soil and Crop Sciences Department
SOP:	Standard operating procedure
TCEQ:	Texas Commission on Environmental Quality
TMDL:	Total maximum daily load
TSSWCB:	Texas State Soil and Water Conservation Board
TWRI:	Texas Water Resources Institute
UPGMA:	Unweighted Pair Group Method with Arithmetic Mean
USDA-ARS:	U.S. Department of Agriculture Agricultural Research Service
UTSPH EP:	University of Texas Health Science Center at Houston School of Public Health, El Paso Campus, Environmental Microbiology Laboratory
WPP:	Watershed protection plan

Executive Summary

In Texas, 255 waterbodies were impaired due to excessive bacteria in 2014. To identify bacterial sources and help address these impairments, Texas established a bacterial source tracking (BST) program circa 2006. To support maintenance, expansion, and use of the Texas BST library and other BST tools, in fiscal year 2015, TWRI, UTSPH EP and AgriLife SCSC collaborated with the TSSWCB to:

(1) Quantify and characterize the possibility of naturalized *E. coli* populations in soil and runoff
(2) Further evaluate and refine the Texas *E. coli* BST library
(3) Continue evaluation and development/refining of source-specific bacterial PCR markers
(4) Support BST efforts in the Arroyo Colorado and other watersheds
(5) Provide outreach regarding BST

Major findings from fiscal year 2015 were as follows:

- Analysis of Riesel soil and water samples and Houston bayou water samples did not support the theory that naturalized *E. coli* populations are significant contributors.
- There was no significant difference between the Quanti-Tray/2000 and EPA Method 1603 modified mTEC medium enumeration methods.
- The current Texas *E. coli* BST Library contains 1,765 isolates from 1,554 known source fecal samples obtained from nearly 4,000 individual known source fecal samples from 18 watersheds. Analysis of the library revealed that:
    - Best matches for many isolates come from their own self-validated local watershed library.
    - 318 isolates (75 human, 39 domestic animals, and 204 wildlife) exhibit extraordinary source specificity and geographical and temporal stability.
    - 65 isolates appear to be cosmopolitan isolates and 40 appear to be transient isolates.
- The Mo Bio kit was found to be more efficient in DNA extraction than the Qiagen kit-based methods currently used for detection of *Bacteroidales* markers from surface water samples.
- Comparison of DNA sequences of PCR products amplified using current *Bacteroidales* BST assays revealed potential targets for the differentiation of deer, hog, and goat *Bacteroidales* PCR amplicons. This will assist in future development of PCR primers for these sources.
- Use of BST in the Arroyo Colorado watershed revealed that wildlife (both non-avian and avian) were the leading contributor of *E. coli* in the Arroyo Colorado while 9% of isolates were human and 13% were domestic animals.
- Outreach included highlighting the BST Program in:
    - The Fall 2015 txH2O reaching over 4,000 subscribers
    - Three Facebook posts
    - Nine tweets
    - Five conferences and two meetings with TSSWCB, EPA, and TCEQ
    - The BST website resulting in 353 visits

Introduction

According to the 2014 *Texas Integrated Report* (303(d) List), 245 streams and rivers, 8 oyster waters, and 2 beaches are impaired due to excessive levels of bacteria. Identifying and assessing sources of these bacteria is critical to target best management practices, develop bacterial total maximum daily loads (TMDLs) or watershed protection plans (WPPs), and assess risk from contact recreation.

Bacterial source tracking (BST) is a valuable tool that can identify, and also rule-out, significant sources of *E. coli* pollution in a watershed. 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 the fecal contamination can be identified. Numerous BST methods are available which use DNA fingerprints and bacterial markers to identify fecal pollution sources. Based on a multi-year study initiated in 2002, Texas selected the two-method approach using ERIC-PCR and RiboPrinting (ERIC-RP), as this approach was found to be the most accurate and cost-effective. Because it provides a direct link with water quality standards, *E. coli* is used as the target bacterium.

For more than a decade, the Texas BST Program has successfully identified sources of *E. coli* in dozens of watersheds across Texas. 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, and (10) Arroyo Colorado. 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 is dynamic, with new isolates being added with each successive BST project. To support maintenance, expansion and use of the library and other BST tools, TWRI, UTSPH EP and AgriLife SCSC collaborated to:

(1) Quantify and characterize the possibility of naturalized *E. coli* populations in soil and runoff
(2) Further evaluate and refine the Texas *E. coli* BST library through data exploration and analysis of presumptive naturalized, cosmopolitan, and transient *E. coli* isolates
(3) Continue work to evaluate and further develop/refine source-specific bacterial PCR markers
(4) Support BST efforts in high priority watersheds
(5) Provide outreach regarding BST

Evaluation of Naturalized *E. coli* Populations

The TSSWCB and TCEQ Bacteria Total Maximum Daily Load Task Force identified several BST- related research areas, including the investigation of environmentally adapted or "naturalized" *E. coli* (Jones, Wagner et al. 2007). A growing number of studies demonstrate the potential for *E. coli* to become naturalized in soils, sediments and water (Byappanahalli, Whitman et al. 2006, Ishii, Ksoll et al. 2006, Walk, Alm et al. 2007, Brennan, Abram et al. 2010, Byappanahalli, Roll et al. 2012, Perchec-Merien and Lewis 2013, Tymensen, Pyrdok et al. 2015), becoming a normal part of environmental microbial communities rather than transient members from fresh fecal deposition. If naturalized soil/sediment associated *E. coli* populations are high at a site, they could potentially contribute to water quality impairments. Studies are needed to isolate presumptive naturalized *E. coli* from selected sites and characterize them via ERIC-RP for comparison to the Texas *E. coli* BST Library to assess the possibility of differentiating "naturalized" *E. coli* populations from those contributed from fresh fecal deposition.

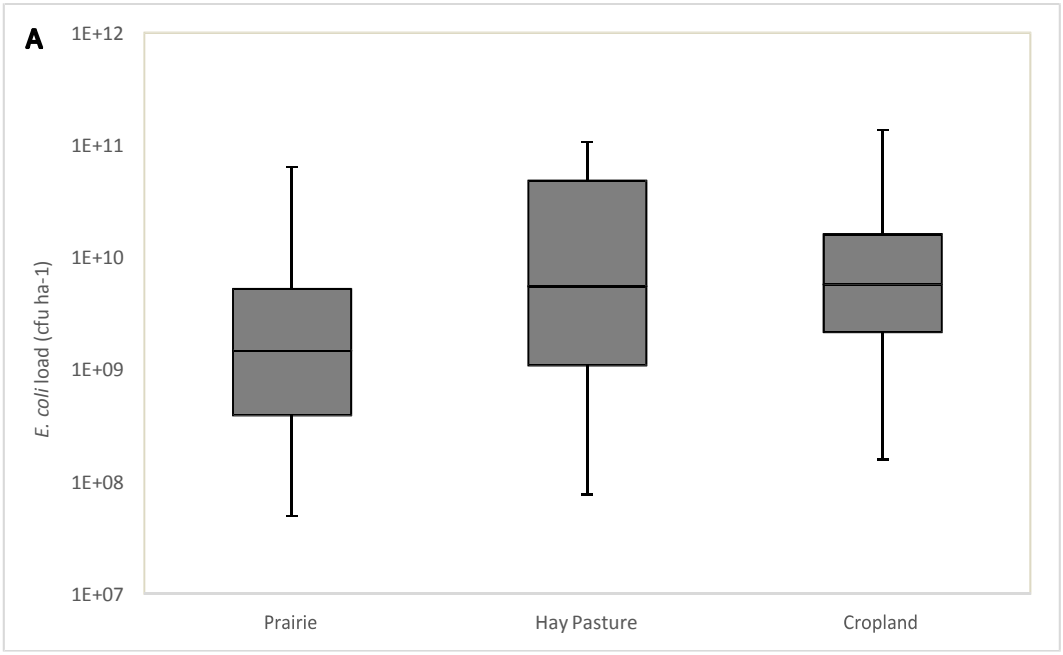
Evaluation of naturalized *E. coli* populations in soil and runoff

To quantify and characterize the possibility of naturalized *E. coli* populations in soil and runoff at the USDA-ARS Grassland Research Center in Riesel, four small exclosures (built from plastic barrels, or similar) were installed in each of 3 designated catchments (un-grazed rangeland, cropland, managed hay pasture). The open end of each exclosure was buried in the soil to exclude inputs of *E. coli* from animals or water. One month after installation, individual soil samples were collected and composited from inside each exclosure. Multiple soil samples were collected in 2014-2015; however, few countable populations of *E. coli* were found in soil samples from the watersheds.

Thus, data and *E. coli* isolates collected from the Riesel watersheds as part of TSSWCB Project #13-56 *Bacteria Growth, Persistence, and Source Assessment in Rural Texas Landscapes and Streams* were used to determine the possibility of naturalized *E. coli* at these sites. *E. coli* loads for each rainfall event were calculated based on *E. coli* concentrations in the runoff water and the runoff volume. Additionally, *E. coli* concentrations in soil were calculated on a per hectare basis for comparison to runoff loads. The ERIC-RP fingerprints generated for BST in TSSWCB Project #13-56 were compared, using a similarity matrix based on the Pearson correlation with UPGMA (Unweighted Pair Group Method with Arithmetic Mean) clustering method, to generate a multi-dimensional scaling (MDS) analysis comparison between the soil and water isolates. In addition, antibiotic resistance was assessed for each of the archived *E. coli* isolates using the Kirby-Bauer disk diffusion method for the following antibiotics: tetracycline (TE-30/ 30 µg), ampicillin (AM-10/ 10 µg), cephalothin (CF-30/ 30 µg), imipenem (IPM-10/ 10 µg), gentamicin (GM-120/ 120 µg), sulfamethoxazole (SXT/ 23.75 µg), and ciprofloxacin (CIP-5/ 5 µg). *Pseudomonas aeruginosa* (ATCC 27853), *Staphylococcus aureus* (ATCC 25923), and *E. coli* (ATCC 25922) were used as controls. The zones of inhibition were then measured using an automated imaging system and the software ImageJ following standard manufacturer parameters for zone diameter. Based on the inhibition zone diameter, the isolates were grouped as resistant or susceptible. The Chi square test for independence was applied to test for significant differences within the dataset. Tested relationships were considered to be significant at $p < 0.05$, or when the Chi square sum was greater than 3.84 at a degree of freedom = 1.

E. coli loads in runoff water and soil

Per rainfall event, median values for *E. coli* loads in runoff were 1.81 x 10^9, 5.5 x 10^9, and 5.71 x 10^9 cfu ha^{-1} for SW12 (Prairie), SW17 (Hay Pasture), and Y6 (Cropland), respectively (Figure 1A). In contrast, median values for soil *E. coli* were only 6.78 x 10^4, 2.26 x 10^5, and 2.41 x 10^5 cfu ha^{-1} for SW12, SW17, and Y6, respectively (Figure 1B), accounting for only 0.004% of the total *E. coli* loading. The large difference observed in the total number of *E. coli* in the runoff versus the top 5 cm of soil suggests that soil-based *E. coli* is not likely a major contributor to *E. coli* loads in runoff from these watersheds.



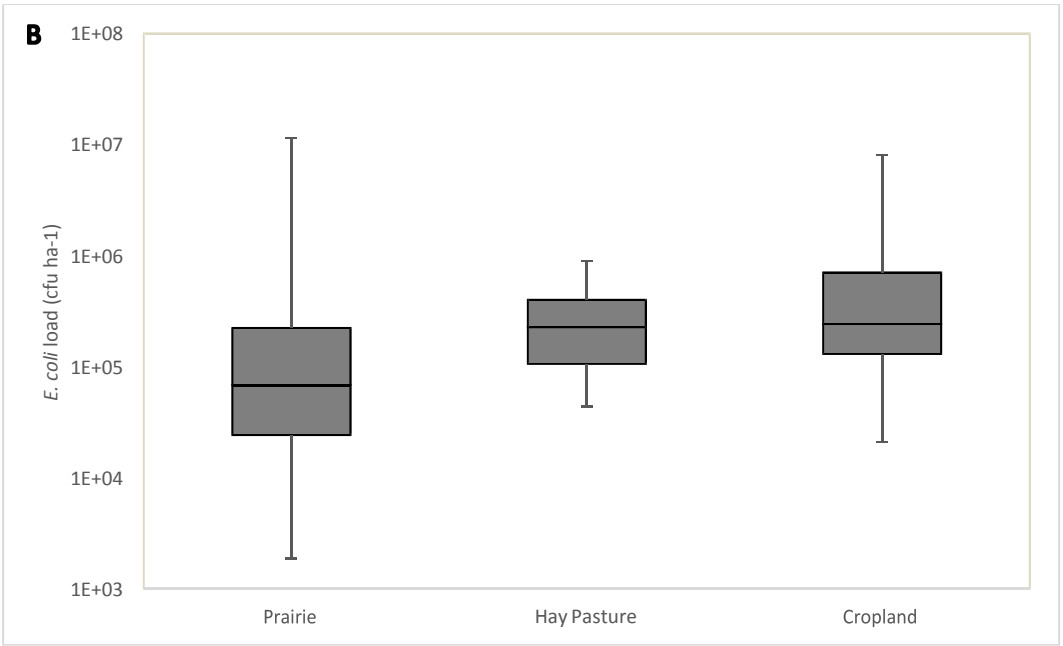


Figure 1. *E. coli* loads in runoff (A) and soil (B) for SW12 (prairie), SW17 (hay pasture), and Y6 (cropland) watersheds. The line within each box indicates the median value.

Genotypes of *E. coli* isolated from runoff water and soil

Multidimensional scaling analysis of the *E. coli* ERIC-RP fingerprints indicated that the soil isolates generally grouped within the runoff isolates (Figure 2). There was some separation between soil and water isolates for SW17, but overall the soil *E. coli* appeared to be similar genotypes to the *E. coli* found in the runoff samples. This would suggest the possibility that soil *E. coli* could be a source of the *E. coli* detected in the runoff, if it were not for the much lower population levels observed in soil. Since several *E. coli* genotypes observed in the runoff were not seen in the soil, this provides additional evidence that most of the *E. coli* within the water samples likely originated from sources other than soil, such as fecal depositions on standing vegetation and on detritus at the soil surface.

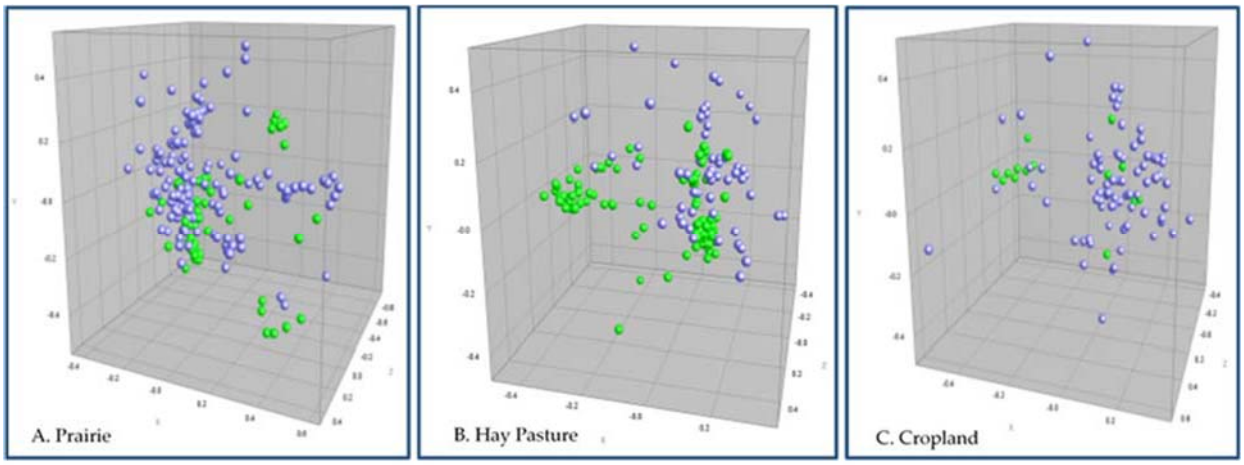


Figure 2. Multidimensional scaling analysis of *E. coli* genotypes isolated from soil (green) and runoff water (purple) from Riesel watersheds SW12 (A), SW17 (B), and Y6 (C).

Antibiotic resistance profiles of *E. coli* isolated from runoff water and soil

Overall, the frequency of antibiotic resistance in the runoff water *E. coli* isolates was slightly higher than that observed for the soil isolates – 95% of the water isolates were resistant to one or more antibiotics compared to 87% of the soil isolates (Figure 3). However, this was largely driven by resistance to a single antibiotic – cephalothin. In total, 93%, 76% and 93% of water isolates and 78%, 98%, and 87% of soil isolates from SW12, SW17, and Y6, respectively, were resistant to cephalothin. There was more variability in the level of resistance to the other antibiotics. For example, a substantial portion of the water isolates (28%, 16% and 22% in SW12, SW17, and Y6, respectively) were resistant to tetracycline; whereas, no resistance to tetracycline was observed for the soil isolates. For the SW12 watershed, levels of tetracycline, cephalothin, and ampicillin resistance were significantly higher in the water isolates than the soil isolates. For the SW17 and Y6, tetracycline resistance was significantly higher for the water isolates than the soil isolates. Overall, none of the water isolates were resistant to gentamycin, ciprofloxacin or imipenem. None of the soil isolates were resistant to tetracycline, cefoperazone, gentamycin, ciprofloxacin, or imipenem. Differences in antibiotic resistance profiles between the water and soil isolates were even more striking when evaluated as the percentage of isolates demonstrating resistance to two or more antibiotics with the water isolates being approximately 10 times more likely to be resistant to multiple antibiotics than the soil isolates were (Figure 4).

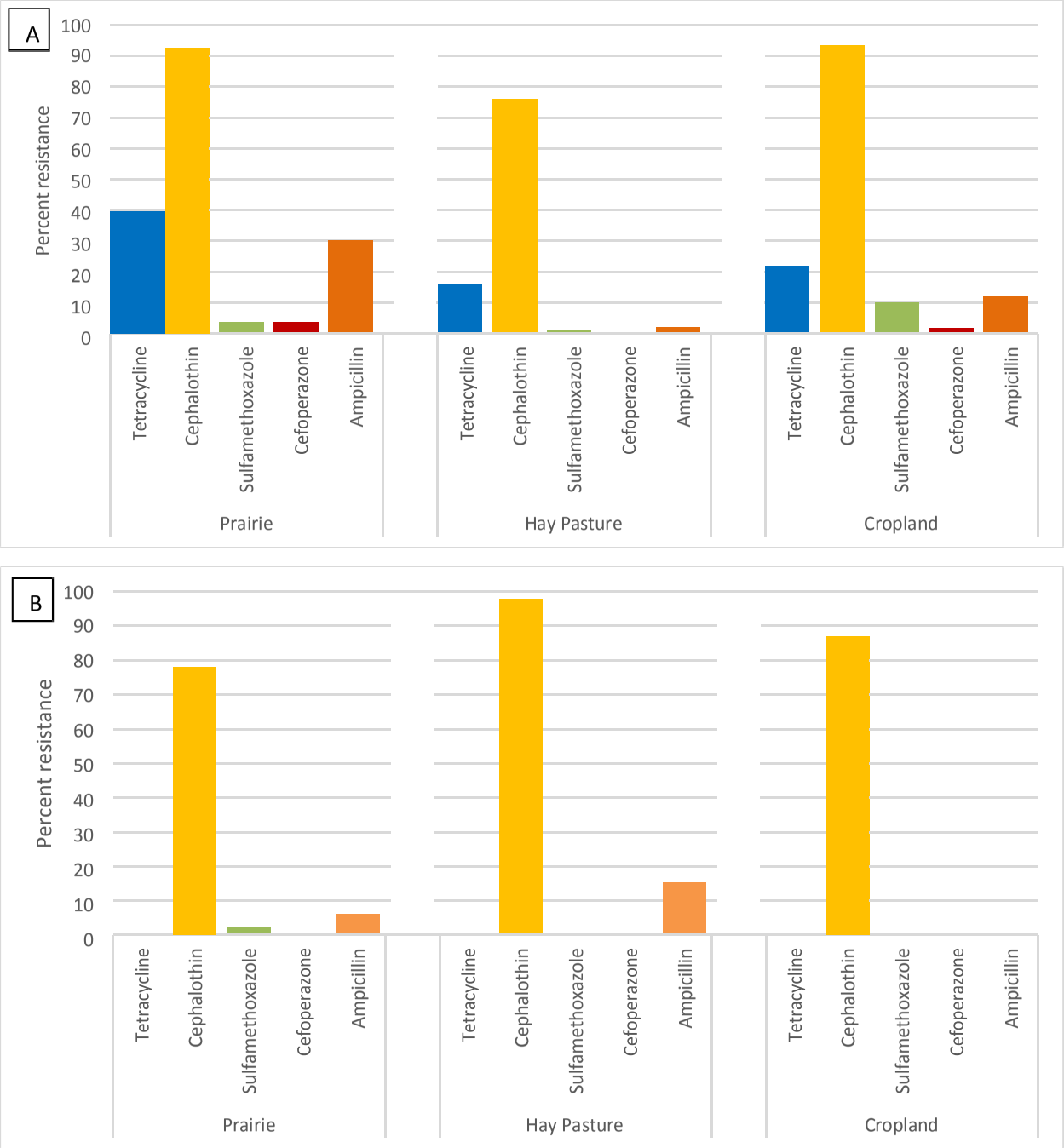


Figure 3. Percentage of *E. coli* isolated from runoff water (A) and soil (B) at Riesel watersheds SW12 (Prairie), SW17 (Hay Pasture) and Y6 (Cropland) resistant to various antibiotics.

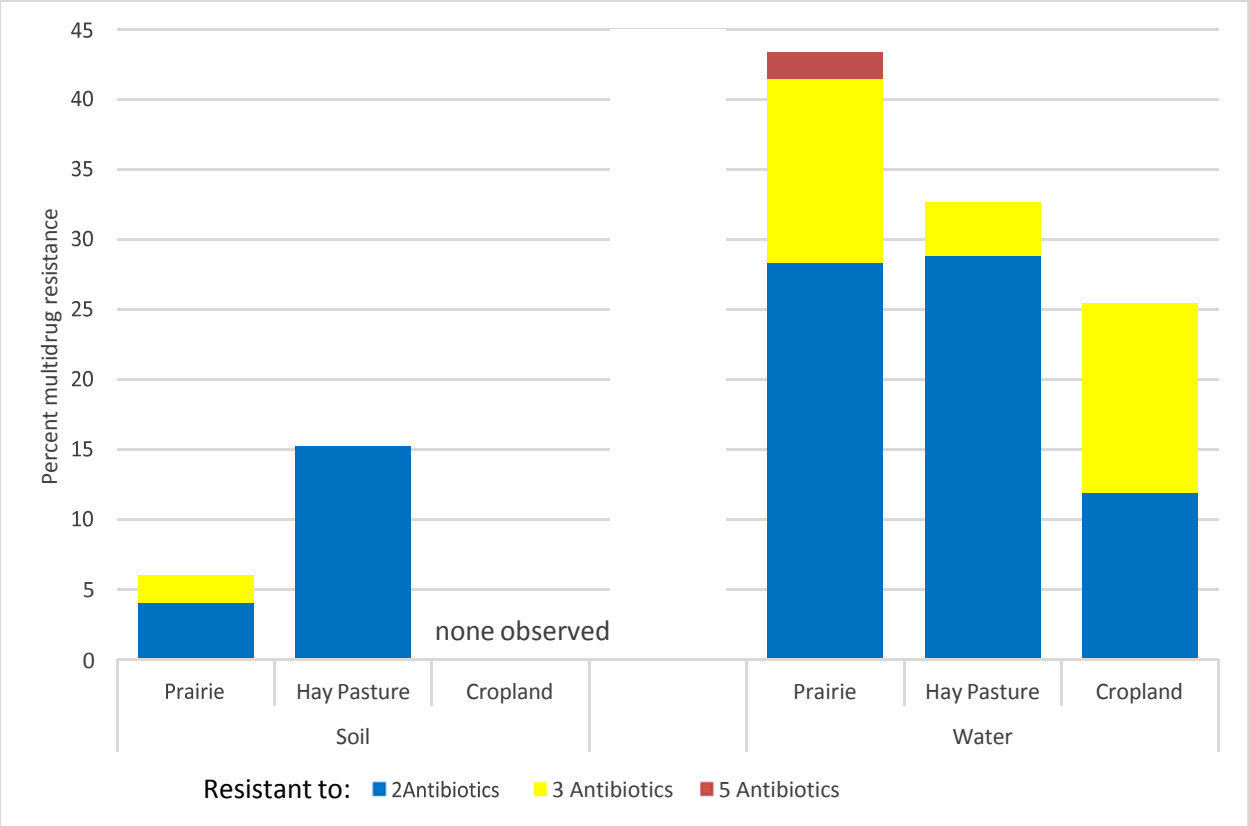


Figure 4. Percentage of *E. coli* isolated from runoff water and soil at Riesel watersheds SW12 (Prairie), SW17 (Hay Pasture) and Y6 (Cropland) resistant to two or more antibiotics.

Conclusions

E. coli loads in soil and runoff water at the tested Riesel watersheds indicate that there are not large populations of soil *E. coli* in the watersheds and that these levels would at most represent a small fraction of *E. coli* in runoff, even if all of the soil *E. coli* were mobilized during a runoff event. The inability to detect *E. coli* during multiple soil sampling events further supports the conclusion that soil *E. coli* are not major contributors to the *E. coli* levels observed in runoff from these watersheds. However, the ERIC-RP fingerprints indicated that the soil *E. coli* were genotypically similar to many of the *E. coli* observed in runoff suggesting that they originated from similar sources. It seems likely that the soil *E. coli* originated from fecal materials deposited on vegetation at the site and/or the surface of the soil. The lower levels of antibiotic resistance seen in the soil *E. coli* suggest that these isolates are either naturalized or possibly becoming naturalized and losing some of their antibiotic resistance in the process. Nevertheless, the low populations of these soil *E. coli*, relative to loads in the runoff, indicate that they are not likely major contributors to the *E. coli* observed in runoff from these watersheds.

Investigating the presence of naturalized *E. coli* in Houston bayous

Levels of *E. coli* exceeding water quality standards are frequently reported for Houston area bayous. Despite various control measures, including the identification of illegal sewage discharges, the City of Houston has continued to be plagued by water quality violations. One research team conducted investigations of *E. coli* and enterococci populations present in bayou sediments and waters and concluded that significant naturalized populations of these bacteria may be present, potentially contributing as much as 90% of the total daily load (Brinkmeyer, Amon et al. 2015). Another Houston area study observed significant diurnal changes in the levels of *E. coli* in bayous (Desai and Rifai 2013). Many of the Houston area bayous receive significant amounts of wastewater treatment plant effluent, and therefore questions still remain with regards to water quality impacts from these effluents, wildlife sources and sediment-derived/naturalized *E. coli*. Given these concerns, the City of Houston partially funded a preliminary evaluation of bayou water sample *E. coli* enumeration methods, *E. coli* strain comparisons, and occurrence of *Bacteroidales* PCR markers.

Methods

Houston bayou water samples were collected and shipped to the UTSPH EP Environmental Microbiology Laboratory by City of Houston and Terracon Consultants, Inc. (Houston, TX) personnel. Samples were shipped directly to the UTSPH EP since *E. coli* enumeration data would only be used for comparison of enumeration methods and not compliance monitoring. *E. coli* was enumerated using EPA Method 1603 modified mTEC medium per SOP #TXBST-02 (Di Giovanni, Casarez et al. 2015) and the Quanti-Tray/2000 system (IDEXX, Westbrook, ME) per manufacturer's instructions. *E. coli* isolates were picked from modified mTEC plates and also recovered from Quanti-Tray/2000 wells which tested positive for *E. coli*. The foil side of trays was wiped with 70% isopropyl alcohol, medium removed from wells using sterile syringes, and samples transferred to sterile microcentrifuge tubes. Samples were streaked onto modified mTEC medium for isolation of *E. coli*. Isolates were fingerprinted by ERIC-PCR and RiboPrinting per SOP #TXBST-04 and #TXBST-05, respectively (Di Giovanni, Casarez et al. 2015). Water sample DNA was extracted and *Bacteroidales* PCR performed per SOP #TXBST-06 and #TXBST-07 (Di Giovanni, Casarez et al. 2015).

Results and discussion

Various methods are approved by EPA and TCEQ for the enumeration of *E. coli* in water. However, there are challenges associated with each of these methods. For example, consortia of non-*E. coli* bacteria can break down substrates in the Quanti-Tray, leading to false positives and elevated counts (Pisciotta, Rath et al. 2002). The City of Houston currently uses the Quanti-Tray/2000 method for enumeration of *E. coli*. Due to the unique aquatic environment in the Houston-Galveston area, enumeration assays were compared to determine the most appropriate for regulatory monitoring. Enumeration of *E. coli* from bayou water samples using the Quanti-Tray/2000 and EPA Method 1603 modified mTEC medium was performed (Table 1). Data analysis using Student's paired t-test with two-tailed distribution indicated there was no significant difference between enumeration methods ($p = 0.36$).

Table 1. Enumeration of *E. coli* from Houston bayou water samples using Quanti-Tray/2000 and EPA Method 1603.

Station ID	Location	*E. coli* Quanti-Tray/2000 MPN/100 mL	*E. coli* modified mTEC cfu/100 mL
15869	Hunting Bayou at Cavalcade St.	6830	4200
11129	Hunting Bayou at Loop 610	410	240
16589	Garners Bayou at Old Humble Rd.	52	20
15862	Halls Bayou at Homestead Rd.	345	580
11155	Vogel Creek at West Little York	96	30
17493	South Mayde Creek at Addicks Reservoir	109	140
11148	Little White Oak Bayou at Trimble St.	1350	440
11148-dup	Little White Oak Bayou at Trimble St.	2130	1670
17976	Sims Bayou at S. Post Oak Rd.	420	500
16661	Berry Bayou at S. Richey St.	1246	1190
18691	Mimosa Ditch Tributary of Brays Bayou	310	1290
11163	South Mayde Creek at Memorial Dr.	74	190

Since there was concern over the potential for naturalized *E. coli*, *Bacteroidales* PCR was also performed on the water samples. *Bacteroidales* bacteria are anaerobic and thus less likely to replicate in the environment. In particular, there was interest in analyzing the samples for the *Bacteroidales* HF183 human marker. If high counts of *E. coli* come from human sources rather than naturalized populations, observation of a corresponding presence of human *Bacteroidales* would be expected. Several bayou water samples were selected for this analysis, representing sites with both high and low *E. coli* counts. Equivalent amounts of water (100 ml) were filtered for each sample and DNA extracted. Undiluted and tenfold serial dilutions of each DNA extract were analyzed by PCR. All samples had at least one of two PCR replicates test positive for the HF183 human marker and the two samples with the highest *E. coli* counts also had the highest levels of HF183 marker (Table 2). These results suggest that there may be domestic sewage sources impacting the bayous. An alternative explanation is that *E. coli* is originating from non-human sources and *Bacteroidales* HF183 human marker abundance is due to treated wastewater treatment plant effluent being discharged into the bayous. Regardless, either scenario warrants further investigation.

Table 2. Analysis of Houston bayou water samples for the *Bacteroidales* HF183 human marker.

Station ID	Location	*E. coli* modified mTEC cfu/100 mL	*Bacteroidales* HF183 human marker			
			undiluted (pos/reps)	10^{-1} (pos/reps)	10^{-2} (pos/reps)	10^{-3} (pos/reps)
15869	Hunting Bayou at Cavalcade St.	4200	2/2	2/2	1/2	0/2
16589	Garners Bayou at Old Humble Rd.	20	1/2	0/2	0/2	0/2
16661	Berry Bayou at S. Richey St.	1190	2/2	0/2	0/2	0/2
11163	South Mayde Creek at Memorial Dr.	190	1/2	0/2	0/2	0/2

E. coli isolates from modified mTEC plates and recovered from Quanti-Tray/2000 wells were fingerprinted using ERIC-RP and compared to the Texas *E. coli* BST Library. A total of 77 *E. coli* isolates were analyzed, 45 from modified mTEC plates and 32 from Quanti-Tray/2000 wells. When compared to known source fecal isolates in the Texas *E. coli* BST Library ver. 5-15 only eight (10%) of the water isolates were unidentified. Sources of the remaining 69 water isolates were identified as 13% human, 22% domestic animals (including pets) and 55% wildlife. Eight of ten isolates identified as human were obtained from water samples having ≥1190 *E. coli* 100 ml^{-1}.

Collectively, results of these investigations do not support the theory that naturalized *E. coli* populations are significant contributors to elevated *E. coli* levels observed in Houston bayous. Rather, it appears that the bayous may be impacted from animal and human fecal pollution sources. Clearly additional investigation is needed to shed light on these issues.

Evaluation of the Texas *E. coli* BST Library

The Texas *E. coli* BST Library has been 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. Currently, the Texas *E. coli* BST Library ver. 5-15 contains 1,765 isolates from 1,554 known source fecal samples. This is the result of screening 6,768 *E. coli* isolates obtained from nearly 4,000 individual known source fecal samples. For detailed isolate information and selection for the library see the TSSWCB Project 13-50 report (Di Giovanni, Casarez et al. 2015).

As described in previous reports, three steps are used to refine the Texas *E. coli* BST Library: de-cloning, self-validation, and cross-validation of isolates. De-cloning compares the ERIC–PCR patterns from up to three isolates per individual known source fecal sample. Isolates which were greater than 80% similar were considered clones (identical strains) and subsequently only one isolate selected for further consideration. All de-cloned isolates from individual source samples were included in their respective local watershed library, independent of their similarity to other library isolates. Self-validation of the local watershed library composite ERIC–RP fingerprints was performed using Jackknife analysis to identify isolates that were correctly classified using a seven-way split of source classes (i.e., human, pets, cattle, other non-avian livestock, avian livestock, avian wildlife, and non-avian wildlife (including feral hogs)). Singleton isolates were defined as those having ERIC–RP fingerprints less than 80% similar to another isolate. In addition to self-validated isolates, singletons were retained as members of their self-validated local libraries. Cross-validation entails a series of watershed/project-inclusive Jackknife analyses on the combined self-validated local libraries to remove all isolates that cross-identified between human, domestic animals, and wildlife source classes with a goal of 100% ARCC using a three-way split. The current and several previous versions of the Texas *E. coli* BST Library were developed using this approach. The current Texas *E. coli* BST Library ver. 5-15 includes known source fecal isolates from 18 different local watershed/project libraries (Table 3).

Developing a statewide BST library using *E. coli* isolates from local watershed libraries allows for time and cost savings. However, this also raises questions regarding the geographic and temporal stability of the library. Several different aspects of the library's geographic and temporal stability were discussed in the TSSWCB Project 13-50 report (Di Giovanni, Casarez et al. 2015). In the current project a closer review of Jackknife results for the Texas *E. coli* BST Library ver. 5-15 revealed that the best match for many isolates comes from their own self-validated local watershed library. Of the 1,765 isolates in the Texas *E. coli* BST Library ver. 5-15, a total of 1,040 isolates found a best match with another isolate from their local watershed cohort (but from a different source sample), while the best match for 400 isolates were from a different watershed study or project. The remaining 325 library isolates were singletons and did not have a match at 80% similarity or greater with another isolate in the library.

Table 3. The number of *E. coli* isolates in the Texas *E. coli* BST Library ver. 5-15 by three-way source class and watershed/project.

Watershed/project	Human Isolates (samples)	Domestic Animal Isolates (samples)	Wildlife Isolates (samples)	Total Isolates (samples)
Arroyo Colorado	23 (16)	4 (4)	32 (28)	59 (48)
Attoyac	0 (0)	34 (34)	23 (23)	57 (57)
Big Cypress	6 (2)	5 (4)	13 (9)	24 (15)
Buck Creek	0 (0)	4 (4)	9 (9)	13 (13)
Lake Granbury	11 (10)	5 (5)	27 (24)	43 (39)
Lampasas	13 (11)	10 (8)	55 (40)	78 (59)
Lampasas—TSSWCB Project 11-51	0 (0)	0 (0)	8 (4)	8 (4)
Little Brazos River	3 (3)	15 (15)	33 (33)	51 (51)
Leon—UTH	9 (8)	24 (18)	43 (30)	76 (56)
Leon—CS	0 (0)	23 (23)	17 (17)	40 (40)
Leon—Infra 2013	0 (0)	12 (11)	12 (8)	24 (19)
Leona	0 (0)	33 (33)	43 (43)	76 (76)
Oyster Creek	0 (0)	69 (69)	61 (61)	130 (130)
Riesel	0 (0)	0 (0)	53 (44)	53 (44)
San Antonio—TCEQ	148 (134)	161 (139)	79 (74)	388 (347)
San Antonio—Infra 2013	0 (0)	0 (0)	109 (67)	109 (67)
Trinity River	9 (9)	8 (8)	30 (30)	47 (47)
Waco-Belton—TSSWCB	162 (137)	125 (120)	202 (185)	489 (442)
Totals	384 (330)	532 (495)	849 (729)	1765 (1554)

To better test the geographic stability of the library, watershed/project-exclusive Jackknife analyses of the Texas *E. coli* BST Library (ver. 5-15) were performed. Isolates were not allowed to match other isolates from their own watershed study or project. Eighteen separate Jackknives were calculated, one for each self-validated local watershed/project library, and the identification results were combined. The average rate of correct classification (ARCC) for the watershed/project-exclusive Jackknife analyses using a three-way split of source classes was 66% with 29% of the isolates left unidentified (Table 4). This approach is a first step in identifying library isolates that appear to be the most source specific and geographically and temporally stable. The analyses identified 828 isolates (131 human, 180 domestic animals, and 517 wildlife) with a correct best match from another watershed or project (which differed in time from its own). Isolates that were left unidentified increased from 325 to 505 when they were not allowed to match a local watershed/project cohort. Although less frequently encountered, these isolates add to the diversity of strains in the library and may still hold value in identifying water isolates. The 432 isolates whose watershed/project-exclusive best match was incorrect may also still be valuable. Most of those isolates are similar to more than one ERIC-RP composite fingerprint above the 80% similarity cutoff.

Although the best match for those isolates was incorrect, many have correct source class matches for their second best match (or lower) down to the 80% similarity cutoff. While it would be a challenging task, further evaluation might allow the identification of library isolates which are "preferentially associated" with specific source classes. For example, those associated only with wildlife and domestic animals but not humans, or those found almost exclusively in wildlife but infrequently in domestic animals.

Table 4. Texas *E. coli* BST Library (ver. 5-15) watershed/project-exclusive rates of correct classification (RCCs) using a three-way split of source classes.

Source Class	Library Composition and Expected Random Rate of Correct Classification*	Calculated Rate of Correct Classification (RCC)	RCC to Random Ratio***	Left Unidentified (unique patterns)
HUMAN	**22%**	**56**	**2.5**	**39**
DOMESTIC ANIMALS	**30%**	**48**	**1.6**	**30**
WILDLIFE	**48%**	**79**	**1.6**	**23**
Overall		**ARCC** ** **= 66**		**29%**

*RARCC, expected random average rate of correct classification. Note different library compositions since watershed studies/projects were excluded one at a time (see Table 3)

**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 2.5-fold greater than random chance.

A second watershed/project-exclusive approach was used to better understand the occurrence of *E. coli* strains in the different watersheds (i.e., geographic distribution) or in some cases the same watershed but another study performed at a different time (i.e., temporal distribution). All library isolates were run against each self-validated local watershed/project library to see if a) a similar strain was found at greater than 80% similarity, and b) if it was correct using a three-way match. This entailed 17 watershed/project-exclusive challenges using each self-validated local watershed/project library and resulted in a tally of the correct, incorrect, and unidentified matches for each of the 1,765 library isolates. The challenges ranged from 1,276 isolates versus a local library of 489 source isolates (the Waco-Belton local library) to 1,757 isolates challenged against a local library of only 8 wildlife avian isolates (TSSWCB Project 11-51, Instream Bacteria Influences from Bird and Bat Habitation of Bridges). It is important to note that not all watershed studies or projects contained isolates from all three source classes. Therefore for a few of the challenges it was not possible for some of the library isolates to have a correct match.

As described earlier for the watershed/project-exclusive Jackknife analysis of the ver. 5-15 library, 505 of 1765 isolates did not have a match greater than 80% similarity to another isolate in any of the other local libraries and were left unidentified. A total of 185 isolates had a match in only one other local library, with 133 (72%) of those matches correct using a three-way split of source classes. At the other extreme, 52 isolates had matches in each of the 17 other local libraries, with 402 of 884 (45%) of all matches correct at a three-way split of source classes (see Figure 5).

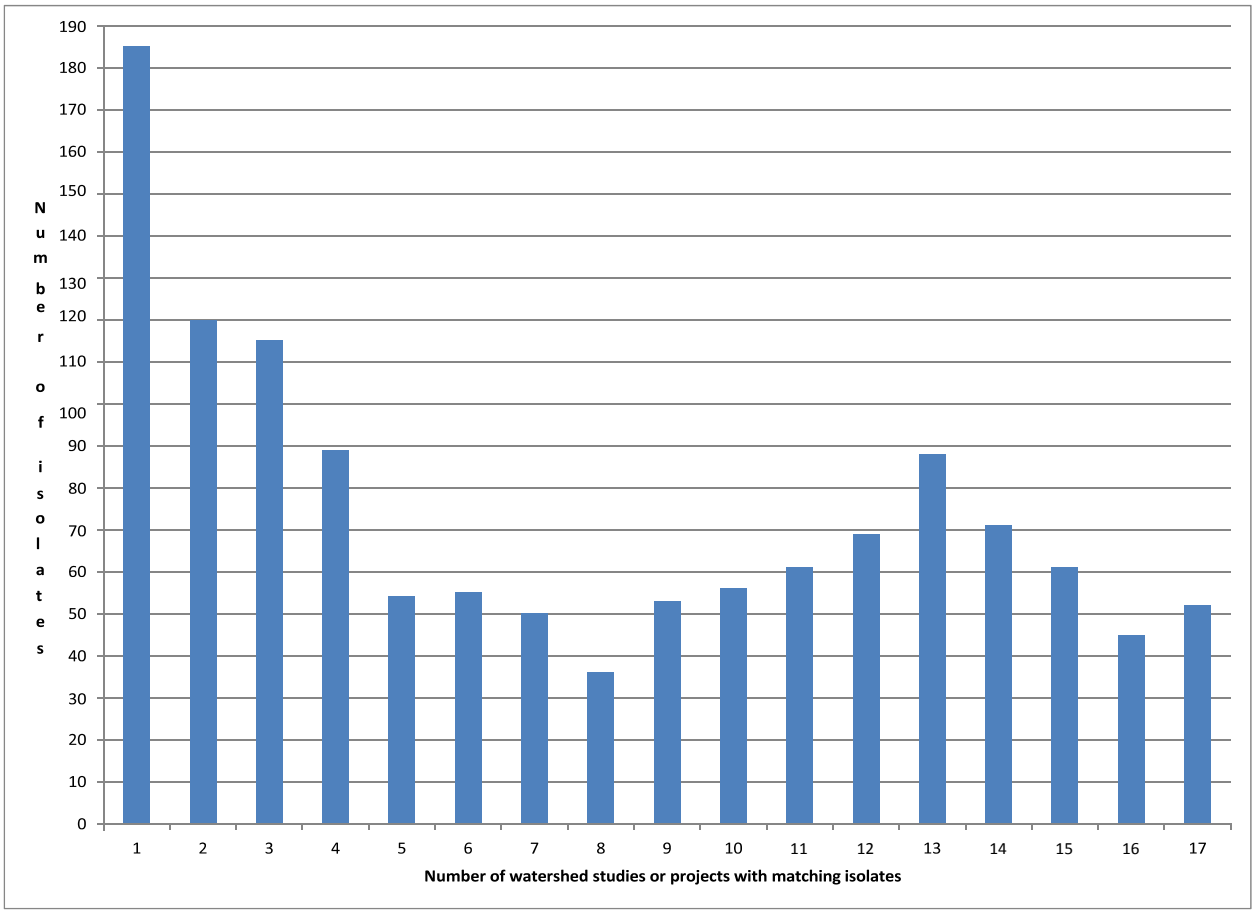


Figure 5. Watershed/project-exclusive occurrence of Texas *E. coli* BST Library (ver. 5-15) isolates based on composite ERIC-RP fingerprints and matches ≥80% similarity to an isolate in one or more of 17 self-validated local watershed/project libraries.

For the development and refinement of the Texas *E. coli* BST Library, isolates demonstrating the highest source specificity and are widely occurring are of most interest. Many of these desirable isolates were identified using the second watershed/project-exclusive analysis approach described. In total, 271 isolates (75 human, 38 domestic animals, and 158 wildlife) had 100% correct best matches to isolates in 1 to 11 other local watershed/project libraries (Figure 6). These isolates represent some of the most geographically and temporally stable members of the state library. A total of 133 isolates had correct best matches in one other local watershed/project library. At the other extreme, 3 wildlife isolates had correct best matches to isolates in 11 of 17 other local watershed/project libraries.

Additional isolates were determined to have varying degrees of geographical occurrence and preferential association with a source-class. For example, 15 isolates had 9 to 13 correct best matches and only 1 incorrect best match to isolates in other local libraries (>90% correct matches). Similarly, 21 other isolates also had only one incorrect best match, but had slightly lower geographic occurrence with best matches to isolates in fewer (7 to 9) other local libraries (>85% correct matches). There were an additional 11 isolates that had between 12 and 15 correct best matches but two incorrect best matches from other local libraries (>85% correct matches). Overall, this detailed analysis allowed the identification of 318 isolates (75 human, 39 domestic animals, and 204 wildlife) with extraordinary source specificity and geographical and temporal stability.

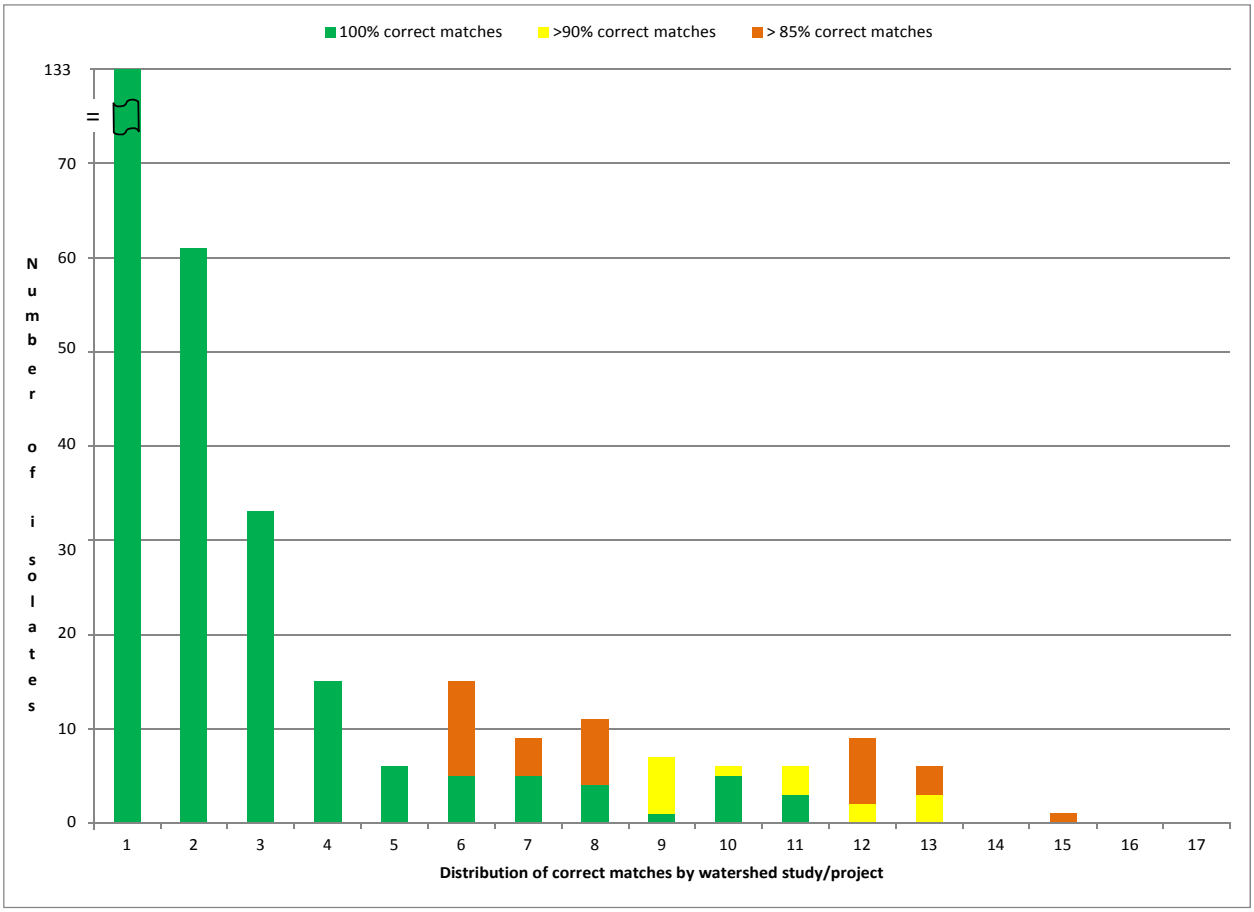


Figure 6. Isolates with greater than 85%, 90%, or 100% correct matches in 1 to 17 other self-validated local watershed/project libraries.

Obviously not all *E. coli* are source specific and those strains found in the feces of many different animals and humans are referred to as "cosmopolitan." In contrast to the library challenges and evaluations performed above to identify the most specific and geographically stable isolates, some cosmopolitan isolates were also identified. It is important to note that these isolates were source specific within the context of their local library of known source fecal samples collected from the same watershed/project. That is, they passed self-validation testing within their local watershed/project library.

These isolates also passed watershed/project inclusive cross-validation for incorporation into the state library. This is because they had a correct best match with another isolate obtained from a different fecal sample from their own local library, despite having the opportunity to have a best match with an isolate from another watershed/project. Recent library evaluations have allowed identification of 65 isolates that appear to be cosmopolitan since they match to more isolates belonging to each of the other source classes in other local libraries than to the source class from which they were originally isolated. Forty (40) isolates were also identified which may be termed "transient". While they appeared source specific in their local watershed study/project library, they match isolates from only one other source class in 2 to 10 other local watershed/project libraries. Rather than removing these cosmopolitan and transient isolates from the library, they will be designated as such to increase the confidence of source identifications for water isolates.

There is still additional evaluation and refinement needed for the Texas *E. coli* BST Library. In particular, further investigation of cross-validation on a watershed/project-exclusive basis when adding isolates to the library in the future and the potential use of statistical probabilities for source identifications when water isolates match library isolates preferentially associated with a source class.

Development and Refinement of Source-Specific Bacterial PCR Markers

Comparison of DNA Extraction Methods

For library-independent BST, a modified Qiagen kit protocol is currently used to extract DNA from water samples as outlined in SOP #TXBST-07 (Di Giovanni, Casarez et al. 2015). The scientific community is increasingly using a kit manufactured by Mo Bio (Carlsbad, CA) that is specifically designed for extraction of microbial DNA from water samples. The Mo Bio kit contains all the supplies needed for these extractions and is more rapid than current methods. Therefore, an experiment was conducted to compare the efficacy of SOP #TXBST-07 with the Mo Bio kit for BST to see if the Mo Bio kit would enhance BST analysis. Water from 12 sites (Table 5) representing a variety of surface water types (creek, rivers, ponds, and runoff) were filtered onto duplicate 47-mm diameter, 0.2-µm pore-size Supor filters (Pall, Port Washington, NY; SOP #TXBST-06). DNA was extracted from each of the duplicate filters using a QIAamp DNA mini kit and the Mo Bio PowerWater® DNA Isolation Kit. Elution volumes on the final step of extraction for both kits was 100 µL. For the QIAamp DNA mini kit, SOP #TXBST-07 was used. For the Mo Bio PowerWater® DNA Extraction kit, extractions were conducted per manufacturer's recommendations. After DNA extractions were completed, *Bacteroidales* PCR and gel electrophoresis were conducted according to SOP #TXBST-07 (Di Giovanni, Casarez et al. 2015).

Table 5. Description of water samples.

Sample number	Date collected	Site collected	Description	Volume of water filtered through each duplicate filter (ml)
1	12/9/15	NAV 11877	River	25
2	12/17/15	LR 13535	River	50
3	12/17/15	LR 16385	River	25
4	1/4/16	NAV 11877	River	25
5	1/5/16	VTA BRAZ IN	Runoff	5
6	1/7/16	VTA ROB IN	Runoff	50
7	1/7/16	VTA ROB OUT	Runoff	50
8	1/8/16	RESEARCH PARK	Pond	50
9	1/11/16	CENTRAL PARK	Pond	50
10	1/11/16	WOLF PEN	Creek	50
11	1/11/16	PARSON'S TANK	Pond	50
12	1/11/16	WASSERMAN POND	Pond	50

With the Qiagen kit, 66.7% of the samples (8 out of 12 samples) were positive for *Bacteroidales* on the agarose gel (Figure 7). Two out of the 8 positive samples showed light or faint bands on the agarose gel. With the Mo Bio kit, 100% of the isolates (12 out of 12) were positive for *Bacteroidales* on the agarose gel using the general primer set (BAC32F-BAC708R) (Figure 7). Three out of the total 12 positive samples exhibited light or faint bands on the agarose gel compared to the others.

Only one sample (NAV11877) showed a stronger band on the agarose gel with the Qiagen Kit (SOP #TXBST-07) when compared to the results from the Mo Bio kit. The runoff samples (VTA BRAZ IN 1/5/16, VTA ROB IN 1/7/16, and VTA ROB OUT 1/7/16) show an equally strong band using both the Qiagen and Mo Bio kit. All of the other samples either showed no band or had a lighter band using the Qiagen kit compared to the Mo Bio kit. These results indicate that the Mo Bio kit is overall more efficient in DNA extraction than the Qiagen kit-based methods currently used in SOP #TXBST-07 (Di Giovanni, Casarez et al. 2015) and results in more sensitive detection of *Bacteroidales* markers from surface water samples.

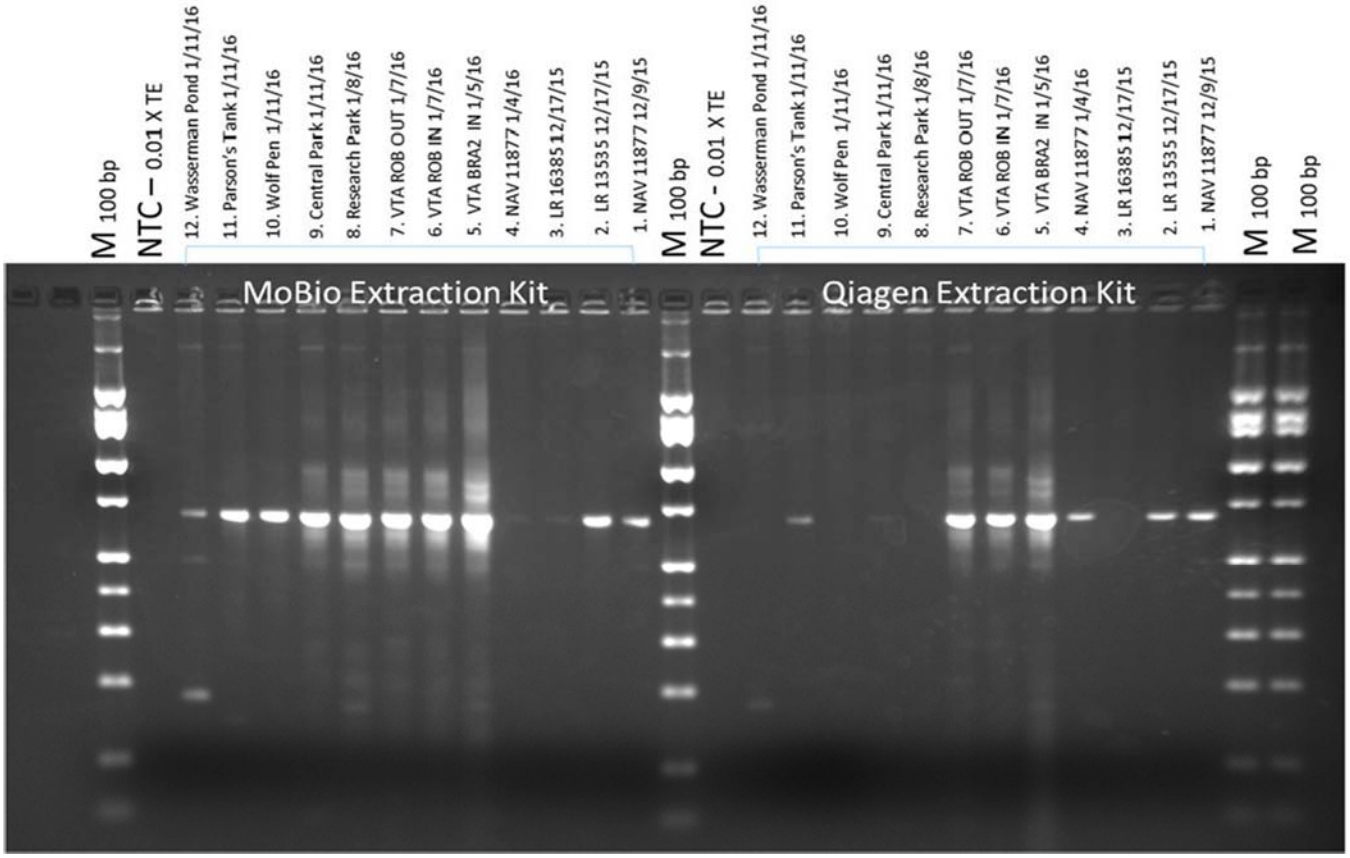


Figure 7. Agarose gel electrophoresis of *Bacteroidales* PCR products from 12 surface water samples extracted with two DNA extraction methods (Mo Bio and Qiagen [#TXBST-07]). M = 100 bp ladder; NTC = no-template control.

Investigation of *Bacteroidales* GenBac PCR products for potential development of a deer-specific marker

As discussed in the TSSWCB Project 13-50 report (Di Giovanni, Casarez et al. 2015), the ability to distinguish between wildlife and livestock sources, particularly deer and cattle, is critical for developing effective BMPs. Unfortunately, the most widely accepted library-independent ruminant specific PCR marker, *Bacteroidales* CF128F, cannot distinguish between cattle and deer (Bernhard and Field 2000).

Previously, the AgriLife SCSC lab used a fecal microbial community approach to identify potential deer-specific BST targets using 454 barcoded pyrosequencing (a next generation sequencing technique). This approach investigated the specific association of bacteria families with deer and identified two promising targets within the *Veillonellaceae* and *Ruminococcaceae* families. Another approach would be to examine the DNA sequences of PCR products amplified using current *Bacteroidales* BST assays. These assays target different bacterial species and strains within the *Bacteroides* and *Prevotella* genera. Therefore, objectives of the current work included DNA sequence analysis of PCR products generated from several deer, feral hog, and goat fecal samples using the general *Bacteroidales* Bac32F and Bac708R (GenBac) primers (Bernhard and Field 2000). A similar approach for the development of chicken and duck-specific *Bacteroidales* primers and qPCR probes was recently reported (Kobayashi, Sano et al. 2013).

Methods

Known source fecal samples were previously collected from animals in the Leon River and San Antonio area watersheds under project TSSWCB 13-50. Samples utilized in this study included those from deer (6), feral hogs (2) and goats (8). The QIAamp DNA Stool Mini Kit (Qiagen, Valencia, CA) was used to extract DNA from approximately 200 mg of each fecal sample per manufacturer's instructions. Extracted DNA was stored at −80°C until analyzed by PCR. *Bacteroides-Prevotella* group 16S rRNA was amplified using the GenBac general *Bacteroidales* primers BAC32F (5'-AACGCTAGCTACAGGCTT-3') and BAC708R (5'-CAA TCG GAG TTC TTC GTG-3') (Bernhard and Field 2000) per SOP #TXBST-07 (Di Giovanni, Casarez et al. 2015). Amplification was performed using a RotorGene 6000 thermal cycler (Qiagen, Valencia, CA). Commercial DNA sequencing of PCR amplicons was performed (SeqWright, Houston, TX) and DNA sequences were analyzed using Kodon (Applied Maths, Houston, TX) and GenBank BLAST (Basic Local Alignment Search Tool) searches (Altschul, Madden et al. 1997).

Results and discussion

Bacteroidales GenBac primers amplify a broad range of species and strains within the *Bacteroides* and *Prevotella* genera. Therefore, the relatively low sequence heterogeneity observed for different samples from the same animal source was somewhat unexpected. Five of six deer, six of eight goat, and both feral hog fecal samples yielded amplicons with high sequence similarity, with the exception of several hypervariable regions. Sequences were aligned and consensus sequences obtained, retaining ambiguous bases (N) for single nucleotide polymorphic positions and some hypervariable regions. Cluster analysis of the aligned sequences is presented in Figure 8 and entries listed with multiple sample IDs representing consensus sequences.

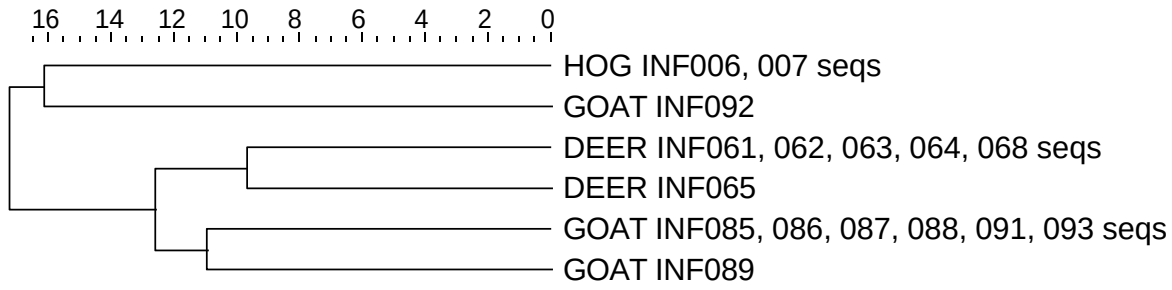


Figure 8. Cluster analysis of *Bacteroidales* GenBac PCR amplicon DNA sequences from deer, feral hog, and goat fecal samples. Entries listed with multiple sample IDs represent consensus sequences. Scale is nucleotide differences per 100 bases.

A closer examination of the aligned sequences allowed the identification of several single nucleotide polymorphic regions and hypervariable regions, even within highly similar sequences from the same animal source (e.g., deer). Comparison of sequences also revealed potential targets for the differentiation of deer, hog, and goat *Bacteroidales* PCR amplicons. Potential forward and reverse priming sites and a hypervariable region that could be targeted using DNA sequence analysis or a qPCR probe (e.g., TaqMan) for the development of a *Bacteroidales* deer-specific PCR marker is presented in Figure 9. BLAST analysis of the resulting 189 bp PCR product revealed top matches with *Bacteroides-Prevotella* sequences. The top five GenBank matches were at 95% to 97% similarity and represented fecal/gut samples obtained from ducks, geese and termites. Although the results are only preliminary and the potential primers and probe sites identified in this work would require further evaluation and refinement, the usefulness of this approach is clear. Therefore, additional sequence analysis of *Bacteroidales* PCR amplicons will be undertaken in the future to gather information and explore the potential development of a deer-specific marker as well as markers for other animal groups.

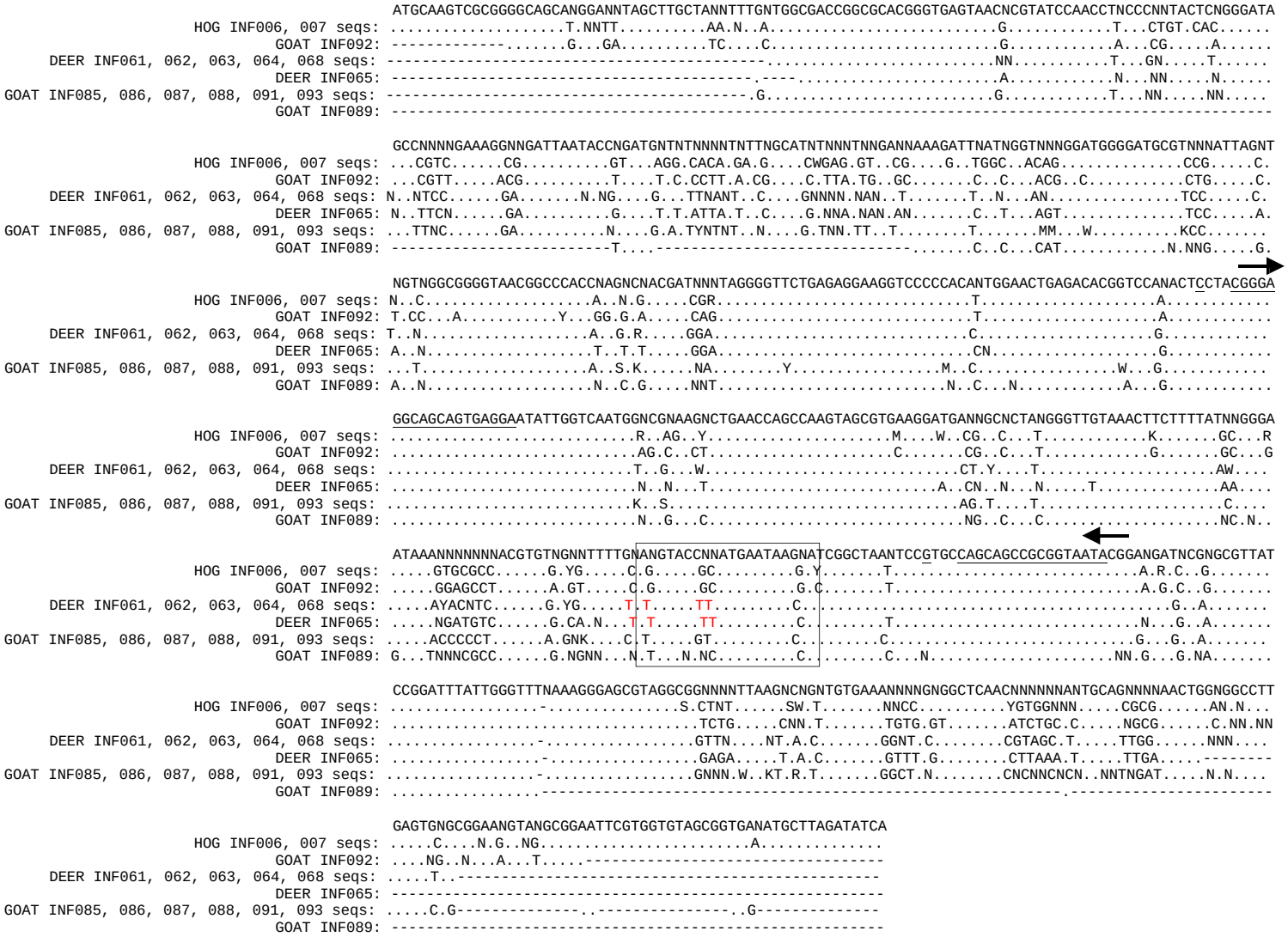


Figure 9. Multiple DNA sequence alignment (partial) of *Bacteroidales* GenBac PCR amplicon DNA sequences from deer, feral hog, and goat fecal samples. Entries listed with multiple sample IDs represent consensus sequences. Underlined sequences and arrows indicate potential forward and reverse priming sites. The boxed region represents a potential site for the differentiation of deer-specific *Bacteroidales* using DNA sequence analysis or qPCR probe (e.g., TaqMan).

Utilization of the Texas *E. coli* BST Library

To identify the human and animal sources of fecal pollution impacting the Arroyo Colorado River, ERIC-RP fingerprints were generated for *E. coli* isolated from river water samples and compared to the Texas *E. coli* BST Library, which was supplemented with known source fecal *E. coli* isolates from the watershed. Monthly water sampling was conducted at ten locations in the Arroyo Colorado River watershed from June 2014 to May 2015. From October 2013 to October 2014, 254 known source fecal samples were also collected from humans and 23 subspecies of animals from the Arroyo Colorado watershed.

Ninety percent of the water isolates were identified using the Texas *E. coli* BST Library (ver. 5-15). Wildlife (both non-avian and avian) was found to be the leading contributor of *E. coli* in the Arroyo Colorado River (Figure 10). Approximately 9% of the isolates were identified as human and another 13% identified as domestic animals.

Full results are not discussed here but can be found in the *Bacterial Source Tracking to Support Adaptive Management of the Arroyo Colorado Watershed Protection Plan* (Berthold et al. 2015). These results are now being used to support watershed protection plan development in the watershed.

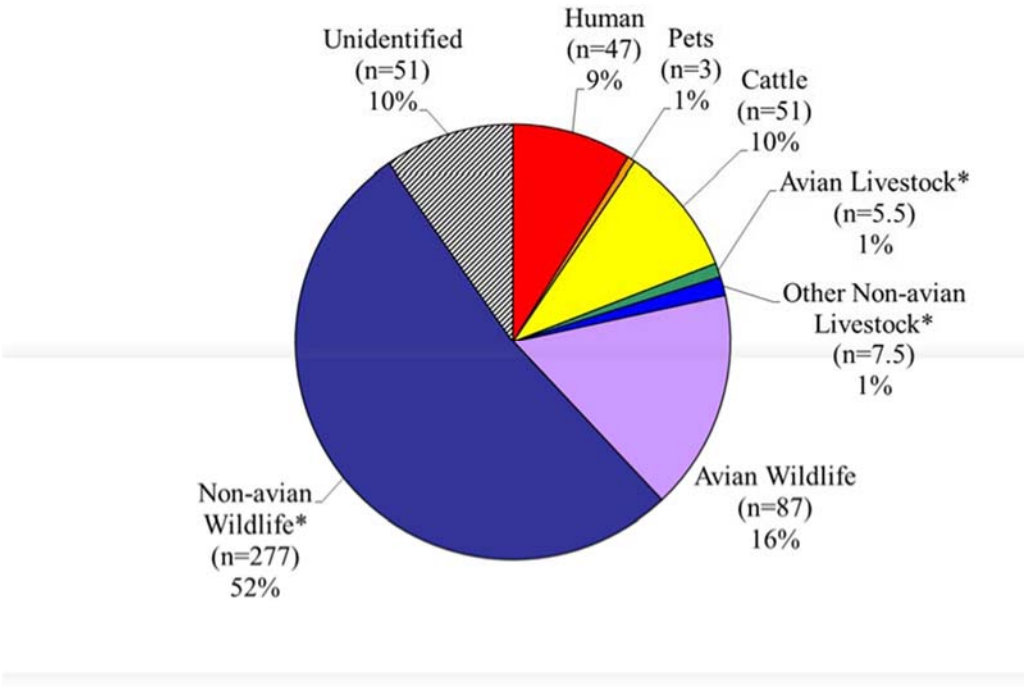


Figure 10. Identification of *E. coli* water isolates from the Arroyo Colorado River watershed using a 7-way split of source classes and an 80% similarity cutoff (n = 529 isolates from 113 samples). One water isolate was equally similar to an "avian livestock" and a "non-avian wildlife" DNA fingerprint, while one other water isolate was equally similar to an "other non-avian livestock" and a "non-avian wildlife" DNA fingerprint. These were considered ties and split between the relevant source classes.

BST Program Outreach

Outreach regarding BST was major focus. The Fall 2015 txH2O highlighted the BST Program in the story titled "A decade of solving water quality mysteries" (Appendix A). Printed copies of the magazine were distributed to 3000 readers and electronic versions were emailed to another 1,158 subscribers. Three subsequent Facebook posts reached 243 readers, with 4 likes, 11 post clicks and 1 share. Additionally, nine tweets resulted in another 2,675 impressions and 28 engagements throughout the term of the project. The BST program was also presented at five conferences and two meetings with TSSWCB, EPA, and TCEQ. AgriLife SCSC gave a presentation on "Resiliency of *E. coli* and Enterococci in poultry litter, and subsequent efficiency of poultry markers, through wetting and drying cycles" at the ASA/CSSA/SSSA Meetings November 2-5, 2014. TWRI presented "What're the sources of bacteria in your watershed? They may not be what you expect" at the 2015 Waste to Worth Conference on March 31, 2015. TWRI UTSPHEP, and AgriLife SCSC promoted the use of and provided resources on BST at the 2015 Environmental Trade Fair and Conference in Austin on May 5-6, 2015. TWRI presented on the Texas BST Program at the 2015 Universities Council on Water Resources (UCOWR) Conference on June 18, 2015 and 70th Soil and Water Conservation Society International Annual Conference on July 28, 2015. TWRI, UTSHP-EP, and AgriLife SCSC presented the BST Program to TSSWCB and EPA on August 25, 2015 and to TCEQ and TSSWCB on December 7, 2015.

Finally, TWRI hosted and maintained the Texas BST Library website. From September 2014 through May 2016, there were 353 visits from 235 visitors (Figure 11). Of the 353 visits, 274 were from the United States and 216 were from Texas (predominantly College Station, Austin, Houston, and San Antonio). The Czech Republic was second to the U.S. in number of visits with 45. There were 992 page views, for a result of 2.81 pages per session. On average, users stayed on the site 2 minutes and 13 seconds. Peak visits occurred in the 5th quarter following the txH2O story highlighting the BST Program.

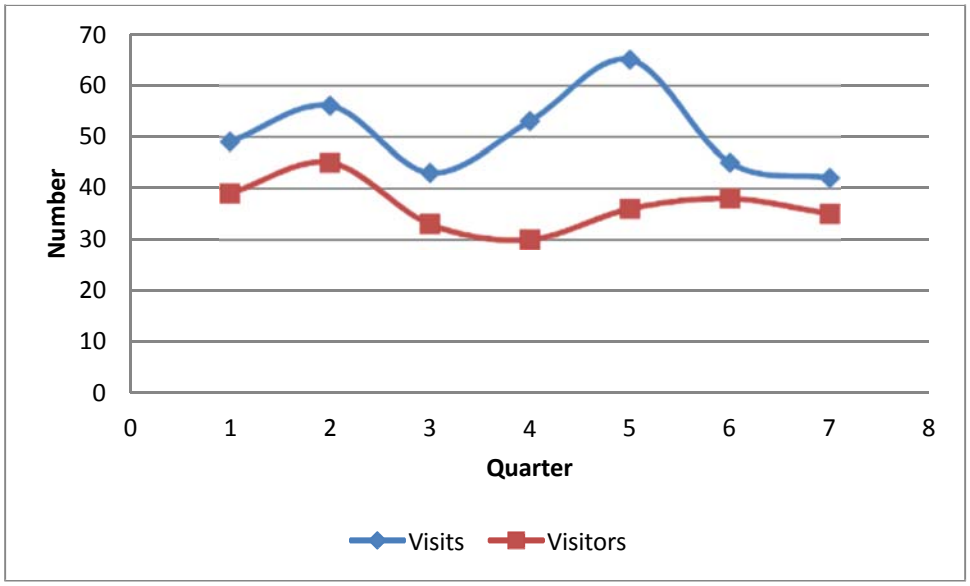


Figure 11. Number of visits and visitors to BST Program website, September 2014 through May 2016.

Literature Cited

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller and D. J. Lipman (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**: 3389-3402.

Bernhard, A. E. and K. G. Field (2000). "A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA." Appl Environ Microbiol **66**(10): 4571-4574.

Berthold, A., G. Di Giovanni, E. Casarez, J. Benavides (2015). Bacterial Source Tracking to Support Adaptive Management of the Arroyo Colorado Watershed Protection Plan. College Station, Texas, Texas Water Resources Institute Technical Report TR-494.

Brennan, F. P., F. Abram, F. A. Chinalia, K. G. Richards and V. O'Flaherty (2010). "Characterization of environmentally persistent *Escherichia coli* isolates leached from an Irish soil." Appl Environ Microbiol **76**(7): 2175-2180.

Brinkmeyer, R., R. M. Amon, J. R. Schwarz, T. Saxton, D. Roberts, S. Harrison, N. Ellis, J. Fox, K. DiGuardi, M. Hochman, S. Duan, R. Stein and C. Elliott (2015). "Distribution and persistence of *Escherichia coli* and Enterococci in stream bed and bank sediments from two urban streams in Houston, TX." Sci Total Environ **502**: 650-658.

Byappanahalli, M. N., B. M. Roll and R. S. Fujioka (2012). "Evidence for occurrence, persistence, and growth potential of *Escherichia coli* and enterococci in Hawaii's soil environments." Microbes Environ **27**(2): 164-170.

Byappanahalli, M. N., R. L. Whitman, D. A. Shively, M. J. Sadowsky and S. Ishii (2006). "Population structure, persistence, and seasonality of autochthonous *Escherichia coli* in temperate, coastal forest soil from a Great Lakes watershed." Environ Microbiol **8**(3): 504-513.

Desai, A. M. and H. S. Rifai (2013). "*Escherichia coli* concentrations in urban watersheds exhibit diurnal sag: implications for water-quality monitoring and assessment." JAWRA Journal of the American Water Resources Association **49**(4): 766-779.

Di Giovanni, G. D., E. A. Casarez, J. A. Truesdale, T. J. Gentry, P. Wanjugi, E. Martin and K. Wagner (2015). Expansion and Evaluation of Texas' Bacterial Source Tracking Program. College Station, Texas, Texas Water Research Institute Technical Report TR-493.

Ishii, S., W. B. Ksoll, R. E. Hicks and M. J. Sadowsky (2006). "Presence and growth of naturalized *Escherichia coli* in temperate soils from Lake Superior watersheds." Appl Environ Microbiol **72**(1): 612-621.

Jones, C. A., K. Wagner, G. D. Di Giovanni, L. Hauck, J. Mott, H. Rifai, R. Srinivasan and G. Ward (2007). Bacteria Total Maximum Daily Load Task Force Final Report, College Station, Texas, Texas Water Resource Institute Technical Report TR-341.

Kobayashi, A., D. Sano, J. Hatori, S. Ishii and S. Okabe (2013). "Chicken- and duck-associated *Bacteroides-Prevotella* genetic markers for detecting fecal contamination in environmental water." Appl Microbiol Biotechnol **97**(16): 7427-7437.

Perchec-Merien, A. M. and G. D. Lewis (2013). "Naturalized *Escherichia coli* from New Zealand wetland and stream environments." FEMS Microbiol Ecol **83**(2): 494-503.

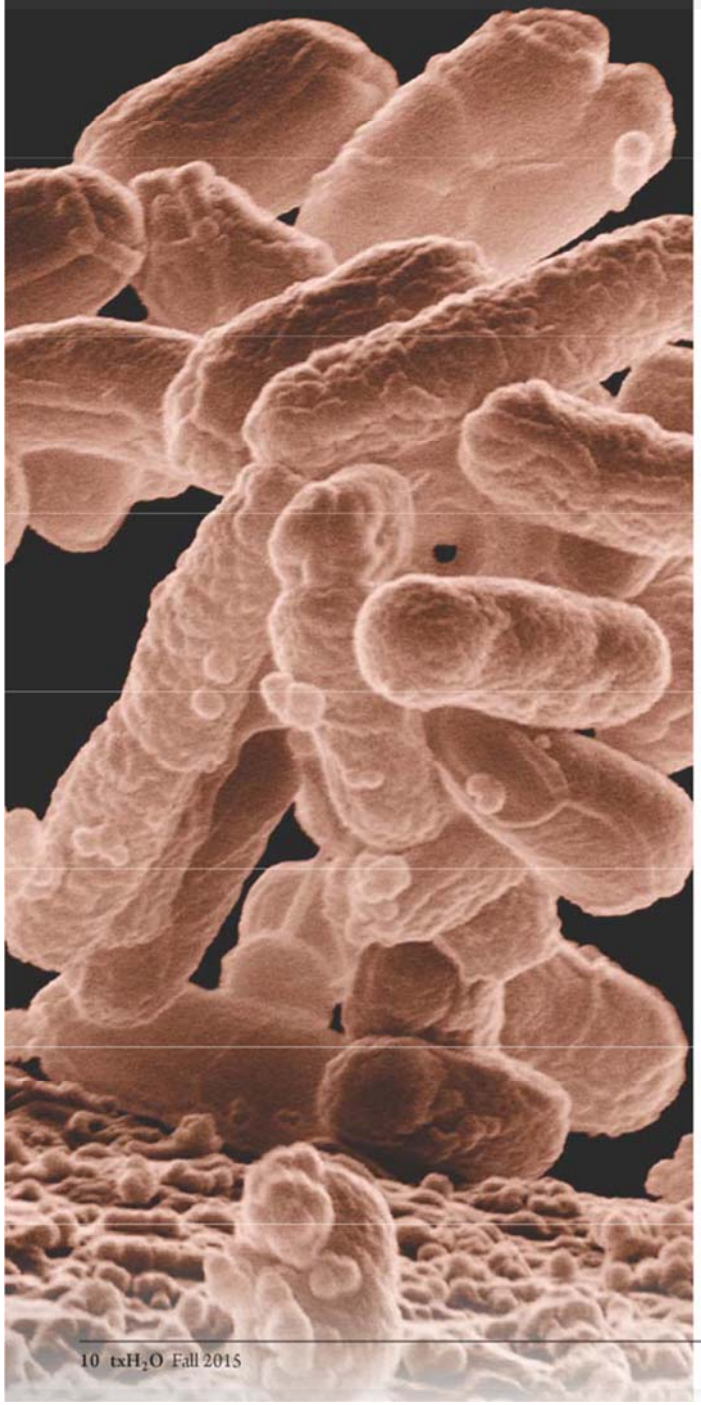
Pisciotta, J. M., D. F. Rath, P. A. Stanek, D. M. Flanery and V. J. Harwood (2002). "Marine bacteria cause false-positive results in the Colilert-18 rapid identification test for *Escherichia coli* in Florida waters." Appl Environ Microbiol **68**(2): 539-544.

Tymensen, L. D., F. Pyrdok, D. Coles, W. Koning, T. A. McAllister, C. C. Jokinen, S. E. Dowd and N. F. Neumann (2015). "Comparative accessory gene fingerprinting of surface water *Escherichia coli* reveals genetically diverse naturalized population." J Appl Microbiol **119**(1): 263-277.

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

Appendix A – txH$_2$O Article

Story by Leslie Lee



Electron micrograph of *E. coli*. Photo by Eric Erbe, digital colorization by Christopher Pooley, USDA Agricultural Research Service.

A decade of solving water quality mysteries

Reflecting on the success of the Texas BST Program

More than 10 years ago, the Texas Bacterial Source Tracking (BST) Program (*texasbst.tamu.edu*) began filling a need in the state's water quality efforts that no other program was pursuing: in-stream measurements of the specific human and animal sources of bacterial nonpoint source pollution in local watersheds.

Before BST technology, water quality restoration projects relied on source surveys and computer models to identify bacterial pollution sources, and these methods oftentimes told an inaccurate or incomplete story.

Source surveys estimated the numbers and distributions of livestock and humans in a watershed, but they could not assess most wildlife species or how bacteria move within waters. Computer modeling addressed bacteria transport issues, but because wildlife populations are rarely known, models were unable to adequately assess wildlife contributions.

However, thanks to a group of researchers from Texas A&M AgriLife Research, the Texas Water Resources Institute (TWRI) and the University of Texas School of Public Health (UTSPH), bacterial pollution sources in watersheds can now be characterized more precisely, and therefore restoration efforts can take more targeted and effective approaches.

"BST is able to evaluate wildlife contributions, along with other major sources, and the impacts of transport because BST uses in-stream water samples for its assessment," said Dr. Kevin Wagner, TWRI associate director. "It has been incredibly helpful in every watershed where we've used it."

How BST science informs restoration projects

When a local water body doesn't meet water quality standards, the most common methods the state of Texas uses for developing plans to restore water quality are either a total maximum daily load (TMDL) paired with an implementation plan or a watershed protection plan (WPP).

TMDLs study and describe the point and nonpoint sources of pollutants affecting a water body; the maximum amount of pollutants the water body can receive daily and still meet standards for its uses; and the reductions needed, if any, from each pollutant source. WPPs are locally developed, comprehensive plans that implement water quality protection and restoration strategies. WPP stakeholders holistically address the causes of impairments and threats to the watershed.

Both routes begin with gaining local stakeholders' involvement and input on potential pollution sources, local water quality problems and possible strategies for restoration.

(left photo) A lab technician works in Dr. George Di Giovanni's lab. Photo courtesy of Dr. George Di Giovanni. (right photo) Lucas Gregory, TWRI project specialist, and a student collect water samples to analyze for *E. coli* and other water quality data. Photo courtesy of Texas Water Resources Institute.

"It's very important to have stakeholder involvement from the beginning," said Dr. George Di Giovanni, professor at the UTSPH El Paso Regional Campus. "That not only helps with community support for the project, but also with sample collection, because accessing many water bodies requires permission to be on private property."

Source tracking field work then begins with frequent monitoring of *E. coli* levels at water monitoring stations throughout the watershed, Di Giovanni said. *E. coli* is the state's indicator bacteria of choice for evaluating the suitability of freshwater water bodies for swimming and other recreation. Samples are usually collected monthly for one to two years to gain a better understanding of the water quality and how it may change seasonally and with a variety of flows, Wagner said.

At the same time, project personnel conduct a survey of the watershed for potential pollutant sources and collect known-source fecal samples in the watershed from wildlife, livestock operations, wastewater treatment plants, septic systems and other sources.

The premise behind BST is that DNA fingerprinting can identify species-specific bacterial strains since each species has different diets and digestive systems with distinct bacterial strains. This distinction allows the original source of the fecal contamination to be identified.

"Landowners are sometimes concerned about source tracking and ask, 'Are you going to tell me that all this *E. coli* bacteria is coming from my cattle and not so-and-so's?'" Di Giovanni said. "But that's not what we're doing; we're looking at general source classifications — cattle, avian wildlife, nonavian wildlife, human, etc."

E. coli bacteria found in water samples from the local water bodies are then cultured in a lab and analyzed using DNA fingerprinting. *E. coli* bacteria from the known-source fecal samples collected in the watershed are also DNA fingerprinted. Then, by comparing those two results, the sources of *E. coli* in that watershed are identified.

"The BST methods that we employ are similar to those used by industry and public health officials to identify microbial sources following process contamination or disease outbreaks," said Dr. Terry Gentry, associate professor in Texas A&M University's Department of Soil and Crop Sciences. "Many BST methods have been developed over the past two ⇨

decades, but we focus on methods that have been both published in the peer-reviewed literature and validated using samples from Texas watersheds."

Pilot project led to trust of BST methods

In 2002, scientists from TWRI and AgriLife Research were tasked with helping address water quality impairments in the Lake Belton and Lake Waco watersheds to support TMDLs being developed there. After getting input from local and regional stakeholders, the researchers began using the BST approach there and collected *E. coli* samples throughout the watershed in 2003.

The research team included Di Giovanni, Dr. Suresh Pillai, AgriLife Research faculty fellow, and Dr. Joanna Mott, now chair of the Department of Life Sciences at Texas A&M University – Corpus Christi. The project was coordinated by the Texas Farm Bureau and funded by the Texas State Soil and Water Conservation Board (TSSWCB), through a U.S. Environmental Protection Agency Clean Water Act grant.

"We used a BST approach to identify the human and animal sources impacting the lakes," Di Giovanni said. "Then we began to build analytical equipment infrastructure, and we started to build a watershed-specific *E. coli* library for that project.

"At the time, one of the weaknesses in most other TMDL approaches was that nonpoint sources of pollution were a significant but uncharacterized component of fecal pollution loading into water bodies," he said.

Nonpoint source pollution is all water pollution that does not come from point sources. Point sources are regulated, end-of-pipe outlets for wastewater or stormwater.

BST technology allowed the researchers to further identify sources within the nonpoint source category. These specific sources include birds, other wildlife, cattle, other livestock, leaking septic systems, wastewater treatment plants and other issues.

"Before BST, many computer models attributed much of the bacterial contributions to cattle, because that was one of the few sources there was good data on," Wagner said. "But BST helped us confirm what many landowners suspected, that cattle were only part of the contributions; on average, cattle contribute about 13 percent of *E. coli* in the rural watersheds studied to date."

One surprising finding of that initial project was that wildlife were a significant contributor of pollution, Di Giovanni said. Their research showed that 40 to 49 percent of the *E. coli* bacteria came from wildlife sources in lakes Belton and Waco watersheds, followed by cattle and then humans.

"And those high wildlife levels have been a finding in almost every study we've done," he said. "Wildlife contributions are much higher than previously thought."

"Having this more complete data helps us increase trust with stakeholders and really helps us with communications during implementation efforts," Wagner said.

Following the success of the Lake Belton and Lake Waco project, in 2006 the Texas Commission on Environmental Quality (TCEQ) and TSSWCB established a joint task force to identify the best, most cost-effective and time-efficient tools for developing bacteria TMDLs. The task force examined BST methods and recommended BST best practices to the state, and TSSWCB began funding the Texas BST Program, led by Di Giovanni, Gentry and Wagner.

BST Library helps improve waters across Texas

Following that initial funding, the *E. coli* bacteria collected and catalogued during the pilot project formed what would become the statewide Texas E. *coli* BST Library, Wagner said.

"The library has grown through subsequent studies," Di Giovanni said. "We've completed 18 studies in 14 different watersheds across the state, and we've certainly expanded and refined the library quite a bit."

The Texas BST Library now contains more than 1,500 *E. coli* isolates obtained from more than 1,300 different domestic sewage, wildlife and livestock fecal samples. These isolates, which represent more than 50 animal subclasses, were selected after screening several thousand isolates from the studies conducted throughout Texas over the past decade.

Di Giovanni's and Gentry's labs oversee and maintain the Texas *E. coli* BST Library data and bacterial culture collections.

In early projects, the BST team could only identify sources through comparisons with known-source fecal samples from that project, but now the library makes the source tracking process more efficient and accurate.

The research team won a 2007 Texas Environmental Excellence Award in Agriculture for its work, as well as the 2014 Texas A&M College of Agriculture and Life Sciences Dean's Outstanding Achievement Award for Interdisciplinary Research.

"One of the reasons the BST Program has been so successful is that it represents a true collaboration between multiple researchers, universities and agencies across Texas," Gentry said. "The extent of this collaboration gives stakeholders greater confidence that the BST results will be helpful for identifying and ameliorating issues in their watersheds."

Watershed restoration projects for Buck Creek, Attoyac Bayou, Leon River, Lampasas River, Bosque River and the Arroyo Colorado have all benefited from source studies conducted by the BST Program, Wagner said.

Moving forward, the program is looking to help with water quality efforts in urban watersheds, he said.

"We're really turning our attention to try to do more of this work in more urbanized settings," he said. "In predominately rural watersheds, wildlife contributes about half of the bacteria. We'll see if that differs in urban settings or not."

"Water quality is a challenge in urban watersheds, and there is much more of a potential impact from human-derived sources," Di Giovanni said. "We need to take a close look at that and see if we can identify controllable sources."

"As we continue to address water quality issues in Texas, the BST Program is always available to provide local entities with guidance and assistance in performing BST for watersheds," Wagner said. "BST has been tremendously helpful in identifying significant bacterial sources throughout Texas."

For more information and resources, visit *txH2O* online at twri.tamu.edu/txH2O.