BACTERIAL SOURCE TRACKING IN IMPAIRED WATERSHEDS: EVALUATION OF CULTURE-DEPENDENT AND -INDEPENDENT METHODS FOR INCREASED SOURCE SPECIFICITY AND IMPROVED MANAGEMENT

A Dissertation

by

EMILY CATHERINE MARTIN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Terry Gentry
Committee Members,	John Brooks
	George Di Giovanni
	Sam Feagley
	Raghupathy Karthikeyan
Head of Department,	David Baltensperger

May 2013

Major Subject: Soil Science

ABSTRACT

Bacterial contamination due to excessive levels of bacteria is a confounding problem and remediation of impaired watersheds relies on the detection of fecal indicator bacteria and then assessing the source of said bacteria. Bacterial source tracking (BST) is an approach for assessing potential sources of this contamination. The purpose of this study was to utilize both cultivation-independent and -dependent methods to improve the ability to track sources of fecal contamination. First, *E. coli* community composition was assessed across three standard water quality assessments including USEPA Methods 1603 and 1604, and Colilert®, to determine their impact on BST library-based performance. Results indicate that the three assessed methods of enumeration and isolation may select for different populations of *E. coli* and standardized methods may be warranted if library-dependent BST is part of a research plan. Next, BST techniques were used to enumerate and characterize *E. coli* communities across various dairy manure management techniques used in the Leon River watershed in central Texas to determine effectiveness of BST efforts in tracking contamination from dairy manure. Results of this study indicated that manure and effluent management strategies which employed means to remove solids from the manure tended to decrease the levels of *E. coli* in the effluent. Some *E. coli* genotypes were found across the managerial treatments even though there were no clear seasonal trends or site groupings among the dataset. The vast majority of the isolates classified using the Texas *E. coli* BST library were correctly classified back to their major source

class, thus increasing confidence in the methods currently being utilized to track dairy fecal contributions in this Central Texas watershed. Finally, deer bacterial fecal communities from south and central Texas were analyzed using 454-pyrosequencing to assess the potential for the development of a deer-specific BST marker. Microbial communities did not cluster by site or year suggesting that deer fecal communities in these Texas regions are stable over time and could be amenable to marker development.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. Terry Gentry, and committee members, Dr. John Brooks, Dr. Sam Feagley, Dr. George Di Giovanni, and Dr. Raghupathy Karthikeyan, for their guidance and support in this endeavor. I owe a huge debt of gratitude to my collaborators in El Paso, Dr. Di Giovanni, Dr. Elizabeth Casarez, and Joy Truesdale, who were instrumental in getting our BST lab setup and were always so giving of their time and expertise. I cannot thank Dr. Emily Hollister enough for being such a wonderful teacher, mentor, and even better friend. Thanks to Ilsa Kantola for being my ever present sounding board and confidant. I am eternally grateful to Heidi Mjelde who was by my side through so very many samples, water and otherwise, early and late, weekday or weekend whose support and friendship I will cherish always. Finally, thanks to my Mom, Dad, and sisters for their love and constant encouragement without whom I would not be the person I am today.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

CHAPTER I

GENERAL INTRODUCTION

Problem and Introduction to Bacterial Source Tracking

The Clean Water Act of 1972 created a platform by which states could regulate the discharge of pollution into US waterways. Following its enactment, many of the point sources of pollution, including industrial sites and municipalities, were identified and subsequent modifications to their waste management have greatly reduced their environmental impact. Also in accordance with the Clean Water Act, states were mandated to develop community driven watershed protection plans (WPP) or regulatory-based total maximum daily loads (TMDL), ultimately implementing plans to reduce remaining impairments. In a 2010 water quality assessment, the Texas Commission on Environmental Quality (TCEQ) listed 621 water bodies in Texas that did not meet state standards. These water bodies were listed on the 303(d) list as impaired watersheds, and bacterial contamination accounted for 51% of those listings (TCEQ, 2010). Remediation and management of bacterially-impaired waters relies on the detection of fecal contamination. Bacterial source tracking (BST) is an approach to determining potential sources of this fecal contamination.

BST is based on the premise that gut communities of various warm blooded hosts are preferentially selected through various factors including diet, pH, and nutrients, and can be host specific (Casarez et al., 2007b; USEPA, 2005b). This specificity can then be exploited through phenotypic and genetics assays to trace fecal contamination back to its

source. Technical approaches to BST are broad and no one approach is universally accepted (USEPA, 2005b). BST approaches can broadly be split into two categories, library (cultivation) dependent and independent methods. Library dependent methods rely on the creation of reference libraries from known sources used to identify water isolates. Many library-dependent BST studies are based on fecal indicators, like *E. coli* and enterococci, due to their regulatory significance and standardized culturing techniques (Casarez et al., 2007b; Field and Samadpour, 2007; USEPA, 2005b). Concerns with library dependent BST include temporal and geographical stability of the library, size and scope of the known-source library, and statistical power of identifications, in addition to time and cost of creating the library (Field and Samadpour, 2007; Stoeckel and Harwood, 2007). Library independent methods generally rely on molecular markers specific to a target of interest. Library-independent characterizations hold the promise for cost effective and rapid estimation of recent contamination without the need for representative indicator library construction, but specificity issues across source classes and a lack of validated marker sets continues to limit its wholesale acceptance in the field. Novel library-independent techniques targeting groups of organisms other than fecal indicators, including *Bacteroidales*, show promise as a way to circumvent many of the culture based techniques, but water quality standards continue to be based on fecal indicators. Thus, BST techniques targeting fecal indicators continue to be a viable and useful resource.

Motivation and Objectives

Even though the science of BST has made great strides over the past decade, many researchers and stakeholders remain skeptical about their usage, especially from a regulatory standpoint. Critical areas of need are ultimately tied to method performance. Library dependent BST methods have several significant factors of consideration that include size of the library, proportionality and representatives of library members, library performance measures, and as well as geographical and temporal stability (Field and Samadpour, 2007; Stoeckel and Harwood, 2007; USEPA, 2005b). But even before library construction and performance is considered, methods of enumeration and isolation should also be considered. There are numerous EPA approved methods to enumerate fecal indicators as well as methods used in clinical settings, but there is a lack of consensus about which method is most appropriate for BST library creation (USEPA, 2005b). Much effort has been made to understanding false positive and negative rates as well as the organisms causing them (Chao et al., 2004; Sercu et al., 2011). Additionally, studies have highlighted the ability or inability to compare data across enumeration techniques (Francy and Darner, 2000; Hamilton et al., 2005). But very little investigation has occurred into what impacts these different media have on the microbial ecology of the bacterial populations isolated and what impacts that may ultimately have on library performance. Once libraries are established, including the dynamic Texas *E. coli* BST library, considerable questions still remain about their geographical and temporal stability (Casarez et al., 2007b). Substantial time and expense are needed to develop a BST known source library and expand it over time, so the goal is to be able to

utilize that library over several years and possibly across a broad geographic region. Confounding results across several studies suggest individual libraries should be examined for performance over time and include as much metadata as possible to attempt to account for shifts in bacterial populations (Casarez et al., 2007b; Gordon, 2001; Hansen et al., 2009; Hartel et al., 2002).

Library independent BST methods also suffer from similar method performance challenges. A suite of source-specific BST molecular markers used to quantify relative proportions of fecal contamination in water could be the panacea of water quality management efforts, and even though considerable progress has been made toward this end, methods to meet these needs are not yet available. Markers have been developed for many of the major source classes, but considerable concerns exist over specificity of the markers (Shanks et al., 2010). From a practical standpoint, stakeholders need the ability to discriminate between three main categories of fecal contamination, humans, domesticated animals, and various forms of wildlife. Currently, there are no accepted methods to distinguish between ruminant animals (Bernhard and Field, 2000). Separation of fecal contamination among ruminants, specifically between cattle and deer, is especially important as TMDL projects and best management practices are developed to alleviate bacterial impairments have traditionally been directed toward livestock management. As molecular means of characterization become available and are validated, it will be imperative to combine new tools with current source tracking resources, including library-dependent *E.coli* based methods, to improve our ability to

both track and prevent fecal contamination in an effort to tailor management practices and remediation schemes to ensure a healthy water supply.

The goal of these studies was to expand BST approaches and attempt to ultimately improve our ability to track sources of fecal contamination. The research objectives were as follows:

(1) Evaluate differences in *E. coli* community composition across three standard water quality assessments to ultimately determine their impact on BST library performance (Chapter III).

(2) Assess the impact of dairy manure management on abundance and diversity of *E. coli* using BST techniques (Chapter IV).

(3) Analyze deer fecal microbial communities using 454 pyrosequencing to evaluate their suitability for development of a deer specific BST marker (Chapter V).

CHAPTER II

EVOLUTION OF BACTERIAL SOURCE TRACKING IN TEXAS

Bacterial source tracking (BST), generally defined for this report as 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 Texas State Soil and Water Conservation Board (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 including both library-dependent methods used to develop the Texas *E. coli* BST library, as well as, library-independent methods utilizing *Bacteroidales*-PCR based approaches.

Copano Bay

An early BST study in Texas was conducted in the Copano Bay watershed north of Corpus Christi. The watershed was very complex in that it had both tidal and non-tidal segments included on the 303(d) list for elevated bacterial levels as well as not meeting water quality for 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 pulse-field gel

electrophoresis (PFGE) which served to confirm the ARA results (Mott and Smith, 2011, USEPA, 2005). 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 two-way (human and non-human), four-way (human, cattle, horse, and wildlife) and six-way classifications (human, cattle, horse, duck, gull, and other wildlife). Ducks and gulls were included as

separate categories based on recommendations from a sanitary survey of the potential sources in the area. The ARCC of cross-validated isolates were highest in the two-way split at 71%, 62% with the four-way split, and 56% in the six-way split. In the six-way split, the largest misclassifications occurred when cattle and horses were identified as sewage. The six-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 six-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).

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 US 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 Trinity River project, isolates were included in a larger library established by IEH from samples collected around the US 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 data set 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.

Method Evaluation

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., 2007b; USEPA, 2005b). 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 TX 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.

Library Design and Performance

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 US 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 1,094 fecal samples total were processed of which 813 were positive for *E. coli*. A group of 100 isolates from South TX 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. Based on that similarity cutoff, one to three isolates were chosen to be included in the library. Isolates whose best match was less than 80% similar were considered unidentified 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.

Results

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 1) 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 1. Jackknife analysis rates of correct classification (%) for individual and four method composite BST methods and the 883 isolate library (Reprinted with permission from Casarez et al., 2007b).

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 1). 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 ERIC-PCR and RP were the two individual methods most similar (91%) to the composite dataset. This project was instrumental in providing a foundation for future BST work in TX. The results were reported to the Texas Farm Bureau and the TSSWCB and resulted in two peer-reviewed publications (Casarez et al., 2007a; Casarez et al., 2007b; Dean, 2006).

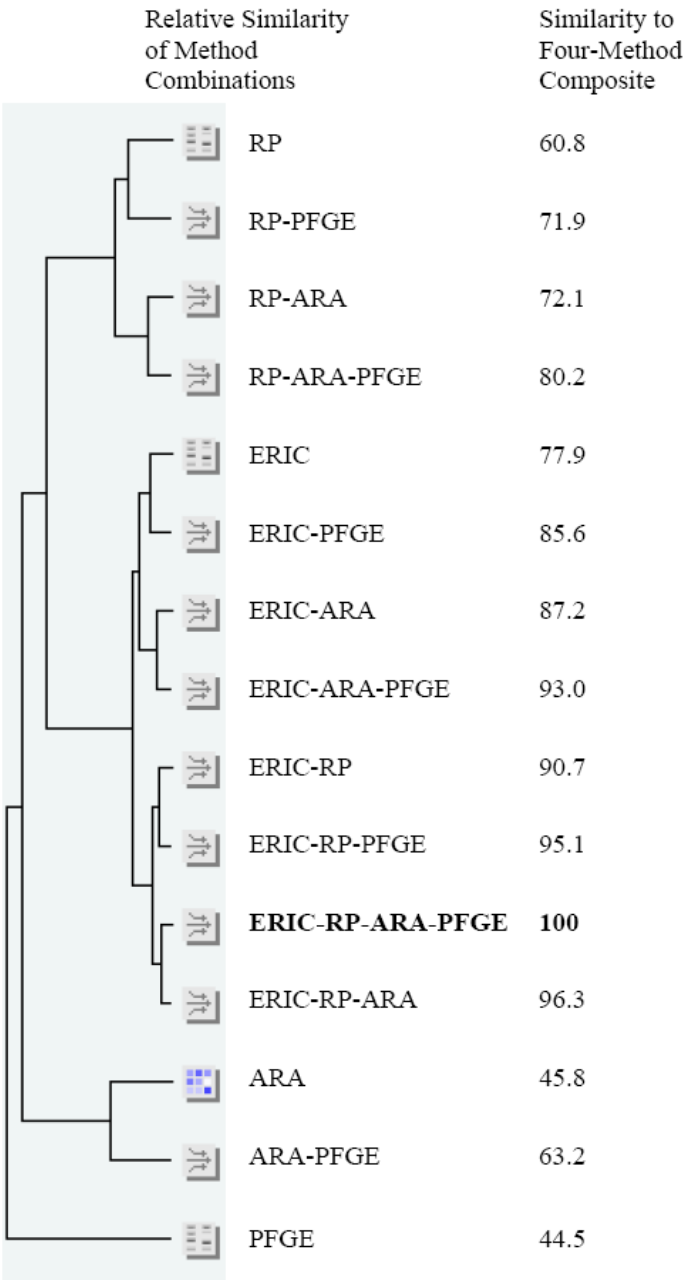

Figure 1. Congruence of individual BST methods and composite datasets (Labels include ERIC, ERIC-PCR, RP-RiboPrinting, ARA-KB-ARA, ERIC-RP-ARA-PFGE-four method composite dataset) (Reprinted with permission from Casarez et al., 2007b).

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 seven or eight way split for this study were 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, 1,008 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 US, 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 interpretation 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 three-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 and 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. The TCEQ and TSSWCB projects discussed previously had 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 served as a guidepost for future TSSWCB and TCEQ TMDL activities (Jones, 2007).

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 Texas AgriLife Research El Paso, but was aimed at expanding BST personnel and expertise to Texas A&M College Station 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 six 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) Buck Creek, (5) Upper Oyster Creek, and (6) 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 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. 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 three-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 a jackknife approach, isolates were removed if their ERIC-RP composite best match at 80% similarity cutoff were not to their specific seven-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 self-validated library included 996 isolates from 884 different known-source samples. Self-validation greatly increased the RCC which averaged 86% for the seven-way split. Cross-identifications were greatest within similar source categories like cattle and other livestock, further solidifying the future use of a less specific three-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 seven-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 three-way split was 86% (Table 2).

Table 2. 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	**1309**	**1185**	**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 the identification and use of cosmopolitan isolates for library construction, temporal and geographical effects on the fitness of the library over time while expanding the library with diverse sources of contamination from watersheds around the state (Di Giovanni et al., 2010).

Other BST Projects in Texas

Principal BST projects leading to the development to statewide projects have been heavily highlighted in this report, but other researchers and source tracking methods have also been utilized across the state. In the Rio Grande River valley, PFGE was used to compare *E. coli* from source irrigation water and sediments. 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 (Lu et al., 2004). 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.

Current and 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 three-way classification level as confidence in the separation of isolates into these categories is greater than for the more stringent six-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 brief overview of relevant findings from three recent BST projects.

Buck Creek WPP (http://buckcreek.tamu.edu/)

Buck Creek is a small creek located Panhandle of Texas in the Red River Basin. A total of 31 known source isolates from 28 samples were added to the expanding state library (ver. 08-09, 1172 isolates from 1044 samples). In total, 79 water samples were analyzed using *Bacteroidales* PCR and 426 isolates were fingerprinted using ERIC-RP and major source class identifies 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 2). 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 non-point pollution in the watershed.

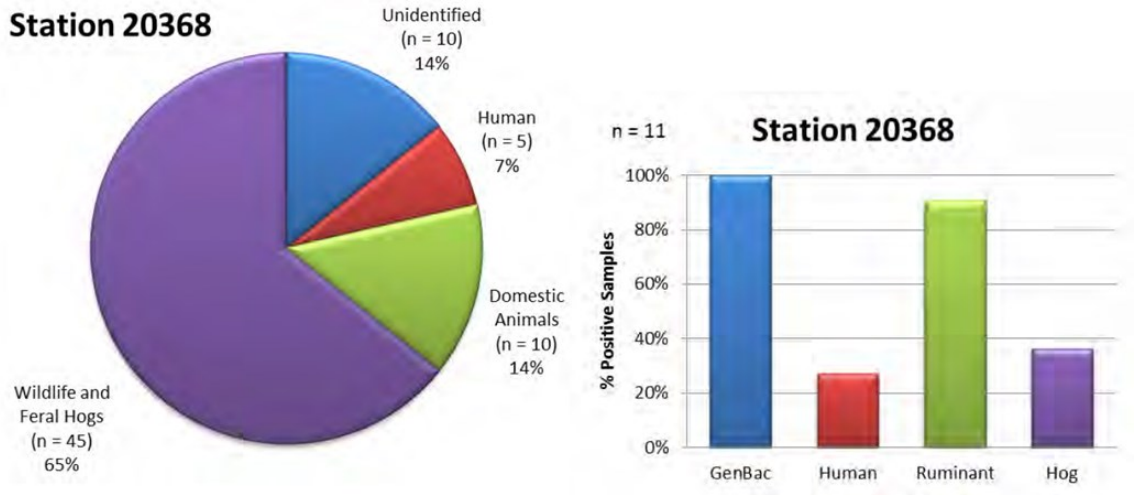

Figure 2. BST results from Buck Creek station 20368. Identification of water isolates (pie chart) using a three-way split for source classification and *Bacteroidales* PCR maker occurrence (bar chart).

Little Brazos River BST (http://lbr.tamu.edu/)

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 3). Using a three-way split, from a total of 69 water isolates classified using the Texas *E. coli* BST library (ver. 12-09, 1196

isolates from 1068 samples), 59% were classified as originating from wildlife with smaller proportions originating from domestic animal (19%) and human sources (6%).

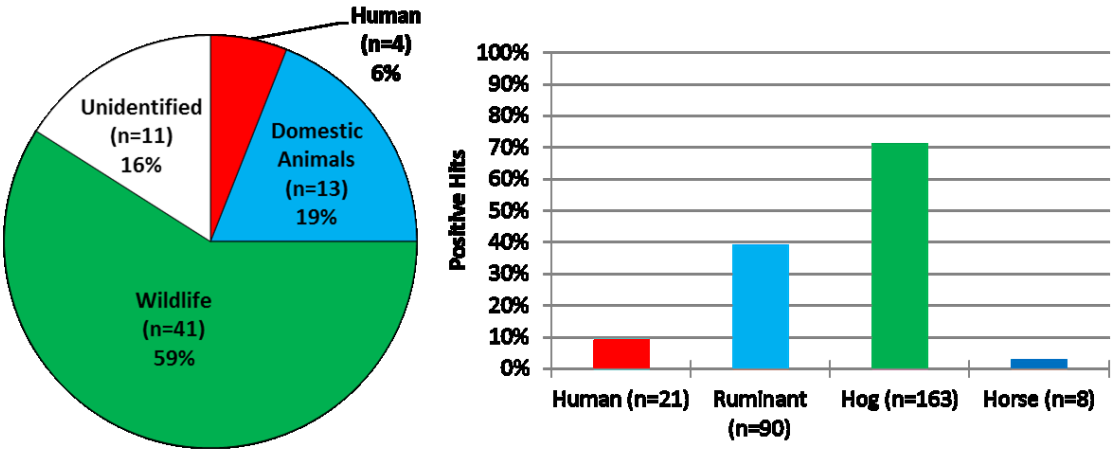

Figure 3. BST results from Little Brazos River. Identification of water isolates (pie chart) from all creek sites using a three-way split for source identifications (n=69) and Bacteroidales PCR marker occurrence (n=259) for human, ruminant, hog, and horse markers.

Big Cypress Creek Modeling and BST (http://bcc.tamu.edu/)

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 facilities (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 4). 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 three-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 4).

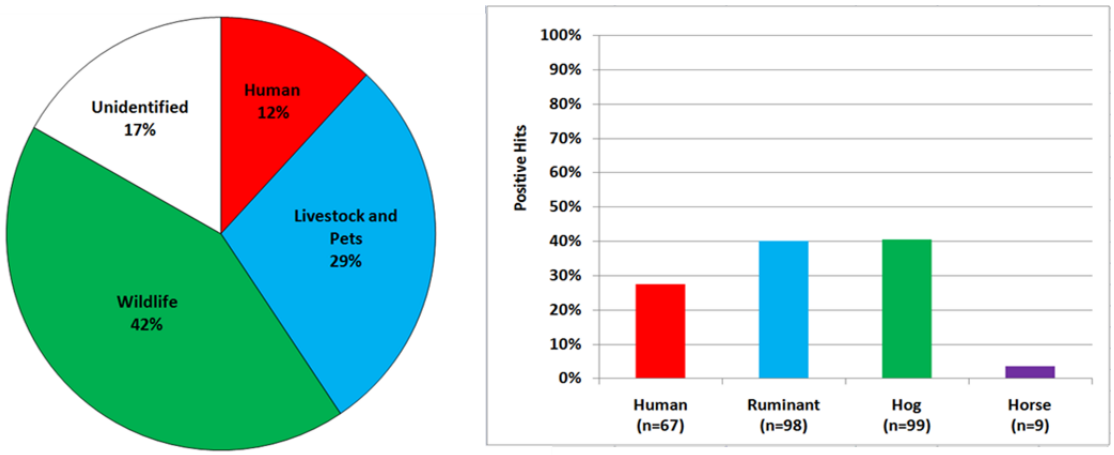

Figure 4. BST results from Big Cypress. Identification of water isolates (pie chart) from all creek sites using a three way-split for source identifications (n=101) and *Bacteroidales* PCR marker occurrence (n=244) for human, ruminant, hog, and horse markers.

Other projects underway which are following the same general approach for BST include Attoyac Bayou (http://attoyac.tamu.edu), Leon and Lampasas (http://leon-

lampasasbst.tamu.edu/, http://www.lampasasriver.org/) and Leona River (http://www.leonariver.org/).

Conclusions

Texas has been a leader in the use of BST as part of a toolbox approach to 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 a watershed, much less a particular sampling site, down to quantitative percentages of fecal contamination are not 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. 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, Leon and Lampasas, and Leona watersheds to the Texas *E. coli* BST Library, from two different laboratories, has the potential to begin to answer long-standing questions about

method performance. The following chapters will begin to address some of these questions:

- Even though stringent quality control measures are in place, are there biases toward fingerprints generated from one laboratory or the other?
- Can other commonly utilized enumeration methods apart from EPA Method 1603 on modified mTEC be utilized to isolate *E. coli* for library-dependent BST analysis? (Chapter II)
- Can we now better assess a more accurate number of potential individual sources as well as number of samples per sources needed to be characterized in a local watershed? Or for that matter, any source in order to gain power to discriminate that source in the library?
- How can cosmopolitan isolates, isolates whose best match is not to their specific source class, be utilized in library building? And can we better predict sources that might need additional samples processed to reach a substantial number of self-validated isolates?
- Is there a geographical bias in source class identification, not just to watershed, but in general? For instance, do beef cattle tend to cluster more with cattle not just from their watershed, but region in Texas? And can we incorporate 'metadata' already at our disposal from the source collections to help answer these questions.
- Are there temporal trends in known-source isolates? Should we be concerned about the time of year the samples are collected? (Chapter III)

- Did adding substantial numbers of additional sources and categories to the library increase our ability to discriminate sources or might we have met our theoretical maximum RCC?
- Can more in-depth bacterial community characterization aid in finding potential targets for new library-independent source tracking methods? (Chapter IV)

CHAPTER III

IMPLICATIONS OF *E. COLI* ISOLATION METHOD ON LIBRARY-DEPENDENT BACERIAL SOURCE TRACKING

Introduction

Water quality standards in the US are largely based on fecal indicator organisms including total coliforms, fecal coliforms, Enterococci, and more specifically *E. coli* (USEPA, 2003a). The USEPA has approved multiple methods for detecting and enumerating these indicators including membrane filtration methods and multiple tube fermentation methods, as well as defined substrate technologies (Olstadt et al., 2007). USEPA Method 1603 (USEPA, 2005a) and USEPA Method 1604 (USEPA, 2004) are chromogenic, membrane filtration methods while Colilert® (IDEXX, Westbrook, ME) 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. USEPA Method 1604 utilizes MI media and can simultaneously detect and enumerate both total coliforms and *E. coli*. The medium utilizes two enzyme substrates, the fluorogen 4-methylumelliferyl-β-D-galactopyranoside (MUGal), and chromogen indoxyl-β-D-glucuronide (IBDG), to detect the enzyme β-galactosidase produced by coliforms and β-glucuronidase produced by *E. coli*. USEPA Method 1603 is used to detect and enumerate *E. coli* only and utilizes a modified mTEC (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 most-probable-number technique

which utilizes a defined substrate medium and the chromogen ortho-nitrophenyl-β-D-galactopyranoside (ONPG) to detect β-galactosidase from total coliforms as well as the fluorogen 4-methylumbelliferyl-β-D-glucoronide (MUG) to detect β-glucuronidase from *E. coli*. These methods have been shown by some to be consistent for enumeration across the various platforms but their impact on *E. coli* community composition has yet to be fully evaluated (Buckalew et al., 2006; Hamilton et al., 2005).

Many water quality monitoring professionals utilize the Colilert® method due to its relatively low cost and ease of use for enumerations versus membrane filtration methods. Further, a majority of monitoring labs have only undertaken the training and cost to become certified in one primary method of enumeration of fecal indicator organisms, and transitioning to other methods is both time consuming and expensive. The Texas *E. coli* BST library is constructed largely from *E. coli* isolates processed using USEPA Method 1603 (Casarez et al., 2007b). *E. coli* was chosen at the target organism for the library due to its link to regulatory standards as well as standardized methods of enumeration. As BST efforts are expanded, this study was developed to evaluate additional water quality assessments, including Colilert® and USEPA Method 1604, and their ability to select similar *E. coli* communities to those isolated using USEPA Method 1603. The objective of this study was to evaluate differences in *E. coli* community composition across three standard water quality assessments including USEPA Method 1603, USEPA Method 1604, and Colilert® to ultimately determine their impact on BST library performance. It was hypothesized that even though all of the methods ultimately detect *E. coli* by the expression of the β-glucuronidase enzyme,

both MI and Colilert® are less selective media than mTEC due to the co-enumeration of both total coliforms as well as *E. coli* and will ultimately select for different communities. The specific growth platforms, i.e. membrane filtration versus defined substrate MPN, chromogens, and incubation temperatures vary across the three methods further solidifying the hypothesis that even though the enumeration data from previous studies have been shown to be comparable, the communities are likely different.

Materials and Methods

Site Description and Sample Collection

Six watersheds from south, central, and eastern Texas were sampled in this study to represent the diverse geography and land uses of the state. Big Iron Ore Creek (TCEQ Station ID 20844) in the Attoyac Bayou is located near Nacogdoches, TX, and is primarily a rural, agricultural watershed in the Piney Woods of East Texas. Burton Creek (TCEQ Station ID 11783) is part of the Navasota River watershed in College Station, TX and is an intermittent stream that is perennial due to wastewater treatment plant output. Campbells Creek (TCEQ Station ID 16395) is in the Little Brazos River watershed and is primarily a rural, agricultural watershed. Moody Creek is located on the Welder Wildlife Refuge near Sinton, TX and is a wildlife management and conservation refuge and cattle operation. Plum Creek (TCEQ Station ID 12640) is located on the southern end of the Plum Creek Watershed in the Guadalupe River Basin near Luling, TX, and is primarily a rural, rangeland and forested area. White Oak Bayou

(TCEQ Station ID 11387) is located in central Harris County, north of downtown Houston and is a channelized urban stream.

Samples were collected in the fall of 2011 and spring of 2012. Stream water was collected into sterile Whirl-Pak® bags and transported on ice back to the laboratory for processing within 3 hours of sample collection. Each sample was split into three aliquots for *E. coli* enumeration and isolation.

Sample Processing

Samples were processed using three standard methods, USEPA Method 1603 (modified mTEC), USEPA Method 1604(MI) and Colilert® per method and manufacturer's instructions. Modified mTEC (BD, Franklin Lakes, NJ) and MI (BD,) media were prepared per label instructions. Colilert® reagents were added to appropriate water dilutions and then sealed into Quanti-Tray®/ 2000 containers (IDEXX, Westbrook, ME). Samples on mTEC were incubated first at 35°C for two hours followed by 22 hours at 44.5°C. Both MI and Colilert® samples were incubated for 24 hours at 35°C. Positive and negative controls were incorporated with all three methods and included *E. coli* ATCC #11775 BioBalls (BTF, Sydney, Australia) and *E. faecalis* ATCC #19433 BioBalls (BTF). Each sample, media combination was processed in triplicate.

Enumerations and Isolations

After 24 hours of incubation, samples were enumerated and isolations initiated. Red-magenta colonies were considered typical *E. coli* on mTEC media while blue colonies were typical on the MI media. Enumeration results were recorded as

CFU/100mls. The Colilert Quanti-Tray®/ 2000 were exposed to long-wave UV light and yellow, fluorescent wells were scored positive. The number of positive wells and the sample dilution were used with the MPN generator program provided by the manufacturer and recorded as MPN/100mls. Media from five positive wells were aseptically removed from the tray, combined, diluted and filter plated following USEPA Method 1603 onto mTEC media for *E. coli* isolation.

Each sample was processed in triplicate and 5 isolates per replicate, for a total of 15 isolates per sample, were transferred onto EC-MUG (EMD, Gibbstown, NJ) media as a secondary screen for β-glucuronidase enzyme activity. Additionally, all of the isolates were also confirmed as *E. coli* following confirmation procedures in USEPA Method 1603 which included being oxidase negative, citrate negative, positive for gas production in EC broth, and being indole positive. A large portion of the isolates required secondary streaking onto EC-MUG media to confirm culture purity, but were all ultimately confirmed to be positive following the secondary confirmations. Isolates were preserved for long-term storage in tryptic soy broth with 20% glycerol and stored at -80°C. Additionally, cell suspensions of each isolate were made using a 1µL loop of cells into 100µL of sterile molecular grade water for downstream fingerprinting applications.

Fingerprinting

E. coli isolates were fingerprinted using the enterobacterial repetitive intergenic consensus polymerase chain reaction (ERIC-PCR) (Versalovic et al., 1991). PCR conditions were previously described by Casarez et al. (2007a). Each 50µL reaction

contained 1X PCR Buffer with 1.5 mM Mg (final) (ABI, Foster City, CA), 200 µM each of dNTP (GE Healthcare Biosciences, Piscataway, NJ), 600nM ERIC Primers 1R and 2 (Invitrogen, Carlsbad, CA), 1.5 µg/µl bovine serum albumin, 2.5 units AmpliTaq Gold (ABI, Foster City, CA), and 5µL of cell suspension (described above). Thermocycling was conducted in an Eppendorf Mastercycler (Hamburg, Germany) under the following conditions: initial denaturation at 95°C for 10 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 1 min, and extension at 72°C for 5 min with a final extension at 72°C for 10 min. Amplification products were stored at -20°C until analyzed by agarose gel electrophoresis. PCR products were loaded onto a 20x25 cm 2% agarose gel prepared with 1X TBE buffer and a 30 tooth, 1mm thick comb (IBI Scientific, Peosta, IA). Electrophoresis occurred in a cold storage room (~8°C) for 1 hour at 100V followed by 4 hours at 200V with buffer recirculation beginning after the first hour. Each gel included a marker ladder (Roche DNA Marker XIV) lane on the outside wells, as well as after every sixth sample. Additionally, a no-template control and a quality control *E. coli* strain (ATCC 51739) were included in every gel to ensure method reproducibility throughout the study. Gels were stained for 20 minutes in 1X TBE buffer containing 0.5 µg/mL of ethidium bromide. Gel images were captured using an UltraLum Omega 10gD Molecular Imaging system (UltraLum, Carlsbad, CA).

E. coli isolates were also fingerprinted using the automated DuPont Qualicon RiboPrinter (RP) system (DuPont Qualicon, Wilmington, DE). RP was performed according to manufacturer's instructions using HindIII restriction endonuclease (NEB, Ipscich, MA) with digestion performed at 37°C for 20 minutes. The RP software

normalizes the banding patterns using 4 marker lanes in each 8 isolate batch as well as reducing noise. The same quality control *E. coli* strain (ATCC 51739) used for ERIC-PCR was also included in every fourth batch. Results were converted to text files and exported for downstream processing.

Fingerprint Processing

BioNumerics software (Applied Maths, Austin, TX) was used to analyze the ERIC-PCR and RP fingerprints. ERIC-PCR fingerprint patterns were imported as tagged image file format (tiff) photos and evaluated using curve-based Pearson's product-moment correlation coefficients. RP files were imported and processed using a script provided by DuPont® and Applied Maths to score and weight the RP patterns. A composite dataset was used with equal weight given to both fingerprint types to analyze the communities. Dendrograms were constructed using the unweighted pair group method with arithmetic mean values (UPGMA) using an 80% similarity cutoff. The cutoff value was based on reproducibility of the fingerprint patterns, both ERIC-PCR and RP, of the quality control *E. coli* strain. Any pattern type or group whose composite similarity was equal to or greater than 80% was considered the same genotype (Casarez et al., 2007b). Once dendrograms and pattern similarities were determined, BioNumerics was also used to calculate community diversity indices.

Source Class Identifications Using the Texas E. coli *BST Library*

To identify sources of the *E. coli*, the fingerprints were queried against thee 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., 2007b). 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 domestic animals (including livestock and pets), domestic sewage and wildlife (three-way split) (Casarez et al., 2007b).

Statistical Analysis

Sigma Plot (Version 11.0) was used to analyze the enumeration data obtained from the six locations across three media types. ANOVA was used to test for differences between the *E. coli* count data from each media type by location. P-values <0.05 were considered to be significant.

Results

Media Type and E.coli *Concentration*

E. coli concentrations obtained from each of the six sites using all three media types are summarized in Table 3. Big Iron Ore Creek showed similar concentrations from the two membrane filtration methods but higher values with the Colilert®. While at Burton Creek, the MI concentrations were significantly higher than the mTEC and Colilert. Campbells Creek followed the Big Iron Ore Creek trend with significantly higher values with the Colilert method. But Moody Creek actually showed significantly lower concentrations with the Colilert method. Plum Creek was the only site that

showed a lack of significant difference between the media types, but high standard error with this data contributed to this lack of difference. White Oak Bayou was the only site which showed the MI concentrations to be significantly lower than the mTEC and Colilert values. The three different methods showed no clear trend in enumeration values across the six sites.

Table 3. Summary of the mean *E. coli* concentrations at each site using USEPA Method 1603 (mTEC), USEPA Method 1604 (MI) and Colilert® (CFU or MPN/100mls +/- SE).

	mTEC	MI	Colilert®
Big Iron Ore Creek	2.7E02 ± 2.3E01[a]	3.7E02 ± 5.8E00[a]	6.3E02 ± 5.9E01[b]
Burton Creek	2.9E03 ± 1.5E02[a]	3.9E3 ± 1.2E02[b]	2.8E03 ± 1.3E02[a]
Campbells Creek	9.9E03 ± 7.5E02[a]	9.5E03 ± 7.5E02[a]	1.4E04 ± 7.3E02[b]
Moody Creek	2.2E02 ± 8.1E00[a]	2.4E02 ± 1.1E01[a]	1.5E02 ± 2.3E01[b]
Plum Creek	5.5E02 ± 1.2E01[a]	5.7E02 ± 8.8E00[a]	7.8E02 ± 2.9E02[a]
White Oak Bayou	7.6E02 ± 5.4E01[a]	1.1E03 ± 4.2E01[b]	8.2E02 ± 3.1E01[a]

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

E. coli *ERIC-RP Diversity*

Community diversity index values, Simpson's and Shannon-Weiner, are summarized in Table 4. Communities from the mTEC and MI media were overall more diverse than the Colilert® communities. Moody Creek exhibited the lowest diversity of any of the sites followed by Plum Creek and White Oak Bayou. Moody Creek had the

lowest *E. coli* counts and the least diverse communities. A large portion of the genotypes, especially in mTEC and MI, contained only unique isolates.

Table 4. *E. coli* diversity estimates (ERIC-RP 80% similarity cutoff).

	mTEC		MI		Colilert®	
	Simpson's	Shannon-Weiner	Simpson's	Shannon-Weiner	Simpson's	Shannon-Weiner
Big Iron Ore Creek	93.33	2.18	91.43	1.99	75.25	1.27
Burton Creek	93.33	2.25	97.14	2.43	88.57	1.84
Campbells Creek	99.33	2.12	96.19	2.40	89.52	1.93
Moody Creek	13.33	0.25	36.19	0.63	13.33	0.25
Plum Creek	79.05	1.71	91.43	2.12	71.43	1.41
White Oak Bayou	90.48	2.03	80.00	1.62	84.76	1.90

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 5). 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).

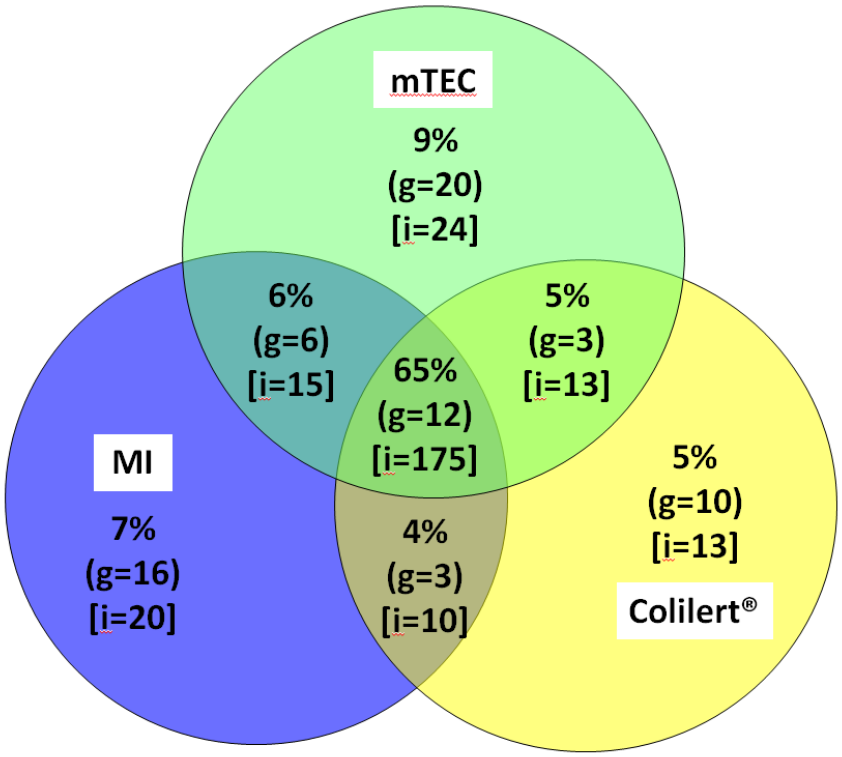

Figure 5. Overall *E. coli* community overlap across all three media types. Data represents the percentage of isolates with unique genotypes detected in each medium and when those isolates occurred in multiple media types. Data in parenthesis represents the number of unique genotypes (g) in each combination while the values in brackets [i]represent the number of isolates in each category.

The fingerprint data was also analyzed by specific site with each site having 45 isolates total, 15 in each media type (Table 5). 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 5.

Source Identifications Using the Texas E. coli *BST Library*

All 270 isolates were queried against the Texas *E. coli* BST library and identified to their closest match in a three-way split including humans, domesticated animals, and wildlife. The identifications were analyzed first broadly by media type alone (Figure 6). The mTEC isolates were identified as originating from wildlife in 54% of the isolates, 28% as domesticated animals, 3% as humans, and 14% were unidentified. Similarly, the MI isolates were identified as originating from wildlife in 64% of the isolates, 21% as domesticated animals, 3% as human and 11% were unidentified. And the Colilert isolates were identified as originating from wildlife in 41% of the isolates, 40% as domesticated animals, 7% as human, and 12% were unidentified. In all three media, wildlife and domesticated animals were classified as the main sources of the contamination. The percentage of unidentified isolates was similar for all three media.

Table 5. Overlap in ERIC-RP genotypes of *E. coli* across media.

	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%

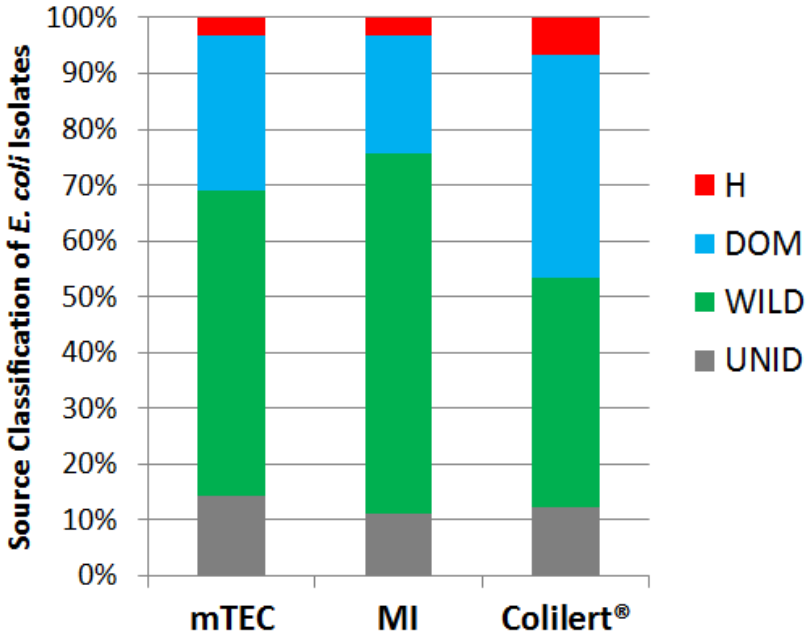

Figure 6. Identification of *E. coli* water isolates sorted by media type using a 3-way split for source classification (H=human, DOM=domesticated animals, WILD=wildlife, UNID=unidentified).

The source class identification results were also analyzed by site and are included in Figures 7 through 9. The results from the specific sites were much less consistent than when viewed broadly. Identifications at Big Cypress 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 7). 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 7). 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 8). Classifications at Moody Creek were not very consistent across media types as all of the isolated from the MI media classified as originating from wildlife (Figure 8). 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 9). 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 9). 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 then either wildlife or domesticated animals (Figure 10).

Discussion

Monitoring fecal indicator bacteria is a necessary requirement for assessing water quality. Understanding what effects various methods have on not only the enumeration values, but also the community makeup as a whole is important for downstream applications especially BST. This study looked to evaluate whether two popular water quality methods, USEPA Method 1604 (MI) and Colilert®, could be incorporated with procedures currently used to construct the Texas *E. coli* BST library, USEPA

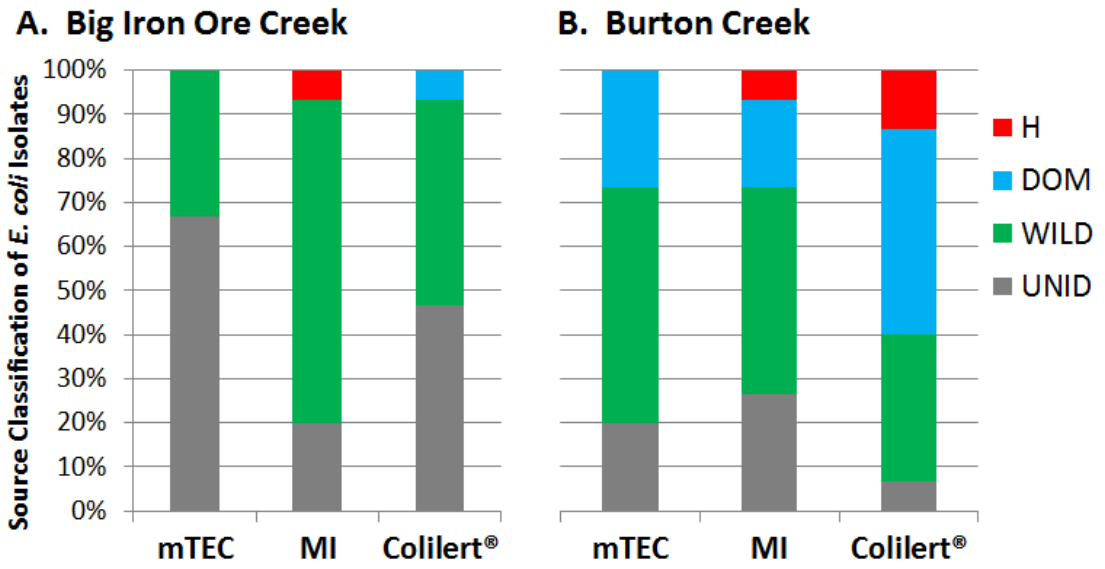

Figure 7. Identification of water isolates at Big Iron Ore Creek (A) and Burton Creek (B) using a 3-way split for source classification (H=human, DOM=domesticated animals, WILD=wildlife, UNID=unidentified).

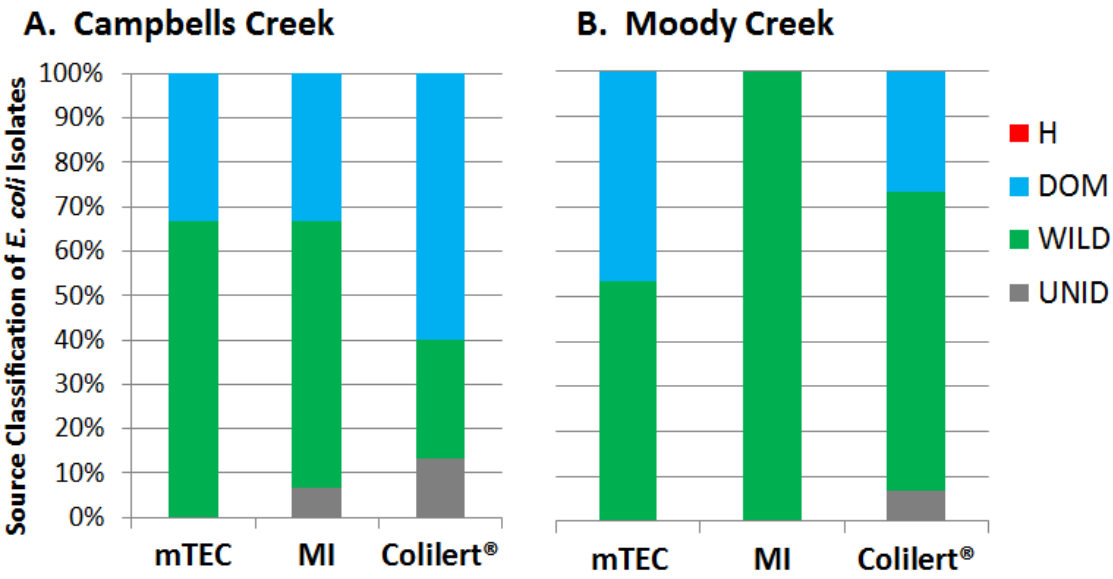

Figure 8. Identification of water isolates at Campbells Creek (A) and Moody Creek (B) using a 3-way split for source classification (H=human, DOM=domesticated animals, WILD=wildlife, UNID=unidentified).

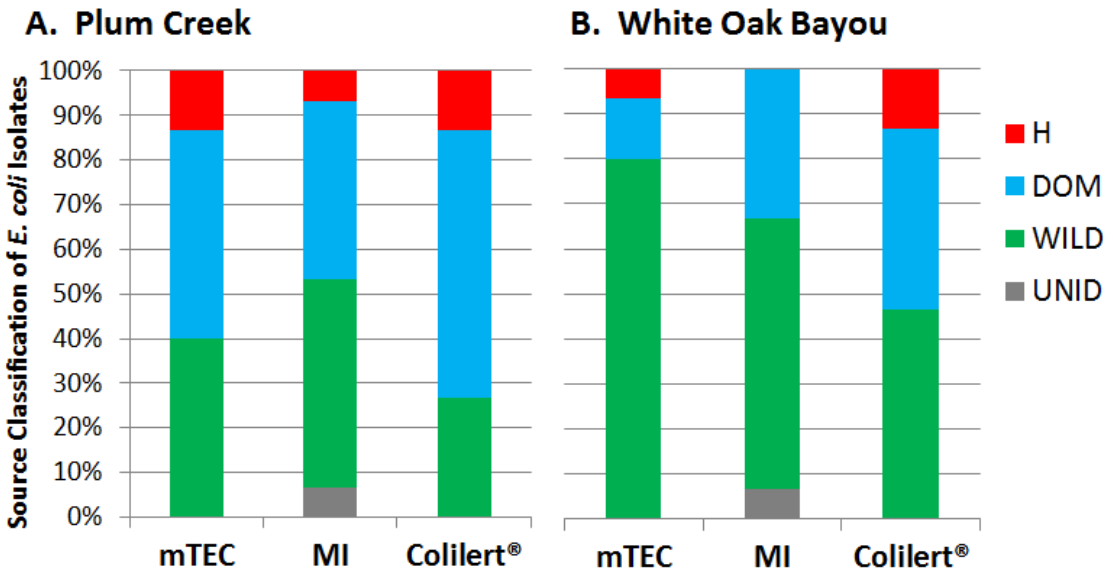

Figure 9. Identification of water isolates at Plum Creek and White Oak Bayou using a 3-way split for source classification (H=human, D=domesticated animals, W=wildlife, U=unidentified).

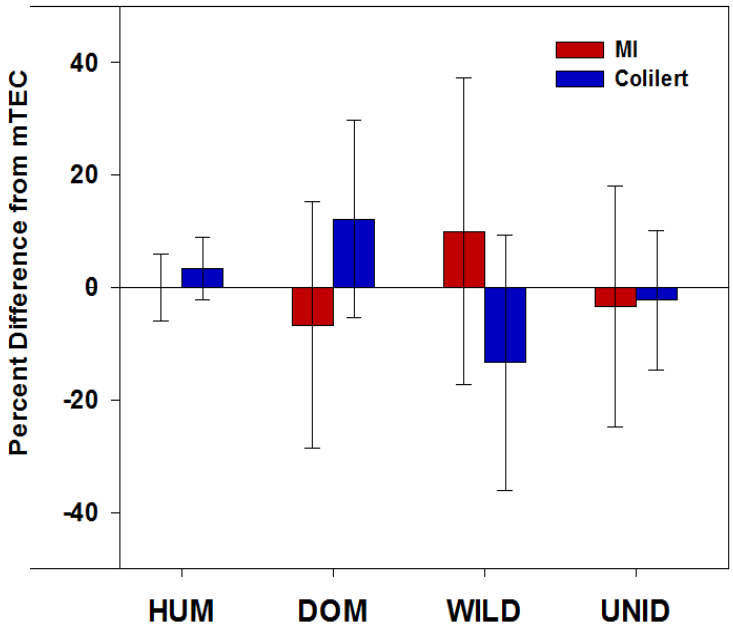

Figure 10. Percentage difference of *E. coli* source classifications from MI and Colilert® compared to mTEC across all sites. Error bars represent standard error of three replicate samples using a 3-way split for source classification (HUM=human, DOM=domesticated animals, WILD=wildlife, UNID=unidentified).

Method 1603 (mTEC). The ability to utilize multiple enumeration techniques for downstream BST applications would greatly simplify water quality monitoring projects by eliminating the need for either duplicate sample processing or switching to the method that was used to construct the Texas *E. coli* BST Library (EPA Method 1603).

Much of the previous research with community characterization and *E. coli* enumeration media involved identifying organisms which may cause false positive *E. coli* reactions due to a lack of specificity in the β-glucuronidase enzyme which also have been found in organisms such as *Salmonella*, *Shigella*, and *Yersinia* sp. (Pisciotta et al., 2002; Rompre et al., 2002). Culture-independent analysis of false-positive *E. coli*

Colilert® wells have shown that Colilert tends to select for a large number of non-target taxa, including *Vibrio* spp. (Sercu et al., 2011). Despite state and government acceptance as well as evidence that some water quality assessment methods yield statistically comparable data, the enumeration values for this data set were not consistent across media type or site (Buckalew et al., 2006; Fricker et al., 1997; Kodaka et al., 2008). Many of the previously published studies did not directly compare the three media types evaluated in this study and caution should be taken to extrapolate those conclusions across other media without confirmation, especially EPA Method 1604 (MI), which is a newer method and none of the reported studies found in a literature search included this medium in their analysis (Olstadt et al., 2007). Even though this was a one-time sampling event at six locations, this data suggests that these methods do not yield consistent enumeration results.

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 EPA Method 1603 confirmation protocol (Brenner et al., 1993). 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.

Community analysis was conducted using the same similarity cutoff of 80% that was used to build the Texas *E. coli* BST library since ultimately the results were queried against the library to identify the major source class of the isolates (Casarez et al., 2007b). This cutoff is the lower boundary for an isolate to be considered a match in the library and served as a complement to the library identifications. The diversity indices were in general agreement at all of the sites with the mTEC and MI communities being the most diverse and the Colilert® community being the least diverse. It was interesting to note that the Moody Creek site which had the lowest average *E. coli* counts and also had the least diverse communities, but the other sites showed comparable diversity values regardless of *E. coli* count.

The main objective of this study was to evaluate whether the additional methods, MI and Colilert, selected for *E. coli* communities similar to those isolated on mTEC which was used to build the Texas *E. coli* BST Library. A similarity analysis was conducted on the data first as a whole with all 270 isolates regardless of site (Figure 5). This interpretation of the data seemed very positive from a community overlap perspective, as 65% of the isolates were seen in all three media types. But a closer examination by site showed a single genotype accounted for 31% of the total isolates. The Moody Creek site was the least diverse and 87% of the isolates from that site fell into this genotype. Even though one genotype accounted for a large number of the

overall isolates, the total diversity of the community was highlighted by the number of singleton genotypes (54% of total) in the dataset (data not shown).

The community overlap, genotypes occurring in multiple media types, across the sites was extremely variable. The overall overlap in the communities ranged from only 16% at Moody Creek up to the 87% at Moody Creek. Plum Creek and White Oak Bayou maintained a high number of overall genotypes, but a few of those accounted for a majority of the overlapping genotypes, where at Burton, Big Iron Ore, and Campbells Creeks less than 30% of the isolates were seen in all three media types. Further, no other combination, much less a complete overlap of the three media accounted for greater than 56% of the total isolates at any site other than Moody Creek further suggesting the selection of different *E. coli* communities by the three media. Both the enumeration and community analysis results at Moody Creek could be explained by the nature of its location. The site is located on a wildlife refuge and cattle operation with little to no human input and would theoretically be the least complex of sampling sites studied which was indicated in the lower *E. coli* counts and more simple communities at this location.

Reasons for the general lack of community similarity could be explained by several factors. MI and Colilert® methods are designed to enumerate both total coliforms and the more specific *E. coli* and are incubated at 35°C where 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 total coliform group. Both the media and temperature

are more specific and thus likely would select for 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. TRFLP analysis of positive Colilert® wells by Sercu et al. (2011) suggests that Colilert may select for large populations of non-target taxa like *Vibrio* and *Clostridiales*. Even though *E. coli* may be present in the well, and therefore it scores positive for the purposes of enumeration, these *E. coli* are potentially competing against the non-target organisms which could certainly explain the lack of diversity.

The source identifications of the dataset followed a similar pattern to the community analysis in that when the dataset was viewed as a whole, the results seemed to indicate 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 EPA 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.

Conclusions

The results of this study indicate that the three evaluated *E. coli* enumeration methods may select for different populations of *E. coli*. Even though the Moody Creek site seemed to contain the least complex *E. coli* community across all three media types, the other sites were characterized as having diverse, but different populations 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 the same organisms, total confirms and *E. coli*, 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.

CHAPTER IV

IMPACT OF MANURE MANAGEMENT ON *E. COLI* ABUNDANCE AND DIVERSITY IN DAIRY MANURE AND LAGOON WATER

Introduction and Rationale

Non-point source agricultural runoff has been implicated as a potential source of fecal contamination across many watersheds in the United States (Meals and Braun, 2006; USEPA, 2003b). Concentrated animal feeding operations like dairies produce copious amounts of animal manure on a daily basis (McGarvey et al., 2004). Cattle manure has been associated with pathogenic organisms including *E. coli* O157:H7, *Salmonella* spp., *Campylobacter* spp., and *Cryptosporidium* spp. (Bicudo and Goyal, 2003; Hill, 2003). Management of these materials can be very advantageous to the producer as a source of nutrients and organic matter for crop and hay production, but requires cautious utilization to limit over-applying nutrients and minimize both nutrient and bacterial movement in runoff to nearby waterways (Cook et al., 2011; Koelsch et al., 2006).

Dairy operations are unique among many livestock production groups in that manure and cleanup is often required more than once per day and thus requires significant time and resource allocation (Fulhage, 1997). Dairy operators select manure management strategies based on factors such as location, size, type and use of cropland, the number of animals on the farm, and type of animal housing (Huber, 2003). Manure management systems are generally classified as solid, slurry, or dilute. Solids are

handled mechanically while slurry and dilute waste waters are handled using pumps and pipelines (Graves, 1994; Huber, 2003). Solids separation is a commonly used treatment process that makes handling of the manure easier using irrigation equipment, allows for recycling of water, especially with dairies utilizing a flushing method to clean manure from barn alleyways. Further, the separation of the bedding material can potentially be re-used or composted and land applied (Graves, 1994). The separation of solids from the manure stream also serve as a means to reduce the organic matter and nutrient content of the manure, prolong the life of downstream storage structures, and minimize odors (Huber, 2003). Most dairies are designed to ultimately collect and transfer manure to a lagoon system, potentially aerobic but most often anaerobic, or storage pond as a liquid or slurry. Lagoons are designed for biological treatment to reduce nitrogen levels in the wastewater but leave high levels of phosphorus and other nutrients in the sludge (Fulhage, 2001).

Even though nutrient management is a significant goal in managing the dairy manure, bacteria and pathogen levels are also of concern for both the health of the dairy herd and dairy personnel as well as potential movement off-farm and into local waterways (Pell, 1997). Assessing the survival of pathogens and fecal indicators in the manure waste stream, and their potential regrowth once land applied are difficult at best due to the myriad of storage and environmental conditions that exist on a daily basis (Bicudo and Goyal, 2003; Pell, 1997). McGarvey et al. (2004) characterized bacteria populations from raw manure, separator pit and lagoon from a free-stall dairy in California. Total bacterial numbers were assessed using brain heart infusion agar and

were estimated to be 2.9 x 10^9 CFU/g raw manure, 1.9 x 10^6 in the separator pit water, and 2.8 x 10^5 in lagoon water. 16S rRNA clone libraries from the same sites showed the manure communities were the most diverse followed by slurry and lagoon communities. The authors proposed that the bacterial populations at the downstream sites, especially the lagoons, were well adapted. Bacterial abundance and community dynamics are not well characterized in dairy manure management waste streams and could be extremely informative to dairymen in their efforts to maximize nutrient and water returns from the waste stream while minimizing their environmental footprint.

Dairy operations have been implicated in the Leon watershed in central Texas as a potential source of fecal contamination. The Leon River below Proctor Lake (Segment 1221) was first placed on State of Texas Clean Water Act §303(d) List of impaired waters in 1996 due to bacterial contamination (TCEQ, 2008). A watershed protection plan has since been completed and is available in draft form (http://www.brazos.org/LeonRiverWPP.asp). Watershed modeling data, as well as BST from the watershed, implicated direct discharge from wildlife and livestock (including waste disposal fields) as substantial sources of bacteria across the watershed. Several concerns arose during the modeling phase, especially the validity of using historical data to predict sources of contamination and their specific load reductions as well as a lack of watershed specific data. Dairy producers in the watershed lack detailed information on *E. coli* abundance throughout their waste management streams. This portion of a larger project aimed to survey commonly used dairy manure management practices in the watershed in order to address several research objectives: (1) identify management

strategies most efficient at lowering *E. coli* abundance in the waste stream, (2) characterize the *E. coli* manure communities across the various management streams using BST techniques to identify whether strains are common across techniques or even dairies, (3) include a temporal group of isolates from each management strategy to examine potential shifts in the *E. coli* community and their possible effects for downstream BST analysis and finally, (4) designate a portion of the isolates fingerprinted to include in the Texas *E. coli* BST library.

Methods

Site Descriptions

Four dairies participated in the study and represented a diverse set of dairy manure management schemes in Central Texas (Figure 11). Dairy A was a traditional dairy that utilized manual barn scraping of manure from the pens into piles prior to land application and a lagoon. Dairy B utilized a flushing system where water was used to flush the barn alley ways into a settling basin to remove some solids and ultimately into a lagoon. Dairy C utilized manual scraping of manure into piles as well as a separator which separated the solids including bedding material and sent the liquid materials down into a lagoon. Dairy D utilized a vacuum to collect raw manure into a slurry pond to gravity separate manure solids prior to liquids flowing into a lagoon; the slurry solids were spread onto an adjacent drying (bed) prior to land application. All of the lagoons were open air with no covering and exposed to daily ambient weather conditions.

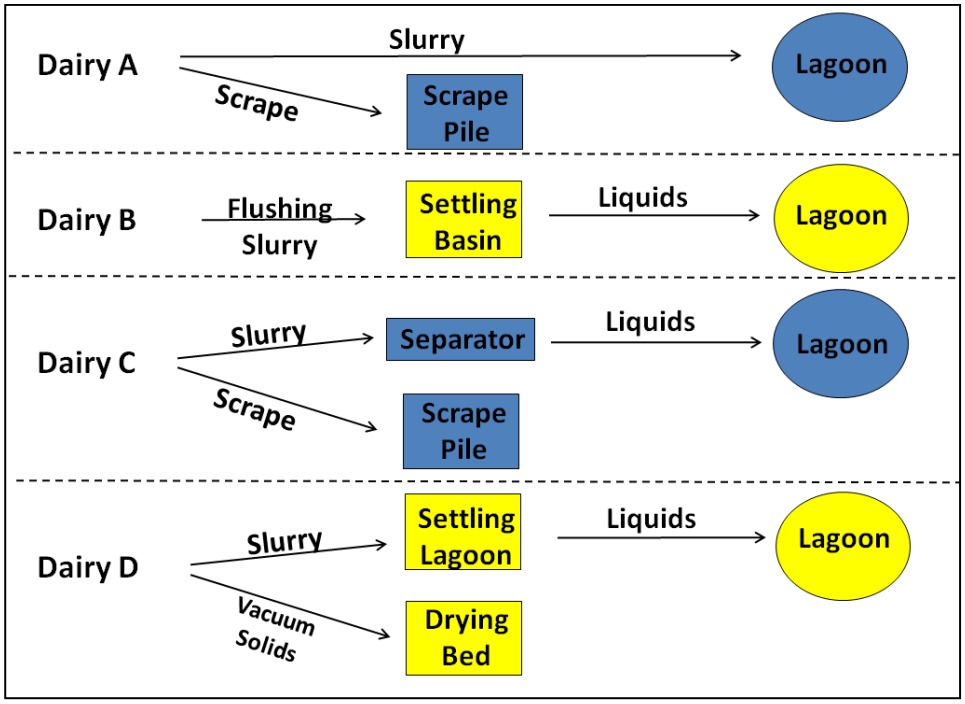

Figure 11. Schematic of dairy manure management schemes utilized at the four dairies.

Monthly Manure Sampling and Initial Processing

Manure samples were collected monthly from all four dairies from January through December of 2009. Dairy A underwent a herd reduction in May of 2009 and stopped actively milking, so scrape sample collection ended in June, but lagoon sampling continued throughout the project. Five to ten representative subsamples were collected at random from each sampling location and mixed to form a composite sample. In order to obtain temporally representative results, including wet and dry conditions and seasonal variation, samples were generally collected during the first week of each month by a combination of Texas AgriLife Stephenville and College Station representatives.

Samples were collected into sterile Whirl-Pak® Bags or autoclaved Nalgene® bottles and were transported on ice back to the lab in Stephenville for processing. Manure processing for initial *E. coli* enumerations was completed with within 6 hours of sample collection following EPA Method 1603 (modified mTEC) per method and manufacturer's instructions (USEPA, 2005a).

Enumerations and Isolations

After 24 hours of incubation, samples were enumerated. Red-magenta colonies were considered typical *E. coli* on mTEC media. Enumeration results were recorded as CFU/gram of wet manure or mL of lagoon wastewater. If not delivered directly, within 48 hours of processing, mTEC plates were shipped on ice overnight to College Station for isolation.

E. coli from quarterly samples were isolated and archived for downstream processing, including the months of January, April, July and September. When possible, from each sample, 5 typical *E. coli* were isolated and transferred from mTEC onto EC-MUG (EMD, Gibbstown, NJ) medium as a secondary screen for β-glucuronidase enzyme activity. Pure culture isolates were preserved for long-term storage in tryptic soy broth with 20% glycerol and stored at -80°C. Additionally, cell suspensions of each isolate were made using a 1µL loop of cells into 100µL of sterile molecular grade water for downstream fingerprinting applications.

Fingerprinting

E. coli isolates were fingerprinted using the enterobacterial repetitive intergenic consensus polymerase chain reaction (ERIC-PCR) (Versalovic et al., 1991). PCR

conditions were previously described by Casarez et al. (2007a). Each 50µL reaction contained 1X PCR Buffer with 1.5 mM Mg (final) (ABI, Foster City, CA), 200 µM each of dNTP (GE Healthcare Biosciences, Piscataway, NJ), 600nM ERIC Primers 1R and 2 (Invitrogen, Carlsbad, CA), 1.5 µg/µL bovine serum albumin, 2.5 units AmpliTaq Gold (ABI, Foster City, CA), and 5µL of cell suspension (described above). Thermocycling was conducted in an Eppendorf Mastercycler (Hamburg, Germany) under the following conditions: initial denaturation at 95°C for 10 min, followed by 35 cycles of denaturation at 94°C for 30 s, annealing at 52°C for 1 min, and extension at 72°C for 5 min with a final extension at 72°C for 10 min. Amplification products were stored at -20°C until analyzed by agarose gel electrophoresis. PCR products were loaded onto a 20x25 cm 2% agarose gel prepared with 1X TBE buffer and a 30 tooth, 1mm thick comb (IBI Scientific, Peosta, IA). Electrophoresis occurred in a cold storage room (~8°C) for 1 hour at 100V followed by 4 hours at 200V with buffer recirculation beginning after the first hour. Each gel included a marker ladder (Roche DNA Marker XIV) lane on the outside wells, as well as after every sixth sample. Additionally, a no template control and a quality control *E. coli* strain (ATCC 51739) were included in every gel to ensure method reproducibility throughout the study. Gels were stained for 20 minutes in 1X TBE buffer containing 0.5 µg/mL of ethidium bromide. Gel images were captured using an UltraLum Omega 10gD Molecular Imaging system (UltraLum, Carlsbad, CA).

Select January 2009 *E. coli* isolates were also fingerprinted using the automated DuPont Qualicon RiboPrinter (RP) system (DuPont Qualicon, Wilmington, DE). RP was performed according to manufacturer's instructions using HindIII restriction

endonuclease (NEB, Ipswich, MA) with digestion performed at 37°C for 20 minutes. The RP software normalized the banding patterns using 4 marker lanes in each 8 isolate batch as well as reduced noise. The same quality control *E. coli* strain (ATCC 51739) used for ERIC-PCR was also included in every fourth batch. Results were converted to text files and exported for downstream processing.

Fingerprint Processing

BioNumerics software (Applied Maths, Austin, TX) was used to analyze the ERIC-PCR and RP fingerprints. ERIC-PCR fingerprint patterns were imported as tagged image file format (tiff) photos and evaluated using curve-based Pearson's product-moment correlation coefficients. RP files were imported and processed using a script provided by DuPont® and Applied Maths to score and weight the RP patterns. ERIC-PCR dendrograms were constructed using the unweighted pair group method with arithmetic mean values (UPGMA) using an 85% similarity cutoff. The cutoff value was based on reproducibility of the ERIC fingerprint patterns of the quality control *E. coli* strain using the single ERIC-PCR fingerprinting method. Any pattern type or group whose composite similarity was equal to or greater than 85% was considered to be the same genotype (Casarez et al., 2007a; Casarez et al., 2007b). Once dendrograms and pattern similarities were determined, BioNumerics was also used to calculate community diversity indices. Relative abundance of genotypes across dairy management schemes was also used to generate nonmetric multidimensional scaling (NMDS) plots using the Bray-Curtis similarity measure in PAST (version 2.05) (Hammer et al., 2001). Graphs were generated using Sigma Plot 11.0.

Source Class Identification Using the Texas *E. coli* BST Library

The January 2009 dairy manure isolates from all four dairies were chosen for screening for potential inclusion into the Texas *E. coli* BST library. Additionally, 9 feral hog and 9 coyote fresh fecal samples were collected by local stakeholders when the animals were killed or trapped on their property. These samples were processed using the same techniques as the dairy manure samples to form a small local watershed library. First, ERIC-PCR was performed on approximately 5 *E. coli* isolates per sample. Isolates within each sample were screened using an 80% cutoff value to identify clonal isolates per state library established SOPs. Any two isolates with greater than 80% similarity were considered clonal in nature and a representative isolate was chosen from that sample to RP. The small local watershed library was used to first self-validate the isolates. To be included in the library, the isolates had to either have a best match of greater than 80% back to the correct 7-way split categories (human, pets, cattle, other avian livestock, other non-avian livestock, avian wildlife, and non-avian wildlife) or their best match in the local library was less than 80% and were left unidentified. At least one self-validated isolate per sample was included in the library (Casarez et al., 2007b).

Results

Manure E. coli *Levels*

Levels of *E. coli* for each of the 11 total samples are summarized in Table 6 and Figure 12. Across dairies and management strategies, average *E. coli* counts were

Table 6. Summary statistics of levels of *E. coli* (CFU/mL or g) at different stages in manure treatment. (Data is the average of samples collected over one year.)

Dairy-Treatment	n	Mean	Std Dev	Std. Error	Min	Max	Range
A-Scrape	5	7.01E+04	5.57E+04	2.49E+04	1.00E+03	1.35E+05	1.34E+05
A-Lagoon	12	1.55E+04	2.96E+04	8.56E+03	6.00E+00	1.00E+05	1.00E+05
B-Settling Basin	12	1.43E+05	3.08E+05	8.88E+04	1.00E+02	1.10E+06	1.10E+06
B-Lagoon	12	3.15E+04	9.10E+04	2.63E+04	6.50E+02	3.20E+05	3.19E+05
C-Scrape	12	3.92E+05	3.08E+05	8.90E+04	1.00E+02	7.50E+05	7.50E+05
C-Separator Solids	12	1.42E+05	1.24E+05	3.59E+04	1.00E+02	3.35E+05	3.35E+05
C-Separator Liquids	10	1.29E+05	1.23E+05	3.90E+04	7.00E+02	4.20E+05	4.19E+05
C-Lagoon	12	4.64E+03	9.66E+03	2.79E+03	5.00E+01	3.50E+04	3.50E+04
D-Drying Bed	12	9.55E+05	1.45E+06	4.19E+05	1.00E+03	5.35E+06	5.35E+06
D-Settling Lagoon	12	2.92E+02	5.27E+02	1.52E+02	1.00E+00	1.85E+03	1.85E+03
D-Lagoon	12	1.54E+02	1.79E+02	5.16E+01	1.30E+01	6.50E+02	6.37E+02

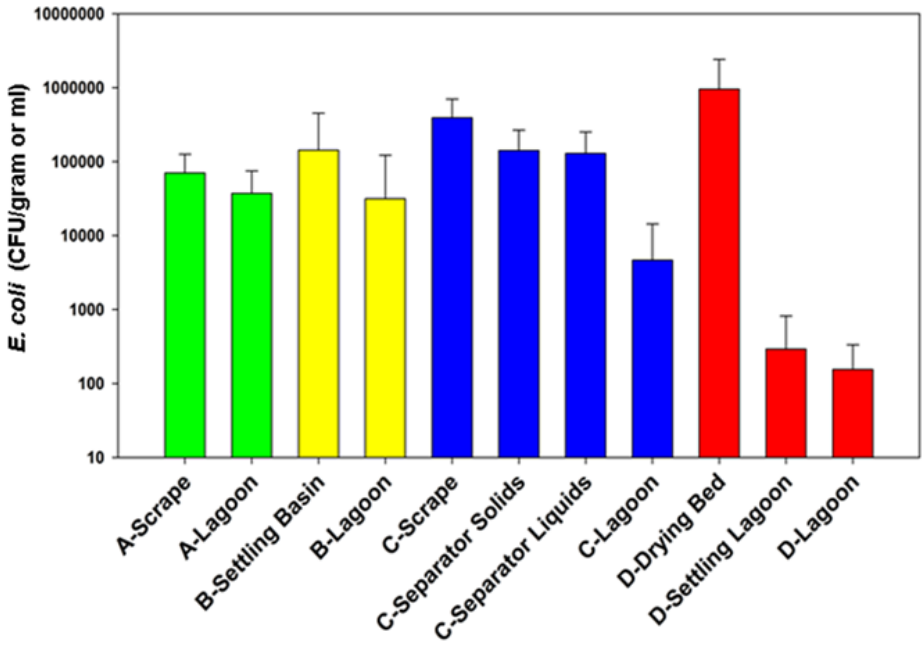

Figure 12. Average *E. coli* abundance (CFU/g or mL) for each of the dairy manure management strategies and their respective treatment stages. (Data represents the average of all samples collected at dairies A-D over one year.)

highest in the scrape piles and continued to drop as the manure migrated through the management stream ultimately into the lagoons. The two-tiered lagoon system from Dairy D yielded the lowest *E. coli* levels seen throughout the project with a mean of 154 CFU/mL of wastewater in the final treatment lagoon. Dairy A Lagoon samples were continued throughout the study, even after the herd reduction, and within two months of the cessation of fresh manure additions (from May to July), *E. coli* levels were at or below the level of detection (10 CFU/mL). Even though many of the management techniques were not significantly different due to variation in the data, practically speaking the level of *E. coli* was reduced as the manure moved through the management

schemes. Manure and effluent handling methods which employed means to remove solids from the manure tended to decrease the level of *E. coli* in effluent.

Total E. coli *Isolates Fingerprinted*

A total of 194 *E. coli* isolates from samples collected each quarter were fingerprinted using ERIC-PCR to assess community structure and are summarized in Table 7. For Dairy A, the scrape isolates were only available for January and April due to the herd reduction. For some months, 5 *E. coli* could not be confirmed or archived and thus could not be fingerprinted. Community diversity indices were calculated for each of the various management strategies and are shown in Table 8. Generally, all of the samples were very diverse, Shannon-Weiner values near 100 and Simpson's greater than zero, even when grouped across the study. Most of the individual samples with 5 isolates fell into greater than 3 unique genotypes per sample. The only exception was the Dairy C Separator Solids isolates from January where 4 of the 5 were from 1 unique genotype.

E. coli *Community Analysis*

NMDS: NMDS analysis (71 total isolates) was performed to visualize the relationship of the *E. coli* ERIC-PCR genotypes across the lagoon communities at all four dairies and time points (Figure 13). No apparent trends were seen in the community groupings by site or by time.

Table 7. Total quarterly *E. coli* isolates fingerprinted using ERIC-PCR.

Dairy-Treatment	Jan09	April09	July09	Sept09	Total
A-Scrape	5	5	0	0	10
A-Lagoon	5	5	0	5	15
B-Settling Basin	5	5	5	5	20
B-Lagoon	4	5	5	4	18
C-Scrape	5	5	5	5	20
C-Separator Solids	5	5	5	4	19
C-Separator Liquids	5	5	5	5	20
C-Lagoon	5	5	5	5	20
D-Drying Bed	5	5	5	5	20
D-Settling Lagoon	5	5	4	0	14
D-Lagoon	5	3	5	5	18
Total	54	53	44	43	194

Table 8. *E. coli* community diversity estimates at different stages of dairy manure treatment (ERIC-PCR 85% Similarity Cutoff).

Dairy-Treatment	Isolates	Shannon-Weiner	Simpson's
A-Scrape	10	93.3	1.90
A-Lagoon	15	90.5	2.00
B-Settling Basin	20	96.3	2.60
B-Lagoon	18	95.4	2.50
C-Scrape	20	91.6	2.21
C-Separator Solids	19	92.4	2.23
C-Separator Liquids	20	93.2	2.37
C-Lagoon	20	97.4	2.70
D-Drying Bed	20	98.4	2.80
D-Settling Lagoon	14	95.6	2.24
D-Lagoon	18	97.4	2.58

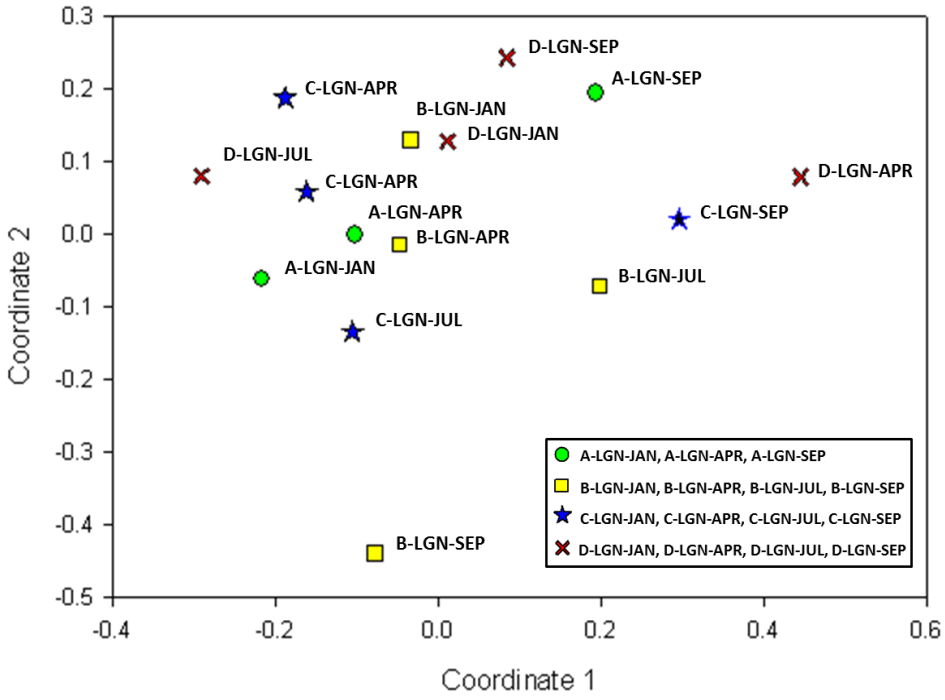

Figure 13. NMDS plot (Bray-Curtis) of relative abundance of ERIC-PCR genotypes across all lagoons and quarterly time points. (Letters designate dairies A, B, C, D; LGN-Lagoon; JAN-January, APR-April, JUL-July, and SEP-September).

The lagoon samples from each of the dairies were also analyzed for community overlap across each dairy (Table 9). There were 71 total isolates from all 4 dairies and they represented 37 unique genotypes using an 85% similarity cutoff. There were a total of 15 possible combinations of overlaps between the 4 dairies. A total of 65% of the genotypes were only found at one individual dairy and did not represent any overlap in the communities analyzed. However, one pattern type that was seen in all four dairies represented 15% of the total isolates.

Table 9. Overlap in *E. coli* ERIC-PCR genotypes across lagoon samples from all four dairies.

	Unique Genotypes		Total Isolates	
Dairy	Number	Percentage	Number	Percentage
A	6	16	9	13
B	6	16	6	8
C	7	19	8	11
D	5	14	6	8
A & B	0	0	0	0
A & C	0	0	0	0
A & D	0	0	0	0
B & C	2	5	4	6
B & D	4	11	9	13
C & D	4	11	8	11
A,B,C	1	3	6	8
A,B,D	0	0	0	0
A,C,D	0	0	0	0
B,C,D	1	3	4	6
A,B,C,D	1	3	11	15
Totals	37	100	71	100

Dairy A: Dairy A had the fewest number of total isolates fingerprinted due to the herd reduction during the project, as well as a lack of lagoon isolates in July, but a total of 25 isolates were fingerprinted (Table 10). Combined, there were 13 unique genotypes in the Dairy A samples with only 32% of the isolates occurring in both the lagoon and scrape samples. There were no clear temporal trends in any of the Dairy A samples across months (data not shown).

Table 10. Overlap in *E. coli* ERIC-PCR genotypes across Dairy A isolates.

	Unique Genotypes		Total Isolates	
Treatment	Number	Percentage	Number	Percentage
Lagoon	6	46	11	44
Scrape	5	38	6	24
Lagoon and Scrape	2	15	8	32
Totals	13	100	25	100

Dairy B: Dairy B had a total of 38 isolates fingerprinted (Table 11). The Dairy B site was very diverse, as 24 of the 38 isolates were unique genotypes. The settling basin and lagoon shared 26% of the isolates, but the majority was unique to each management type. There were no clear temporal trends in the Dairy B lagoon or settling basin isolates as none of the isolates or genotypes were seen in a substantial portion of isolates in any of the seasons (data not shown).

Table 11. Overlap in *E. coli* ERIC-PCR genotypes across Dairy B isolates.

	Unique Genotypes		Total Isolates	
Treatment	Number	Percentage	Number	Percentage
Lagoon	10	42	13	34
Settling Basin	10	42	15	40
Lagoon and Settling Basin	4	16	10	26
Totals	24	100	38	100

Dairy C: Isolates at Dairy C exhibited a higher percentage of overlap than the other dairies (Table 12). Even though 58% of the unique genotypes were from individual samples, 42% of the total number of isolates came from 3 genotypes that were seen in all four sampling sites at Dairy C. Another 13% of the isolates overlapped in the separators and lagoon. When the sampling sites were characterized by date, there were no clear temporal trends or overlaps with time of sampling (data not shown).

Dairy D: Dairy D had 30 unique genotypes from 52 individual isolates from the three management techniques utilized (Table 13). Isolates from individual sampling sites made up 64% of the isolates. But the lagoon and drying bed as well as the settling lagoon and drying field also accounted for 26% of the overlapping genotypes and 42% of the total isolates characterized.

Addition of Isolates to the Texas E. coli *BST library*

The January 2009 isolates were chosen for potential inclusion into the Texas *E. coli* BST library. A small watershed local library of 30 total samples, including 12 dairy manure samples, 9 coyote and 9 feral hog samples was created. A total of 146 isolates were ERIC fingerprinted and following de-cloning 72 isolates were also RP fingerprinted. Using the small local library of 72 isolates total for the self-validation, 58 total isolates (36 dairy, 13 coyote, 9 feral hog) were self-validated against the small local library and added to the Texas *E. coli* BST Library (ver. 10-11+LRSV; 1365 isolates from 1215 samples). Since the local library was limited to just the two major source classes, cattle and wildlife, the isolates that failed the self-validation test were misidentified to the opposite source class. Of the 36 dairy isolates fingerprinted, 32

Table 12. Overlap in *E. coli* ERIC-PCR genotypes across Dairy C isolates.

	Unique Genotypes		Total Isolates	
Treatment	Number	Percentage	Number	Percentage
Lagoon	5	18	5	6
Scrape	5	18	7	9
Separator Liquids	5	18	5	6
Separator Solids	1	4	1	1
Lagoon and Scrape	2	7	6	8
Lagoon and Separator Liquid	1	4	2	3
Lagoon and Separator Solids	0	0	0	0
Scrape and Separator Liquids	0	0	0	0
Scrape and Separator Solids	1	4	2	3
Lagoon and Separator Solids	0	0	0	0
Separator Liquids and Separator Solids	1	4	2	3
Lagoon, Scrape and Separator Liquids	0	0	0	0
Lagoon, Scrape and Separator Solids	1	4	3	4
Lagoon, Separator Liquids, and Separator Solids	2	7	10	13
Scrape, Separator Solids, and Separator Liquids	1	4	3	4
Lagoon, Scrape, Separator Solids and Liquids	3	11	33	42
Totals	28	100	79	100

Table 13. Overlap in *E. coli* ERIC-PCR genotypes across Dairy D isolates.

	Unique Genotypes		Total Isolates	
Treatment	Number	Percentage	Number	Percentage
Lagoon	8	27	9	17
Settling Lagoon	5	17	5	10
Drying Bed	6	20	6	12
Lagoon and Settling Lagoon	1	3	2	4
Lagoon and Drying Bed	4	13	10	19
Settling Lagoon and Drying Bed	4	13	12	23
Lagoon, Settling Lagoon, and Drying Bed	2	7	8	15
Totals	30	100	52	100

(89%) were self-validated as their closest match was to another dairy isolate in the local library. Further, 90% (28 of the 32) of the isolates best match was to a different dairy. There were no clear trends as to whether they matched back to a similar level of treatment or not (liquid versus solid).

Discussion

As increasing numbers of streams are listed on the Texas Water Quality Inventory and 303(d) list for bacterial contamination, the need to adequately assess sources and loads of bacteria will become even more important. Bacterial impairments are often addressed using a TMDL based largely on modeling efforts. Confirmation in addition to watershed specific details on source loading and allocations are critical for decreasing the uncertainty associated with these types of analysis (Guber et al., 2011; Harmel et al., 2010). This study aimed to increase the knowledge of manure

management strategies in the Leon watershed and their ability to reduce the abundance of fecal indicator organisms and thus the potential for watershed contamination via runoff from application fields. Further, this study offered a unique opportunity to incorporate established BST techniques to further characterize *E. coli* communities from these various management techniques in order to assess similarities and differences across the waste streams as well as get some indication for potential temporal shifts in those communities which might confound downstream BST efforts.

Monthly Manure Sampling

This study utilized the same methodological approach used by surface water quality managers to enumerate *E. coli* in manure to produce directly comparable results to recreational water quality standards (USEPA, 2005a) as well data needed for TMDL models (Teague et al., 2009). All four of the dairies utilized different levels of manure management from the more traditional Dairy A which utilized a manual scrape and lagoon system, to Dairy B which used the flushing system with a settling basin and ultimately a lagoon, and then more complex Dairy C which employed a separator prior to lagoon, and finally Dairy D which a vacuum system followed by a two-tiered lagoon.

The overall counts were highest in the more raw manures (scrape) followed by the secondary treatments (separators and settling basins) and the lowest in the lagoons with significant reductions at Dairies C and D. Even though direct comparisons with previous studies are difficult to interpret side-by-side due to differences in enumeration methods, these results mirror results from previous studies from dairies in San Joaquin Valley California where aerobic plate counts, anaerobic plate counts, and coliform

bacteria dropped two orders of magnitude after transitioning from raw manure into the slurry and then another order of magnitude, down to an average of 2.8 x 10^5, in the lagoon wastewater (McGarvey et al., 2004). At Dairy C, the lagoon samples averaged two orders of magnitude less (10^3 CFU/mL) than the scrape piles (10^5 CFU/g) as well as the separator liquids (10^5 CFU/mL) and solids (10^5 CFU/mL) prior to treatment. At Dairy D, both the settling lagoon (10^2 CFU/ml) and lagoon (10^2 CFU/mL) averaged three orders of magnitude lower than the drying field (10^5 CFU/g).

The *E. coli* counts at the dairy A lagoon were not significantly different from the scrape piles sampled through the first five months of the study. It was positive to note the drop in *E. coli* abundance to below levels of detection 60 days after the herd reduction and cessation of milking activities (data not shown). Dairy B overall showed the greatest variation in *E. coli* counts, but that may have been attributed to the nature of their management system. The settling basin was often a mixture of solids and liquids depending on when the last flush had occurred, and the lagoon at this dairy may have seen more total water volume flux due to the flushing and this re-using of the water than the other dairies which yielded wider ranges of counts. This same reasoning may attribute to the lack of statistical difference between counts at the settling basin and lagoon. Dairy C scrape and separator solids and liquids means were statistically the same even though the separator means were numerically lower than the raw scrape values. This is logical since the separator takes the raw materials from the barns and separates the solid materials to be composted and land applied and sends the liquid materials straight to the lagoon where a significant reduction was seen in the *E. coli*

abundance. The lagoons at Dairy D yielded the lowest *E. coli* data in the study. As the settling basin filled, the solids were spread onto an adjacent drying field to desiccate and dry prior to land application. The drying bed received a mixture of slurry solids and fresh manure and consistently yielded the highest *E. coli* counts seen across the study. Re-growth when spread onto the field prior to desiccation is not beyond the realm of possibility as the counts were significantly higher in the drying field than the slurry pond (Sinton et al., 2007).

There were no clear seasonal trends in the monthly count data (data not shown) that might account for the variability in a portion of the *E. coli* enumeration data. Dairies produce fresh manure every day, 365 days per year, and the manure stream is in constant flux from routine activities. Dungan et al. (2012) noted no significant differences in abundance of cultivated *E. coli* from dairy wastewater over a three month spring and summer study. Continuous addition of manure may have masked any temperature-driven trends as fresh manure is created daily and all of the sampling sites are thus in constant flux (Hutchison et al., 2005). Variability in these results can be attributed to a large number of factors that ultimately also alter the fate and transport of the bacteria in the environment including pH, moisture, nutrient content, and temperature as well as age of the manure, age and diet of the cattle (Callaway et al., 2009; Dungan et al., 2012; Franz et al., 2005; van Elsas et al., 2011). Since attempting to monitor that many factors was not feasible in a field study, the sampling strategy was to sample at the same time each month over a calendar year to hopefully capture a normal breadth of routine conditions manure at these dairies normally undergoes.

E. coli *ERIC-PCR Genotype Analysis*

The goal of fingerprinting the *E. coli* communities from various management schemes over time was to utilize previously established methods used to build the Texas BST *E. coli* library (Casarez et al., 2007a; Casarez et al., 2007b; Di Giovanni et al., 2010) in order to attempt to assess diversity in the fingerprint communities as well as potential patterns or shifts in the *E. coli* community based on dairy manure management schemes and time of sample collection.

There were no clear seasonal or site grouping trends seen among the lagoon ERIC-PCR genotypes across all of the dairies (Figure 13). Nor were there any strong grouping trends when assessing individual dairies across treatment and time (data not shown). This approach looked at relative abundance of the genotypes across dairy-treatment-time to see whether the communities were similar or not. Although analysis of a greater number of isolates for each dairy-treatment-time combination would likely have increased the ability to detect overlap in the various *E. coli* communities, the results based upon small-to-moderate-sized isolate collections suggest a large amount of variability in *E. coli* genotypes at each site over time. Others have likewise reported variability in *E. coli* fecal genotypes. For example, Jenkins et al. (2003) assessed various *E. coli* ribotypes in feedlot steers over a one-year period and concluded that only 8% of the ribotypes could be considered resident ribotypes. Further, Anderson et al. (2005) used antibiotic resistance and ribotyping to analyze *E. coli* communities from horses, cattle, and humans and found that most of the host populations were not stable over time.

Overall, the ERIC-RP genotype analysis showed commensurate levels of diversity throughout the study. For a majority of the individual samples 3 or 4 of 5 of the isolates represented a unique genotype which even at an 85% similarity cutoff is comparable to previous known-source sample characterizations. The same approach that has been used to build the Texas *E. coli* BST Library was used in this case, fingerprinting 5 isolates per sample, but clonal isolates were not removed for this study as the aim was to look at genotype-level diversity across the manure management schemes .

The community overlap analysis combined the dairy-treatment combinations over time to increase total number of isolates assessed at each combination. The results showed some overlap in the communities across management schemes. For the lagoon samples overall, 65% of the unique genotypes were from individual dairies. But 60% of the total isolates were seen in some combination of the four dairies with 15% seen in all four dairies. Dairy C showed the most commonality of all of the dairies. All four of the management schemes at this dairy shared 3 unique genotypes and 42% of the total isolates.

The majority of all samples, including dairy samples, currently represented in the state library are from individual animals. Lagoon samples would presumably be representative of the entire waste stream as well as the dairy manure most likely to be land applied and thus most apt to enter waterways. Similarly, human waste is generally represented in the library from domestic sewage collected from wastewater treatment plants or septic tanks for the same reasoning. The NMDS plots did not indicate the

communities were very similar, but when combining the dataset and looking strictly at genotypes that occurred in all of the dairies, there was one common genotype across all four dairies that accounted for 15% of the total isolates. Furthermore, genotypes that shared at least two dairies accounted for another 44% of the total isolates characterized. These results show promise for using lagoon samples versus individual cow samples from dairy operations in the future. McGarvey et al. (2004) showed relative abundance of dominant bacterial taxa shifted in the manure at various stages and that the lagoon communities were less diverse than the raw manure and separator communities. The authors deemed these communities to be 'adapted to the environment'. Lagoons are essentially large biological reactors used to recycle nutrients in the manure and even though their microbial community may undergo turnover with season or cleanout, it would not be surprising to learn that there is a 'core bacterial community' that includes *E. coli* populations that are relatively stable, including the most abundant cultivable portion of the community utilized in library-dependent BST. These samples would be more representative of the herd as a whole as well as represent the portion of the manure stream most likely to be land applied and thus have more possibility of reaching surface water.

The January isolates were chosen for screening and inclusion into the state library. Those isolates were de-cloned (80% similarity cutoff) and the unique isolates also RP fingerprinted. The more stringent 85% cutoff was used in the ERIC-PCR genotype analysis due to the use of only one fingerprinting technique, instead of the ERIC-RP combination, which has a reproducibility cutoff of 85% for the QC strain for

E. coli. A more liberal cutoff value would have decreased the diversity and increased the amount of overlap seen in the communities. Per current statewide library building methods, a small watershed library was built with available fecal samples and included the dairy manure samples as well as coyotes and feral hogs. The isolates could only be included in the library if their best match was at least 80% similar to their 6-way source specific category or were less than 80% similar to any of the local library members and left unidentified.

Of the 42 total dairy isolates (from twelve individual dairy samples) that were ERIC-RP fingerprinted, 36 (86%) were self-validated by the local known-source library. Even more impressive from a library-performance perspective was that 89% of these isolates best-matches in the local library were to *E. coli* isolated from a different dairy. This is a good indication that despite the variability in *E. coli* genotypes across and within the different dairies over time, that the use of library-dependent BST and the local library was sufficiently robust, both geographically and temporally, to correctly classify the vast majority of dairy *E. coli* isolates from this project.

Conclusions

Results of this study showed that the surveyed dairy manure management systems are successful at reducing *E. coli* loads in their waste streams. The two dairies with the most advanced waste management systems, Dairies C and D, showed the greatest reductions in *E. coli* abundances. The two-tiered system at Dairy D yielded the lowest abundances of *E. coli* seen throughout the study, but the separator system prior to

movement into the lagoon at Dairy C was also effective. Although there were no clear seasonal or site grouping trends seen among the *E. coli* genotypes across all of the dairies, some genotypes were common across managerial treatments and the communities across the dairies generally showed as much similarity as within each dairy. Furthermore, successful classification of a vast majority of the dairy *E. coli* isolates in the local library indicated the reliability of the library-dependent BST approaches currently being used to track *E. coli* contributions from dairy manure in Central Texas watersheds.

CHAPTER V

MICROBIAL CHARACTERIZATION OF DEER FECAL COMMUNITIES IN TEXAS

Introduction

Molecular advances have allowed for a tremendous expansion of library-independent BST methods. These approaches offer the promise for a more rapid and cost-effective means of assessing fecal contamination back to a particular source or source-group in a culture-independent manner (USEPA, 2005b; Wuertz et al., 2011). Intestinal communities of warm-blooded hosts have been shown to exhibit both coevolution and codiversification which strongly supports the hypothesis that there are source-specific bacterial lineages (Ley et al., 2008; Ley et al., 2006; Wuertz et al., 2011). A large portion of library-independent methods have looked to develop genetic markers for host-specific bacterial populations (Wuertz et al., 2011). Many of these reported host-specific BST markers come from members of the *Bacteroidales* order and assays (both presence/absence as well as quantitative) have been developed targeting humans, ruminants, cattle, hogs, geese and ducks, and horses (Bernhard and Field, 2000; Dick et al., 2005; Hamilton et al., 2006; Layton et al., 2006; Seurinck et al., 2005; Shanks et al., 2008; Shanks et al., 2006). Even though host-specific markers are being developed, they are generally still limited, especially for wildlife which severely restricts our ability to specifically track this important group. Many of these markers have come under substantial criticism generally due to a lack critical performance evaluation of sensitivity

and specificity, as well as a lack of data quality assurance (Harwood and Stoeckel, 2011; Shanks et al., 2010).

Several approaches have been taken to identify host-specific BST markers including length-heterogeneity PCR and terminal restriction fragment length polymorphism (Bernhard and Field, 2000), suppression subtractive hybridization (Hamilton et al., 2006) and genome fragment enrichment (Shanks et al., 2006). But advances in next generation sequencing technology, including 454 pyrosequencing, have substantially enhanced our ability to characterize entire communities at a reduced cost and allowed a more broad look at the entire 16S rRNA community instead of targeting specific groups. In Texas, wildlife sources, such as deer and feral hogs, have been implicated as major contributors of bacterial impairment, but our fundamental knowledge of wildlife gut communities and thus ability to track them as specific contamination sources is lacking. Library-independent means to track deer specifically 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. The objective of this study was to use 454 barcoded pyrosequencing to characterize deer fecal communities in Texas in an effort to evaluate their suitability for development of a deer-specific BST marker.

Materials and Methods

Sample Collection

Deer fecal pellets were obtained directly from the lower section of the large intestine at time of field dressing of recently killed animals. Samples were collected at the Welder Wildlife Refuge Foundation, near Sinton Texas, in both 2008 and 2009 during annual youth hunts. Welder staff members assisted in weighing the animals prior to field dressing and approximated the age of the animals. Leon samples from Comanche County, TX were collected by stakeholders in the watershed during the winter of 2009. Samples were kept on ice during handling and transport to College Station and stored at -80°C.

Bacterial Tag-Encoded Amplicon Pyrosequencing

Bacterial community DNA was extracted in triplicate from each sample using a Power Soil DNA isolation kit (Mo Bio Laboratories, Inc., Carlsbad, CA) per manufacturer's instructions. DNA samples were purified using illustra MicroSpin™ G-25 Columns (GE Healthcare Biosciences, Pittsburg, PA) and quantified using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE). Extracts were stored at -20°C for downstream applications. Community DNA was submitted to the Research and Testing Laboratory (Lubbock, TX) for tag-pyrosequencing. Samples were amplified using primers 27F and 519R and sequenced using Roche titanium chemistry (Acosta-Martinez et al., 2008).

Sequence Analysis and Community Comparisons

Sequence libraries were analyzed using a combination of The Ribosomal Database Project (RDP) (Cole et al., 2009) (accessed 17 May 2011) and MOTHUR (version 1.18.1) (Schloss et al., 2009). Using MOTHUR, sequencing primers and tags were removed, the database was quality checked, and chimeras removed prior to downstream processing. The RDP pipeline was used to assign taxonomic identities to the quality-screened, final 454 sequence data. Each sequence was classified down to the genus level, but if an organism could not be classified with at least 80% confidence in RDP, it was named Unclassified at the previous scientific classification level.
The dist.seqs function in MOTHUR was used to create distance matrices and then assign sequences to operational taxonomic units (OTUs, 97% similarity). Diversity estimates were calculated including Shannon's and Simpson's diversity indices and Chao I richness estimates. For community comparisons, the samples were grouped into three treatments by location (Welder and Leon) and year (2008 or 2009). Phylogenetic structure of the libraries was assessed using the Yue-Clayton index (Theta-YC) as this approach utilizes both incidence and relative abundance of OTUs and thus is not sensitive to sample size. Parsimony test, analysis of molecular variance (AMOVA), and homogeneity of molecular variance (HOMOVA) were conducted using the Theta-YC similarities. Parsimony, ANOVA, and HOMOVA tests with P-values <0.05 were considered to be significant (Schloss, 2008). Relative abundance of OTUs across each sample was also used to generate nonmetric multidimensional scaling (NMDS) plots using PAST (version 2.05) (Hammer et al., 2001). Graphs were generated using Sigma

Plot 11.0. Nearest-neighbor joining trees were created in MEGA (version 5.10) to showcase overlapping and the most abundant OTUs using representative sequences from the deer fecal communities and their closest GenBank hits (Altschul et al., 1990) (accessed 18 May 11).

Results

Deer Physical Characteristics

The physical characteristics of the deer samples collected are summarized in Table 14. Detailed information was available for the Welder samples from both 2008 and 2009, but was not available for the Leon samples. The Welder Wildlife refuge is home to large herds of deer and the samples collected varied in age and weight for both years.

Table 14. Deer physical descriptions.

Site	Sample ID	Sex	Age (yrs)	Live Weight (kg)
Welder 2008	81	Male	4.5	47
	84	Female	5.5	42
	87	Female	7.5	46
	88	Female	1.5	23
Welder 2009	91	Female	4.5	44
	92	Female	3.5	44
	93	Female	4.5	48
	95	Male	1.5	38
Leon	L1	N/A	N/A	N/A
	L2	N/A	N/A	N/A
	L3	N/A	N/A	N/A

Community Composition, Diversity, and Estimated Richness

A total of 32,163 amplicon sequences were utilized in the analysis with an average sequence library of 2,923 ± 610bp (mean ± sd) (Table 15). The sequence libraries ranged in size from 1,948 sequences in sample 91 to 3825 sequences in sample 81. The samples contained between 525 OTUs in sample 93 and 1559 OTUs in sample 92. Chao richness estimates suggest that the sequencing efforts captured approximately half of the diversity within the samples and additional sequencing would most likely yield additional OTUs in each sample. Shannon and Simpson diversity index values suggest similar diversity across the samples except for sample 93 which was the least diverse of all the samples.

Table 15. Summary of sequence library size, OTUs, and diversity and richness estimates

Sample	Sequence Library Size	Number of OTUs	Chao I Richness Estimate	Shannon H'	Simpson D
81	3825	1388	2595	6.63	0.99
84	2304	1168	2893	6.59	0.99
87	2825	1358	3092	6.73	0.99
88	3262	1298	2916	6.55	0.99
91	1948	903	2014	6.30	0.99
92	3508	1559	3593	6.78	0.99
93	2047	525	1269	4.97	0.97
95	3368	1145	2301	6.17	0.99
L1	3198	1327	3029	6.51	0.99
L2	2752	1019	1929	6.33	0.99
L3	3126	1148	2317	6.42	0.99
Overall	32163	8956	-	-	

Community Structure

The parsimony test showed no significant difference in the community structure overall between the Welder samples in 2008, 2009 and the Leon samples (P=0.457). Similarly, the analysis of molecular variance (AMOVA) also showed no significant differences in the three communities (P=0.129). The test for homogeneity of molecular variance (HOMOVA) did show significant differences between the three communities (P=0.007). Pairwise comparisons showed significant differences between Welder 2009 and Leon samples (P<0.001) but no significant difference between Welder 2008 and 2009 (P=0.135) and Leon and Welder 2008 (P=0.054). NMDS plots of all 11 samples show the communities grouping across one axis but not necessarily by location or time (Figure 14). However, the three Leon samples grouped together more prominently than did the Welder samples from either year. The Yue-Clayton similarity values were very high (ranging from 0.81 to 0.991) indicating that the bacterial communities in all of the samples were very similar (Yue-Clayton estimator is scored on a 0 to 1 scale with 0 representing complete dissimilarity and 1 representing complete similarity) (Table 16).

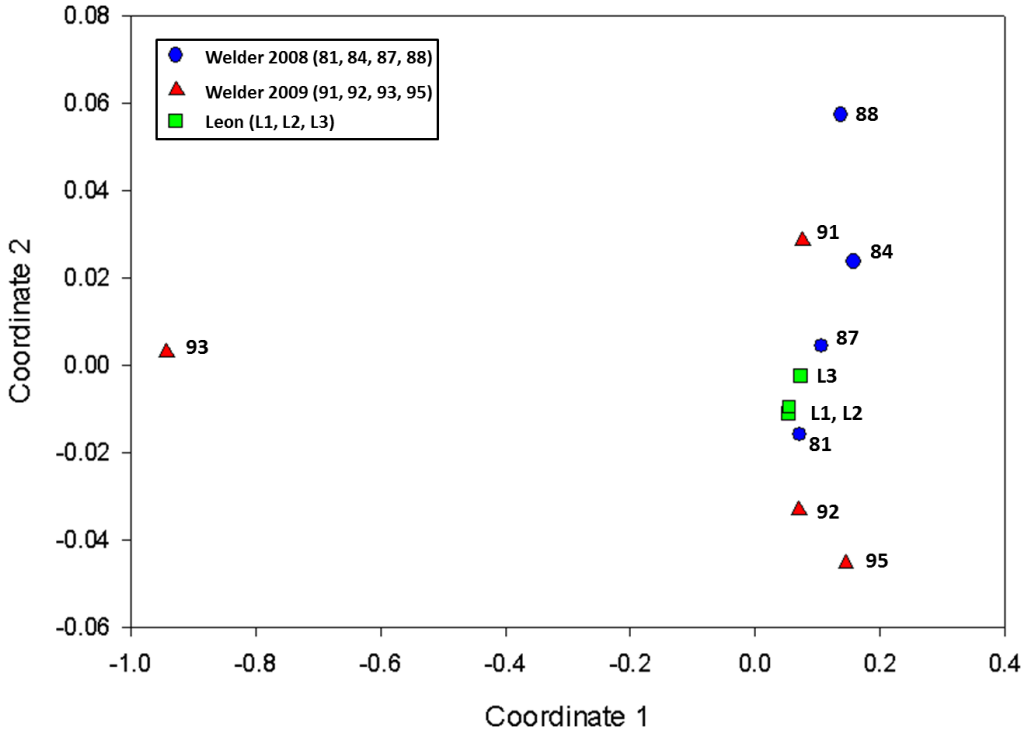

Figure 14. NMDS plot (Bray-Curtis) of 11 deer bacterial fecal communities based on relative abundance of all OTUs (97% similarity).

Table 16. Yue-Clayton similarities based on OTUs (97% similarity) of all 11 deer bacterial fecal communities.

Sample	81	84	87	88	91	92	93	95	L1	L2	L3
81											
84	0.9722										
87	0.9216	0.9096									
88	0.9667	0.9459	0.9210								
91	0.9510	0.8937	0.9536	0.9608							
92	0.9198	0.9718	0.9565	0.9751	0.9658						
93	0.9825	0.9892	0.9718	0.9844	0.9841	0.9808					
95	0.8732	0.8761	0.9662	0.9851	0.9689	0.9771	0.9906				
L1	0.9500	0.9718	0.9545	0.9605	0.9502	0.9402	0.9793	0.9714			
L2	0.9112	0.8266	0.9188	0.9568	0.9391	0.9321	0.9685	0.9698	0.8989		
L3	0.9477	0.9140	0.9449	0.9576	0.9356	0.9455	0.9637	0.9753	0.9026	0.8212	

Phylogenetic Classifications

The entire sequence database (32,163 sequences) was identified to their highest potential taxonomic level using RDP classifier. To account for variable numbers in each sequence library, relative abundances of each particular taxonomic level were calculated. At the order level, *Clostridiales* and *Bacteroidales* dominated all eleven of the samples accounting for 90 to 97% of the total community (Figure 15). *Clostridiales* ranged from 56 to 93% of the community composition across samples while *Bacteroidales* ranged from 4 to 41% (Figure 15). At the family level, abundances for major taxonomic levels are shown as a heat map relative to the average with hierarchical clustering by site (Figure 16). The sites did not appear to cluster at the family level by particular location or year. Again, Welder sample 93 appeared to be the most distant of the group due to an enrichment of *Bacteroidales* members and a depletion of *Ruminococcaceae* (Figure 16).

Abundant and Overlapping OTUs and GenBank Hits

Of the 8,956 total overall OTUs, 2 were seen in all 11 communities. Referencing the RDP classifier data, these OTUs were identified as members of the *Ruminococcaceae* and *Veillonellaceae* families. Representative sequences from these OTUs were chosen and a neighbor joining tree was created using their best GenBank hits (Figure 17). The five most abundant OTUs represented 12% of the total OTUs across samples and included members of the *Rikenellaceae*, *Porphyromonadaceae*, *Bacteroidales* (Unclassified), and *Ruminococcaceae* families (Figure 18). The most commonly shared and abundant OTUs along with their top GenBank hit and maximum identity are summarized in Table 17. *Ruminococcaceae* family members dominated

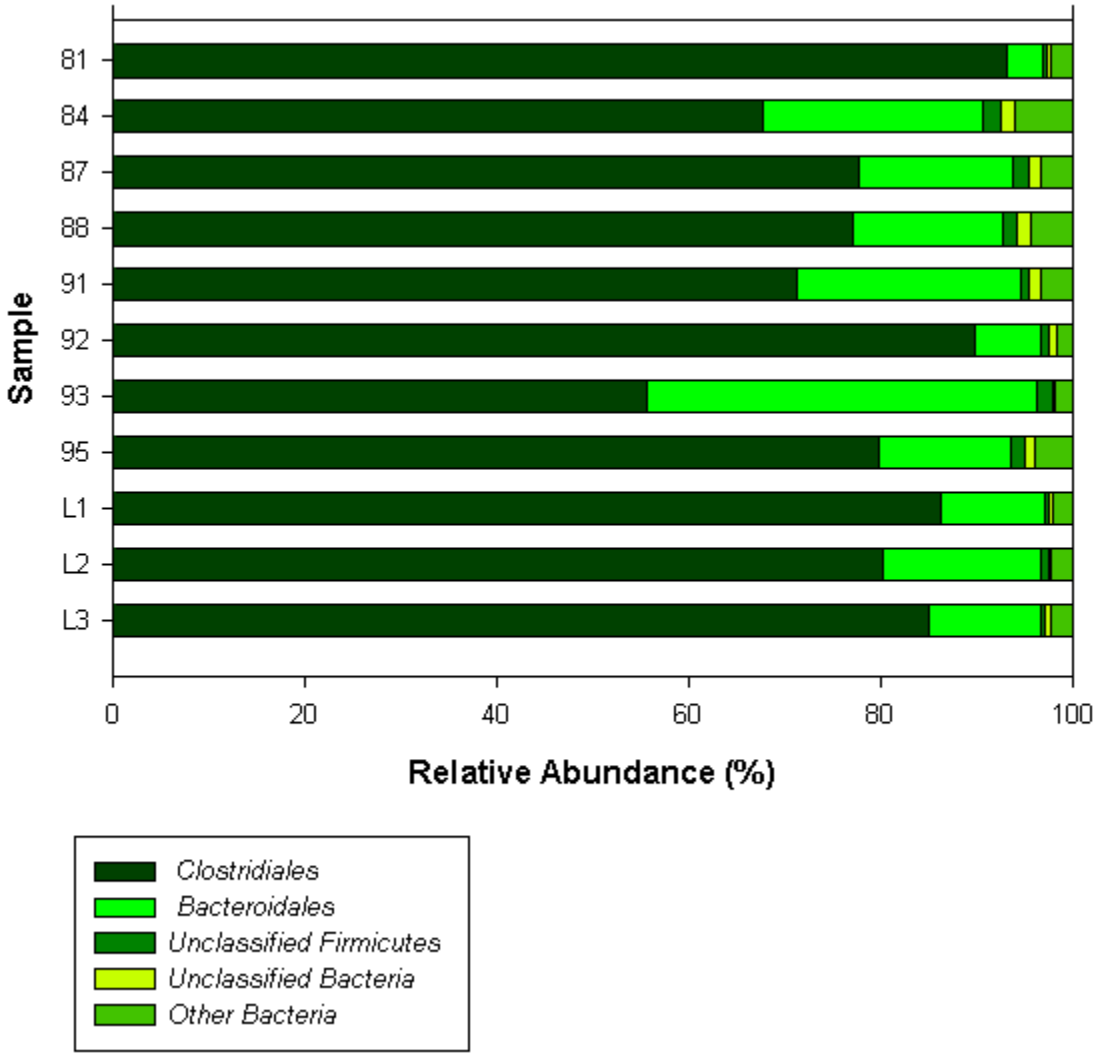

Figure 15. Bacterial composition across all 11 deer fecal samples at the order level.

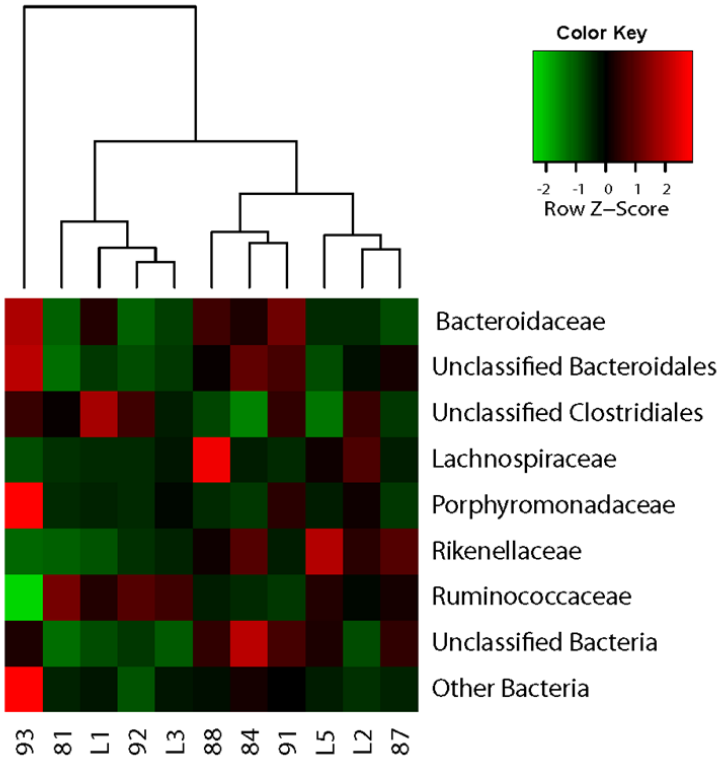

Figure 16. Heatmap depicting family-level taxonomic relative abundance across all 11 deer fecal communities. Abundances for each taxonomic group (row) were scaled relative to the mean across all samples and are depicted by color (red=above average, green= below average).

both the common and most abundant OTUs in the dataset (Table 17). All but one (OTU_2593) of the OTUs top hits were to uncultured bacterial clones originating from fecal communities. The top GenBank hits for the *Ruminococcaceae* family OTUs were to dairy cattle, humans, and a miniature gazelle.

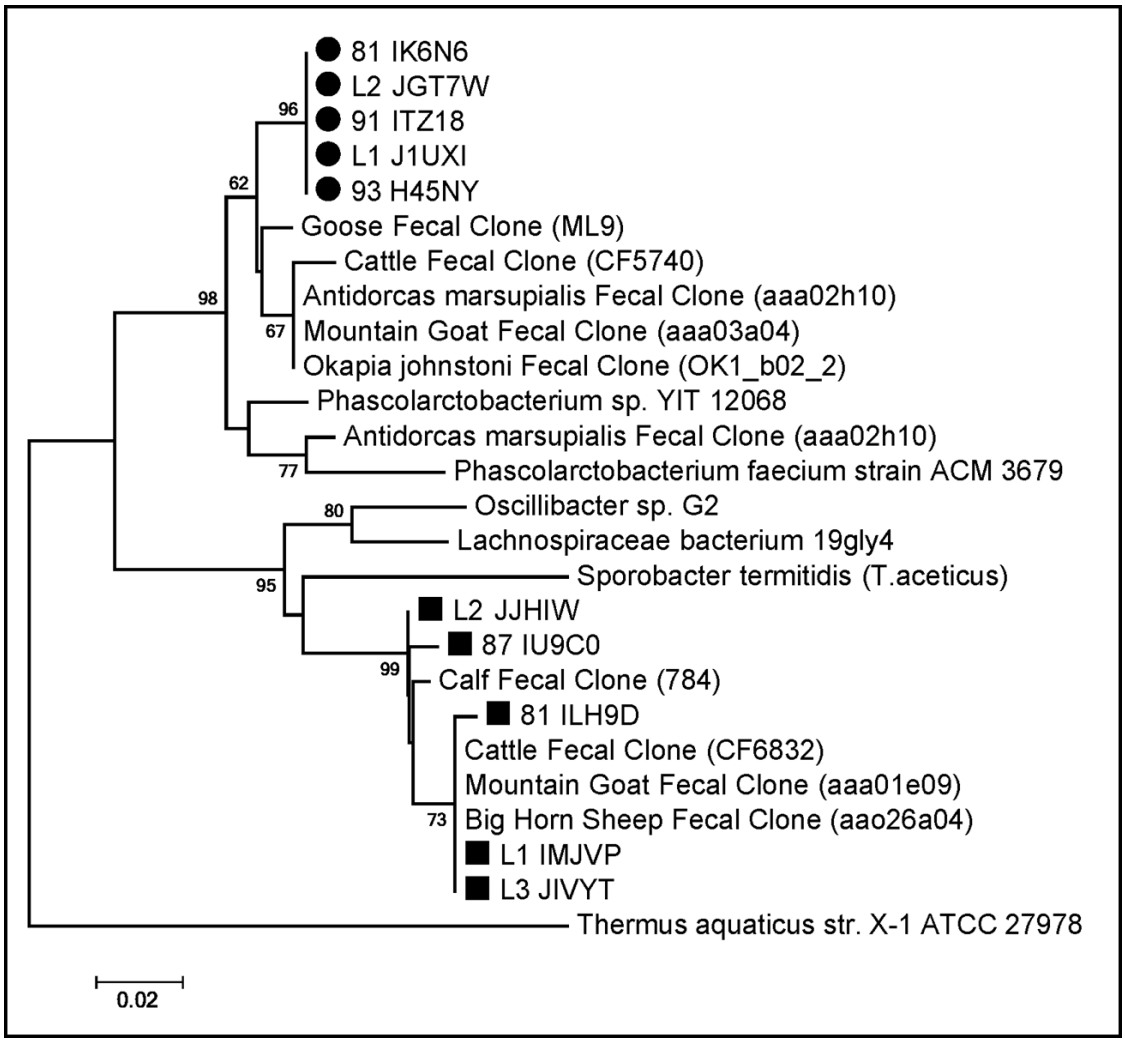

Figure 17. Neighbor-joining phylogenetic tree of the two OTUs represented in all 11 of the deer fecal communities and their top GenBank hits. Deer fecal community sequences representing the two OTUs are indicated with either a black circle representing *Ruminococcaceae* or a black square representing *Veillonellaceae* followed by their site name and five digit individual sequence identity code.

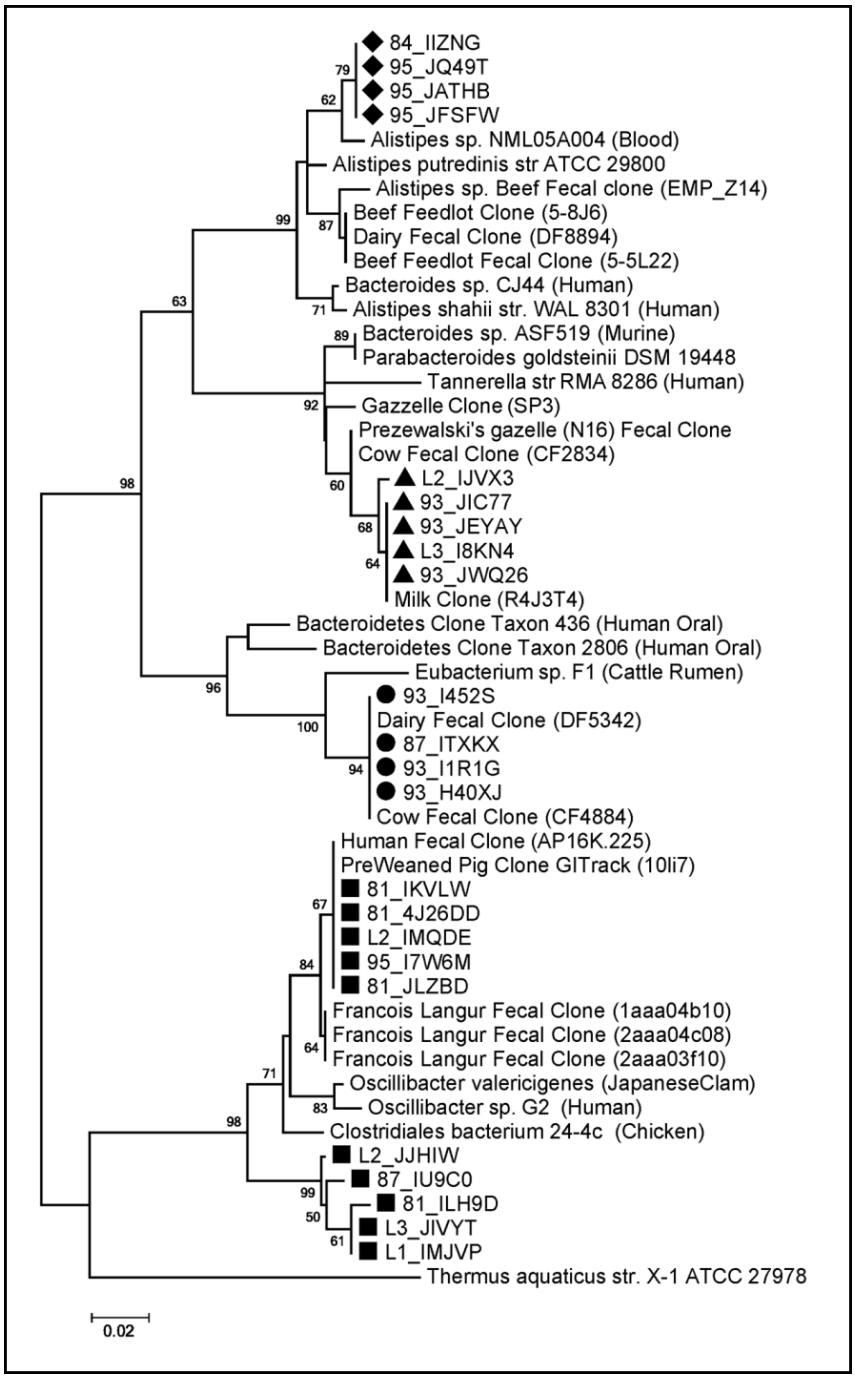

Figure 18. Neighbor-joining phylogenetic tree of the five most abundant OTUs found within the deer fecal communities collectively and their top GenBank hits. Deer fecal community sequences representing the five OTUs are indicated with a black circle representing *Bacteroidales,* a black square representing *Ruminococcaceae*, black triangle representing *Porphyromonadaceae*, and black diamond representing *Rikenellaceae* followed by their site name and five digit individual sequence identity code.

Table 17. The five most shared and abundant OTUs found in the deer fecal communities. The top GenBank hit with the maximum identity is listed along with a description of the hit and its maximum percentage identity with the OTU.

OTU	OTU Taxonomy	Top GenBank Hit Accession#	Top GenBank HIT Description (16s rRNA gene)	GenBank Hit Source	GenBank Max Identity
OTU_36[1]	Veillonellaceae	EU778779.1	Uncultured bacterium clone SBSD_aaa02h10_1	Springbok antelope feces	95%
OTU_111[1, 3]	Ruminococcaceae	GU611449	Uncultured bacterium clone DF3272	dairy cow feces	98%
OTU_207[2]	Ruminococcaceae	EU468955	Uncultured bacterium clone SP2_h05	Speke's gazelle feces	97%
OTU_2512[3]	Ruminococcaceae	FJ651134	Uncultured Firmicutes bacterium clone OB_425	human feces	98%
OTU_2593[2]	Clostridiales	JX109040	Uncultured bacterium clone MID39_30977	dairy cow uterus	98%
OTU_3135[3]	Rikenellaceae	GU617071	Uncultured bacterium clone DF8894	dairy cow feces	97%
OTU_3604[3]	Bacteroidales	GU613519	Uncultured bacterium clone DF5342	dairy cow feces	97%
OTU_3630[3]	Porphyromonadaceae	EU469137	Uncultured bacterium clone SP3_a11	bighorn sheep feces	97%
OTU_4560[2]	Ruminococcaceae	GU604899	Uncultured bacterium clone CF4911	cow feces	96%

[1]OTUs common in all 11 samples, [2]OTUs common in 10 of 11 samples, [3]Five most abundant OTUs

Discussion

This survey aimed to better characterize deer fecal communities as a first step toward potentially developing deer-specific BST markers. The bacterial communities

were characterized using an OTU approach and then further classified taxonomically. Chao I richness estimates showed that larger sequence libraries, approximately double, from each sample would be necessary to capture the breath of diversity across the fecal communities (Table 15). Shannon and Simpson diversity indices were consistent across the samples with the exception of Welder sample 93 which was the least diverse of all the samples and proved to be somewhat of an outlier.

Hypothesis testing was utilized to examine overall community structure and ask whether the communities were significantly different than would be expected by chance. The samples were grouped into three treatment categories for testing, Welder 08, Welder 09, and Leon. The global parsimony test showed no significant differences between the three treatments. AMOVA is a non-parametric analog of a traditional analysis of variance and tests the hypothesis that genetic diversity within two populations is not significantly different from that which would result from pooling the two populations. This test also showed no significant difference between the communities. Finally, HOMOVA is a non-parametric analog of Bartlett's test for homogeneity of variance. There was a significant difference between variance across the groups. Using the Yue-Clayton distance measure, the central communities from all three groups were statistically the same, but the Leon samples displayed greater variance within samples. Diversity within individual groups was greater than between them. The significant difference in variances between samples likely masked any potential power to see differences in the overall communities (Schloss, 2008; Schloss et al., 2009). For ease of interpretation, the OTUs were converted to a relative abundance per sample basis and

were plotted using NMDS (Figure 14). All of the communities lined up on one axis except for Welder sample 93. There was little clustering of samples across years.

The phylogenetic analysis yielded much lower perceived taxonomic diversity than the OTU-based analysis, but could have been expected. The OTU-based analysis does not depend on a pre-defined taxonomy, and since many of the sequences could not be classified down to the genus level in RDP with confidence, they were left at the family level or higher. Many of the individual OTUs classified to the same families. The communities were dominated by two phyla, *Firmicutes* and *Bacteroidetes*. These phyla have previously been shown to constitute the majority of gut-associated bacteria in other mammals (Durso et al., 2010; Ley et al., 2006; Shanks et al., 2011) and each of the overlapping and abundant OTUs fell into these two phyla. *Proteobacteria*, including *E. coli*, averaged 0.5% across all 11 samples. The relative abundances of the family level taxonomic classification did not appear to cluster by site location or year (Figure 16). The Welder, 93, sample was the least diverse and exhibited a shift from the *Firmicutes* into the *Bacteroidetes* overall as seen in the depletion of *Ruminococcaceae*. Previous studies have shown diet and geographical location cause shifts in gut and fecal microbial populations (Ley et al., 2008; Shanks et al., 2011). The deer communities examined in this study would seem to fit the description of being geographically distant, approximately 350 miles apart and are in two completely different ecoregions of the state in the gulf coast region at Welder versus the prairie and lakes region where Comanche County is located. But surprisingly at the family-level taxonomy none of the samples tended to cluster by location or by year. Both the OTU analysis and taxonomic

classifications suggest the deer fecal communities in these two parts of Texas are similar and stable over time.

Two 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). A majority of the OTUs overall were singletons (56%). Eckburg et al. (2005) noted a similar trend assessing diversity of the human intestinal microbial flora where 60% of the genera were recovered only once. Further, the five most abundant OTUs only represented 12% of the total. The overlap or percentage of abundant OTUs overlapping at either Welder or Leon also did not represent over 10% of the total OTUs. The five most abundant taxa were classified as *Ruminococcaceae*, *Veillonellaceae*, *Rikenellaceae*, *Porphyromonadaceae*, and *Bacteroidales* (Unclassified). The top GenBank hits for representative sequences from all of the OTUs were from fecal communities, except for the dairy cow uterus hit. 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.

Conclusions

The goal of this project was to utilize 454 pyrosequencing to better characterize deer bacterial fecal communities in Texas with the aim of finding organisms that were common across geographic regions and time in order to serve as a starting point for future research toward development of a deer-specific BST marker. The microbial communities were not significantly different from an overall OTU (97% cutoff) standpoint and did not cluster by site or year and suggesting that the deer fecal bacterial communities, at least in south and central Texas, were stable over time which bodes well for the potential of a temporal and geographically stable source-specific marker. At least two of these OTUs, OTU_36 and OTU_4560, appear to be potentially deer-specific with their closest non-deer matches in GenBank being only 95 and 96% similar, respectively. Future work will focus on primer design and screening these potential OTUs against non-target sources in order to verify their suitability as deer-specific BST markers.

CHAPTER VI

SUMMARY AND FUTURE BST

Summary

BST aims to identify sources of fecal contamination in water in an effort to tailor best management practices to ameliorate this contamination. The studies in this dissertation aimed to enhance BST method performance and lay the groundwork to expand BST in the future and are summarized below in addition to some thoughts on BST in the future.

The method comparison study (Chapter III) looked to evaluate differences in *E. coli* community composition across three standard water quality assessments, including USEPA Method 1603 (mTEC), USEPA Method 1604 (MI), and Colilert® to ultimately determine their impact on BST library performance. Results indicated that the enumeration data was not consistent across the six sites evaluated across Texas. Genotypic analysis revealed that generally, mTEC selected for the most diverse *E. coli* populations followed by MI and finally Colilert® which displayed the least amount of diversity across isolates at all locations. One site, Moody Creek, displayed the simplest *E. coli* communities across all three media types, but the other five sites were characterized as having diverse and different *E. coli* communities from each media type. The isolates were queried against the Texas *E. coli* BST library which is based almost exclusively on isolates using EPA Method 1603 (mTEC). When viewed on a site by site basis, results were extremely varied and showed no clear pattern or trend based on media

type. Even though the three media assessed in this study ultimately enumerated the same groups of organisms (total coliforms and *E. coli*) the differences in the media composition, incubation temperature, and growth platform appear to have a strong selective influence on the populations of *E. coli* isolated using these media. This study suggests that standardized methods of enumeration and isolation may be warranted if stakeholders anticipate the possibility of using library-dependent BST.

Dairy farms have been implicated in the Leon River basin as a source of non-point source pollution and this study aimed at using BST tools to enumerate and characterize *E. coli* communities across various dairy manure management techniques to evaluate current BST methods to track this potential contamination source (Chapter IV). Stakeholders in the watershed mainly utilized four general dairy manure management strategies and were evaluated at four dairies including manual scraping, separators, settling basins, and lagoons. Results of this study indicated that manure and effluent management strategies which employed means to remove solids from the manure tended to decrease the levels of *E. coli* in the effluent. The study aimed to evaluate relative abundance of genotypes across dairy treatment and time to assess potential BST library effects, but there was no clear seasonal or site grouping trends seen among the lagoon ERIC-PCR genotypes across all of the dairies. Nor were there any strong grouping trends when assessing individual dairies across treatment and time. But when each dairy-manure management scheme treatment was evaluated as a whole across the entire study, there was overlap in genotypes across treatments where 60% of total isolates were seen across some combination of the four dairies and 15% were seen in all four dairies.

Using ERIC-PCR and RP combination, isolates from all four dairies and management styles were also screened for inclusion into the Texas *E. coli* BST library. Of the 42 total local library isolates that were ERIC-RP, 36 were self-validated and 89% of those isolates best match (highest similarity match at 80% cutoff value for identification) were to an isolate from another dairy. Despite the variability seen in the ERIC-PCR genotype analysis across dairies and time, the use of library-dependent BST and the local library was sufficiently robust, both geographically and temporally, to correctly classify the vast majority of dairy *E. coli* isolates from this project.

Culture-independent or marker-based BST has the potential to significantly reduce labor and cost associated with use of BST. Even though host-specific markers have been and continue to be developed, they are generally still limited, especially for wildlife which severely restricts our ability to specifically track this important group. Library-independent means to track deer specifically 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. The deer study aimed to use 454 barcoded pyrosequencing to characterize deer fecal communities in Texas in an effort to evaluate their suitability for development of a deer-specific BST marker (Chapter V). Deer fecal samples from Welder Wildlife refuge were collected over two years as well as from the Leon watershed. A parsimony test showed no significant difference between the samples collected from both years at Welder or Leon. The fecal communities were dominated by

two phyla, *Firmicutes* and *Bacteroidetes.* Two 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. Future work will focus on primer design and screening these potential OTUs against non-target sources in order to verify their suitability as deer-specific BST markers.

BST – Thoughts on the Future

Expansion of BST is currently focused on library-independent means mainly thru source specific marker development. But those efforts continue to be confounded when challenged with large datasets as well as more and more stringent quality control measures warranted for regulatory purposes. Unlike most library-independent techniques, library-dependent efforts often utilize fecal indicators like *E. coli* and thus have a more direct correlation to regulatory standards. These efforts can be tailored to

the watershed in question where watershed-specific sources of contamination can be used to develop a library, whereas library-independent methods are reliant on source-specific markers which may or may not be relevant or even source-specific in the watershed (Mott and Smith, 2011). For these reasons, for the foreseeable future, I feel there will be place and utility for library-based techniques. One suggestion might include increased use of automated fingerprinting to decrease labor and processing costs and speed library development.

However, the ability to characterize entire communities, through next generation sequencing, more and more affordably has opened up considerable areas of study across the scientific community. The idea of a single marker gene that is specific to an animal over time and large geographic regions may not ultimately be feasible, but with the current and growing sequencing capabilities we have opportunity to answer those questions. Massive sequencing efforts in the human microbiome project and many others are looking to match health effects with the microbial community (Turnbaugh et al., 2007). Further, Fierer et al. (2010) showed individuals could be traced to their own keyboards based on the DNA they left behind. Bowers et al. (2011) recently characterized air samples across the US and then used previously sequenced potential known -source communities to assign source categories to the environmental communities. These types of studies are shedding light on the use of next generation sequencing to fields besides medicine. Reflecting on the resources we invest in a fecal or water sample to study only one or two organisms of interest, such as *Bacteroidales* or *E. coli*, we should recognize that more thorough characterization of the entire bacterial

community could open doors to other avenues of tracking sources of fecal contamination.

REFERENCES

Acosta-Martinez, V., Dowd, S., Sun, Y. and Allen, V. (2008) Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. Soil Biology & Biochemistry 40(11), 2762-2770.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. Journal of Molecular Biology 215(3), 403-410.

Anderson, M.L., Whitlock, J.E. and Harwood, V.J. (2005) Persistence and differential survival of fecal indicator bacteria in subtropical waters and sediments. Applied and Environmental Microbiology 71(6), 3041-3048.

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.

Bicudo, J.R. and Goyal, S.M. (2003) Pathogens and manure management systems: a review. Environmental Technology 24(1), 115-130.

Bowers, R.M., Sullivan, A.P., Costello, E.K., Collett, J.L., Knight, R. and Fierer, N. (2011) Sources of bacteria in outdoor air across cities in the midwestern United States. Applied and Environmental Microbiology 77(18), 6350-6356.

Brenner, K.P., Rankin, C.C., Roybal, Y.R., Stelma, G.N., Scarpino, P.V. and Dufour, A.P. (1993) New medium for the simultaneous detection of total coliforms and Escherichia coli in water. Applied and Environmental Microbiology 59(11), 3534-3544.

Buckalew, D.W., Hartman, L.J., Grimsley, G.A., Martin, A.E. and Register, K.M. (2006) A long-term study comparing membrane filtration with Colilert® defined substrates in detecting fecal coliforms and Escherichia coli in natural waters. Journal of Environmental Management 80(3), 191-197.

Callaway, T.R., Carr, M.A., Edrington, T.S., Anderson, R.C. and Nisbet, D.J. (2009) Diet, Escherichia coli O157:H7, and cattle: a review after 10 years. Current Issues in Molecular Biology 11, 67-79.

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

Casarez, E.A., Pillai, S.D., Mott, J.B., Vargas, M., Dean, K.E. and Di Giovanni, G.D. (2007b) Direct comparison of four bacterial source tracking methods and use of composite data sets. Journal of Applied Microbiology 103(2), 350-364.

Chao, K.K., Chao, C.C. and Chao, W.L. (2004) Evaluation of Colilert-18 for detection of coliforms and Eschericha coli in subtropical freshwater. Applied and Environmental Microbiology 70(2), 1242-1244.

Cole, J.R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R.J., Kulam-Syed-Mohideen, A.S., McGarrell, D.M., Marsh, T., Garrity, G.M. and Tiedje, J.M. (2009) The ribosomal database project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37, D141-D145.

Cook, K.L., Bolster, C.H., Ayers, K.A. and Reynolds, D.N. (2011) Escherichia coli diversity in livestock manures and agriculturally impacted stream waters. Current Microbiology 63(5), 439-449.

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., 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.

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.

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

Dungan, R.S., Klein, M. and Leytem, A.B. (2012) Quantification of bacterial indicators and zoonotic pathogens in dairy wastewater ponds. Applied and Environmental Microbiology 78(22), 8089-8095.

Durso, L.M., Harhay, G.P., Smith, T.P.L., Bono, J.L., DeSantis, T.Z., Harhay, D.M., Andersen, G.L., Keen, J.E., Laegreid, W.W. and Clawson, M.L. (2010) Animal-to-animal variation in fecal microbial diversity among beef cattle. Applied and Environmental Microbiology 76(14), 4858-4862.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E. and Relman, D.A. (2005) Diversity of the human intestinal microbial flora. Science 308(5728), 1635-1638.

Field, K.G. and Samadpour, M. (2007) Fecal source tracking, the indicator paradigm, and managing water quality. Water Research 41(16), 3517-3538.

Fierer, N., Lauber, C.L., Zhou, N., McDonald, D., Costello, E.K. and Knight, R. (2010) Forensic identification using skin bacterial communities. Proceedings of the National Academy of Sciences of the United States of America 107(14), 6477-6481.

Francy, D.S. and Darner, R.A. (2000) Comparison of methods for determining Escherichia coli concentrations in recreational waters. Water Research 34(10), 2770-2778.

Franz, E., van Diepeningen, A.D., de Vos, O.J. and van Bruggen, A.H.C. (2005) Effects of cattle feeding regimen and soil management type on the fate of Escherichia coli O157:H7 and Salmonella enterica serovar typhimurium in manure, manure-amended soil, and lettuce. Applied and Environmental Microbiology 71(10), 6165-6174.

Fricker, E.J., Illingworth, K.S. and Fricker, C.R. (1997) Use of two formulations of Colilert and QuantiTray™ for assessment of the bacteriological quality of water. Water Research 31(10), 2495-2499.

Fulhage, C., and Hoehne, J. (2001). Planning and evaluation of manure storage. Lesson 20. Midwest Plan Service, Iowa State University, Ames.

Fulhage, C.D. (1997) Manure management considerations for expanding dairy herds. Journal of Dairy Science 80(8), 1872-1879.

Gordon, D.M. (2001) Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiology 147, 1079-1085.

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

Graves, R.E. (1994). Dairy manure handling. G-72. Penn State Agricultural and Biological Engineering Extension, University Park.

Guber, A.K., Pachepsky, Y.A., Yakirevich, A.M., Shelton, D.R., Sadeghi, A.M., Goodrich, D.C. and Unkrich, C.L. (2011) Uncertainty in modelling of faecal coliform overland transport associated with manure application in Maryland. Hydrological Processes 25(15), 2393-2404.

Hamilton, M.J., Yan, T. and Sadowsky, M.J. (2006) Development of goose- and duck-specific DNA markers to determine sources of Escherichia coli in waterways. Applied and Environmental Microbiology 72(6), 4012-4019.

Hamilton, W.P., Kim, M. and Thackston, E.L. (2005) Comparison of commercially available Escherichia coli enumeration tests: implications for attaining water quality standards. Water Research 39(20), 4869-4878.

Hammer, O., Haper, D.A.T. and Ryan, P.D. (2001) PAST: paleontological statistics software package for education and data analysis. Palaeontologica Electronica 4(1), 9.

Hansen, D.L., Ishii, S., Sadowsky, M.J. and Hicks, R.E. (2009) Escherichia coli populations in Great Lakes waterfowl exhibit spatial stability and temporal shifting. Applied and Environmental Microbiology 75(6), 1546-1551.

Harmel, R.D., Karthikeyan, R., Gentry, T. and Srinivasan, R. (2010) Effects of agricultural management, land use, and watershed scale on E. coli concentrations in runoff and stream flow. Transactions of the American Society of Agricultural and Biological Engineers 53(6), 1833-1841.

Hartel, P.G., Summer, J.D., Hill, J.L., Collins, J.V., Entry, J.A. and Segars, W.I. (2002) Geographic variability of Escherichia coli ribotypes from animals in Idaho and Georgia. Journal of Environmental Quality 31(4), 1273-1278.

Harwood, V.J. and Stoeckel, D.M. (2011). Performance criteria microbial source tracking: methods, applications, and case studies. In: Hagedorn, C., Blanch, A.R. and Harwood, V.J. (Eds), Microbial Source Tracking: Methods, Applications, and Case Studies. Springer, New York. pp. 7-30.

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

Hill, V.R. (2003) Prospects for pathogen reductions in livestock wastewaters: a review. Critical Reviews in Environmental Science and Technology 33(2), 187-235.

Huber, D., Joern, B., Sutton, A., Healy, J., McCloud, P. and Wilxocx, J. (2003). Dairy manure management planning. Purdue University Cooperative Extension Service and Natural Resources Conservation Service, West Lafayette.

Hutchison, M.L., Walters, L.D., Moore, A. and Avery, S.M. (2005) Declines of zoonotic agents in liquid livestock wastes stored in batches on-farm. Journal of Applied Microbiology 99(1), 58-65.

Jenkins, M.B., Hartel, P.G., Olexa, T.J. and Stuedemann, J.A. (2003) Putative temporal variability of Escherichia coli ribotypes from yearling steers. Journal of Environmental Quality 32(1), 305-309.

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.

Kodaka, H., Mizuochi, S., Saito, M. and Matsuoka, H. (2008) Evaluation of a new medium for the enumeration of total coliforms and Escherichia coli in Japanese surface waters. Journal of Applied Microbiology 104(4), 1112-1118.

Koelsch, R.K., Lorimor, J.C. and Mankin, K.R. (2006) Vegetative treatment systems for management of open lot runoff: review of literature. Applied Engineering in Agriculture 22(1), 141-153.

Layton, A., McKay, L., Williams, D., Garrett, V., Gentry, R. and Sayler, G. (2006) Development of Bacteroides 16S rRNA gene TaqMan-based real-time PCR assays for estimation of total, human, and bovine fecal pollution in water. Applied and Environmental Microbiology 72(6), 4214-4224.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R. and Gordon, J.I. (2008) Evolution of mammals and their gut microbes. Science 320(5883), 1647-1651.

Ley, R.E., Peterson, D.A. and Gordon, J.I. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell 124(4), 837-848.

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.

McGarvey, J.A., Miller, W.G., Sanchez, S. and Stanker, L. (2004) Identification of bacterial populations in dairy wastewaters by use of 16S rRNA gene sequences

and other genetic markers. Applied and Environmental Microbiology 70(7), 4267-4275.

Meals, D.W. and Braun, D.C. (2006) Demonstration of methods to reduce E. coli runoff from dairy manure application sites. Journal of Environmental Quality 35(4), 1088-1100.

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.

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

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.

Olstadt, J., Schauer, J.J., Standridge, J. and Kluender, S. (2007) A comparison of ten USEPA approved total coliform/E. coli tests. Journal of Water and Health 5(2), 267-282.

Pell, A.N. (1997) Manure and microbes: public and animal health problem? Journal of Dairy Science 80(10), 2673-2681.

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

Rompre, A., Servais, P., Baudart, J., de-Roubin, M.R. and Laurent, P. (2002) Detection and enumeration of coliforms in drinking water: current methods and emerging approaches. Journal of Microbiological Methods 49(1), 31-54.

Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A., Oakley, B.B., Parks, D.H., Robinson, C.J., Sahl, J.W., Stres, B., Thallinger, G.G., Van Horn, D.J. and Weber, C.F. (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology 75(23), 7537-7541.

Schloss, P.D. (2008) Evaluating different approaches that test whether microbial communities have the same structure. International Society for Microbial Ecology Journal 2(3), 265-275.

Sercu, B., Van De Werfhorst, L.C., Murray, J.L.S. and Holden, P.A. (2011) Cultivation-independent analysis of bacteria in IDEXX Quanti-tray/2000 fecal indicator assays. Applied and Environmental Microbiology 77(2), 627-633.

Seurinck, S., Defoirdt, T., Verstraete, W. and Siciliano, S.D. (2005) Detection and quantification of the human-specific HF183 Bacteroides 16S rRNA genetic marker with real-time PCR for assessment of human faecal pollution in freshwater. Environmental Microbiology 7(2), 249-259.

Shanks, O.C., Kelty, C.A., Archibeque, S., Jenkins, M., Newton, R.J., McLellan, S.L., Huse, S.M. and Sogin, M.L. (2011) Community structures of fecal bacteria in cattle from different animal feeding operations. Applied and Environmental Microbiology 77(9), 2992-3001.

Shanks, O.C., White, K., Kelty, C.A., Sivaganesan, M., Blannon, J., Meckes, M., Varma, M. and Haugland, R.A. (2010) Performance of PCR-based assays targeting Bacteroidales genetic markers of human fecal pollution in sewage and fecal samples. Environmental Science & Technology 44(16), 6281-6288.

Shanks, O.C., Atikovic, E., Blackwood, A.D., Lu, J.R., Noble, R.T., Domingo, J.S., Seifring, S., Sivaganesan, M. and Haugland, R.A. (2008) Quantitative PCR for detection and enumeration of genetic markers of bovine fecal pollution. Applied and Environmental Microbiology 74(3), 745-752.

Shanks, O.C., Santo Domingo, J.W., Lamendella, R., Kelty, C.A. and Graham, J.E. (2006) Competitive metagenomic DNA hybridization identifies host-specific microbial genetic markers in cow fecal samples. Applied and Environmental Microbiology 72(6), 4054-4060.

Sinton, L.W., Braithwaite, R.R., Hall, C.H. and Mackenzie, M.L. (2007) Survival of indicator and pathogenic bacteria in bovine feces on pasture. Applied and Environmental Microbiology 73(24), 7917-7925.

Stoeckel, D.M. and Harwood, V.J. (2007) Performance, design, and analysis in microbial source tracking studies. Applied and Environmental Microbiology 73(8), 2405-2415.

TCEQ. (2010). Executive Summary - 2010 Texas integrated report for clean water act sections 305(b) and 303(d). Texas Commission on Environmental Quality, Austin.

TCEQ. (2008). Executive Summary – 2008 Texas water quality inventory and 303(d) list. Texas Commission on Environmental Quality, Austin.

Teague, A., Karthikeyan, R., Babbar-Sebens, M., Srinivasan, R. and Persyn, R.A. (2009) Spatially explicit load enrichment calculation tool to identify potential E. coli sources in watersheds. Transactions of the American Society of Agricultural and Biological Engineers 52(4), 1109-1120.

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.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R. and Gordon, J.I. (2007) The human microbiome project. Nature 449(7164), 804-810.

USEPA. (2005a). Method 1603: Escherichia coli (E. coli) in water by membrane filtration using modified membrane-thermotolerant Escherichia coli agar (modified mTEC). EPA-821-R-06-011. United States Environmental Protection Agency, Washington, DC.

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

USEPA. (2004). Method 1604: total coliforms and Escherichia coli in water by membrane filtration using a simultaneous detection technique (MI medium). EPA-821-R-02-024. United States Environmental Protection Agency, Washington DC.

USEPA. (2003a). Baceterial water quality standards for recreational (fresh and marine waters). EPA-823-R-03-008. United States Environmental Protection Agency Washington, DC.

USEPA. (2003b). National management measures to control nonpoint source pollution from agriculture. EPA-841-B-03-004. United States Environmental Protection Agency, Washington, DC.

van Elsas, J.D., Semenov, A.V., Costa, R. and Trevors, J.T. (2011) Survival of Escherichia coli in the environment: fundamental and public health aspects. International Society for Microbial Ecology Journal 5(2), 173-183.

Versalovic, J., Koeuth, T. and Lupski, J.R. (1991) Distribution of repetitive DNA-sequences in Eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Research 19(24), 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.

Wuertz, S., Wang, D., Reischer, G.H. and Farnleitner, A.H. (2011). Library-independent bacterial 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. 61-112.