SUPPORT ANALYTICAL INFRASTRUCTURE AND FURTHER DEVELOPMENT OF A STATEWIDE BACTERIAL SOURCE TRACKING LIBRARY

September 2013

Funding provided by the Texas State Soil and Water Conservation Board through the State Nonpoint Source Grant Program

Texas Water Resources Institute TR-448

SUPPORT ANALYTICAL INFRASTRUCTURE AND FURTHER DEVELOPMENT OF A STATEWIDE BACTERIAL SOURCE TRACKING LIBRARY

STATE NONPOINT SOURCE GRANT PROGRAM

TSSWCB PROJECT 10-50

Prepared for:

TEXAS STATE SOIL AND WATER CONSERVATION BOARD

Prepared by:

GEORGE D. DI GIOVANNI
ELIZABETH A. CASAREZ

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

TERRY J. GENTRY
EMILY C. MARTIN

TEXAS A&M UNIVERSITY

LUCAS GREGORY
KEVIN WAGNER

TEXAS WATER RESOURCES INSTITUTE

SEPTEMBER 2013

TEXAS WATER RESOURCES INSTITUTE TR-448

TABLE OF CONTENTS

TABLE OF CONTENTS i
LIST OF FIGURES iii
LIST OF TABLES v
LIST OF ACRONYMS vi
EXECUTIVE SUMMARY vii
INTRODUCTION 1
Goals of This Study 3
TEXAS *E. COLI* BST LIBRARY 4
Expansion of the Texas *E. coli* BST Library 5
Refinement of the Texas *E. coli* BST Library 7
Source Isolate Needs 10
Geographic and Temporal Stability of the Library 12
Environmental *E. coli* strains 13
Summary and Conclusions 13
EVOLUTION OF BACTERIAL SOURCE TRACKING IN TEXAS 14
Copano Bay 14
Oyster Creek 16
Trinity River 17
Assessment of Bacterial Sources Impacting Lake Waco and Belton Lake 17
Upper and Lower San Antonio River, Salado Creek, Peach Creek, and Leon River 23
Bacterial Total Maximum Daily Load Task Force Report 24
Lake Granbury 25
Increased Analytical Infrastructure and Development of a Statewide BST Library 26
Other BST Projects in Texas 28
Ongoing BST Projects in Texas 29
Conclusions 32
IMPLICATIONS OF *E. COLI* ISOLATION METHOD ON LIBRARY-DEPENDENT BST 33
Introduction and Rationale 33
Methods 33
Results 34
Discussion and Conclusions 39
ENUMERATION OF *E. COLI* IN KNOWN-SOURCE FECAL SAMPLES 43

STANDARD OPERATING PROCEDURE UPDATES 46
DEVELOPMENT AND EVALUATION OF NEW SPECIES SPECIFIC BACTERIAL MARKERS FOR LIBRARY-INDEPENDENT BST 47
EXPANDED DELIVERY OF BST EDUCATION AND OUTREACH 49
State of the Science Website 49
Promotion of BST 49
State of the Science Conference 49
LITERATURE CITED 51
APPENDIX A: 54
APPENDIX B: 58

LIST OF FIGURES

Figure 1. Texas *E. coli* BST Library (ver. 1-13) library composition by 7-way split of source classes 9

Figure 2. Texas *E. coli* BST Library (ver. 1-13) library composition by 3-way split of source classes 9

Figure 3. Location of known source fecal sample collection efforts supporting the development of the Texas *E. coli* BST Library. 15

Figure 4. Congruence of individual BST methods and composite datasets (Labels include ERIC, ERIC-PCR, RP-RiboPrinting, ARA-KB-ARA, ERIC-RP-ARA-PFGE-four method composite dataset) 22

Figure 5. BST results from Buck Creek station 20368. Identification of water isolates (pie chart) using a 3-way split for source classification and *Bacteroidales* PCR maker occurrence (bar chart) 30

Figure 6. BST results from Little Brazos River. Identification of water isolates (pie chart) from all creek sites using a 3-way split for source identifications (n=69) and *Bacteroidales* PCR marker occurrence (n=259) for human, ruminant, hog, and horse markers 31

Figure 7. BST results from Big Cypress. Identification of water isolates (pie chart) from all creek sites using a 3-way split for source identifications (n=101) and *Bacteroidales*PCR marker occurrence (n=244) for human, ruminant, hog, and horse markers. 32

Figure 8. Overall *E. coli* genotype overlap across all three media types. Data represents the percentage of isolates with unique fingerprint patterns detected in each media type and when those isolates occurred in multiple media types. Data in parenthesis represents the number of unique genotypes (g) in each combination while the values in brackets represent the number of isolates {i} in each category. 36

Figure 9. Identification of water isolates sorted by media type using a 3-way split for source classification (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified) 38

Figure 10. Identification of water isolates at Big Iron Ore Creek (A) and Burton Creek (B) using a 3-way split for source classification (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified) 39

Figure 11. Identification of water isolates at Campbells Creek (A) and Moody Creek (B) using a 3-way split for source classification (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified) 40

Figure 12. Identification of water isolates at Plum Creek and White Oak Bayou using a 3-way split for source classification (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified) 40

Figure 13. Comparison of MI and Colilert® BST results to mTEC across all sites using a 3-way split for source classification. Error bars represent standard error of three replicate samples (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified). 41

LIST OF TABLES

Table 1. Effort for Sample Collection, Fingerprinting, and Screening for the Texas *E. coli* BST Library (10-12 PRE) 6

Table 2. Texas *E. coli* BST Library (ver. 1-13, cross-library validation) composition and rates of correct classification (RCCs) 8

Table 3. Jackknife analysis rates of correct classification (%) for individual and four method composite BST methods and the 883 isolate library 20

Table 4. Texas *E. coli* BST Library (ver. 8-10) composition and rates of correct classification (RCCs). 28

Table 5. Summary of the mean *E. coli* concentrations from six sites across the three media types. (CFU or MPN/100mls +/- SE) 35

Table 6. *E. coli* diversity estimates (ERIC-RP 80% similarity cutoff). 35

Table 7. Overlap in *E. coli* genotypes across media types 37

Table 8. Summary of the mean *E. coli* concentrations from known-source fecal samples (n=424). 44

LIST OF ACRONYMS

ARA	antibiotic resistance analysis
ARCC	average rate of correct classification
BMPs	best management practice
BST	bacterial source tracking
CFU	coliform forming units
CS	College Station
E. coli	*Escherichia coli*
EP AREC	AgriLife Research and Extension Center - El Paso
EPA	United States Environmental Protection Agency
ERIC	enterobacterial repetitive intergenic consensus sequence
ERIC-RP-ARA-PFGE	ERIC-PCR-RiboPrint-antibiotic resistance analysis-pulse field gel electrophoresis composite fingerprints
HRM	high resolution melt
IBDG	chromogen indoxyl-b-d-glucuronide
IEH	Institute for Environmental Health
KB-ARA	Kirby-Bauer antibiotic resistance analysis
MPN	most probable number
mTEC	membrane thermotolerant *E. coli*
MUGal	fluorogen 4-methylumbelliferyl-β-D-galactopyranoside
MUG	fluorogen 4-methylumbelliferyl-β-D-glucuronide
NA-MUG	nutrient agar with 4-methylumbelliferyl- β-D-glucuronide
NPS	nonpoint source
ONPG	chromogen ortho-nitrophenyl-b-d-glactopyranoside
OTU	operational taxonomic units
PCR	polymerase chain reaction
PFGE	pulsed-field gel electrophoresis
RARCC	random average rate of correct classification
RCC	rate of correct classification
RP	automated ribosomal ribonucleic acid (RNA) genetic fingerprinting
SCSC	Texas A&M Department of Soil and Crop Sciences
SOPs	standard operating procedures
TAMU-CC	Texas A&M University - Corpus Christi
TCEQ	Texas Commission on Environmental Quality
TIAER	Texas Institute for Applied Environmental Research
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 means
UTSPH-EP	University of Texas Health Science Center At Houston School of Public Health El Paso Regional Campus
WPP	watershed protection plan
WWTP	wastewater treatment plant

EXECUTIVE SUMMARY

The project titled *Support Analytical Infrastructure and Further Development of a Statewide Bacterial Source Tracking Library* funded by the Texas State Soil and Water Conservation Board was established to provide needed resources to expand the application and utility of bacterial source tracking (BST) while simultaneously advancing the science of BST. In some form, BST has been applied in Texas for more than a decade; however, method differences, inconsistent approaches, and limited geographical coverage of the Texas *E. coli* BST Library have caused concern over the applicability of BST results between watersheds. Corollary to these concerns is a general lack of readily available information on the subject. Subsequently, the applicability and utility of BST are often misunderstood. Additionally, the formation of the Bacteria Total Maximum Daily Load (TMDL) Task Force and the recommendations it produced outlined needed advancements in BST research and development. This project was established to allay these concerns by focusing on increasing application, capacity and coverage of BST resources available in the state and to accomplish several of the recommended research and development needs.

Operationally, two laboratories in Texas have conducted the bulk of BST work in the state and have the equipment and personnel capacity to effectively perform large volumes of BST. The laboratory originally established at the Texas A&M AgriLife Research and Extension Center at El Paso (EP AREC) recently relocated to the University of Texas Health Science Center at Houston School of Public Health, El Paso Regional Campus (UTSPH EP) is one of these facilities. The other is the Texas A&M AgriLife Research, Soil and Aquatic Microbiology Laboratory in the Soil and Crop Sciences Department at Texas A&M University in College Station (AgriLife SCSC). Investments in equipment and personnel at each of these entities require continued support; which was provided through this project.

The application and approaches utilized for BST studies conducted in the state since the early-2000's have changed over time. To provide a clear picture of this evolution, AgriLife SCSC conducted an extensive literature review and summarized the number of samples processed, methods and materials utilized, results, and performing entities for each of the majority of BST projects completed in Texas prior to the start of this project. A brief description of these findings is presented here to illustrate both the similarities and differences between studies.

Through this project, AgriLife SCSC and UTSPH EP researchers were able to effectively expand the Texas *E. coli* BST Library. A total of 504 fecal source samples were collected with 943 *E. coli* isolates screened for clonality, resulting in 579 isolates being included in the local watershed libraries. Additional screening for source self-validation resulted in the addition of 406 *E. coli* isolates to the Texas *E. coli* BST Library. Additionally, the isolates from the six watershed projects greatly increased the geographic distribution of the library. A refinement process was also employed to increase source specificity. Collectively, 259 cosmopolitan isolates were identified and removed, improving the average rate of correct classification (ARCC) by fifteen percentage points. An assessment of *E. coli* sources collected and screened for the library was also completed and used to identify additional *E. coli* source isolate needs.

E. coli can be enumerated using a variety of approved methods; however, only samples processed using a membrane filter method developed by the U.S. Environmental Protection Agency (USEPA) that selects exclusively for *E. coli* (USEPA Method 1603) has been used by the majority of previous Texas BST studies. Unfortunately, this one method is less commonly utilized by typical water quality laboratories for compliance monitoring, effectively precluding these samples from being further processed for use in the Texas *E. coli* BST Library. Through this project, water samples were collected from six locations and processed using three separate methods. Comparisons of *E. coli* enumeration counts as well as BST results were made between samples and between sampling sites to determine if the methods produced statistically significant similarities in the results reported. The *E. coli* enumeration, strain diversity and BST classifications were not consistent between methods, and therefore, did not provide support for the use of alternative processing methods for samples that are to be subjected to BST analysis. In addition, standard operating procedures (SOPs) for BST analysis procedures developed by AgriLife SCSC and UTSPH EP were reviewed to ensure that any advancements or changes in technique were appropriately described.

Also of use in watershed planning efforts is the enumeration of *E. coli* in known-source fecal material. Through this project, 424 known-source fecal samples were received, processed, and enumerated to yield a count of *E. coli* per wet gram of feces. In many cases, *E. coli* levels below the detection limit (100 cfu/g) were observed; however, the bulk of samples produced viable counts. These data will greatly improve the available information base that can be used for estimating *E. coli* loads from evaluated species.

Dissemination of information to practitioners, agency personnel and academia was also a need addressed through this project. To accomplish this objective, a project website was developed and maintained, discussions with focus groups were held, and promotional flyers were developed and distributed. A ‘state of the science’ conference was also held which brought national BST experts to Texas to provide insight on recent developments in the science to meeting attendees. In total, nearly 120 attendees from 13 states participated in the conference.

Collectively, this project made great strides in advancing the science of BST and addressed many of the BST related research and development needs identified in the Bacteria TMDL Task Force report.

INTRODUCTION

Protection of our water resources is one of the most significant environmental challenges of the new millennium. Nonpoint sources (NPS) of pollution are often considered to be the largest contributors of pollutants to receiving waters and can greatly impact water quality. One key component in effectively implementing a NPS pollution abatement program is the identification and assessment of sources of fecal pollution. Proper evaluation of these sources is needed to target best management practices (BMPs) and develop bacterial total maximum daily loads (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 illness do not colonize nonhuman hosts. The 2012 303(d) List identified 257 contact recreation use impairments (waterbody-pollutant combinations) and 15 oyster water use impairments due to excessive bacteria (*E. coli*, *Enterococcus* spp., or fecal coliforms). These bacteria impairments account for approximately 48% of all impairments on the 2012 303(d) List.

Fecal coliform bacteria have extensively been used as an indicator of fecal pollution and the potential presence of other pathogenic microorganisms in water. It has been established that the fecal coliform bacterium *E. coli* is more closely associated with fecal pollution than other fecal coliform bacteria, which may normally reside and multiply in the environment. *E. coli* is a common inhabitant of animal and human intestines and recent studies have shown that isolates from humans and various host animals (e.g., cattle, chickens, and pigs) may differ genetically and phenotypically. Use of genetic and biochemical tests may allow the original host animal to be identified and is referred to as bacterial source tracking (BST).

The premise behind BST is that genetic and phenotypic tests can identify bacterial strains that are host specific so that the original host animal and source of the fecal contamination can be identified. Often *E. coli* or *Enterococcus* spp. are used as the bacteria targets in source tracking, as this provides a direct link with water quality standards which are usually based on one of these two indicators. BST of *E. coli* has the advantages of direct regulatory significance and availability of standardized culturing techniques for water samples, such as EPA Method 1603 (EPA 2005).

BST is a valuable tool for identifying human and animal sources of fecal pollution. A Task Force on Bacteria TMDLs was jointly established by the TSSWCB and the TCEQ in fall 2006. In the Task Force's Report, a strategy to address current and future bacterial TMDLs and Implementation Plans (I-Plans) was outlined. The Task Force recommended the usage of BST and the TSSWCB and TCEQ adopted the general process laid out by the Task Force including the use of BST. Comprehensive BST has been completed by UTSPH-EP (formerly the team at EP AREC) for (1) the Lake Waco and Belton Lake watersheds, (2) several San Antonio area watersheds and (3) the Lake Granbury watershed. The application of BST to assess water quality impairments has expanded across the State to include projects with UTSPH EP and AgriLife SCSC in the Big Cypress Creek, Buck Creek, Leon River, Leona River, Lampasas River, Attoyac Bayou, the Little Brazos River tributaries, Upper Trinity River, and Upper Oyster Creek watersheds. A statewide library of known source *E. coli* isolates was developed and has been expanded based on known source samples from these studies. The use of this library, called the

Texas *E. coli* BST Library, will provide for significant cost and time savings for the identification of NPS pollution in the development of TMDLs and WPPs.

The Task Force recommended using library-independent methods such as *Bacteroidales* PCR for preliminary qualitative analyses and library-dependent methods (e.g., ERIC-PCR and RiboPrinting) if more quantitative data are required. To help achieve the recommendations of the Task Force and to support the current and anticipated need for BST studies in Texas, this project addressed several issues including: 1) the appropriate maintenance of the established statewide BST analytical infrastructure including both equipment and skilled personnel, 2) periodic reviews and updates to BST SOPs to ensure they are current with applicable methods, technologies, and markers, and 3) expansion of the Texas *E. coli* BST Library to include additional known source isolates from different Texas watersheds and different animal hosts.

In light of the increased application of BST across the state, the validity of isolates processed using methods other than the modified mTEC method (EPA 1603) for use in BST analysis has come into question. The current approach of UTSPH EP/SCSC utilizes only isolates processed using the EPA 1603 method; however, the majority of water quality labs across Texas do not currently utilize this method. Potential cost savings, the ability to utilize more samples due to holding time constraints and the potential to advance the science of BST all warrant the evaluation of the EPA 1604 and Colilert® methods of *E. coli* enumeration/isolation to produce statistically similar types and counts of *E. coli* isolates as those produced in using the EPA 1603 method.

There have been significant developments in library-independent BST methods, including bacterial genetic markers specific to different animal sources and humans. While library-independent methods are cost-effective, rapid, and potentially more specific than library-dependent methods, concerns include uncertainties regarding geographical stability of markers and difficulty in interpreting results in relation to regulatory water quality standards and microbial risk. More importantly, these library-independent methods can only detect a limited range of pollution sources and are currently only semi-quantitative. Despite these limitations, library-independent methods may be very useful for the rapid and inexpensive assessment of the possible sources of fecal pollution impacting a water body. Expanding the discriminatory ability of this approach to detect a broader spectrum of species, specifically poultry, cattle, feral hogs, and deer, will greatly enhance the applicability of this approach. Future research will be conducted and peer reviewed markers from other researchers will be evaluated for possible inclusion into Texas' standard approach to conducting BST.

Lastly, the state of BST science, methodologies, application and confidence has evolved greatly in the past few years. A host of new information is currently available, yet not readily distributed or known to state and federal agency personnel. To remedy this lack of information transfer, TWRI as part of this project developed printed media and a website to advance the science and application of BST in Texas and nationally. A state of the science workshop was also held targeting academia involved in BST analysis; state, federal and regional agency personnel; elected officials; and other interested persons.

Goals of This Study

The primary objectives of this project were to support anticipated volume of bacterial source tracking (BST) studies across the State and advance the application and utility of BST in water quality restoration efforts statewide. To accomplish this, the project was designed with the following project goals in mind:

1) to provide continued personnel support, operation and maintenance of BST analytical infrastructure at the UTSPH-EP and AgriLife SCSC laboratories

2) to continue the development, updating and implementation of statewide BST standard operating procedures (SOPs) for ERIC-PCR, RiboPrinting, and *Bacteroidales* PCR along with coordination amongst other BST conducting entities in the state to standardize methodologies employed in this process

3) to provide for the development of educational and informational materials that give an overview of BST activities in Texas to date and that promote the use, capabilities, and applicability of BST and the services provided by the state-supported analytical labs, and to deliver these materials and presentations to local, state and national stakeholder audiences

4) to continue the development of the Texas *E. coli* BST Library by incorporating additional known source fecal sample isolates from Texas and around the U.S., and to quantify *E. coli* production in known sources of fecal material

5) to further development of suitable species-specific bacteria markers for library independent BST

6) to conduct comparisons between isolation methods to evaluate their potential usefulness in future BST

7) to develop and deliver agency-specific workshops focusing on BST applications, advancements and usability and also coordinate, plan and deliver a state of the science conference on BST technologies and capabilities.

TEXAS *E. COLI* BST LIBRARY

BST is a valuable tool for identifying human and animal sources of fecal pollution in watersheds. The premise behind BST is that genetic and phenotypic tests can identify bacterial strains that are host-specific so that the original host animal and source of the fecal contamination can be identified. Two of the molecular tools for BST that provide sensitive characterization and discrimination of bacterial isolates (in this case *E. coli*) are automated ribosomal ribonucleic acid (RNA) gene fingerprinting (RiboPrinting) and enterobacterial repetitive intergenic consensus sequence polymerase chain reaction (ERIC-PCR). Since both of these DNA fingerprinting techniques are "library-dependent" methods, it is necessary to construct a reference library of DNA fingerprints for *E. coli* isolated from known fecal sources (e.g., humans, livestock, pets and wildlife). By matching the fingerprints of *E. coli* isolated from water samples with the fingerprints in the source library, the likely animal or human fecal sources of surface water contamination can be determined. While library-independent BST methods can be cost-effective, rapid, and potentially more specific than library-dependent methods, the library-independent methods can only detect a limited range of pollution sources and typically target non-regulated fecal bacteria. Therefore, interpretation of library-independent BST data is challenging with respect to identifying all contributing sources of fecal pollution and relationship to regulatory water quality standards based on *E. coli*.

Library-dependent bacterial source tracking methods such as ERIC-PCR and RP require large and diverse collections of *E. coli* for the most reliable source identification of water isolates. It is not feasible due to time and resource constraints to create such comprehensive local known source libraries for every watershed. Because the overriding factors of the host enteric environment should remain similar, it may be possible to combine small collections of known source isolates into a larger library that can be used to identify unknown water isolates collected from different watersheds in future studies. We are therefore developing a Texas *E. coli* BST Library, a statewide reference library of *E. coli* isolates and their genetic fingerprints from known source samples collected from BST projects in watersheds across the state. Previous studies have shown that using the Texas *E. coli* BST Library: 1) supplements difficult-to-obtain wildlife samples, 2) decreases by 10-fold (from approximately 1,000 to 100) the number of known source samples that need to be collected and processed from each local watershed, saving significant time and money, and 3) decreases the number of water isolates left unidentified, allowing for more accurate identification of the sources contributing to the bacterial loading. The use of the Texas *E. coli* BST Library will provide for significant cost and time savings for the identification of NPS pollution and the development of TMDLs and WPPs.

The development of the Texas *E. coli* BST Library is a dynamic process. To represent the different potential human and animal sources of fecal contamination impacting Texas watersheds, as well as the diversity of *E. coli* associated with these different sources, additional known source isolates need to be collected and included in the Texas *E. coli* BST Library. In addition, the library is being refined through cross-validation of source isolates to remove cosmopolitan (non-specific) strains and de-cloning (removal of identical strains). Together, these steps will increase the accuracy and utility of the Texas *E. coli* BST Library.

Expansion of the Texas *E. coli* BST Library

Version 1-10 of the Texas *E. coli* BST Library contained 1,190 isolates from 1,063 samples from the Waco-Belton, San Antonio, Lake Granbury, Buck Creek, and Oyster Creek and Trinity River watersheds. This was the result of collecting 2,838 samples, of which 2,135 samples tested positive for *E. coli,* yielding 7,226 *E. coli* isolates for archival. Of the archived isolates, 5,085 were screened for clones by ERIC-PCR and 2,551 isolates were RiboPrinted. A total of 2,415 working isolates from 2,092 samples were selected for these local libraries. Self-validation Jackknife screening of these six local watershed libraries resulted in the selection of 1,307 isolates from 1,185 samples for further evaluation and the final inclusion of 1,190 isolates from 1,063 samples in the Texas *E. coli* BST Library (ver. 1-10).

As of October 2012, an additional 504 samples have been collected from six additional watershed projects (TSSWCB projects 06-07, 09-55, 10-51, 09-10, 09-52 in Leon River, Big Cypress Creek, Leon and Lampasas Rivers, Attoyac Bayou, and the Little Brazos River tributaries, respectively). Of these, 384 samples were positive yielding 1586 additional *E. coli* isolates for the archive. Screening for clones using ERIC-PCR was completed for 943 of these isolates and 582 were RiboPrinted. These projects produced a total of 579 working isolates from 382 samples for their local libraries, close to the goal of 600 isolates to be considered for the Texas *E. coli* BST Library. After self-validation Jackknife screening for source specificity, 406 isolates from 299 samples have been added so that the Texas *E. coli* BST library version 10-12 (PRE) has 1,713 isolates from 1,484 samples representing over 50 animal subclasses. Table 1 shows the breakdown from samples collected to the level of self-validated sample by watershed. Additional samples from Buck Creek and the ongoing project in the Leona watershed have recently been analyzed and additional isolates will be screened and added to the library.

Table 1: Effort for sample collection, fingerprinting, and screening for Texas *E. coli* BST Library (10-12 PRE)

Watershed	# of total samples collected	# of (+) samples	# of isolates archived	# of isolates ERIC-PCR	# of isolates RiboPrinted	# of isolates local library	# of samples local library	# of isolates self-validated	# of samples self-validated
San Antonio	1,013	786	3,330	2,107	947	932	778	457	403
Waco-Belton	1,143	834	3,224	2,275	1,079	958	813	537	481
Lake Granbury	74	59	198	173	80	80	59	60	48
Oyster Creek	355	298	292	286	286	286	286	166	166
Trinity River	193	130	129	128	128	128	128	67	67
Buck Creek	60	28	53	53	31	31	28	20	20
Little Brazos River	75	66	166	63	85	85	66	66	57
Leon (CS)	30	30	146	146	72	72	30	58	27
Leon (UTH)	95	71	323	204	133	132	71	85	60
Lampasas	118	85	384	244	145	143	83	97	67
Big Cypress	30	19	73	73	34	34	19	28	16
Attoyac	156	113	494	113	113	113	113	72	72
TOTAL	3,342	2,519	8,812	6,028	3,133	2,994	2,474	1,713	1,484

Refinement of the Texas *E. coli* BST Library

The process for selecting known source isolates for inclusion into the state BST library has recently been refined and was applied to the six watersheds mentioned earlier. All de-cloned isolates from individual source samples (up to 3) were included in the local watershed library, independent of their similarity to other library isolates. Jackknife analysis of the local watershed library ERIC-RP fingerprints was used to identify the isolates that were 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, including feral hogs). Singletons (isolates which were left unidentified using an 80% similarity cutoff and therefore have unique fingerprints) were also included to create the local self-validated libraries as described above.

The premise behind BST is that different strains of *E. coli* have adapted to conditions in the guts of their specific animal hosts, resulting in strains that are specifically associated with that species or closely related species. There can also be cosmopolitan or transient strains of *E. coli* which can be found in many different hosts. These are not helpful for BST, but will be present in source samples and water samples nonetheless. Performing the stringent 7-way source class Jackknife self-validation screening at the local watershed level removes obvious cosmopolitan strains without temporal or geographical confounding factors. To determine whether isolates that were singletons in their local watershed could also be cosmopolitan strains, Jackknife analysis on the Texas *E. coli* BST library was then used to screen out any singleton isolates that were incorrectly matching isolates from other watersheds using a 3-way split of source classes (human, domestic animals, and wildlife). Isolates that were still unique (left unidentified using an 80% similarity cutoff; singletons) were left in the library. When this refinement approach was used, 81 isolates were removed and the overall ARCC went from 77% with 17% unidentified to 83% with 18% unidentified using a 7-way split of source classes and from 84% to 89% ARCC using a 3-way split of source classes. While removing previously unvalidated isolates did improve the library statistics, there were still a few isolates that cross identified at a 3-way split of source classes.

The next less conservative approach to address these cosmopolitan strains was to run a series of Jackknife analyses 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 3-way split of source classes. We began with the 1,713 isolates of the combined self-validated local watershed libraries, as described in Table 1. After the first Jackknife analysis, 228 isolates were removed leaving 1,485 isolates. When Jackknife analysis was run again, 27 isolates cross-identified and were removed, leaving 1,458 isolates. Another round of Jackknife analysis removed four additional misidentified isolates. When the Jackknife analysis was run on the remaining 1,454 isolates, it resulted in 100% ARCC with a 3-way split of source classes and a 92% ARCC using the 7-way split of source classes. A total of 20% of the isolates were identified as singletons (unique fingerprints). See Table 2 for the statistics on this cross-watershed validated version of the Texas *E. coli* BST Library (ver. 1-13) and Figures 1 and 2 for graphical representation of the library's 3- and 7-way split source composition. On average, approximately 15% of the original self-validated isolates per source class were identified as cosmopolitan strains and removed. The percentage was highest for the pet source class where 28 of the 111 original self-validated isolates (25%) were identified as cosmopolitan and removed.

Table 2. Texas *E. coli* BST Library (ver. 1-13, 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

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	**364**	**315**	**25%**	**100**	**4.0**	**22**
DOMESTIC ANIMALS	**497**	**442**	**34%**	**100**	**2.9**	**21**
Pets	83	73	6%	88	15.4	41
Cattle	220	192	15%	94	6.2	12
Avian Livestock	93	80	6%	88	13.8	27
Other Non-Avian Livestock	101	97	7%	89	12.8	16
WILDLIFE	**593**	**534**	**41%**	**100**	**2.4**	**19**
Avian Wildlife	232	214	16%	85	5.3	21
Non-Avian Wildlife	361	320	25%	91	3.7	19
Overall	**1,454**	**1,291**		**ARCC** = 3-way :100% 7-way: 92%**		**20%**

**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 4.0-fold greater than random chance.

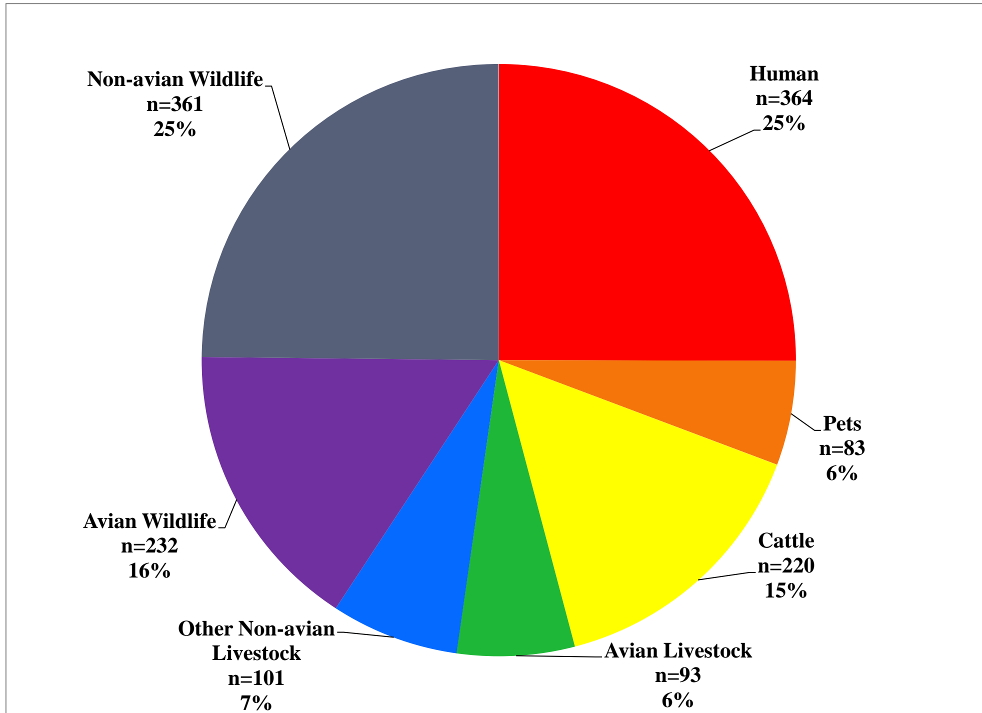


Figure 1. Texas *E. coli* BST Library (ver. 1-13) library composition by 7-way split of source classes (1,454 isolates from 1,291 different fecal source samples).

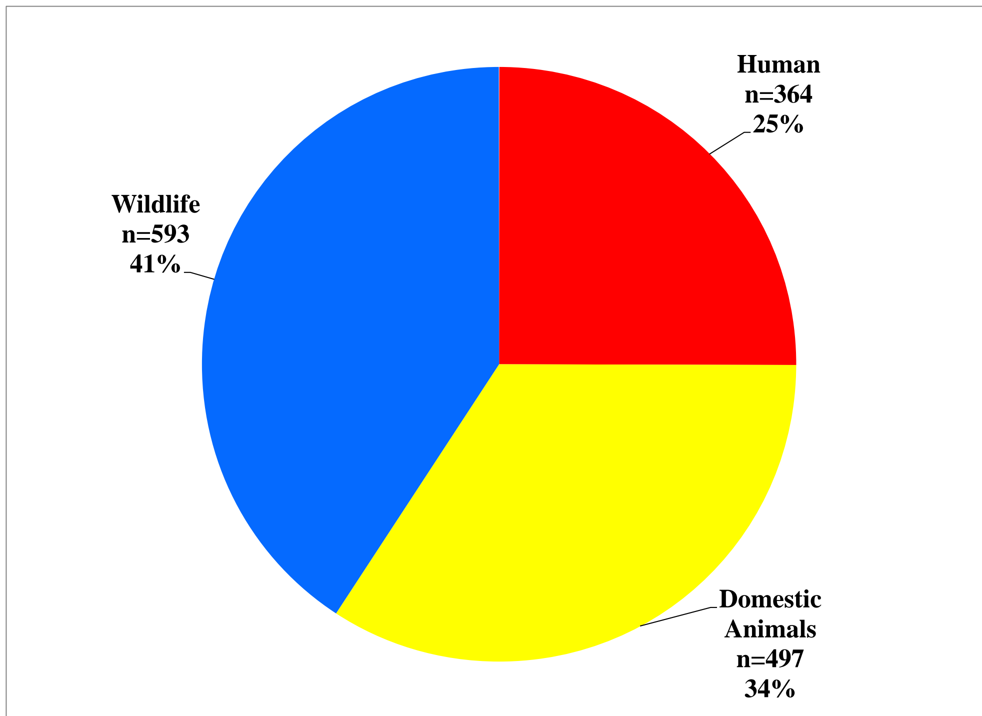


Figure 2. Texas *E. coli* BST Library (ver. 1-13) library composition by 3-way split of source classes (1,454 isolates from 1,291 different fecal source samples).

To determine the effects of removing these "cosmopolitan" strains on identification and the percentage of water isolates left unidentified, the cross-watershed validated Texas *E. coli* BST Library (ver. 1-13) was used to re-identify the isolates from the recent Leon and Lampasas watershed studies. Identification of Leon and Lampasas watershed isolates was previously performed using the Texas *E. coli* BST Library (ver. 10-12) which had only unvalidated singleton source isolates removed. Re-identification with the ver. 1-13 library revealed no significant differences in identifications at either the 3-way or 7-way split of source classes, with only a maximum of 3% difference per host class. The number of isolates left unidentified also remained low at 12%. Although removing cosmopolitan strains from the library does not seem to make a significant difference in these cases, it should be noted that strain removal most affected the pets and other non-avian livestock source classes. These source classes were not significant contributors of fecal pollution in the Leon or Lampasas watersheds, so the insignificant effect on water isolate identifications was not surprising. Further study of the occurrence of cosmopolitan strains in fecal and water samples is needed. Characterization of this group of strains could be used to develop a new "cosmopolitan" source class for the Texas *E. coli* BST Library.

Source Isolate Needs

Since a major concern in Texas is the effect of feral hogs on watershed quality, one of the greatest needs identified for the further development of the Texas *E. coli* BST Library was to increase the number of feral hog isolates included in the library. With respect to feral hogs, version 1-10 of the library was based on the screening of 54 feral hog fecal samples, resulting in the inclusion of 31 self-validated feral hog isolates. As of October 2012, an additional 72 feral hog fecal samples had been collected for an overall total of 126 samples. The 126 samples yielded a total of 407 archived isolates, 258 which were screened by ERIC-PCR, and 96 self-validated isolates from 81 samples included in the ver. 10-12 library. An interesting observation is that *E. coli* strains from feral hogs tend to group with wildlife source isolates and seem distinct from non-avian livestock source isolates, including domestic swine.

From the 12 completed source tracking projects, 3,342 individual source samples have been collected from 142 source subclasses. Included in these counts are 101 samples collected from 58 exotic zoo animals which are not usually used in building the identification library. Individual species were not used for the calculation of the number of source subclasses. For example, all ducks were considered as single source subclass, as opposed to separating species (e.g., Mallard vs. Muscovy). Therefore, the inventory library is even more diverse with respect to animal species represented.

An inventory of the known source samples and isolates collected to build the Texas *E. coli* BST Library is provided as Appendix A. The inventory is presented as a 7-way split of source classes followed by source subclasses. The 7-way and 3-way split source compositions of the current Texas *E. coli* BST Library (ver. 1-13) are presented as Figures 1 and 2, respectively.

After review of the inventory data, several observations were made which may help guide future efforts to address weaknesses in library composition and which may lead to more efficient and

realistic expectations for sample collection and processing. General observations by source class are presented below.

Human source class
The majority of the human samples come from wastewater treatment plants, although they are not always designated as raw sewage or effluent. One concern is that leaking septic tanks may be a more likely contamination contributor than regulated treatment plants. Although it is not likely that we will be able to distinguish whether *E. coli* came from sewage or septage, additional septage samples may increase the diversity of fingerprints in the library and therefore aid source identification. More research is needed to characterize septage and sewage *E. coli* populations.

Cattle source class
A large number of cattle samples have been collected over the course of the 12 watershed projects. It is interesting to note that the percentage of unique fingerprint patterns (singletons) is lowest for the cattle source class compared to other source classes. This may be a reflection of large number of cattle source samples and adequate representation of *E. coli* strains in the library. Dairy and beef cattle, as well as cattle with undesignated production purposes, are represented in the library. It does not appear possible to confidently distinguish between *E. coli* strains derived from dairy and beef cattle.

Avian livestock source class
A total of 207 samples from 7 subclasses of avian livestock have been collected for the library. The majority of avian livestock samples collected for the library were from chickens, with small numbers of samples from domestic farm ducks and geese. More than 25% of the avian livestock *E. coli* fingerprint patterns in the library are unique, suggesting more sampling is needed to better represent the diversity of *E. coli* from this source class.

Other non-avian livestock source class
A total of 335 samples from 9 subclasses of non-avian livestock (besides cattle) have been collected for the library. It should be noted that self-validation screenings of local watershed libraries using a 7-way split of source classes resulted in significant removal (>50%) of horse, sheep, and domestic swine isolates due to cross identification with cattle and occasionally pets. However, these isolates would correctly be identified as derived from domestic animals using a 3-way split of source classes. Therefore, for studies which do not need cattle as a separate source class, we may wish to reconsider these isolates for inclusion in the library.

Pet source class
A total of 306 samples from 12 subclasses of pets have been collected for the library. The greatest numbers of samples were obtained from cats and dogs, which also represent the greatest potential impact on watershed quality. It should be noted that less than 60% of the cat samples collected tested positive for *E. coli*. Dogs and the pet source class overall seem to have a high occurrence (approximately 25%) of cosmopolitan *E. coli* strains, even when using the 3-way split of source classes. Greater than 40% of the pet *E. coli* fingerprint patterns remaining in the library are unique, suggesting more sampling is needed to better represent the diversity of *E. coli* from this source class. However, in most watersheds, dogs and cats are not likely significant

contributors to fecal pollution. Collection of additional pet samples, in particular from dogs and cats, should be addressed on a study-by-study basis.

Avian wildlife source class
A total of 609 samples from 31 subclasses of avian wildlife have been collected for the library. An important finding is that only 65% of the collected samples tested positive for *E. coli*, which is the lowest percentage for any source class. For fecal samples obtained from swallows, only 33 of 96 (34%) samples tested positive for *E. coli*. Further research is needed to determine if swallow feces contain *E. coli* strains which are not detectable using our isolation protocols and if they are contributing to *E. coli* levels in watersheds. Avian fecal samples appear to dry quickly, and dried samples often test negative for *E. coli*. A specialized protocol for collection and isolation of *E. coli* from avian samples may need to be developed. Despite the challenges of working with avian fecal samples, avian wildlife has often been identified as a major contributor to fecal contamination in watersheds studies across the state. Our BST studies to date have been for inland watersheds, and the current library has very few *E. coli* from waterfowl such as gulls, herons, egrets, kingfishers, and pelicans. Sample collection from these subclasses of avian wildlife should be the priority for watershed projects which include appropriate wildlife habitat.

Non-avian wildlife source class
A total of 659 samples from 23 subclasses of non-avian wildlife, including feral hogs, have been collected for the library. Coyote, deer, feral hog, and raccoon source samples represent nearly two-thirds of all collected samples and 70% of the samples that tested positive for *E. coli*. Smaller animals, such as mice, rats, and nutria, are poorly represented with only 7 samples collected, 9 isolates archived, and only two *E. coli* isolates from rats included in the ver. 1-13 library. Non-avian wildlife has also frequently been identified as a major contributor to fecal contamination in watersheds studies across the state. The impact of small non-avian wildlife on fecal loading and water quality is uncertain. Consequently, efforts at expanding the library with *E. coli* isolates from these subclasses of non-avian wildlife should be a priority.

Bats
No fecal samples from bats have been collected in any of the previous BST studies. Bats often roost under bridges and therefore may contribute to fecal loading and bacterial counts in watersheds. It is recommended that bat fecal samples be collected in future watershed studies where sanitary surveys identify them as potential contributors.

Geographic and Temporal Stability of the Library

Concerns for library-dependent bacterial source tracking include the geographic and temporal stability of the library. These are particularly relevant to our statewide library that has been developed with *E. coli* isolates from multiple watersheds studies collected over a number of years. As of October 2012, our group has completed twelve watershed projects across the state of Texas. Four of these have covered sections of the Leon River from 2004 to 2012. This provides us with an opportunity to compare the *E. coli* isolates collected from this waterbody and watershed over time. Recently completed studies of the Leon and Lampasas watersheds were done concurrently by the same teams and may serve as a temporal constant for evaluation of

geographical differences. Assessment of geographic and temporal stability of the library is included as an objective in TSSWCB Project 13-50, the Statewide Bacterial Source Tracking Program for FYs 2013-2014.

Environmental *E. coli* strains

Recent research by several different groups of investigators suggests there are innocuous (non-pathogenic) strains of *E. coli* that persist in soil and water environments that appear genetically distinct from other *E. coli* strains. Some water *E. coli* isolates from our previous studies which could not be identified using the Texas *E. coli* BST Library may represent environmentally adapted strains. If so, fingerprints of these isolates may represent another "source class" for the Texas *E. coli* BST Library that could aid in developing best management practices and watershed protection plans. Assessment of strain diversity in the library is included as an objective in TSSWCB Project 13-50, the Statewide Bacterial Source Tracking Program for FYs 2013-2014.

Summary and Conclusions

The Texas *E. coli* BST Library has been significantly expanded and includes isolates from 12 watershed projects from across the state. The culture collection (archive) currently contains 8,812 isolates from 2,519 individual known fecal source samples representing humans and over 130 animal source subclasses. In addition, thousands of *E. coli* isolates from water samples archived from the completed BST projects may be useful in identifying environmentally adapted strains. Three steps have been identified to improve the library: 1) focus future source sample collection on small wildlife (e.g., rats, mice, nutria, rabbits, and bats) which are not well represented in the library to fill the gaps revealed by this inventory, 2) continue to implement new library refinements that increase source class specificity and allow a better understanding of cosmopolitan strains, and 3) use the recently completed studies of the Leon and Lampasas watersheds to better understand the temporal and geographical stability of the library.

EVOLUTION OF BACTERIAL SOURCE TRACKING IN TEXAS

Use of BST, generally defined in this report as the suite of methods designed to identify sources of fecal contamination in environmental waters, in Texas dates back to the early-2000s with both coastal and mainland projects. Many of these projects have been funded through the TCEQ and the TSSWCB using Clean Water Act §319(h) funds and employed BST in the TMDL process to complement monitoring and modeling activities. These projects initially spanned a wide range of methods, but ultimately have led to a more standardized set of procedures utilized across the state today that include both library-dependent methods used to develop the Texas *E. coli* BST library, as well as, library-independent methods utilizing *Bacteroidales*-PCR based approaches. There are a few BST projects that have been performed in Texas which are not described. However, the emphasis of this section is to describe BST studies which have contributed to the development of the Texas *E. coli* BST Library, as well as provide examples of other BST projects. Figure 3 illustrates the location of each watershed where known source fecal sample collection has been focused.

Copano Bay

An early BST study in Texas was conducted in the Copano Bay watershed north of Corpus Christi. The watershed was complex in that it had both tidal and non-tidal segments included on the 303(d) list for elevated bacterial levels which caused it to not meet its designated uses for contact recreation and oyster harvesting. The Texas General Land Office funded a study by Dr. Joanna Mott at Texas A&M Corpus Christi (TAMU-CC) to use BST applications to identify sources of fecal contamination in the watershed. Library-dependent antibiotic resistance analysis (ARA) was used as well as pulsed-field gel electrophoresis (PFGE) which served to confirm the ARA results (Mott and Smith, 2011; USEPA, 2005a). Even though fecal coliforms served as the current water quality standard for oyster waters, the more specific *E. coli* was used as the indicator of choice for this project. Water samples were processed using EPA Method 1103.1 onto mTEC media. The isolates were confirmed for culture purity onto Rainbow® agar plates and confirmed as *E. coli* using the MicroLog™ Microbial Identification System (Biolog, 1999). Known-source samples were collected using fresh fecal samples or swabs from freshly killed animals. Known-source *E. coli* isolates were collected and confirmed with the same procedures as the water isolates. ARA analysis was performed using a standard Kirby-Bauer disk diffusion method with a panel of 20 antibiotics. Zones of inhibition were scored using BIOMIC® software and discriminant analysis was used to differentiate the various source results and calculate the average rates of classification (ARCC). A portion of the known-source isolates were then analyzed by PFGE as a secondary confirmation of the source classifications. The isolates' DNA extracts were digested with restriction enzyme, *Not*I, separated using a CHEF-DRI III Gel Electrophoresis Unit (Bio-Rad, Hercules, CA) and processed using the Quantity One program (Bio-Rad, Hercules, CA).

The ARA library was constructed and analyzed using discriminatory analysis to classify resistance data from 1,058 total isolates into several different source category groupings. This study divided the source classifications into 2-way (human and non-human), 4-way (human, cattle, horse, and wildlife) and 6-way classifications (human, cattle, horse, duck, gull, and other wildlife). Ducks and gulls were included as separate categories based on recommendations from

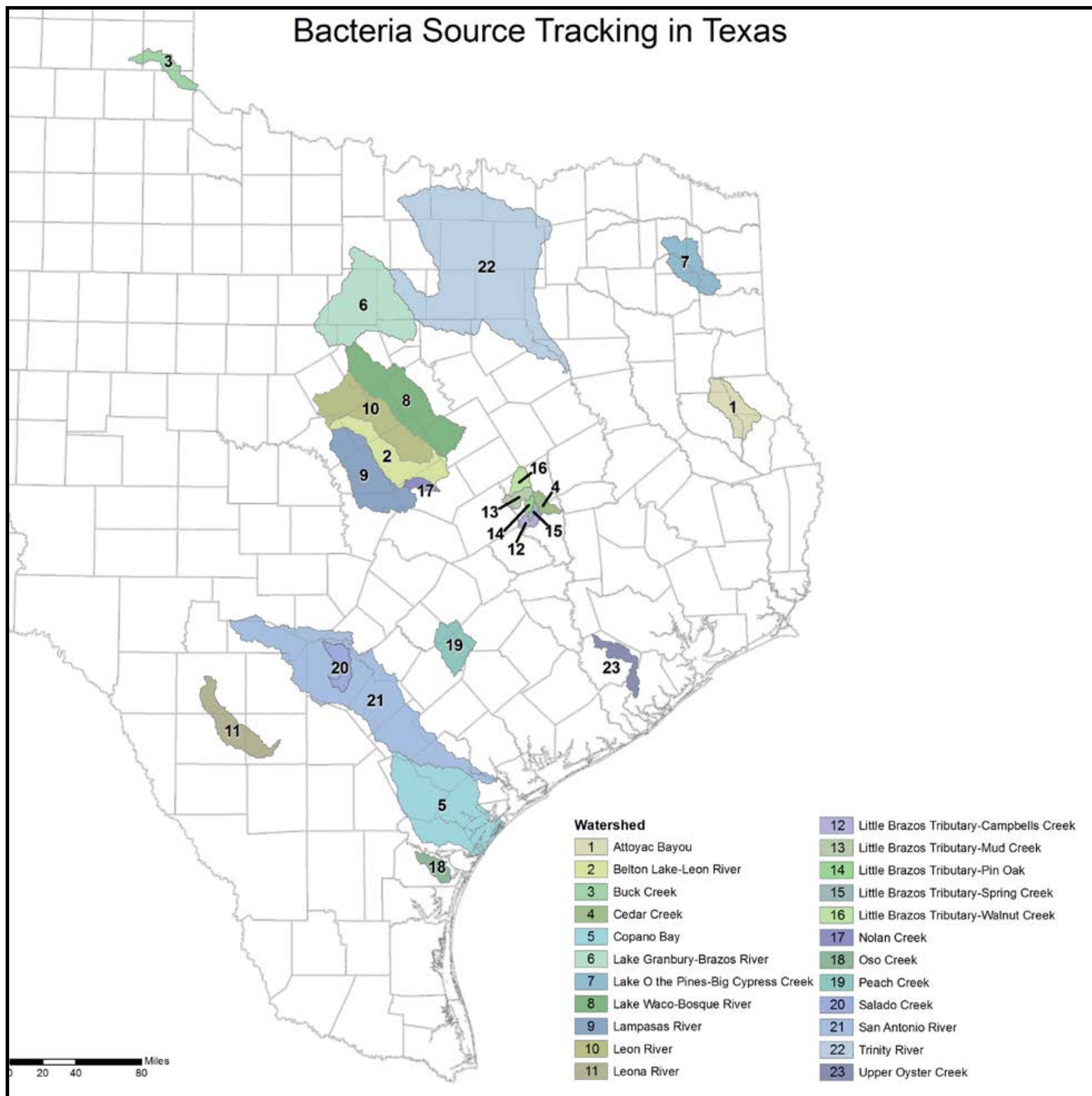


Figure 3. Location of known source fecal sample collection efforts supporting the development of the Texas *E. coli* BST Library.

a sanitary survey of the potential sources in the area. The ARCC of cross-validated isolates were highest in the 2-way split at 71%, 62% with the 4-way split, and 56% in the 6-way split. In the 6-way split, the largest misclassifications occurred when cattle and horses were identified as sewage. The 6-way split was ultimately used to classify the isolates in the study. In total, 2,811 isolates from across the study were fingerprinted from 14 stations across 8 sampling events including normal flow and storm flow events. Some stations had considerably more isolates per site due to a lack of *E. coli* during sampling events at several of the locations. But overall, using a 6-way split, 22% of the isolates were characterized as being from human contributions, 35%

from horses, 21% from ducks, 20% from cattle, and 1% from gulls and wildlife. As possibly expected, there were considerable differences in source category allocations across the 14 sample locations. The PFGE analysis was meant as a confirmation of source categorization and 1,077 isolates were also fingerprinted and source identifications characterized based on cluster analysis of the PFGE banding patterns. Overall, 63% of the human isolates, 27% of the cattle isolates, 18% of the horse isolates, and 9% of the duck isolates were classified to the same source category using PFGE and ARA. The results showed some promise as the human source category showed greater congruence between the two methods, but many questions remained as to the size and scope of the source library needed to more accurately distinguish between source contributors and even whether these methods would ultimately be able to do so. Further, some of the fecal samples from horses and ducks were collected in a different time period than the water sample collection possibly introducing confounding, temporal differences in the known-source fecal communities used in the study (Mott, 2005).

A point to note from this study was that watershed stakeholders largely disagreed with final project results. Percentages allocated to each source class were viewed as inappropriate due to their disparity with actual animal number in the watershed.

Oyster Creek

BST was conducted on the Upper Oyster Creek in the Brazos River Basin, located southwest of Houston, in 2004. The watershed includes several incorporated municipalities including Fulshear, Sugar Land, Stafford, and Missouri City. Significant hydrologic modification occurs at several locations in the watershed where water is relocated for irrigation, industrial, and public drinking supplies outside the watershed. The project was funded by the TCEQ through a contract with the Texas Institute for Applied Environmental Research (TIAER) and BST was conducted by the Institute for Environmental Health (IEH) in Seattle, Washington. A TMDL has been completed, and the watershed is currently in the implementation phase overseen by the Houston-Galveston Area Council. The project utilized library-dependent BST using ribotyping of *E. coli*. Strains of *E. coli* from both water and known-source samples were digested with two restriction enzymes, *Eco*RI and *Pvu*II, resolved by agarose gel electrophoresis, and subsequently processed using southern hybridization to create specific restriction fragment length polymorphism patterns or ribotypes. The ribotypes were scored using an alpha-numeric pattern where bands within 3 mm of each other enumerated (1, 2 or 3) and scored as a group and any band or group of bands greater than 3mm distance from another was scored as a separate entry in the code. Banding patterns that scored exactly the same code but were visually shifted were considered the same ribotype. Isolates with the same *Eco*RI and *Pvu*II ribotypes were considered to be members of the same ribogroup. Only isolates with two identical ribotypes were grouped together, and only isolates with an exact match were classified into a particular source category. Quality control was tested through a blind study of 60 isolates (20 isolates in triplicate) where the precision was 100%, all 60 of the isolates yielded the same ribotype when repeated, and 100% accurate identification occurred down to source species.

A sanitary survey characterized potential sources of contamination in the watershed and guided source selection of the 501 known-source *E. coli* isolates used to build the watershed-specific ribotype library. These isolates were included in a larger library established by IEH from

samples collected around the U.S. which was used to identify water isolates back to their source. Specific details of makeup of the entire library were not included in the technical report. Ribotypes that were not source-specific were characterized as 'transient' but were included in the library, therefore, water isolates that were considered unidentified may have been so labeled because there were no ribotypes in the library or they were not host-specific. The authors of this report classified the known sources into six major source categories (humans/sewage, livestock, mammalian wildlife, avian wildlife, pets, and unknown) but results were also presented down to a single source. The water analysis included 6 core monitoring stations and 12 different events, including runoff events, with over 120 isolates ribotyped from each core station for a total of 1,136 isolates. Overall and when analyzed by site, there was no significant difference in the source characterizations between the runoff and non-runoff events. Specific site source characterizations were similar to the overall results, and when they did differ, specific site characterizations explained the results, e.g., livestock contributions were slightly higher in the more rural portions of the watershed. Wildlife was the largest source contributor in the dataset representing 43% of the isolates, with 23% of the total from avian wildlife and 20% from mammalian wildlife. Livestock were the next largest contributor at 19% followed by humans at 14% and pets at 9% with 15% of the isolates unidentified (Hauck, 2006).

Trinity River

An urban BST project was sponsored by TCEQ in the Trinity River Basin in Dallas in 2005. The TMDL project is currently close to the implementation plan phase, but BST was conducted in the early stages of the TMDL to supplement modeling activities and was directed by the Institute for Environmental Health. The BST methods and library construction were similar to those used in the Oyster Creek project as described previously. Quality control was tested through a blind study of 30 isolates where the precision was 100%, all 30 of the isolates yielded the same ribotype when repeated, and 100% accurate identification occurred down to source species.

A sanitary survey guided investigators to collect fecal samples, isolate *E. coli* and build a known-source ribotype library from 522 watershed specific isolates. Similar to the Oyster Creek project, isolates were included in a larger library established by IEH from samples collected around the U.S. which was used to identify water isolates back to their source. Specific details of makeup of the entire library were not included in the technical report. Overall, 550 water samples were collected from 10 different stations with approximately two isolates from each water sample ribotyped for a total of 1,135 isolates. Overall, no one source category was characterized across the watershed as being a dominant contributor due to the diversity of and variability in sources of *E. coli* detected at each station. The only consistencies seen in the dataset were in dominant sources seen in each major source category. Non-waterfowl species dominated the avian wildlife, bovine and horses in livestock, rodents in mammalian wildlife, and dogs in pets (Texas Institute for Applied Environmental Research, 2006).

Assessment of Bacterial Sources Impacting Lake Waco and Belton Lake

The Lake Waco and Belton Lake study was a significant collaboration of the Texas Farm Bureau, TSSWCB, City of Waco, and Brazos River Authority with EP AREC, TAMU, TAMU-CC, and Parsons Water and Infrastructure, Inc. to assess potential sources of fecal contamination

in the watersheds after concerns were raised over possible contamination from agricultural activities in the area. This study was also designed to evaluate several promising BST methods and identify the most appropriate methods for future work in Texas.

Four library-dependent methods were evaluated including Kirby-Bauer antibiotic resistance analysis (KB-ARA), enterobacterial repetitive intergenic consensus sequence PCR (ERIC-PCR), PFGE, and automated ribotyping (RP) (Casarez et al., 2007a; USEPA, 2005a). These methods were chosen due to their previous use in other BST studies, their range of capacity to discriminate between bacterial strains, as well as cost and labor considerations. KB-ARA and PFGE had been used in previous BST applications and were established techniques used at TAMU-CC and TAMU, respectively (Lu et al., 2004). ERIC-PCR, a type of rep-PCR, was chosen as an additional method of screening based on its discriminatory capabilities and relatively inexpensive cost. The PCR amplification of adjacent ERIC elements which are variable among bacterial strains yields a number of different sized fragments that are resolved on an agarose gel creating a banding pattern or fingerprint used to differentiate different strains of *E. coli*. Manual ribotyping, as used by IEH in previous Texas BST studies, was simplified and standardized by automating the process using the DuPont Qualicon RiboPrinter Microbial Characterization system (RiboPrinting; RP). The initial investment was high and consumable costs for RP are the highest of any of the four methods, but automation and reproducibility of the data was advantageous. Further, the construction of this library was meant to stand as the foundation for a potential statewide BST library to be used in future studies and for expansion of these techniques around the state.

Several pivotal technical approaches were implemented in this study. *E. coli* was chosen as the target for library construction due to its direct link to fecal contamination and regulatory standards, as well as the availability of standardized culturing techniques designed especially for environmental water samples. Water and known-source fecal samples were processed using EPA Method 1603 on modified mTEC media. This medium is designed for its simplicity and specificity to enumerate *E. coli* using a chromogen, 5-bromo-6-chloro-3-indolyl-β-D-glucuronide to detect β-D-glucuronidase. All isolates were also streaked for culture purity onto NA-MUG media to confirm glucuronidase activity. The use of the automated DuPont system for RP enabled the use of standardized reagents with a robotic workstation to increase the reproducibility of results and thus comparability with work performed at other labs using the same methods. Also, ribotyping completed in previous studies conducted by IEH used two restriction enzymes, *Eco*RI and *Pvu*II. However, there was not a consensus regarding the best enzymes to use for BST with various projects across the U.S. using a variety of different methods. Based upon available information regarding specificity, cost, and detection sensitivity, it was decided to use a single restriction enzyme, *Hin*dIII, for RP of isolates for the Texas *E. coli* BST Library.

Library structure was a significant consideration when the project was designed. Depending on the assay, the size of the library could have a significant impact on the ability to identify sources of contamination especially if identical strains from the same source were included in the analysis, so the investigators looked to maximize the number of unique strains of *E. coli* that would be included in the analysis. Another significant hurdle in BST research was how to analyze the fingerprint data, so BioNumerics software (Applied Maths, Austin, TX) was chosen

due to its ability to process multiple fingerprint techniques as well as the ability to create composite datasets to identify methods or combinations of methods that would yield the most positive outcomes. For both ERIC-PCR and RP fingerprints, curve-based Pearson-product similarity coefficients were used to compare the banding patterns, which use both the position and the intensity of the bands to make comparisons. Finally, the unweighted pair group method with arithmetic means (UPGMA) was used to construct dendrograms to depict relationships between the isolates. Quality control strains were used in all four methods to measure the reproducibility of the methods and to determine the minimum similarity values needed to categorize the patterns as different types.

A sanitary survey helped identify major source classes of potential fecal contamination in the area and in total, 1,094 fecal samples were processed of which 813 were positive for *E. coli*. A group of 100 isolates from South Texas wildlife sources collected from a previous study by Mott at TAMU-CC was also included. In order to build a more diverse library, three isolates from each sample were fingerprinted using ERIC-PCR and any isolates that were greater than 80% similar were considered identical or 'clonal'. The similarity cutoff value was based on reproducibility of a quality control strain over time. To create a diverse fingerprint library, one to three isolates per sample were selected to be included in the library in a dynamic process of comparing their ERIC-PCR fingerprints to those already in the library. Isolates whose best match was less than 80% similar were considered unique and included in the library. Also, if the best match was to a single isolate, it was also selected to make sure that clusters of isolates had a minimum of two members. At least one isolate from each sample was included in the analysis even if the ERIC-PCR type was already present in the library to include common and abundant strains from different samples in the library, but not clonal isolates from individual samples. After ERIC-PCR screening, 883 isolates from 745 different sources were ultimately analyzed by all four BST methods and used to construct the known-source library.

In total, 11 different water monitoring stations were sampled over a 10-month period during which many of the samples did not have detectable levels of *E. coli*. At the beginning of the project it was noted that the geometric means at several of the locations tested were well below the geometric mean criterion for recreational water quality. Ultimately, 650 water samples were collected, 412 samples were positive for *E. coli*, between 1 and 5 isolates were isolated and archived per sample and 555 total water isolates were analyzed using all four BST methods.

Quality control was tested using a blind analysis of 30 test isolates (10 triplicates). All four of the methods were able to identify the replicate isolates (100% precision). Method accuracy ranged from 70-90% accuracy in identifying each isolate back to a single library isolate and correct source class. KB-ARA, which was analyzed using both a best match and discriminant analysis approach, was less successful using discriminant analysis with 40% precision (identification of the replicates) and 50% for method and source identification. When combined, the four method composite data set identified all of the replicates and identified all of the strains and their sources correctly for 100% precision and accuracy. Jackknife analysis was used to evaluate RCC for the library using a best match approach. Isolates whose best match was below the minimum similarity cutoff for each method were considered unidentified. The minimum similarity cutoffs values were 85% for ERIC-PCR, RP, and KB-ARA, 70% for PFGE, and 70% for the composite dataset (all 4 methods combined). These cutoffs were based on replication of a

quality control strain over time in each of the methods. The 70% cutoff was used in the combined dataset to allow for variation in the individual methods and to strike a balance between increasing RCC and the proportion of isolates left unidentified. The composite dataset equally weighed the four methods and gave an average of the similarities of all of the methods. Isolates were identified back to a single isolate, but were classified back only to one of seven major source classes which included domestic sewage, pets, cattle, other livestock avian, other livestock non-avian, wildlife avian, and wildlife non-avian. PFGE had the highest RCC (95%) but also left the highest percentage of isolates unidentified across all of the methods. So, even though the ability to classify isolates back to a source category was high, a very large percentage of isolates using only this method could not be identified using a library of this size.

The composite dataset (Table 3) using all four methods had RCC's ranging from 22% in the other livestock avian category up to 83% in the domestic sewage. This dataset also was able to identify a larger percentage of the isolates (81%) than any single method. A cross-validation study was conducted to identify specific source classes that might be implicated in cross-identifications. Overall, the largest percentage of the identified isolates from each source category was to the correct source and was 3 to 7 times greater than would be identified by random chance.

Table 3. Jackknife analysis rates of correct classification (%) for individual and four method composite BST methods and the 883 isolate library (from Casarez et al., 2007a).

Source class	Random*	PFGE	ERIC-PCR	RiboPrinting	KB-ARA using best matching	KB-ARA using discriminant analysis	Four-method composite data set
Domestic sewage	26	95 (35)†	64 (29)	60 (2)	60 (3)	43 (0)	83 (15)
Pet	5	54 (69)	19 (38)	17 (0)	17 (0)	27 (0)	33 (14)
Cattle	17	80 (60)	46 (13)	43 (4)	41 (0)	27 (0)	61 (3)
Other livestock avian	3	0 (60)	10 (20)	0 (0)	8 (4)	36 (0)	22 (8)
Other livestock nonavian	10	55 (55)	30 (20)	16 (3)	24 (2)	10 (0)	40 (8)
Wildlife avian	16	74 (52)	37 (27)	40 (6)	35 (2)	41 (0)	48 (11)
Wildlife nonavian	24	84 (49)	55 (17)	47 (5)	60 (0)	44 (0)	66 (11)

*Random is the percentage of isolates from each source class represented in the library of 883 source isolates.
†The number in parentheses is the percentage of isolates for that source class left unidentified after Jackknife analyses (<85% similarity for ERIC, RiboPrinting and KB-ARA best match, <70% similarity for PFGE and the composite data set). There is not an unidentified classification or a minimum similarity in discriminant analysis.
ERIC-PCR, enterobacterial repetitive intergenic consensus sequence polymerase chain reaction; KB-ARA, Kirby-Bauer antibiotic resistance analysis; PFGE, pulsed-field gel electrophoresis; BST, bacterial source tracking.

Using the composite dataset for source identifications, there was a wide variety of source contributors at each watershed site with no single source category being the dominant contributor. However, wildlife, cattle, and domestic sewage were generally the major sources of contamination. At the Lake Waco and North Bosque sites, wildlife (23% wild birds and 17% non-avian wildlife) were characterized as the source for estimated 40% of the isolates followed by 29% from livestock, 17% from human sewage, and 3% from pets. Source category could not be identified for 11% of the isolates from these locations. The combined Belton Lake and Leon River isolates indicated that 49% of the isolates originated from wildlife, 28% from birds and 21% from non-avian wildlife, followed by 32% from livestock, 11% from human sewage, and

3% from pets. Source category could not be identified for 5% of the isolates from these sites. Previous speculation in the watershed had implicated cattle and other livestock sources to be the major contributors of *E. coli* in the watershed, but at each of the 11 stations at either lake, cattle were attributed to less than or equal to 25% at any one station. Of particular note in the study, a site with high sewage contributions (27%) was detected in Lake Waco at a site near the dam which is close to a drinking water treatment intake.

The results of this study highlighted the discriminatory capabilities of the four methods with KB-ARA being the least discriminatory, followed by ribotyping and ERIC-PCR and finally PFGE having the highest discriminatory power (Figure 4). A sanitary survey of the watershed should help determine the level of discriminatory capability needed in a BST method in a particular watershed. The scope of this project allowed for a comparison of the various methods and how well they corresponded. This was especially important as it most likely would not be feasible in either cost or time to use all four methods used in this study. Congruence measurements showed that a two-method composite of ERIC-PCR and RP (ERIC-RP) was 90.7% to the 4-way method composite dataset (Figure 4). This project was instrumental in providing a foundation for future BST work in Texas. The results were reported to the Texas Farm Bureau and the TSSWCB (Dean et al, 2006) and resulted in two peer-reviewed publications (Casarez et al., 2007a; Casarez et al., 2007b).

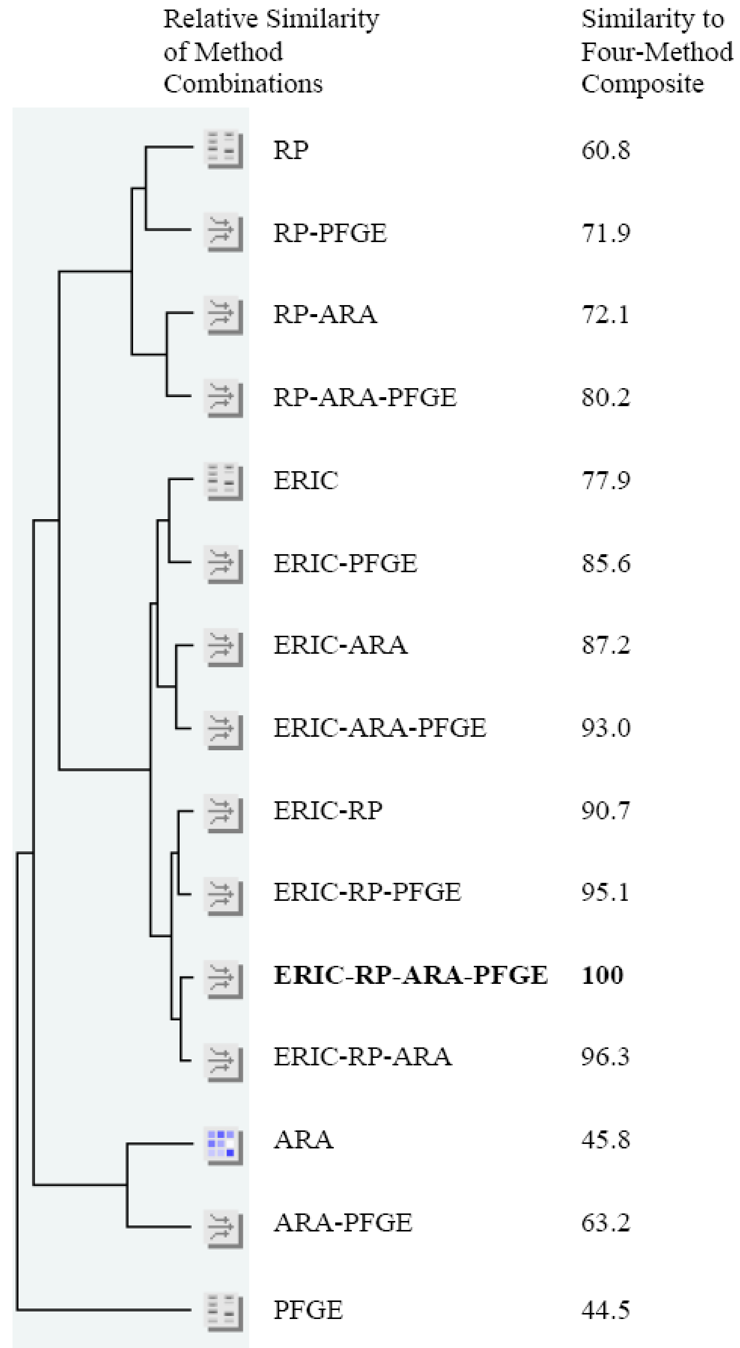


Figure 4. Congruence of individual BST methods and composite datasets (Labels include ERIC: ERIC-PCR, RP; RiboPrinting, ARA: KB-ARA, ERIC-RP-ARA-PFGE: four method composite dataset) (Reprinted with permission from Casarez et al., 2007a).

Upper and Lower San Antonio River, Salado Creek, Peach Creek, and Leon River

Concurrent to the Lake Waco and Belton Lake study, BST efforts conducted by EP AREC and sponsored by the TCEQ, were also underway in the San Antonio River, Salado Creek, Peach Creek, and Leon River watersheds. These watersheds were ultimately broken into four separate projects all of which are in various stages of TMDL or WPP development. Based on results from the TSSWCB Lake Waco and Belton Lake study, ERIC-PCR and RP were used as BST methods to assess sources of fecal contamination in the watersheds. Like in the previous study, a large library of known-source samples was collected following a sanitary survey in the area which for this project also included zoo animals in addition to the potential sources in the previously described seven-way source classification. The zoo isolates were only used in the source identifications of the water isolates in the watershed in which the zoo animals were identified as potential sources. Samples were collected and processed using the same standardized techniques as described from the Lake Waco study. Further, the library was built using the same initial ERIC-PCR screening technique to limit including clonal isolates from the same sample, in an effort to build a diverse library.

In total, 797 known-source samples were positive for *E. coli*, 2,152 isolates were screened using ERIC-PCR and excluding the 100 zoo isolates, a total of 847 known-source samples were analyzed with ERIC-RP and included in the 'TCEQ library'. In an effort to increase the diversity of *E. coli* used for source identifications and to assess the geographical stability of the library, these isolates were combined with 980 isolates from the concurrent TSSWCB Lake Waco project and used to identify water isolates from the watershed. The individual TCEQ library, as well as the combined TCEQ+TSSWCB library was used to identify source classifications from water isolates obtained from the watershed. The same best-match approach was used with an 80% similarity cutoff to classify sources into eight categories including domestic sewage, pets, cattle, other livestock avian, other livestock non-avian, avian wildlife, non-avian wildlife, and zoo animals. Jackknife analysis was used again to evaluate library fitness. The RCC at the 7- or 8-way split for this study was lower than with the 4 method composite dataset from the TSSWCB study, but were still 2 to 5 times higher than random. The combined TCEQ+TSSWCB library had the least number of unidentified isolates and greatly increased the RCC for non-avian wildlife. The RCC ranged from 9% in the zoo isolates up to 66% in the domestic sewage. The zoo isolates had very low RCC as they tended to match more closely to wildlife and domestic sewage. Further, there was some cross-identification of cattle and non-avian livestock and the power to separate domesticated animals into three separate classes was considered a limitation of the constructed library.

The watershed analyzed was geographically very large, so the results were shown for each individual sampling site with the number of isolates at each site ranging from less than 100 to over 300. In total, 1008 water isolates were ERIC-RP fingerprinted and identified. Wildlife was characterized as being a significant source contributor to the watershed as a whole with 39% of the total isolates identified as either avian or non-avian wildlife. Animal agriculture including cattle and poultry operations had been suspected of being major contributors in the watershed, and even though cattle and other livestock were identified in the watershed, they were not the leading source found. From a human health perspective, it was problematic that domestic sewage was found to be the source of 15% of the total isolates, ranging from 11% up to 18% at some locations (Di Giovanni et al., 2006).

Bacterial Total Maximum Daily Load Task Force Report

In 2006, the TCEQ and TSSWCB tasked a group of water research professionals along with expert advisors with evaluating current trends in TMDL developments around the U.S., including modeling as well as BST approaches, and to recommend appropriate cost and time-effective approaches to developing TMDLs in Texas and suggesting the potential research objectives needed to reduce ambiguity in bacterial assessment across the state. The BST portion of the report was coordinated by Drs. George Di Giovanni (EP AREC) and Joanna Mott (TAMU-CC). The report described methods being used for BST efforts in the state including KB-ARA, ERIC-PCR, ribotyping, PFGE, and carbon substrate utilization (CSU). The report highlighted the results from the Lake Waco and Belton Lake study described previously. Several key data interpretations and expectations were given in the report. First, identification of fecal pollution sources down to the level of individual species is desired, but not scientifically justified with current BST methods. Rather, the RCC values are much more acceptable when categorizing the potential fecal contaminants into 3-way split categories including human, domesticated livestock, and non-domestic animals. Further, significant numbers of water isolates would need to be characterized from each particular sampling station over a sustained period of time in order to identify specific sources of pollution at individual sites, and for that reason, library-dependent BST project results have been previously reported on a watershed basis due to these cost and time constraints. Library-dependent results were semi-quantitative at best and did not readily fit into quantitative modeling TMDL approaches. Lastly, sampling site selection was impressed as a significant consideration factor as BST results only identify potential sources of contamination and not their entry pathway. The report stressed that no one method should be relied upon solely for any BST effort and that the choice of methods should be made based on a combination of needed discrimination in the watershed as well as cost and expertise constraints.

Newly developed library-independent BST methods targeting source-specific *Bacteroidales* molecular markers were recommended as an alternative to more time consuming library-dependent analyses, with several caveats. These methods have the potential to be an effective and rapid estimation of recent contamination events without the need for library construction, but specificity issues across source classes, a lack of validated marker sets, and a lack of direct link to regulatory water quality standards are problematic. TCEQ and TSSWCB projects discussed previously built a strong foundation for library-dependent work in the state of Texas and the task force recommended expanding upon that foundation in several ways: (1) expand the current TCEQ+TSSWCB known-source libraries with additional watersheds from around the state, (2) refine the library to increase BST accuracy, (3) expand BST infrastructure, including personnel and equipment, to increase BST capabilities, and (4) continue to utilize and refine BST SOPs used across the state to maximize potential BST applications. Research and development needs were also included in the task force report and included: 1) further refinement of reasonable expectations for BST results, 2) investigating the expansion of library-independent methods and their most appropriate incorporation with TMDL activities, 3) investigating the geographic and temporal stability of a statewide BST library, and 4) further refinement of appropriate sampling schemes to yield the most statistically sound BST results.

Finally, the task force recommended a three-tiered approach to bacterial TMDL development. BST would be used in the early stages of TMDL development using mainly library-independent

methods in addition to limited library-dependent applications if initial models were not sufficient in characterizing the watershed and identifying attainable bacterial load reductions in a Tier 2 analysis. For *Bacteroidales* gene screens, 50 to 100 samples would be tested using a presence/absence approach for human, ruminant, horse, and swine sources. Additionally, if funds were available, 50 to 100 water isolates would be characterized using the statewide library to asses sources of contamination with additional known-source samples collected from the watershed if less than 80% of the water isolates could not be identified. In a Tier 3 analysis, generally used for I-Plans or particularly controversial watersheds where a very detailed characterization of sources of fecal contamination is warranted, 100-200 water isolates from approximately 40 separate sampling locations will be characterized using the statewide library that has been supplemented with isolates from at least 100 various known-source fecal samples from the watershed. This task report was published in 2007 by the Texas Water Resource Institute (TWRI) and was meant to serve as a guidepost for future TSSWCB and TCEQ TMDL activities (Jones et al., 2007).

Lake Granbury

Lake Granbury, on the Brazos River in north Central Texas, is a vital water resource for the region, providing drinking water for approximately 250,000 residents. When monitoring sites detected consistently high levels of *E. coli* in man-made coves in 2007, the Lake Granbury BST project, sponsored by the Brazos River Authority (BRA), became the first BST project to implement the recommendations in the Bacteria TMDL Task Force Report and use the TCEQ-TSSWCB known source library with only a small supplement of known source isolates from the local watershed. In total, 94 known source fecal samples were collected from wildlife, domestic septage/sewage, pets, and livestock. These sources were specifically targeted since the most likely pollution sources for Lake Granbury were believed to be domestic sewage due to the high density of housing in the coves dependent on aging septic systems, as well as runoff from nearby agricultural range and croplands, wildlife, and possibly pet waste. Indeed, fecal pollution modeling of Lake Granbury performed by consultants indicated that 99% of the *E. coli* in the Port Ridglea East cove water was derived from leaking septic systems (Brazos River Authority, 2008). BST was done by EP AREC (Di Giovanni et al. 2009; Farnleitner et al. 2011) as part of a preliminary assessment using several BST tools, including ERIC-PCR and RiboPrinting composite fingerprinting using the TCEQ-TSSWCB library, PCR detection of *Bacteroidales*, and two other library-independent PCR methods for *Methanobrevibacter smithii* (Ufnar et al. 2006) and human polyomavirus (McQuaig et al. 2009) that specifically detect human fecal pollution.

Water samples were collected monthly for 6 months (October 2007-April 2008), mostly representing routine, low-flow conditions from five sites on Lake Granbury: Lake Granbury at Highway 377 (11861); Sky Harbor (18015); Waters Edge (18018); Indian Harbor (20215); and Port Ridglea East (18038). For all sampling locations combined, a total of 233 water *E. coli* isolates and 36 *Bacteroidales* water samples were analyzed following EP AREC protocols. Known source fecal samples were used to evaluate the distribution of *Bacteroidales* host-specific markers in the watersheds and for *E. coli* library development. After de-cloning, a total of 80 *E. coli* isolates from 59 fecal samples were added to supplement the state library.

E. coli and *Bacteroidales* BST results suggested that the Lake Granbury Port Ridglea East site was impacted primarily by animal-derived (wildlife) fecal pollution. These findings were surprising since it was assumed that the site was highly impacted by human fecal pollution from leaking septic systems. 45% of the *E. coli* isolates were identified as originating from wildlife sources, while only 15% were identified originating from human sources. Further, none of the six monthly water samples were positive for the *Bacteroidales* human marker, while all were positive for the ruminant marker.

As a follow-up, more intensive sampling was performed at this site. Two sets of samples were collected approximately two weeks apart from ten different locations within the Port Ridglea East cove for *Bacteroidales* analysis and *E. coli* enumeration. In addition, *Methanobrevibacter smithii* and human polyomavirus PCR was performed for the detection of human source pollution. *Bacteroidales* PCR results again revealed the presence of animal fecal pollution and the absence of human source pollution, despite some of the samples having *E. coli* levels up to 2400 CFU/100 ml. In addition, only one of the follow-up water samples (and its field duplicate) tested positive for human polyomavirus, and none tested positive for human *M. smithii*. This shows consistency between the library-independent BST results and the results from the library-dependent methods using the statewide library supplemented with local watershed isolates.

The results from the multiple BST approaches did not agree with the pollution source modeling, however. The model did not account for subsurface flow which may bring in wildlife-derived pollution from adjacent undeveloped land. Still, the lack of a significant human pollution signature in the BST results was puzzling. While there is anecdotal evidence of backed-up septic systems during periods of high lake water levels, a septic tracer dye study of 44 systems in the area found only two minor leaks with on-ground pooling in 2 locations, and no dye observed entering the canals, indicating that the systems tested were not significantly contributing to the high bacteria levels detected at the time of the study (Brazos River Authority, 2010). Stakeholders decided to move ahead with plans to construct a sanitary sewer system. Since human fecal pollution may contain a variety of human pathogens, this is a prudent course of action with respect to protecting human health. Nevertheless, it will be interesting to see if the *E. coli* levels in the cove show a significant decrease after the improvements.

Increased Analytical Infrastructure and Development of a Statewide BST Library

This project was funded by the TSSWCB in 2008 to increase the statewide capabilities to conduct BST research and refine, validate, and expand the statewide BST library. The project was led by Dr. Di Giovanni and his team at EP AREC, but was aimed at expanding BST personnel and expertise to AgriLife SCSC with Dr. Gentry. Using methods refined in the previous TSSWCB and TCEQ BST projects, known-source ERIC-RP fingerprints from six BST studies were combined into the Texas *E. coli* BST Library. The studies, previously discussed in this report, include (1) Lake Waco and Belton Lake, (2) Upper and Lower San Antonio River, Salado Creek, and Peach Creek, (3) Lake Granbury sponsored by Brazos River Authority, (4) Upper Oyster Creek, and (5) Trinity River.

Except for the Oyster Creek and Trinity River isolates, all of the known-source samples were collected and processed using the same procedures. These SOPs were implemented with the

Lake Waco/Belton Lake as well as San Antonio River studies. In short, sanitary surveys and collaboration with stakeholders helped guide the collection of as many unique known source samples as possible. Fresh fecal samples including WWTP raw influent were collected and processed on mTEC media generally following EPA Method 1603 which was also used to process the water samples. Isolates were streaked for culture purity and to confirm glucuronidase activity on NA-MUG and stored in glycerol stocks at -80°C for long-term applications. The isolates from the previous TCEQ projects at Oyster Creek and Trinity Creek were originally isolated using clinical media which did not screen for glucuronidase activity and were considered less likely to produce library matches to the isolates obtained using the regulatory media. These isolates were secondarily screened for glucuronidase activity on NA-MUG media and only positive cultures were used in library construction.

Known-source isolates were screened via ERIC-PCR and isolates were chosen to build each local library as described above from the Lake Waco/Belton Lake study. ERIC-PCR was used to screen isolates from individual samples to identify clonal or identical isolates using an 80% similarity cutoff, but at least one isolate from each individual sample was included in the library even if the ERIC-PCR type was already represented from another sample. This approach sought to increase the diversity of the library while including abundant or common strains from various animals. Isolates chosen for local library construction were then RP fingerprinted and composite datasets were created using BioNumerics software. The first version of the dynamic statewide library combined isolates from the TCEQ and TSSWCB projects and consisted of 1,793 isolates from 1,505 fecal samples. For identification purposes, the known-source samples were divided into seven management-related groups including domestic sewage, pets, cattle, other avian livestock, other non-avian livestock, avian wildlife, and non-avian wildlife. The library was made up of 26% human isolates, 10% pets, 15% cattle, 6% avian livestock, 11% non-avian livestock, 15% avian wildlife, and 17% non-avian wildlife. Separating the domesticated animals into separate categories (cattle, pets, avian and non-avian livestock) as seen in the TCEQ study greatly decreased the accuracy of source classifications, so a less specific 3-way split was proposed to include humans, domesticated animals, and wildlife which increased the accuracy of source characterizations while maintaining general management delineations needed to develop best management practices for remediation.

As first mentioned in the TCEQ study, it was expected that some *E. coli* isolates were not source specific. Using jackknife analysis, isolates were removed if their ERIC-RP composite best matches within their own library, were not to their specific 7-way source category. Isolates with a best match of less than 80% were considered unidentified but were left in the library as they were unique, diversified the library, and could be helpful in identifying water isolates. This resulting self-validated library included 996 isolates from 884 different known-source samples. Self-validation greatly increased the RCC which averaged 86% for the 7-way split. Cross-identifications were greatest within similar source categories like cattle and other livestock, further solidifying the future use of a less specific 3-way split. Individual watershed local libraries were used as challenge isolates against the self-validated library to see how well they could identify those isolates, and they performed roughly equally. The results highlighted the need to self-validate the source specificity of any isolates ultimately being included in the library as large portions of the challenge isolates from Lake Granbury, Oyster Creek and Trinity River were cosmopolitan isolates and thus incorrectly identified in a Jackknife analysis to their correct

7-way split source. Ultimately, a statewide self-validated library was compiled using all of the aforementioned known-source isolates and named the Texas *E. coli* BST Library. The library is dynamic in nature as each new iteration and addition of validated isolates changes the overall makeup of the library as well as the RCC for the various source categories. Average RCC for ver. 8-10 for a 3-way split was 86% (Table 4).

Table 4. Texas *E. coli* BST Library (ver. 8-10) composition and rates of correct classification (RCCs).

Source Class	Number of Isolates	Number of Samples	Library Composition and Expected Random Rate of Correct Classification	Calculated Rate of Correct Classification (RCC)	Left Unidentified (unique patterns)
Human	374	327	29%	89%	19%
Livestock and Pets	462	424	35%	83%	20%
Wildlife	473	434	36%	86%	18%
Overall	**1,309**	**1,185**	**RARCC* = 33%**	**ARCC = 86%**	**19%**

*RARCC, expected random average rate of correct classification

The creation of the library and further refinement yielded important results and raised important concerns and needs moving forward with the development and enhancement of a statewide library. Even though the ARCC were similar with the composite library versus the local libraries, use of the larger data set yielded less unidentified isolates and the composite dataset could identify isolates from discrete watersheds. The results suggested that local watershed isolates were needed to supplement the larger statewide library to aid in representing any geographic variability seen in the watershed. Using a large, diverse statewide library but including small local watershed additions serves as a significant cost savings for conducting library-dependent BST studies rather than having to build a large database for each watershed. Another concern has been managing the potential number of isolates that may need to be screened to ultimately gain 7-way split source specific isolates, especially for sources that seem to be dominated by cosmopolitan isolates, such as coyotes. Ongoing library refinement challenges identified in this report include: 1) the identification and use of cosmopolitan isolates for library construction, 2) temporal and geographical effects on the fitness of the library over time, and 3) the need to expand the library with *E. coli* from underrepresented sources and watersheds from around the state (Di Giovanni et al., 2010).

Other BST Projects in Texas

In addition to the BST projects and methods already highlighted in this report, a handful of other source tracking methods have also been used across Texas. In the Rio Grande River valley,

PFGE was used to compare *E. coli* from source irrigation water and sediments (Lu et al., 2004). The results showed that there was significant diversity among the 50 fingerprinted isolates and persistent strains could be seen, but laboratory studies of PFGE patterns over time in these surviving persistent strains exhibited a range of genetic relatedness from >95% to <83%. It was concluded that the extreme resolving power of PFGE may prove prohibitive for BST efforts as an extremely large library would be needed to identify source isolates. Moussa and Massengale (2008) utilized a combination of carbon substrate utilization profiles and ARA to build a 600 member BST library to identify sources of contamination in the South Bosque River. The authors reported RCC upwards of 85% for up to a six-way source classification split. Graves et al. (2009) used carbon substrate utilization patterns with the BIOLOG system to characterize *Enterococcus* strains in both fresh and dry cattle, horse and sheep manure. The authors reported some shift in population in dry versus fresh manure, but overall the relative proportion of the two dominant strains of *Enterococcus* was similar among all three animal groups in dry and fresh manure.

Wagner (2011) evaluated the ability of the AllBac (general bacteria)/BoBac (ruminant-specific) marker sets to accurately assess the percentage of bovine-associated fecal loading, as well as their correlation to regulatory fecal indicator bacteria, in small watersheds used for grazing livestock. Neither AllBac nor BoBac concentrations were correlated with grazing management or annual stocking rate, but were significantly correlated with percentage of runoff events occurring during either stocked or de-stocked sites indicating utility of this marker to detect recent fecal contamination events. An additional significant finding from this study was that the correlation between AllBac and BoBac gene copy numbers and fecal indicators was greatest in the watershed where fecal samples had been collected to produce the standard curves for analysis thus suggesting potential geographical variability in the creation of these standards.

Ongoing BST Projects in Texas

Several WPPs sponsored by TSSWCB from across the state have incorporated BST, along with modeling efforts, in order to identify sources of bacterial contamination in watersheds. Generally, the BST involved in most of the projects was and continues to be conducted based on the recommendations from the Bacterial Task Force and include both library-dependent and -independent methods. The majority of watershed specific projects include screening for presence/absence of the source specific *Bacteroidales* markers for humans, ruminant, horses, and swine in approximately 250 water samples. In addition, approximately 100 *E. coli* isolates from sampling sites across the watershed are characterized using the Texas *E. coli* BST library which for most projects also included the addition of known-source isolates from the local watershed. Results are presented to stakeholders during stakeholder meetings during the watershed planning phases as well as and technical reports submitted the TSSWCB. The *Bacteroidales* analysis results have been reported as a percentage of positive hits in bar graph format from the overall watershed as well as individual sampling sites to identify possible ‘hot spots’ of contamination in the watershed that require more in-depth examination. Generally, fewer water isolates are identified per sampling site and these results are presented in total across the watershed.

A majority of the studies also include the addition of known-source samples from the local watershed and a breakdown of total processed samples, number of isolates fingerprinted using

ERIC-PCR, isolates ultimately ERIC-RP fingerprinted, and finally the number of isolates which are self-validated and added to the Texas *E. coli* BST Library. Generally, at the conclusion of a project or projects, the library is updated to a new version with the inclusion of the local isolates. New library metrics, including rates of correct classification, are calculated and included in the results (as seen in Table 2). To date, results have only been reported to stakeholders at the 3-way classification level as confidence in the separation of isolates into these categories is greater than for the more stringent 6-way source classification. A critical goal of the expansion of the Texas *E. coli* BST Library continues to be adding known-source samples from underrepresented or low-confidence groups of animals, including pets and poultry, in order to improve the ability to delineate these sources of contamination. The following is a brief overview of relevant findings from three recent BST projects.

Buck Creek WPP

Buck Creek is a small creek located Panhandle of Texas in the Red River Basin that was impaired with elevated levels of *E. coli*. A total of 31 known source isolates from 28 samples were added to the expanding state library (ver. 08-09, 1,172 isolates from 1,044 samples). In total, 79 water samples were analyzed using *Bacteroidales* PCR and 426 water isolates were fingerprinted using ERIC-RP and classified using the Texas *E. coli* BST library. Overall, the majority of bacteria present at Buck Creek were derived from wildlife sources (including feral hogs) (Figure 5). A hot spot of potential human contamination was also identified using this approach and was investigated by stakeholders. The Buck Creek watershed has recently been highlighted by the USEPA as a success story since the stream was removed from the 303(d) list in 2010 due in large part to extensive efforts of the local stakeholders to input best management practices to reduce nonpoint pollution in the watershed.

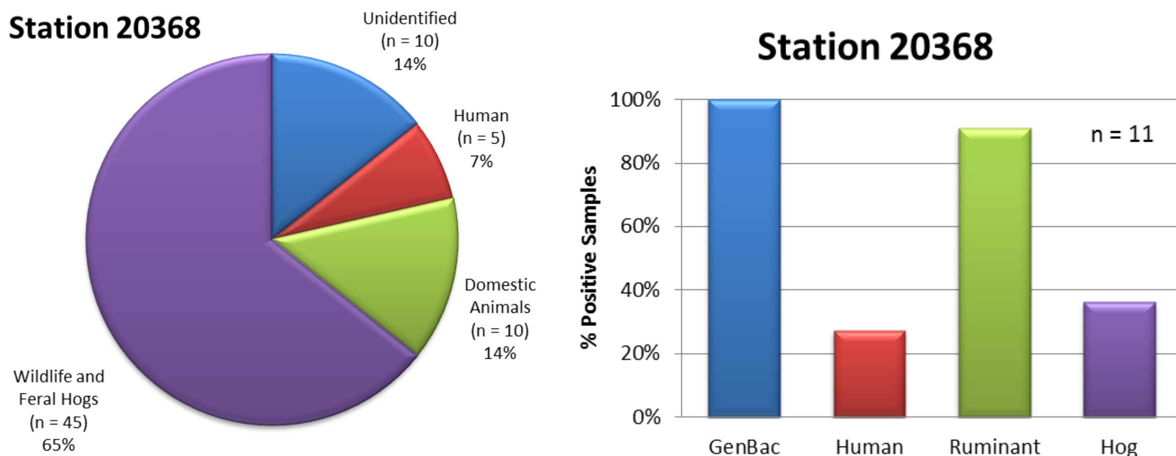


Figure 5. BST results from Buck Creek station 20368. Identification of water isolates (pie chart) using a 3-way split for source classification and *Bacteroidales* PCR maker occurrence (bar chart).

Little Brazos River BST

The Little Brazos River tributaries studied are located in the Little Brazos River Basin in Robertson County, Texas. The hog marker (71% of positive hits) was the most commonly detected marker across the entire study of 259 samples followed by the ruminant marker (39% of positive hits) (Figure 6). Using a 3-way split, from a total of 69 water isolates classified using the Texas *E. coli* BST library (ver. 12-09, 1,196 isolates from 1,068 samples), 59% were classified as originating from wildlife with smaller proportions originating from domestic animal (19%) and human sources (6%).

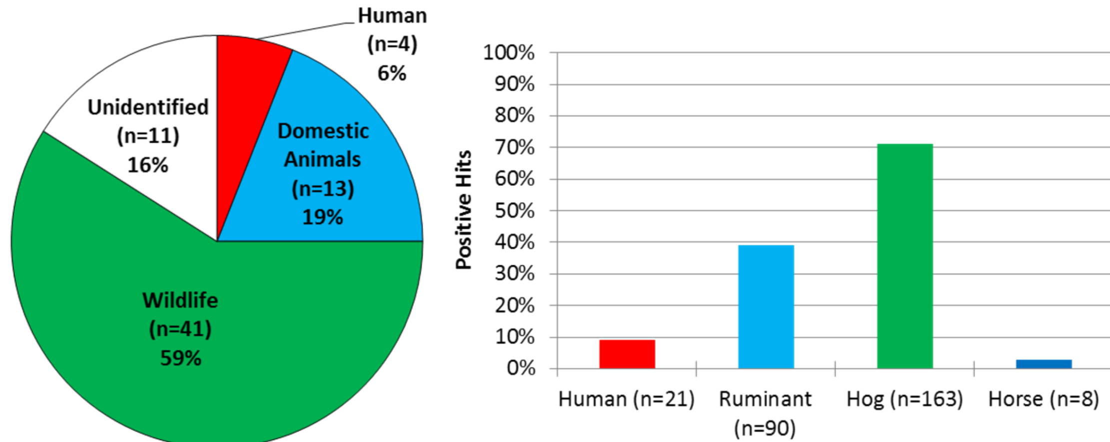


Figure 6. BST results from Little Brazos River. Identification of water isolates (pie chart) from all creek sites using a 3-way split for source identifications (n=69) and *Bacteroidales* PCR marker occurrence (n=259) for human, ruminant, hog, and horse markers.

Big Cypress Creek Modeling and BST

Big Cypress Creek and its tributaries are located in the Cypress Creek Basin in northeastern Texas and encompasses approximately 445 square miles in Camp, Morris, Titus and Upshur Counties. A total of 28 self-validated isolates from wastewater treatment plants (6), beef cattle (1), poultry litter (7), deer (4), ducks (7) and raccoons (3) were added to the Texas *E. coli* BST Library (ver. 10-11+BigCypSV; 1335 isolates from 1201 samples). Ruminant (40% of positive hits) and hog (41% of positive hits) markers were most commonly detected across all samples 244 samples (Figure 7). A total of 101 *E. coli* isolates were classified into main source categories using ERIC-RP and the Texas *E. coli* BST library. Using a 3-way split, the majority of isolated *E. coli* were classified as originating from wildlife (42%) or livestock and pets (29%) while isolates originating humans only constituted 12% of the isolates (Figure 7).

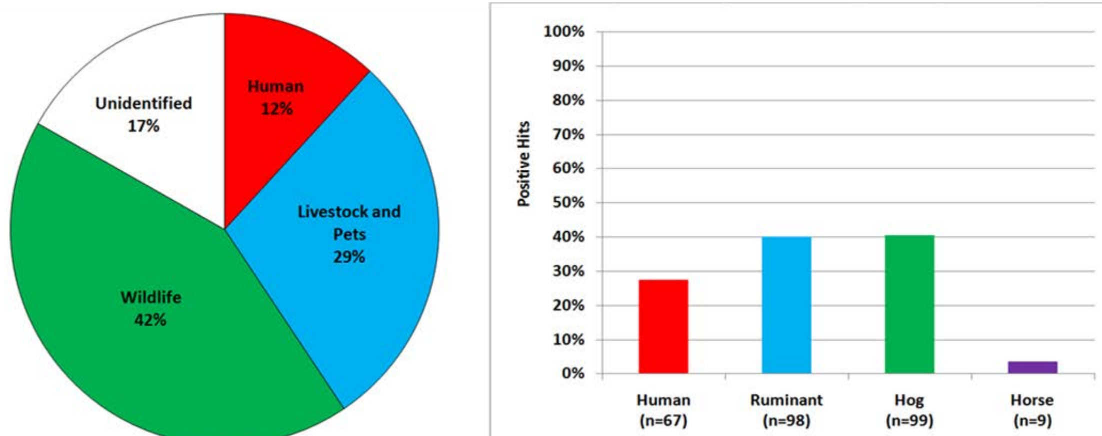


Figure 7. BST results from Big Cypress. Identification of water isolates (pie chart) from all creek sites using a three way-split for source identifications (n=101) and *Bacteroidales* PCR marker occurrence (n=244) for human, ruminant, hog, and horse markers.

Other Projects

Several other projects which followed the same general approach for BST include Attoyac Bayou, Leon and Lampasas Rivers, and the Leona River.

Conclusions

Texas has been a leader in the use of BST as part of a toolbox approach in the development of TMDLs and WPPs. To date, BST results have been met with mixed review from stakeholders and governmental agencies; often with cause. Unlike modeling efforts, current methodologies tend to be more qualitative than quantitative as BST can identify relative sources of fecal contamination but the ability to resolve source contributions across an entire watershed, much less a particular sampling site, down to quantitative percentages of fecal contamination are not yet scientifically available. Many researchers abandoned library-dependent BST when library-independent markers began being developed in hopes of short-cutting the need for extensive library development and considerable concerns over library performance. But molecular marker-based approaches have come under great scrutiny due to a lack of sensitivity and specificity and limited availability of markers for many animal species. The approach taken with BST in Texas is to use BST tools as a means of providing lines of evidence toward understanding fecal contamination in a watershed. As the Texas *E. coli* BST Library is expanded and library-independent methods are improved it will be important to keep a strong pulse on new and emerging technologies to shape future BST efforts.

The continued development of BST approaches and including recent known-source isolate additions from the Attoyac, Little Brazos River, Big Cypress, Leon and Lampasas, and Leona watersheds to the Texas *E. coli* BST Library. These additions, and subsequent explorations of the library, have the potential to help answer long-standing questions about method performance, especially geographical and temporal stability over time and enhance our ability to more specifically identify sources.

IMPLICATIONS OF *E. COLI* ISOLATION METHOD ON LIBRARY-DEPENDENT BST

Introduction and Rationale

The Texas *E. coli* BST library has been constructed with *E. coli* that have been enumerated and isolated primarily using USEPA Method 1603 (USEPA, 2005b). Many standard methods are widely accepted and utilized to enumerate *E. coli* and are often used in combination or interchangeably for enumeration purposes. USEPA Method 1603 (mTEC) and USEPA Method 1604 (MI) (USEPA, 2004) are chromogenic, membrane filtration methods while Colilert® is a defined substrate technology in a most-probable-number (MPN) format. All three methods rely on end-point screening for enzymes specific to the groups of interest. EPA Method 1604 utilizes MI media and can simultaneously detect and enumerate both total coliforms and *E. coli*. The medium utilizes two enzyme substrates, fluorogen 4-methylumbelliferyl-β-D-galactopyranoside (MUGal), and chromogen indoxyl-β-D-glucuronide (IBDG), to detect the enzyme β-galactosidase produced by total coliforms and β-glucuronidase produced by *E. coli*. EPA Method 1603 is used to detect and enumerate *E. coli* only and utilizes a modified mTEC medium containing the chromogen, 5-bromo-6-chloro-3-indolyl-β-D-glucuronide, to also detect β-D-glucuronidase. Colilert® on the other hand, is a MPN technique which utilizes a defined substrate medium utilizing the chromogen ortho-nitrophenyl-β-D-galactopyranoside (ONPG) to detect β-galactosidase from total coliforms as well as a fluorogen 4-methylumbelliferyl-β-D-glucuronide (MUG) to detect β-glucuronidase to also enumerate *E. coli*. Enumeration with these methods has been shown to be statistically comparable (Hamilton et al., 2005), but specific effects on the community composition of detected *E. coli* have not been evaluated. Even though all three methods utilize the same end-point enzyme to enumerate *E. coli*, different media compositions and growth conditions as well as growth platforms (MPN versus membrane filtration) may cause considerable variation in the *E. coli* communities isolated.

In light of the increased application of BST across the state, the validity of isolates processed using methods other than the USEPA Method 1603 for use in BST analysis has come into question. The current approach of UTSPH EP and AgriLife SCSC utilizes only isolates processed using the USEPA 1603 method; however, the majority of water quality labs across Texas do not currently utilize this method. Potential cost savings, the ability to utilize more samples due to holding time constraints and the potential to advance the science of BST all warrant the evaluation of other methods such as Colilert® and USEPA Method 1604 to produce statistically similar types and counts of *E. coli* isolates as those produced using USEPA 1603 Method. The objective of this study was to evaluate differences in *E. coli* community composition across three standard water quality assessments including EPA Method 1603, EPA Method 1604, and Colilert® to ultimately determine their impact on BST library performance.

Methods

Six different watersheds from across south, central and eastern Texas were sampled in this study including Big Iron Ore Creek (TCEQ Sampling Station 20844) in the Attoyac Bayou near Nacogdoches, Campbells Creek (TCEQ Sampling Station 16395) in the Little Brazos River watershed near Hearne, Moody Creek on the Welder Wildlife Refuge near Sinton, Plum Creek

(TCEQ Sampling Station 12640) watershed near Lockhart, White Oak Bayou (TCEQ Sampling Station 11387) in Houston, and Burton Creek (TCEQ Sampling Station 11783) in College Station. These sampling sites represent both rural and agricultural watersheds as well as urban and suburban areas. Each water sample was collected and processed using all three methods - USEPA Method 1603 (mTEC), USEPA Method 1604 (MI), and Colilert® - per method and/or manufacturer's instructions as well as UTSPH EP BST SOPs. Colilert® quanti-trays were enumerated after 24 hours and positive wells were combined and processed using USEPA Method 1603 via dilution and filter plating for isolation of *E. coli*. Samples were processed in triplicate and five isolates per replicate were isolated onto EC-MUG media to confirm culture purity as well as a secondary screen for β-glucuronidase enzyme activity, for a total of 15 isolates per media type at each of 6 sites. Further, all isolates were secondarily verified as *E. coli* using the verification protocol in EPA Method 1603.

Isolates were then fingerprinted using the ERIC-RP UTSPH EP BST protocols. Both ERIC-PCR and RP were performed as previously described by Casarez et al. (2007). *E. coli* isolates were first DNA fingerprinted using ERIC-PCR (Versalovic et al., 1991). Following ERIC-PCR analysis, *E. coli* isolates were riboprinted using the automated DuPont Qualicon RiboPrinter® system and the restriction enzyme *Hind*III. Analysis of composite ERIC-RP DNA fingerprints was performed using Applied Maths BioNumerics software. To identify potential sources, genetic fingerprints of *E. coli* from the water samples were compared to fingerprints of known-source *E. coli* isolates in the Texas *E. coli* BST Library (ver. 3-12; consisting of fingerprint patterns from 1,459 *E. coli* isolates from 1,285 different human and animal samples). The ERIC-RP composite patterns were compared to the library using a best match approach and an 80% similarity cutoff (Casarez et al., 2007a). If a water isolate was not at least 80% similar to a library isolate, it was considered to be unidentified. Although fingerprint profiles are considered a match to a single entry, identification is to the host source class, and not to the individual animal represented by the best match. Water isolates were identified to human, domesticated animals (including livestock and pets), and wildlife (3-way split).

Analysis of variance was used to determine significant differences between enumeration means using Sigma Plot 11.0 at a significance cutoff p-value of 0.05. *E. coli* diversity was further characterized using Applied Maths BioNumerics software using the same ERIC-RP composite fingerprint, maintaining the 80% similarity cutoff values, to calculate number of different patterns types across the media types and sites as well as to calculate Shannon-Wiener and Simpson's diversity indices.

Results

Media Type and E. coli Concentration

There were no clear trends in enumeration values across sites or media types (Table 5). Generally speaking, the mTEC and MI concentrations were more similar than the Colilert® concentrations. Even though previous studies, as well as monitoring directives from TCEQ, have indicated any of these methods could be used for monitoring activities, the counts on this particular study were not consistent across these three methods at the six sites evaluated.

ERIC-RP Diversity

E. coli diversity index values, Simpson's and Shannon-Wiener, indicated the mTEC and MI communities were the most diverse followed by Colilert® (Table 6). Moody Creek exhibited the lowest diversity of any of the sites followed Plum Creek and White Oak Bayou. A large portion of the genotypes, especially in mTEC and MI, contained only unique isolates.

Table 5: Summary of the mean *E. coli* concentrations from six sites across the three media types. (CFU or MPN/100mls +/- SE)

	mTEC	MI	Colilert®
Big Iron Ore Creek	277 ± 23[a]	370 ± 6[a]	628 ± 59[b]
Burton Creek	2,900 ± 153[a]	3,900 ± 116[b]	2,846 ± 133[a]
Campbells Creek	9,900 ± 751[a]	9,567 ± 406[a]	14,221 ± 733[b]
Moody Creek	221 ± 8[a]	239 ± 11[a]	149 ± 23[b]
Plum Creek	550 ± 12[a]	573 ± 9[a]	776 ± 291[a]
White Oak Bayou	763 ± 54[a]	1,077 ± 42[b]	822 ± 31[a]

Different letters indicate significant differences between media type at each site, $p < 0.05$

Table 6: *E. coli* diversity estimates (ERIC-RP 80% similarity cutoff)

	mTEC		MI		Colilert®	
	Simpson's	Shannon-Wiener	Simpson's	Shannon-Wiener	Simpson's	Shannon-Wiener
Big Iron Ore Creek	93.3	2.18	91.4	1.99	75.3	1.27
Burton Creek	93.3	2.25	97.1	2.43	88.6	1.84
Campbells Creek	99.3	2.12	96.2	2.40	89.5	1.93
Moody Creek	13.3	0.25	36.2	0.63	13.3	0.25
Plum Creek	79.1	1.71	91.4	2.12	71.4	1.41
White Oak Bayou	90.5	2.03	80.0	1.62	84.8	1.90

For BST purposes, it was important to look not only at overall diversity of the *E. coli* communities isolated, but also where those communities overlapped to identify methods or combinations of method that were similar. Similarity analysis was conducted by broadly examining the communities each medium tended to select across locations. There were 70 unique genotypes among the 270 total isolates and 12 of which were seen in all three media types totaling 65% of the isolates (Figure 8). The mTEC detected the greatest number of unique genotypes, 20, but this only represented 9% of the total isolates. Of the 70 total unique genotypes, 38 or 54% were singletons with only one isolate represented (data not shown).

The fingerprint data was also analyzed by specific site with each site having 45 isolates total, 15 in each media type (Table 7). Genotypes seen in all three media ranged from only 16% at Big Iron Ore Creek to 87% at Moody Creek. Plum Creek and White Oak Bayou had high numbers

of unique genotypes, but a small number of those genotypes made up over 50% of their communities. Big Iron Ore Creek, Burton Creek, and Campbells Creek had greater numbers of unique genotypes in the media themselves and very few in common across all three media types. Moody Creek was the least diverse and appeared to select for a very simple *E. coli* community in all three of the media types with only 5 different unique genotypes, one of which represented 87% of the 45 isolates. Further, that same genotype was seen at all six sites and accounted for 36% of the overall isolates driving the large percentage of genotypes seen to overlap across all media (Figure 8).

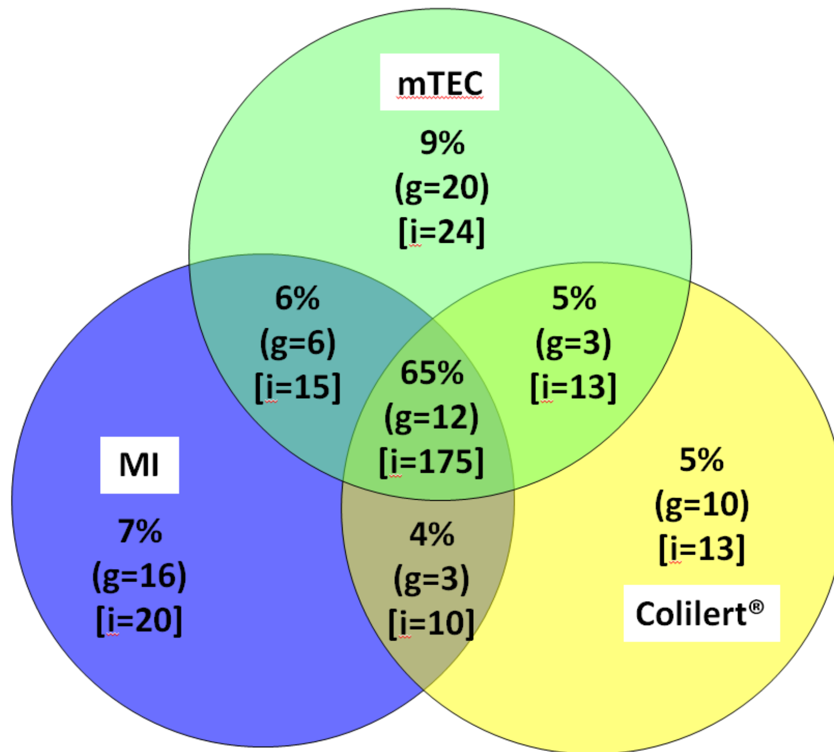


Figure 8: Overall *E. coli* genotype overlap across all three media types. Data represents the percentage of isolates with unique fingerprint patterns detected in each media type and when those isolates occurred in multiple media types. Data in parenthesis represents the number of unique genotypes (g) in each combination while the values in brackets represent the number of isolates [i] in each category.

Table 7: Overlap in *E. coli* genotypes across media types.

	Big Iron Ore Creek		**Burton Creek**		**Campbells Creek**		**Moody Creek**		**Plum Creek**		**White Oak Bayou**	
Media	Patterns (n=16)	% Total Isolates (n=45)	Patterns (n=22)	% Total Isolates (n=45)	Patterns (n=20)	% Total Isolates (n=45)	Patterns (n=5)	% Total Isolates (n=45)	Patterns (n=19)	% Total Isolates (n=45)	Patterns (n=19)	% Total Isolates (n=45)
mTEC	6	17%	7	18%	3	7%	1	2%	5	11%	5	13%
MI	4	16%	6	15%	7	16%	1	2%	8	20%	3	7%
Colilert®	1	11%	1	2%	3	9%	1	2%	4	13%	5	13%
mTEC + MI	2	16%	2	9%	1	4%	1	7%	0	0%	1	5%
mTEC + Colilert®	1	13%	1	7%	3	24%	0	0%	0	0%	0	0%
MI + Colilert®	1	11%	3	20%	1	13%	0	0%	0	0%	2	9%
mTEC + MI + Colilert®	1	16%	2	29%	2	27%	1	87%	2	56%	3	53%

Source Identifications using the Texas E. coli BST Library

Community analysis of the dataset suggested that the various media types selected for diverse *E. coli* communities, but the ultimate goal of this study was to see what impact these additional media, MI and Colilert® would have on our ability to identify *E. coli* to their major source class. When analyzing the data strictly from a media standpoint, the 3-way split shows some differences between the additional media types and mTEC, but not overwhelming changes (Figure 9). In all three media, wildlife and domesticated animals were classified as the main sources of the contamination. Domesticated animals increased 12 percentage points with the Colilert® media from 28% to 40%, while wildlife increased with the MI media from 54% to 64%. Water isolates identified to humans were less than 10% in any of the three media. The percentage of unidentified isolates was similar for all three media.

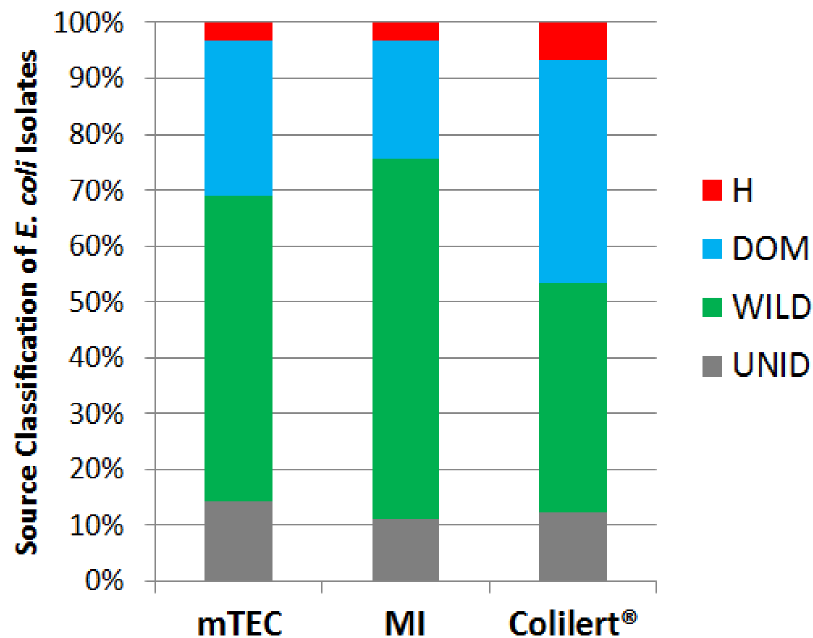


Figure 9: Identification of water isolates sorted by media type using a 3-way split for source classification (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified).

The source class identification results were also analyzed by site and are included in Figures 10 through 12. The results from the specific sites were much less consistent than when viewed broadly. Identifications at Big Iron Ore Creek had a high percentage of unidentified isolates in all three media types. There was a considerable shift to wildlife sources with the MI media but the Colilert® identifications were similar to the mTEC (Figure 10). At Burton Creek, Colilert® selected for a greater percentage of domesticated animal sources versus the mTEC or MI media. Isolates characterized as being from human sources were highest for the Colilert® isolates, but were still less than 10% of the total (Figure 10). Campbells Creek isolates from mTEC and MI were in general agreement with wildlife and domesticated animals being the primary sources, but Colilert® showed a shift toward domesticated animals from mTEC (Figure 11). Classifications at

Moody Creek were not very consistent across media types as all of the isolates from the MI media classified as originating from wildlife (Figure 11). All three media types were in greater agreement at Plum Creek than any other site, but MI conflicted with the mTEC and Colilert® with wildlife rather than domesticated animals being the main contributor (Figure 12). White Oak Bayou isolates maintained the same ranking of dominant source contributors with wildlife leading with all media, but the relative percentage of those was different (Figure 12). When comparing the classification results back to the mTEC communities, there was no consistent trend in identifying contributing source class. The human identifications were the least variable, but also accounted for a much smaller portion of the overall isolates than either wildlife or domesticated animals (Figure 13).

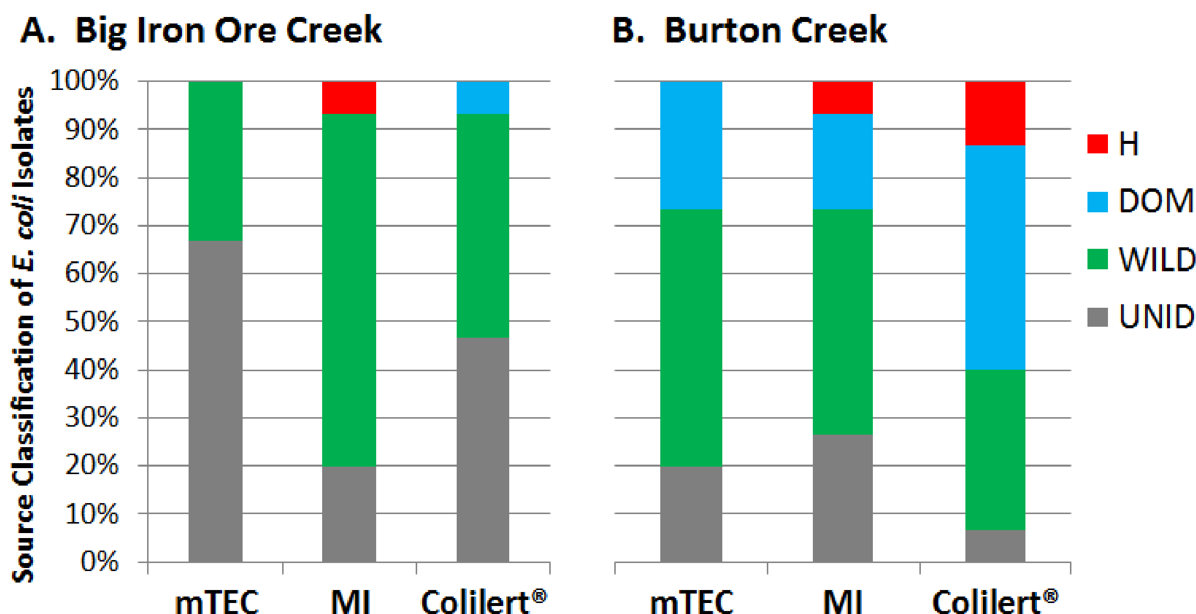


Figure 10: Identification of water isolates at Big Iron Ore Creek (A) and Burton Creek (B) using a 3-way split for source classification (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified).

Discussion and Conclusions

The Texas *E. coli* BST library is built largely of isolates obtained using USEPA Method 1603 on mTEC media. This study was performed to evaluate whether additional methods like USEPA Method 1604 on MI and Colilert® which are used by water quality managers throughout the state could be easily amendable to BST projects. The results of this study indicate that the three evaluated *E. coli* enumeration methods may select for different *E. coli* communities.

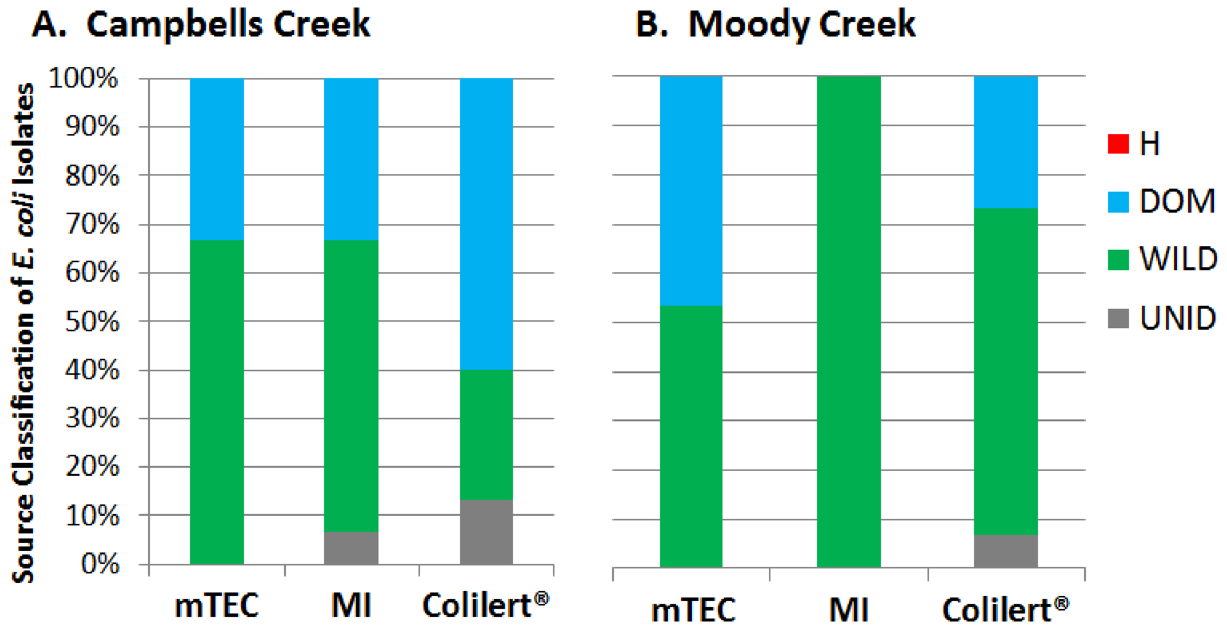


Figure 11: Identification of water isolates at Campbells Creek (A) and Moody Creek (B) using a 3-way split for source classification (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified).

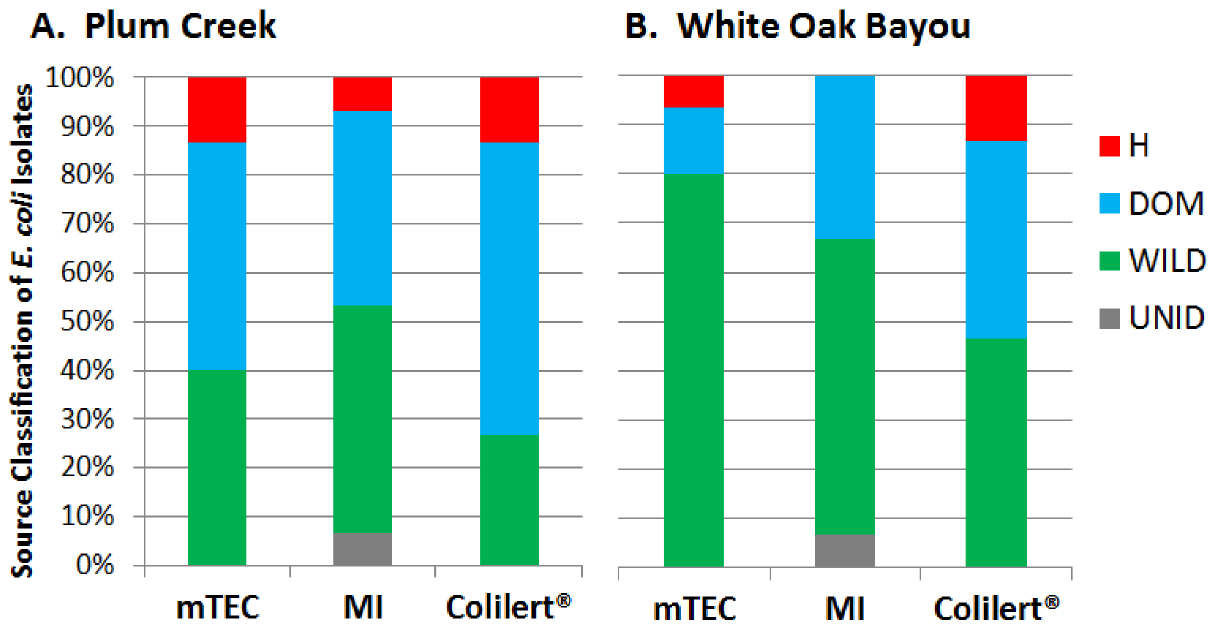


Figure 12: Identification of water isolates at Plum Creek and White Oak Bayou using a 3-way split for source classification (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified).

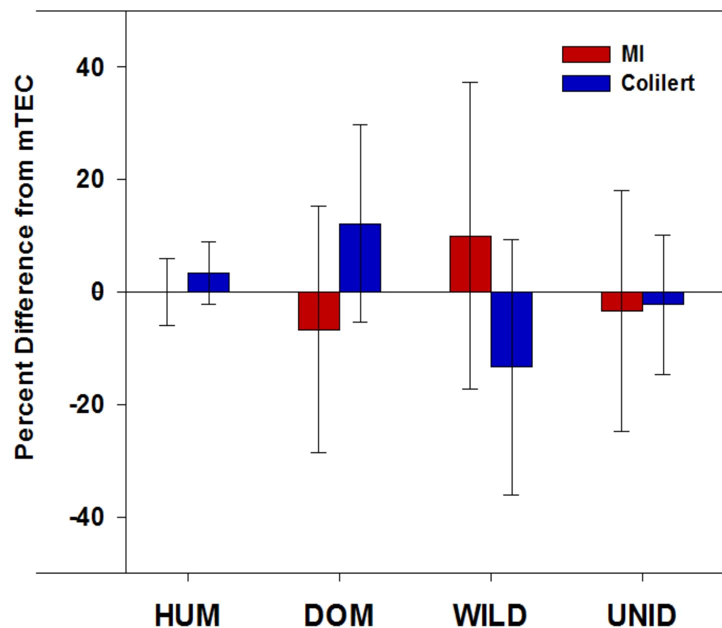


Figure 13: Comparison of MI and Colilert® BST results to mTEC across all sites using a 3-way split for source classification. Error bars represent standard error of three replicate samples (H=Human, D=Domesticated Animals, W=Wildlife, U=Unidentified).

Reasons for the general lack of *E. coli* similarity could be explained by several factors. MI and Colilert® methods are designed to enumerate both total coliforms and *E. coli* and are incubated at 35°C; whereas, mTEC media is designed to only enumerate thermotolerant *E. coli* and is incubated at 44.5°C. The increased temperature aids in adding selection pressure against non-specific taxa including other members of the coliform group. Both the media and increased temperature likely result in the selection of different *E. coli* populations. Further, even though all three of the media ultimately screen for the same enzyme to detect *E. coli*, they all three utilize different enzyme substrates and chromogens adding an additional layer of potential differentiation. Colilert® utilizes a completely different growth platform in the MPN format rather than a membrane filtration. The liquid culture versus solid medium offers a completely different growth habitat and would logically select for different populations. Additionally, the Colilert® communities underwent a secondary selection on mTEC media in order to get them isolated in pure culture which may have contributed to the lack of diversity seen in those communities.

It is also important to note several other laboratory observations when evaluating the utilization of these additional methods for BST purposes. Growth of non-target taxa on the MI medium made even enumerating the organisms difficult and isolating the *E. coli* in pure culture problematic. Also, pin-point blue-fluorescent colonies were seen on the MI media, as noted by a previous study, but were confirmed as *E. coli* through the USEPA Method 1603 confirmation protocol. Even though the strains were ultimately all confirmed as *E. coli*, the need for extra

streaking and isolations certainly increased the time and labor involved in isolating these organisms. The Colilert® method is touted as being a one-stop method with no need for secondary confirmation steps, but unlike the membrane filtration method, the product is in an MPN (liquid) format and requires an additional step to get the positive wells filtered on a solid medium in order to get physical colonies isolated. So, the use of this method with library-dependent BST, albeit, easier for basic enumerations, will ultimately mean an additional labor step for downstream processing.

The source identifications of the dataset indicated that the *E. coli* from all three media were identified to similar source categories with wildlife and domesticated animals being the dominant contributors. But, when each of the sites was evaluated individually no common trend could be seen across the various sites. There was tremendous variation across locations as to a particular selection toward one source classification or another. The communities selected for at each site across the different media types appear to have no specific site tendencies. Since the Texas *E. coli* BST Library is constructed largely from isolates processed using USEPA Method 1603 on mTEC media there was some concern that the number of unidentified isolates would be greater in both the MI and Colilert® isolates, but that was not the case as the percentage of unidentified isolates was essentially the same overall. The highest percentage of unidentified isolates from any media type came from the Big Iron Ore Creek samples (67% on mTEC). At the time of library screening, this particular area of Texas was not well represented in the statewide library. All of the other sites had lower numbers of unidentified isolates (0-13%) and, with the exception of White Oak Bayou, either had watershed specific isolates included in the library or were geographically near represented watersheds thus indicating the benefit of including local known-source samples in the Texas *E. coli* BST library.

The Moody Creek site from the Welder Wildlife Refuge was unique in this dataset as all three media types seemed to select for similar populations of *E. coli*. The majority of those isolates identified back to wildlife and domesticated animals as would be expected based upon the rural location of the site. This site also produced the lowest numbers and diversity of *E. coli* suggesting that there was a less-complex suite of *E. coli* sources at the site. This was likely at least partially responsible for the greater congruency of the source classifications across the three methods at this site as compared to the other locations that had greater numbers and diversity of *E. coli*.

The goal of any BST project is to accurately assess the main contributors of fecal contamination and their relative abundance so stakeholders can implement best management practices to improve water quality conditions. However, the results of this study indicate that using different methods to isolate *E. coli* may not provide consistent results. Even though the three media assessed in this study ultimately are designed to select the same organisms, *E. coli* (and in some cases total coliforms), the differences in the media composition and incubation temperature should give researchers pause in using them interchangeably when community characterization is a goal. This study suggests that a standardized method of enumeration and isolation may be warranted if stakeholders anticipate the possibility of using library-dependent BST.

ENUMERATION OF *E. COLI* IN KNOWN-SOURCE FECAL SAMPLES

A secondary objective of collection of known-source fecal samples for addition to the statewide Texas *E. coli* BST library was to enumerate *E. coli* in these samples. Many TMDLs and WPPs utilize modeling as a major component to assessing bacterial loads and often are forced to utilize *E. coli* fecal concentrations that are outdated or from other parts of the country. To this end, known-source fecal samples collected from various BST projects across the state were analyzed using USEPA Method 1603 and UTSPH EP BST SOPs for isolation as well as enumeration of *E. coli*. A majority of all known-source samples collected for five different BST projects were included with the exception of wastewater treatment plant samples and samples with less than sufficient material for enumeration. In total, 424 samples collected were included from five BST projects and one county where samples were also collected: Attoyac Bayou (127), Big Cypress Creek (28), Leon River (16), Leon County (20), Leona River (219), and Plum Creek (14). Results are shown in Table 8 and expressed as CFU/gram of wet fecal material. The limit of detection was 100 CFU/gram of wet fecal material and any sample yielding no cultivable *E. coli* was estimated at the level of detection.

Domesticated animals (cattle and poultry litter) as well as wildife (deer, feral hogs, and ducks) were the most intensively sampled animals due to their potential as sources of fecal contamination in Texas waterseds. Enumerated *E. coli* values ranged from the limit of detection,100, up to 10^8/gram of wet fecal material. Overall, there was a considerable range of *E. coli* values detected across all animals and watersheds. All of the samples were collected as fresh excrement and processed within three days of collection, so samples could be compared across watersheds and time. The extreme ranges of values obtained highlight the large variability of *E. coli* concentrations in fecal samples even across the same species of animal. This variabiilty should be considered when utilizing this type of data for downstream applications.

Table 8. Summary of the mean *E. coli* concentrations from known-source fecal samples (n=424).

Animal	Number	Arithmetic Mean	Geometric Mean	Std. Dev	Std. Error	Min	Max	Median	% of Samples Below Limit of Detection
		(CFU/gram wet fecal material)							
Cattle									
Cattle, Beef	59	5.4E+06	3.1E+04	2.2E+07	2.8E+06	1.0E+02	1.5E+08	3.2E+04	8%
Cattle, Beef (Feedlot)	14	2.5E+06	6.7E+05	4.7E+06	1.2E+06	2.0E+04	1.8E+07	6.5E+05	0%
Cattle, Dairy	4	1.2E+06	6.6E+05	1.2E+06	5.8E+05	1.7E+05	2.2E+06	1.2E+06	0%
Other Livestock Non-Avian									
Horse/Donkey/Mule	22	2.5E+05	3.1E+04	4.9E+05	1.0E+05	1.0E+02	1.9E+06	2.9E+04	5%
Goat	16	2.3E+06	9.4E+03	6.1E+06	1.5E+06	1.0E+02	1.9E+07	2.6E+03	13%
Sheep	13	2.1E+05	9.8E+03	5.5E+05	1.5E+05	3.0E+02	2.0E+06	4.0E+03	0%
Hog, Domesticated	1	7.1E+08	7.1E+08	7.1E+08	7.1E+08	7.1E+08	7.1E+08	7.1E+08	0%
Other Livestock Avian									
Poultry, Chicken Litter[A]	35	2.9E+06	4.9E+04	6.5E+06	1.1E+06	1.0E+02	3.3E+07	1.7E+05	26%
Yard Chickens, Ducks, Geese	16	9.5E+07	6.0E+06	1.9E+08	4.8E+07	2.0E+03	6.1E+08	9.0E+06	0%
Pets									
Dog	11	6.0E+06	5.2E+04	1.4E+07	4.3E+06	1.0E+02	4.8E+07	2.1E+05	36%
Cat	4	2.3E+05	3.6E+04	3.7E+05	1.9E+05	1.5E+03	7.8E+05	6.2E+04	0%
Rabbit	1	5.0E+02	5.0E+02	5.0E+02	5.0E+02	5.0E+02	5.0E+02	5.0E+02	0%

Wildlife Non-Avian									
Feral Hog	63	7.5E+07	2.4E+06	1.6E+08	2.1E+07	3.0E+02	8.3E+08	3.3E+06	0%
Deer	36	2.0E+07	6.8E+04	7.0E+07	1.2E+07	1.0E+02	3.9E+08	6.4E+04	8%
Coyote	19	7.3E+07	6.4E+06	1.1E+08	2.6E+07	1.7E+04	3.4E+08	7.2E+06	0%
Raccoon	11	8.0E+07	3.5E+06	1.1E+08	3.2E+07	1.0E+02	3.5E+08	6.2E+07	9%
Turkey	8	3.2E+06	4.4E+03	8.8E+06	3.1E+06	1.0E+02	2.5E+07	2.0E+03	50%
Fox	7	1.9E+08	2.0E+07	3.1E+08	1.2E+08	3.0E+05	8.5E+08	1.5E+07	0%
Squirrel	7	1.4E+06	4.0E+03	3.4E+06	1.3E+06	1.0E+02	9.0E+06	1.0E+02	57%
Armadillo	2	2.7E+04	4.0E+03	3.8E+04	2.7E+04	3.0E+02	5.4E+04	2.7E+04	0%
Bobcat	1	1.9E+06	1.9E+06	0.0E+00	0.0E+00	1.9E+06	1.9E+06	1.9E+06	0%
Wildlife Avian									
Duck	50	4.4E+06	1.6E+04	1.7E+07	2.5E+06	1.0E+02	1.2E+08	1.8E+04	32%
Small Birds	24	1.7E+06	3.2E+03	4.9E+06	1.0E+06	1.0E+02	2.1E+07	1.0E+02	54%

[A] Litter samples from wetter portions of poultry houses were specifically targeted in order to obtain *E. coli* isolates for BST. This likely resulted in higher numbers of detected *E. coli* than would have been obtained if the litter had been randomly sampled.

STANDARD OPERATING PROCEDURE UPDATES

Techniques and procedures for performing BST analysis can evolve quite rapidly. As such, periodically reviewing and updating the SOPs developed to outline the process for performing ERIC-PCR, RP and *Bacteriodales* PCR is necessary. Through this project, reviews of these SOPs were completed to ensure that the procedures described were updated as new techniques were developed and accepted for use.

As SOPs are updated with new methods and techniques, laboratory staff must also receive training on updated approaches to ensure that changes are properly implemented. When updates to SOPs were made, laboratory staff received the needed training as soon as feasible.

DEVELOPMENT AND EVALUATION OF NEW SPECIES SPECIFIC BACTERIAL MARKERS FOR LIBRARY-INDEPENDENT BST

The specificity of library-independent BST markers is critical to their proper use and interpretation of results. During our BST studies in Buck Creek and Lake Granbury we identified 16 wildlife fecal samples which cross-reacted with the human HF183 *Bacteroidales* PCR marker. Preliminary characterization of the PCR products (amplicons) from several of these cross-reacting samples and human fecal control samples was performed using high resolution melt (HRM) analysis. HRM analysis uses fluorescence-based chemistry and allows the precise determination of PCR amplicon melt temperatures (Tm) and characteristics. The temperature at which an amplicon melts depends on its DNA sequence. As a result, if there is DNA sequence variation among amplicons then different HRM curves will be observed. Preliminary HRM results for the cross-reacting samples and human fecal control samples suggested that the PCR amplicons were highly similar. Additional analysis of these samples and additional wildlife fecal samples was included in TSSWCB Project 13-50, the Statewide Bacterial Source Tracking Program for FYs 2013-2014.

In addition, there are currently limited library independent BST markers for wildlife. This severely restricts our ability to specifically track this important group. Library-independent methods to specifically track deer are hindered by the fact that the most widely accepted ruminant specific marker, CF128F, cannot distinguish between cattle and deer (Bernhard and Field, 2000). The ability to distinguish between wildlife and livestock sources is critical to developing best management practices to reduce fecal contamination from these respective sources. To this end, 454 barcoded pyrosequencing was utilized to characterize deer fecal communities in Texas in an effort to evaluate their suitability for development of a deer-specific BST marker. Deer fecal samples from Welder Wildlife Refuge were collected over two years as well as from the Leon River watershed. A parsimony test showed no significant difference between the samples collected from both years at either location. The fecal communities were dominated by two phyla, *Firmicutes* and *Bacteroidetes.* Two operational taxonomic units (OTUs) were shared across all 11 samples and were classified as *Ruminococcaceae* and *Veillonellaceae*. An additional 3 OTUs occurred in 10 of the 11 samples, two of which were also *Ruminococcaceae* and the other *Clostridiales* (Unclassified). The top GenBank hits for representative sequences from all of the OTUs were from fecal communities, except for one. The top GenBank hit for the *Veillonellaceae* OTU_36 was to feces from Springbok antelope which is a ruminant like deer and cattle. The GenBank maximum identity to all of the common and abundant OTUs was less than 100% indicating uniqueness in the database. The two strongest candidates for potential marker development are OTU_36 and OTU_4560. The *Veillonellaceae* OTU_36 has the lowest identity match (95%) and was common across all of the samples and the *Ruminococcaceae* OTU_4560 also has a low maximum identity (96%) and was found in 10 of the 11 samples. This study laid the groundwork for future research on primer design and screening these potential OTUs against non-target sources in order to verify their suitability as deer-specific BST markers.

One particular ‘wildlife’ concern for Texas is feral hogs (defined as wildlife for BST purposes). Preliminary evaluation of a modified hog PF163 *Bacteroidales* PCR protocol for the detection of pigs/swine and feral hogs was performed. We have a collection of 89 feral hog fecal samples

from 6 different areas of Texas to use for evaluation of the protocol. Preliminary results indicated that the modified hog PF163 *Bacteroidales* PCR protocol can successfully be used to detect feral hog fecal pollution. However, future research is needed to determine the prevalence of the marker in different feral hog populations around the state and to investigate its ability to differentiate between domestic and feral swine. Additional research was included in TSSWCB Project 13-50, the Statewide Bacterial Source Tracking Program for FYs 2013-2014.

Finally, although we use ERIC-PCR as one of our library-dependent BST tools to identify sources of *E. coli*, very little is known about the PCR amplicons generated with this method. In particular, it is not clear whether these amplicons contain meaningful DNA sequences which may be used as source-specific molecular markers similar to those for *Bacteroidales*. As a starting point for this work, human-specific isolates in the current Texas *E. coli* BST Library were identified for further analysis. Additional research including DNA sequence analysis of ERIC-PCR amplicons from these isolates was included in TSSWCB Project 13-50, the Statewide Bacterial Source Tracking Program for FYs 2013-2014.

EXPANDED DELIVERY OF BST EDUCATION AND OUTREACH

One of the primary objectives of this project was to expand the delivery of education and outreach on BST. Misconceptions and a general lack of information about BST have raised questions on the credibility of BST results from some. This project provided a means to allay those concerns through published materials, web-based resources and in person meetings and workshops.

State of the Science Website

To disseminate information to as broad an audience as possible, a project specific website was developed and used to house project related information such as project reports, promotional materials, and information on the 'State of the Science Conference.' This website is available to anyone at any time and provides a sound basis of information related to BST in general and in Texas specifically.

http://texasbst.tamu.edu

Promotion of BST

Promoting the use of BST across Texas and in general was also an objective of this project. To accomplish this, several avenues were employed. First, a historical look at the application of BST in Texas (discussed earlier) was developed to describe the extent of BST work conducted to date in Texas and specifically focused on comparing and contrasting methodologies and results. Second, two promotional flyers were developed and distributed to 423 individuals representing elected officials, councils of government, special interest groups, utilities, regional water planning groups, agencies and analytical laboratories. One brochure was geared toward the layman while the other toward laboratory technicians. Each provided BST information relative to the target audience. Third, a workshop entitled "Bacterial Detection and Tracking Symposium" was held at the 2012 Land Grant and Sea Grant National Water Quality Conference in Portland, Oregon. This symposium highlighted some of the recent work in the BST field and provided national exposure for work being conducted in Texas. Fourth, a project one pager was developed and distributed as appropriate to agencies, special interest groups, practitioners and others as appropriate. Additionally, conference calls with agency representatives and academia were held to discuss advancements and application of BST as appropriate; however, these calls were met with limited attendance on all fronts. Appendix B includes copies of the flyers and project one-pager.

State of the Science Conference

As is illustrated throughout this report, the science of BST is constantly evolving. Widely distributing this information is often difficult and rarely does it trickle down to agency personnel who are tasked with implementing and also creating the backbone of water quality management policy. To remedy this information transfer deficiency, the '2012 Bacterial Source Tracking – State of the Science Conference' was held as a platform to directly engage people with a vested interest in BST application and utility.

Advertisements for the conference were sent out through focused media outlets including the Oklahoma State University Water Programs Listserv, USEPA's NPSINFO Listserv, the American Society for Microbiology Listserv, the Houston-Galveston Area Council Listserv, the Soil Science Society of America Listserv, TSSWCB's *Conservation News* e-newsletter, TWRI's *Conservation Matters* e-newsletter, TCEQ's *News from the Texas TMDL Program* e-newsletter, Texas A&M AgriLife's media outlet *AgriLife Today,* as well as Texas A&M University System's website and newswire. Using these avenues, academia involved in BST analysis; state, federal, and regional agency personnel; elected officials; and other interested persons were invited to attend the conference.

Prior to the conference, conference organizers compiled a list of websites, presentations, documents, and publications and sent them to registered conference participants as a source of background information on BST. These materials included general information on BST and detection techniques, overviews, advantages and disadvantages, applications and case studies.

On February 28th and 29th, nearly 120 participants from 13 states arrived at the T Bar M Resort and Conference Center in New Braunfels, Texas for a day and a half of presentations from experts in the field as well as agency representative who provided their agency's view of BST as a tool for water quality management planning. Content for the conference focused on two primary content areas:

- demonstrate how BST can be used as a tool to aid stakeholders and agencies in assessing fecal pollution, developing TMDLs and WPPs, and in solving water pollution issues
- describe how the state of BST science, methods, application and confidence has evolved recently and highlight lingering deficiencies that need continued advancements

A call for posters was also announced and provided students an opportunity to give informal presentations on their research during networking breaks. The conference planning committee selected seven poster abstracts to be presented at the conference. The abstracts for each poster are included in section 4 of the Conference Proceedings which is available online at: http://twri.tamu.edu/reports/2012/tr427.pdf.

The conference proceedings provide copies of all presentations given at the conference, poster abstracts, a conference participant list, the conference primer materials and speaker biographies.

LITERATURE CITED

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

Brazos River Authority (2008). Lake Granbury Water Quality Modeling In: http://www.brazos.org/gbWPP/12-3-2008_LG_Modeling.pdf Retrieved January 21 2010.

Brazos River Authority and Espey Consultants Inc. (2010) Lake Granbury Watershed Protection Plan Final Report.2010.

Casarez, E.A., S.D. Pillai, J.B. Mott, M. Vargas, K.E. Dean and G.D. Di Giovanni. (2007a). Direct comparison of four bacterial source tracking methods and use of composite data sets. J. Appl. Microbiol. 103:350-364.

Casarez, E.A., Pillai, S.D. and Di Giovanni, G.D. (2007b) Genotype diversity of *Escherichia coli* isolates in natural waters determined by PFGE and ERIC-PCR. Water Research 41(16), 3643-3648.

Dean, K., Vargas, M, Casarez, E. A. and Di Giovanni, G.D. (2006) Assessment of bacterial sources impacting Lake Waco and Belton Lake. Texas Farm Bureau, Waco, and Texas State Soil and Water Conservation Board, Temple.

Di Giovanni, G.D. and Casarez, E.A. (2006) Upper and lower San Antonio River, Salado Creek, Peach Creek and Leon River Below Lake Proctor bacterial source tracking project. Texas Commission on Environmental Quality, Austin.

Di Giovanni, G. D., J. Truesdale, K. Barrella, and E. A. Casarez. (2009). Preliminary Assessment of Fecal Pollution Sources Impacting Lake Granbury as Determined by Bacterial Source Tracking (BST). Brazos River Authority, Waco, Texas, 16 pages.

Di Giovanni, G.D., Casarez, E.A., Gentry, T.G. and Martin, E.C. (2010) Increased analytical infrastructure and further development of a statewide bacterial source tracking library. Project Number 08-50 and 08-51. Texas State Soil and Water Conservation Board, Temple.

Farnleitner, A. H., G. H. Reischer, H. Stadler, D. Kollanur, R. Sommer, W. Zerobin, G. Blöschl, K. M. Barrella, J. A. Truesdale, E. A. Casarez, and G. D. Di Giovanni. (2011). Agricultural and rural watersheds, p. 399-431. *In* A. R. B. C. Hagedorn, V. J. Harwood (ed.), Microbial Source Tracking: Methods, Applications, and Case Studies. Springer Science+Business Media, LLC, New York.

Graves, A., Weaver, R.W. and Entry, J. (2009) Characterization of enterococci populations in livestock manure using BIOLOG. Microbiological Research 164(3), 260-266.

Hamilton, W.P., M. Kim and E.L. Thackston. (2005). Comparison of commercially available *Escherichia coli* enumeration tests: Implications for attaining water quality standards. Water Res. 39:4869-4878.

Hauck, L.M. and Du, B. (2006) Technical Support Document: Upper Oyster Creek (Segment 1245) Bacteria TMDL. Texas Commission on Environmental Quality, Austin.

Jones, C.A., Wagner, K., Di Giovanni, D., Hauck, L., Mott, J., Rifai, H., Srinivasan, R., and Ward, G. (2007) Bacteria total maximum daily load task force final report. Texas Water Resources Institute Technical Report 341. Texas A&M University, College Station.

Lu, L.G., Hume, M.E., Sternes, K.L. and Pillai, S.D. (2004) Genetic diversity of *Escherichia coli* isolates in irrigation water and associated sediments: implications for source tracking. Water Research 38(18), 3899-3908.

McQuaig, S. M., T. M. Scott, et al. (2009). Quantification of human polyomaviruses JC Virus and BK Virus by TaqMan quantitative PCR and comparison to other water quality indicators in water and fecal samples. Appl Environ Microbiol 75(11): 3379-88.

Mott, J.B. and Lehman, R.L (2005) Bacterial source tracking in Copano Bay, Phase II, Final Report. Texas General Land Office, Austin.

Mott, J. and Smith, A. (2011). Library-Dependent Source Tracking Methods. In: Hagedorn, C., Blanch, A.R. and Harwood, V.J., (Eds), Microbial Source Tracking: Methods, Applications, and Case Studies. Springer, New York. pp. 31-59.

Moussa, S.H. and Massengale, R.D. (2008) Identification of the sources of *Escherichia coli* in a watershed using carbon-utilization patterns and composite data sets. Journal of Water and Health 6(2), 197-207.

Texas Institute for Applied Environmental Research, IEH, Inc., Parsons Water & Infrastructure Inc. and James Miertschin & Associates, Inc. (2006) Monitoring report for bacterial source tracking segments 0806, 0841, and 0805 of the Trinity River bacteria TMDL. Texas Commission on Environmental Quality, Austin

TWRI. (2012). 2012 Bacterial Source Tracking – State of the Science Conference: Conference Proceedings. Texas Water Resources Institute, Technical Report 427.

Ufnar, J. A., S. Y. Wang, et al. (2006). Detection of the nifH gene of Methanobrevibacter smithii: a potential tool to identify sewage pollution in recreational waters. J Appl Microbiol 101(1): 44-52.

USEPA. (2004). Method 1604: Total Coliforms and *Escherichia coli* in water by membrane filtration using a simultaneous detection technique (MI Medium). Environmental Protection Agency, Washington DC.

USEPA. (2005a). Microbial source tracking guide document. EPA-600-R-05-064. United States Environmental Protection Agency, Washington, DC.

USEPA. (2005b). Method 1603: *Escherichia coli* (*E. coli*) in water by membrane filtration using modified membrane-thermotolerant *Escherichia coli* agar (Modified mTEC). Environmental Protection Agency, Washington, DC.

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

Wagner, K..L. (2011) Evaluation of methods to assess and reduce bacterial contamination of surface water from grazing lands. Doctor of Philosophy Dissertation, Texas A&M University, College Station.

APPENDIX A:

INVENTORY OF KNOWN SOURCE SAMPLES AND ISOLATES COLLECTED FOR THE TEXAS *E. COLI* BST LIBRARY (AS OF JANUARY 2013)

Animal	# of Watersheds (out of 12)	# of Total Samples Collected	# of *E. coli* Positive Samples	# of *E. coli* Isolates Archived	# of Isolates for ERIC-PCR	# of Isolates for RiboPrinting	# of Local Library Self-Validated Isolates	# of Local Library Self-Validated Samples	TX Library ver. 1-13 (Cross- Validated) Isolates	TX Library ver. 1-13 (Cross- Validated) Samples
HUMAN (H)										
Human--septage	5	35	26	72	64	34	24	21	19	18
Human--sewage	9	586	457	1717	1263	639	397	338	345	297
HUMAN--ALL	**9**	**621**	**483**	**1789**	**1327**	**673**	**421**	**359**	**364**	**315**
WILDLIFE (WILD)										
NON-AVIAN WILDLIFE (WN)										
Armadillo	5	15	12	45	26	16	12	9	8	8
Aoudad (wild sheep)	1	1	0	0	0	0	0	0	0	0
Beaver	2	3	1	2	2	1	1	1	1	1
Bobcat	4	18	17	41	29	18	13	12	12	11
Buffalo	1	1	0	0	0	0	0	0	0	0
Coyote	9	94	81	288	201	120	73	57	58	50
Coyote or Fox*	1	1	1	5	3	1	1	1	0	0
Deer	8	95	75	262	168	94	51	40	44	35
Feral hog	10	126	111	407	258	144	96	81	84	72
Mouse or Rat*	1	1	0	0	0	0	0	0	0	0
Fox	4	20	14	42	30	17	13	11	13	11
Javelina	1	10	10	10	10	10	6	6	5	5
Mouse	2	8	1	1	1	1	0	0	0	0
Oppossum	6	57	54	101	79	58	45	42	39	37
Porcupine	1	2	2	7	7	2	2	2	2	2
Prairie dog	2	7	5	9	7	6	3	3	2	2
Rabbit	6	20	9	26	19	9	6	6	4	4
Raccoon	9	124	100	279	207	124	77	68	63	58
Rat	3	6	3	3	3	3	3	3	2	2
Ringtail	1	1	1	4	3	1	0	0	0	0
Skunk	3	16	13	38	27	16	11	9	11	9
Squirrel	6	30	16	63	31	18	13	13	12	12
Squirrel, Rat, or Nutria*	1	1	1	5	3	1	0	0	0	0
Wolf	1	2	1	5	3	1	1	1	1	1
NON-AVIAN WILDLIFE--ALL	**12**	**659**	**528**	**1643**	**1117**	**661**	**427**	**365**	**361**	**320**

* Species identification of found source sample limited—does not affect source classification

Animal	# of Watersheds (out of 12)	# of Total Samples Collected	# of *E. coli* Positive Samples	# of *E. coli* Isolates Archived	# of Isolates for ERIC-PCR	# of Isolates for RiboPrinting	# of Local Library Self-Validated Isolates	# of Local Library Self-Validated Samples	TX Library ver. 1-13 (Cross- Validated) Isolates	TX Library ver. 1-13 (Cross- Validated) Samples
WILDLIFE (WILD)										
AVIAN WILDLIFE (WA)										
Blackbird or Starling*	2	10	9	17	12	8	7	7	4	4
Bittern, least	1	1	1	1	1	1	0	0	0	0
Blackbird	1	7	3	13	9	5	3	3	3	3
Blue jay	1	1	1	5	3	1	1	1	1	1
Buzzard	3	4	3	13	9	6	4	3	4	3
Cardinal	1	3	2	9	6	3	0	0	0	0
Crow	2	4	2	7	6	3	1	1	1	1
Double crested cormorant	1	1	1	2	2	1	1	1	1	1
Dove	4	22	15	43	30	17	10	10	9	9
Duck	7	179	143	397	273	160	93	88	86	83
Duck or Goose*	1	1	1	1	1	1	0	0	0	0
Egret	3	37	24	82	54	29	19	18	18	17
Egret or Heron*	1	2	1	1	1	1	1	1	1	1
Goose	5	34	30	100	63	34	18	16	17	15
Grackle	4	44	26	93	62	31	14	13	11	10
Gull	2	3	2	6	4	2	1	1	1	1
Hawk	1	1	0	0	0	0	0	0	0	0
Heron	3	15	11	42	27	16	9	8	9	8
Killdeer	3	5	3	6	5	5	4	3	4	3
Martin	1	7	1	5	3	1	0	0	0	0
Meadow lark	1	4	2	10	6	4	0	0	0	0
Mockingbird	2	2	2	10	6	4	1	1	1	1
Owl	2	2	0	0	0	0	0	0	0	0
Pelican	1	9	7	34	21	11	5	5	5	5
Pigeon	3	48	33	117	81	43	20	18	16	14
Pigeon or Grackle*	1	1	1	1	1	1	1	1	1	1
Quail	1	1	0	0	0	0	0	0	0	0
Scissortail fly catcher	1	1	0	0	0	0	0	0	0	0
Sparrow	3	9	4	10	8	4	1	1	1	1
Starling	3	17	9	22	19	13	7	6	5	4
Swallow	7	96	33	107	82	40	23	19	22	18
Turkey	3	28	18	82	53	22	7	6	6	5
Vulture	2	8	8	28	18	8	6	6	5	5
Werbler, yellow	1	1	0	0	0	0	0	0	0	0
Wren, Bewicks	1	1	1	1	1	1	0	0	0	0
AVIAN WILDLIFE--ALL	**9**	**609**	**397**	**1264**	**867**	**476**	**257**	**237**	**232**	**214**

* Species identification of found source sample limited—does not affect source classification

Animal	# of Watersheds (out of 12)	# of Total Samples Collected	# of *E. coli* Positive Samples	# of *E. coli* Isolates Archived	# of Isolates for ERIC-PCR	# of Isolates for RiboPrinting	# of Local Library Self-Validated Isolates	# of Local Library Self-Validated Samples	TX Library ver. 1-13 (Cross- Validated) Isolates	TX Library ver. 1-13 (Cross- Validated) Samples
DOMESTIC ANIMALS (DOM)										
CATTLE (C)										
Cattle--beef	8	187	154	638	394	179	87	78	71	67
Cattle--dairy	5	96	88	381	274	136	96	66	80	65
Cattle--undesignated	6	180	141	487	335	164	78	70	69	60
CATTLE--ALL	**12**	**463**	**383**	**1506**	**1003**	**479**	**261**	**214**	**220**	**192**
OTHER NON-AVIAN LIVESTOCK (OLN)										
Bison	1	2	2	2	2	2	0	0	0	0
Donkey	5	14	12	30	19	11	5	5	5	5
Goat	8	78	60	213	141	60	30	30	26	26
Horse	9	118	91	295	193	109	39	35	27	24
Llama	3	5	3	11	9	4	3	3	3	3
Rabbit	3	3	1	1	1	1	1	1	1	1
Sheep	4	40	35	147	101	44	16	13	12	11
Swine--domestic	5	75	60	141	106	68	33	33	27	27
Penned deer	1	2	2	10	6	0	0	0	0	0
OTHER NON-AVIAN LIVESTOCK--ALL	**9**	**337**	**266**	**850**	**578**	**299**	**127**	**120**	**101**	**97**
AVIAN LIVESTOCK (OLA)										
Chicken	8	181	138	585	356	165	103	84	87	74
Duck--farm	2	7	7	31	21	7	2	2	2	2
Goose--farm	3	4	4	12	8	6	2	2	2	2
Guinea fowl	3	10	7	11	7	7	1	1	1	1
Turkey	1	3	3	14	9	6	1	1	1	1
Pigeon--state fair	1	1	1	5	3	1	0	0	0	0
Grackle--state fair	1	1	1	5	3	3	0	0	0	0
AVIAN LIVESTOCK--ALL	**9**	**207**	**161**	**663**	**407**	**195**	**109**	**90**	**93**	**80**
PETS (P)										
Cat	6	86	50	186	125	63	30	22	26	19
Dog	6	201	161	530	354	181	78	74	55	52
Rabbit--pet	3	4	2	7	6	2	1	1	0	0
Ferrett	4	5	0	0	0	0	0	0	0	0
Gerbil	1	1	0	0	0	0	0	0	0	0
Guinea pig	1	1	0	0	0	0	0	0	0	0
Hamster	1	1	0	0	0	0	0	0	0	0
Parakeet	1	1	1	1	0	0	0	0	0	0
Parrot	1	3	0	0	0	0	0	0	0	0
Duck--pet	1	1	1	1	1	1	1	1	1	1
Opossum--pet	1	1	1	1	1	1	0	0	0	0
Turtle--pet	1	1	1	3	3	2	1	1	1	1
PET--ALL	**7**	**306**	**217**	**729**	**490**	**250**	**111**	**99**	**83**	**73**

APPENDIX B:

BST PROMOTIONAL FLYERS AND PROJECT ONE-PAGER

Moving Forward

For future WPP and TMDL development projects, an assessment phase using a "toolbox" approach is recommended. The assessment phase should include targeted monitoring of suspected pollution sources, use of library-independent and dependent methods to identify the presence of domestic sewage pollution and screening of water isolates from the new watershed against the existing library to determine the need for collection of local source samples and expansion of the library.

Decision on which method to utilize can be assisted with the use of the matrix provide in Chapter 2 of the EPA Microbial Source Tracking Guide. It is critical to follow the same analytical protocols for comparability of BST data sets.

The state BST laboratories (UTSPH – El Paso Regional Campus; Texas A&M Soil and Aquatic Microbiology Laboratory) can provide detailed BST protocols. In addition, the sharing of bacterial isolates and BST data between the state laboratories and others is welcomed.

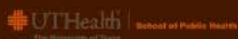
texasbst.tamu.edu

AgriLIFE RESEARCH & EXTENSION
Texas A&M System

Texas Water Resources Institute
make every drop count

BST

Bacterial Source Tracking



Developed by the Texas Water Resources Institute and funded through a State General Revenue nonpoint source grant from the Texas State Soil and Water Conservation Board

Cover photo: © Lynn McBride

EM-110

An introduction for laboratories and public agencies to the foremost tool for identifying sources of fecal pollution

The Need

According to the 2010 Texas Integrated Report, there are 303 bacterially impaired waterbodies in Texas. Nonpoint sources (NPS) of pollution greatly affect water quality. Identifying and assessing sources of fecal pollution is a key component in effectively implementing a NPS pollution management program. Proper evaluation of these sources is needed to properly assess risk in contact recreation, target best management practices, and develop effective watershed protection plans (WPPs) and bacterial total maximum daily loads (TMDLs).

Genotypic (molecular) tools appear to hold promise for BST, providing the most conclusive characterization and level of discrimination for isolates. Of the molecular tools available, automated ribosomal ribonucleic acid (RNA), gene fingerprinting (RiboPrinting), repetitive element polymerase chain reaction (rep-PCR), and pulsed-field gel electrophoresis (PFGE) are emerging as a few of the versatile and feasible BST techniques.

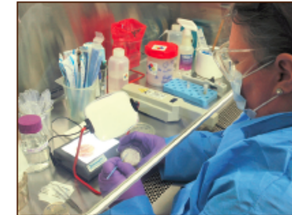
BST Technologies

The premise behind BST is that genetic and phenotypic tests can identify bacterial strains that are host-specific so that the original host animal and source of the fecal contamination can be identified. Often *E. coli* or *Enterococcus* spp. are used as the bacteria targets in BST, as this provides a direct link with water quality standards, which are usually based on one of these two indicators. The technologies used for BST have evolved greatly in the past few years.

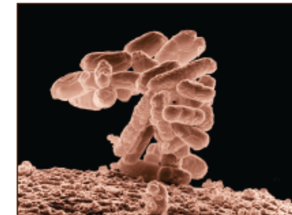
Identification libraries consisting of thousands of isolates obtained from thousands of animal and human fecal samples collected in different geographical regions of Texas have already been established for ERIC-PCR, PFGE, RiboPrinting, CSU and KB-ARA patterns. In addition, several thousand more *E. coli* isolates from source samples have been archived and are available to researchers.

About the Texas *E. coli* BST Library

The Texas *E. coli* BST library currently contains 1,393 *E. coli* isolates obtained from 1,201 different domestic sewage, wildlife, livestock and pet fecal samples. Isolates were selected after screening several thousand isolates from nine different studies throughout Texas.



Library development is one of the most costly components of BST studies. Currently, Dr. George Di Giovanni, at the University of Texas School of Public Health – El Paso Regional Campus, and Dr. Terry Gentry, at the Texas A&M University Soil and Aquatic Microbiology Laboratory, are cross-validating the libraries generated in Texas BST studies in an attempt to explore issues of geographical and temporal stability of BST libraries, refine library isolate selection and determine accuracy of water isolate identification.



By selecting *E. coli* source isolates that are correctly identified from multiple watersheds, the BST library hopes to find more geographically stable and host-specific isolates, resulting in more accurate source tracking. Library-independent methods are also currently being explored, based on Texas Commission on Environmental Quality and Texas State Soil and Water Conservation Board Bacterial Total Maximum Daily Load Task Force recommendations.

Relative comparison of several bacterial source tracking techniques

Technique	Target organism(s)	Basis of characterization	Accuracy of source identification	Size of library needed for water isolate IDs	Capital cost	Cost per sample (reagents and consumables only)	Ease of use	Hands on processing time for 32*** isolates	Time required to complete processing 32 isolates
ERIC-PCR	*E. coli* and *Enterococcus* spp.	DNA fingerprint	Moderate	Moderate	$20,000 ($15,000 BioNumerics software, $5,000 equipment)	$8	Moderate	3 h	24 h**
RP	*E. coli* and *Enterococcus* spp.	DNA fingerprint	Moderate	Moderate	$115,000 ($100K RiboPrinter, $15K BioNumerics software)	$40	Easy	1 h	24 h
PFGE	*E. coli* and *Enterococcus* spp.	DNA fingerprint	High	Large	$30,000	$40	Difficult	10 h	5 days
KB-ARA	*E. coli* and *Enterococcus* spp.	Phenotypic fingerprint	Moderate*	Moderate	$35,000	$15	Easy	3 h	24 h**
CSU	*E. coli* and *Enterococcus* spp.	Phenotypic fingerprint	Moderate	Moderate	$15,000	$10	Easy	4 h	24 h**
Bacteriodales PCR	Bacteriodales species	Genetic marker presence or absence (not quantitative)	Moderate to high for only human, ruminant, horse, and pig sources	N/A	$5,000	$8	Easy to moderate	3 h	8 h**
E. faecium esp marker	*E. faecium*	Genetic marker presence or absence (not quantitative)	High for only human	N/A	$8,000	$8 to $12	Easy to moderate	3 to 6 h	8 to 24 h**
ERIC-RP	*E. coli*	DNA fingerprints	Moderate to high	Moderate	$120,000	$48	Moderate	4 h	24 h
ERIC-ARA	*E. coli*	DNA and phenotypic fingerprints	Moderate to high	Moderate	$55,000	$23	Moderate	6 h	24 h
ARA-CSU	*E. coli* and *Enterococcus* spp.	Phenotypic fingerprints	Moderate to high	Moderate	$50,000	$23	Easy to moderate	7 h	24 h

†A manual ribotyping version is also used by some investigators (i.e. Dr. M. Samadpour with IEH Laboratories and Consulting Group in Seattle), but no detailed information is available for comparison. ‡A variation of this technique using replica plating and +/- scoring of growth on media with different concentrations of antibiotics, called ARA, has been used extensively in Virginia for TMDLs. *This technique is better for distinguishing broader groups of pollution sources. For example, "wildlife" and "livestock" as opposed to "avian wildlife", "non-avian wildlife," "cattle," etc. **With sufficient personnel, up to approximately 150 isolates can be analyzed in 24 h. ***Thirty two isolates selected for comparison because it is the maximum throughput per day of the RiboPrinter, which is the only automated system described.

AgriLIFE RESEARCH & EXTENSION
Texas A&M System

Texas Water Resources Institute
make every drop count

texasbst.tamu.edu

BST

Bacterial Source Tracking

UTHealth School of Public Health

TSSWCB

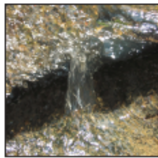
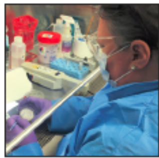
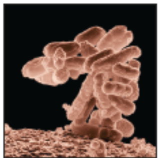
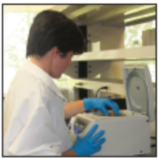
Texas Water Resources Institute

AgriLIFE RESEARCH & EXTENSION
Texas A&M System

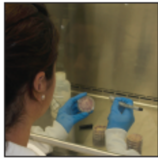
Developed by the Texas Water Resources Institute and funded through a State General Revenue nonpoint source grant from the Texas State Soil and Water Conservation Board

Cover photo: © Lynn McBride EM-111

Learn why Bacterial Source Tracking is the foremost tool for identifying sources of fecal pollution

Proper evaluation of nonpoint sources is needed to accurately assess risk in contact recreation, target best management practices, and develop effective watershed protection plans.



The Need

According to the 2010 Texas Integrated Report, there are 303 bacterially impaired waterbodies in Texas. Nonpoint sources (NPS) of pollution greatly affect water quality. Identifying and assessing sources of fecal pollution is a key component in effectively implementing a NPS pollution management program.

Proper evaluation of these sources is needed to properly assess risk in contact recreation, target best management practices, and develop effective watershed protection plans (WPPs) and bacterial total maximum daily loads (TMDLs).

The freshwater contact recreation use criterion used to determine impairment includes both a geometric mean for *Escherichia coli* (*E. coli*) of 126 colonies per 100 ml and. The saltwater contact recreation use criterion includes both a geometric mean for enterococci of 35 colonies per 100 ml. The oyster water use criterion includes a median fecal coliform concentration of 14 colonies per 100 ml and no more than 10% of samples may exceed 43 colonies per 100 ml.

BST Technologies

The premise behind BST is that genetic and phenotypic tests can identify bacterial strains that are host-specific so that the original host source of the fecal contamination can be identified.

Often *E. coli* or *Enterococcus* spp. are used as the bacteria targets in BST, as this provides a direct link with water quality standards which are usually based on one of these two indicators. The technologies used for BST have evolved greatly in the past few years.

Identification libraries consisting of thousands of isolates obtained from thousands of animal and human fecal samples collected in different geographical regions of Texas have already been established. In addition, several thousand more *E. coli* isolates from source samples have been archived and are available to researchers.

About the Texas *E. coli* BST Library

The Texas *E. coli* BST library currently contains 1,393 *E. coli* isolates obtained from 1,201 different domestic sewage, wildlife, livestock and pet fecal samples. Isolates were selected after screening several thousand isolates from nine different studies throughout Texas.

Library development is one of the most costly components of BST studies. Currently, Dr. George Di Giovanni, at the University of Texas School of Public Health – El Paso Regional Campus, and Dr. Terry Gentry, at the Texas A&M University Soil and Aquatic Microbiology Laboratory, are cross-validating the libraries generated in Texas BST studies in an attempt to explore issues of geographical and temporal stability of BST libraries, refine library isolate selection and determine accuracy of water isolate identification.

By selecting *E. coli* source isolates that are correctly identified from multiple watersheds, project partners hope to find more geographically stable and host-specific isolates, resulting in more accurate source tracking. Library-independent methods are also currently being explored, based on Texas Commission on Environmental Quality and Texas State Soil and Water Conservation Board Bacterial Total Maximum Daily Load Task Force recommendations.

Moving Forward

For future WPP and TMDL development projects, an assessment phase using a "toolbox" approach is recommended. The assessment phase should include targeted monitoring of suspected pollution sources, use of library-independent and dependent methods to identify the presence of domestic sewage pollution and screening of water isolates from the new watershed against the existing library to determine the need for collection of local source samples and expansion of the library.

Decision on which method to utilize can be assisted with the use of the matrix provide in Chapter 2 of the EPA Microbial Source Tracking Guide. It is critical to follow the same analytical protocols for comparability of BST data sets. The state BST laboratories (UTSPH – El Paso Regional Campus; Texas A&M Soil and Aquatic Microbiology Laboratory) can provide detailed BST protocols. In addition, the sharing of bacterial isolates and BST data between the state laboratories and others is welcomed.

AgriLIFE RESEARCH & EXTENSION
Texas A&M System

TEXAS A&M UNIVERSITY

texas bacterial source tracking library

texasbst.tamu.edu

Escherichia coli (*E. coli*) bacterial source tracking (BST) is a valuable tool for identifying human and animal sources of fecal pollution.

Protection of our water resources is one of the most significant environmental challenges of the new millennium. According to the *2010 Texas Water Quality Inventory and 303(d) List*, there are 318 bacterially impaired water bodies in Texas. Nonpoint sources (NPS) of pollution greatly affect water quality. Identifying and assessing sources of fecal pollution is a key component in effectively implementing a NPS pollution management program. Proper evaluation of these sources is needed to target best management practices and develop bacterial total maximum daily loads (TMDLs). This information may also be useful to properly assess risk in contact recreation, as many waterborne pathogens causing human illness do not colonize nonhuman hosts.

The Texas *E. coli* BST library has expanded through this project and currently contains 1,393 *E. coli* isolates obtained from 1,232 different domestic sewage, wildlife, livestock and pet fecal samples. Expansion of the library to include additional known source isolates from different Texas watersheds and different animal hosts is still needed. The use of an expanded Texas *E. coli* BST library to identify NPS and development of TMDLs would provide significant cost and time savings.

The technologies used for BST have evolved greatly in the past few years. A host of new information is currently available, yet not readily distributed or known to state agency personnel.

Objectives

- Support and maintain analytical infrastructure at BST laboratories
- Develop and update statewide BST standard operating procedures
- Expand statewide known-source library
- Develop and evaluate new species-specific bacteria markers
- Conduct state of the science workshop for Texas on BST technologies and capabilities

Accomplishments

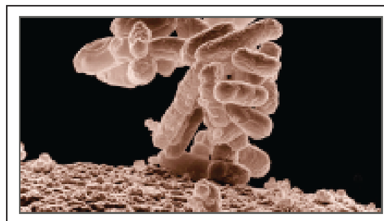
- Nearly 120 participants from 13 states participated in the 2012 Bacterial Source Tracking–State of the Science Conference in February 2012 to hear discussions on BST and current practices, scientific advances and improvements in application.

Collaborators

- Texas AgriLife Research
- Texas Water Resources Institute

Funding Agency

- Texas State Soil and Water Conservation Board



Texas Water Resources Institute
make every drop count

Texas Water Resources Institute
1500 Research Parkway, Suite 110, 2260 TAMU
College Station, TX 77843-2260
979.845.1851
twri@tamu.edu
twri.tamu.edu

4/2012