Bacteria Source Tracking on the Mission and Aransas Rivers

Publication CBBEP – 56
Project Number – 630
December 2008

Principal Investigators

Joanna Mott, Ph.D.
Roy Lehman, Ph.D.
and
Amanda Smith

Department of Life Sciences
Texas A&M University-Corpus Christi,
6300 Ocean Drive,
Corpus Christi, TX 78412

Report prepared by:
Jace Tunnell , CBBEP Project Manager
Coastal Bend Bays and Estuaries Program, Inc.

Submitted to:
Coastal Bend Bays & Estuaries Program
1305 N. Shoreline Blvd., Suite 205
Corpus Christi, TX 78401

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

Funding for this project was provided by the Coastal Bend Bays and Estuaries Program. Field sampling and laboratory analyses were conducted by field and laboratory personnel in the Environmental Microbiology Laboratory at Texas A&M University-Corpus Christi. Bridgette Froeschke, Christine Bowen, Jessica Akins, Megan Hoch, and Margaret Spoo were the lead personnel responsible for performing laboratory analyses. Supervision and oversight was provided by Amanda Smith, Research Specialist, and the P.I.s, Dr. Mott and Dr. Lehman. Dr. Blair Sterba-Boatwright provided technical expertise in statistical analyses.

Bacteria Source Tracking on the Mission and Aransas Rivers, Tidal

Joanna Mott, Ph.D.
and Roy Lehman, Ph.D.

EXECUTIVE SUMMARY

The 2004 Texas Water Quality Inventory and 303(d) list includes Segment 2001, Mission River Tidal (from the confluence with Mission Bay in Refugio County to a point 7.4 kilometers [4.6 miles] downstream of U.S. 77 in Refugio County), and Segment 2003, Aransas River Tidal (from the confluence with Copano Bay in Aransas/Refugio County to a point 1.6 kilometers [1.0 mile] upstream of US 77 in Refugio/San Patricio County). Both segments are listed for bacteria with Non-Point Sources contributing to the impairment. Both are listed as Category 5c with a rank of D, meaning the water bodies do not meet applicable water quality standards and additional data must be collected before a total maximum daily load (TMDL) is scheduled. Rankings are based on the current understanding of the causes of the non-support of the water quality standards and the sources of pollution, the importance of the resource, the severity of the impact, and the likelihood of TMDL success. As these segments are tidal, bacteriological water quality is assessed using the saltwater indicator, enterococci. However, both rivers flow into Copano Bay, a shellfish-harvesting bay, which is assessed using fecal coliforms (Texas Administrative Code, 2000, §307.7(b)(3)(B)).

In this study, bacteria source tracking (BST) was used to evaluate the sources of fecal contamination in the Mission and Aransas River segments and to provide additional data for assessment of sources of contamination into Copano Bay, the water body into which both segments empty. The BST methods utilized in this study, antibiotic resistance analysis (ARA), carbon source utilization (CSU), and pulsed-field gel electrophoresis (PFGE), are library-dependent methods, which require the use of a library of known source isolates. The existing ARA *E. coli* library at TAMU-CC consists of fecal isolates from humans, livestock animals (cow and horse), and wildlife (eg. raccoon, seagulls, deer, duck). The current PFGE library consists of fecal isolates from cow, horse, duck and human. Additional fecal isolates (cow, horse, duck, hog, and human) from the Mission and Aransas Rivers watershed were collected and analyzed for expansion of the current libraries. In addition, ARA library isolates collected under previous projects were analyzed with CSU to provide a library of known source CSU profiles, and ARA and CSU profiles were combined to form a composite dataset.

Water samples were collected from the Mission and Aransas Rivers during three sampling events – two during dry weather and one following rainfall. The samples were transported directly to the TAMU-CC Environmental Microbiology Laboratory, where they were immediately analyzed via membrane filtration for *E. coli* (EPA Method 1103.1).

After isolation on mTEC agar and Rainbow Agar O157®, all known source (fecal samples) and unknown source (water) *E. coli* isolates were confirmed as *E. coli* through carbon source utilization profiles generated with the MicroLog™ Microbial Identification System. Only isolates confirming as *E. coli* with this system were used for ARA. Additionally, CSU analysis provided profiles for comparison with known sources via discriminant analysis using SPSS software (Release 15.0, 2006).

Procedures for ARA followed the Kirby-Bauer disk diffusion method outlined by the Clinical and Laboratory Standards Institute (CLSI). An automated plate analyzer, BIOMIC®, was used to automatically measure diameters of zones of inhibition. Zone diameters were analyzed via discriminant analysis and compared to those of the library of known isolates.

Pulsed-field gel electrophoresis (PFGE) analysis followed Bio-Rad Methodology and Standards as described in Bio-Rad Laboratories manual, CHEF-DR III Pulsed Field Electrophoresis Systems: Instruction Manual and Applications Guide (1995, Hercules, California). PFGE was used to obtain genetic 'fingerprints' for approximately 200 known source isolates to add to the existing library and 150 unknown source isolates for comparison with results from ARA and CSU.

The known source library used for comparison with unknown sources is considered to be representative of the possible sources contributing to elevated bacterial levels in the Mission and Aransas tidal segments. The known source libraries for ARA and PFGE were composed of isolates collected from various animal and human sources throughout the Mission and Aransas watershed over several years, spanning multiple projects. Isolates previously analyzed with ARA were re-analyzed with CSU to provide profiles to create a CSU library. Profiles generated from both ARA and CSU were combined to form a composite library. Additional isolates were collected under this project to increase underrepresented sources in the existing libraries. Over 1000 known source profiles were utilized in the statistical analysis for this project.

The dataset with the most accurate classification for ARA and CSU was the composite dataset combining both types of profiles. This composite library produced an average rate of correct classification (ARCC) of 93.2% with three-way classification, indicating that the known sources present in the library are classified correctly at an average rate of 93.2%.

The majority of unknown source isolates collected from water samples at the five sampling stations along the Mission and Aransas tidal segments were classified as human source, utilizing the three bacterial source tracking techniques: antibiotic resistance analysis (ARA), carbon source utilization (CSU) analysis, and pulsed-field gel electrophoresis (PFGE). Overall, 63.7-66.9% of unknown source isolate profiles from the composite (ARA+CSU) dataset were classified as treated human sources (originating from treated wastewater effluent). PFGE profile analysis confirmed this finding with 40.5% of isolates having the highest similarity to human sources. The remaining unknown source isolates were classified as livestock animals and wildlife,

with cow, horse and duck contributions accounting for the majority of the animal sources in both the composite dataset and PFGE profiles.

When grouped by sampling event, analysis of unknown source isolates using the composite library showed a difference in sources for the event following rainfall. For both dry weather sampling events, more than 70% of the unknown source isolates classified as human (treated) with the three-way classification. In contrast, only 38.8% isolates from the sampling event following rainfall were attributed to treated human wastewater. This difference was consistent for both the four-way and seven-way classifications. With fewer unknown source isolates classifying as human, a greater proportion of Event 3 (wet event) isolates were categorized as livestock and wildlife sources. The number of isolates classifying as livestock (horse/cow) was more than double following rainfall (Event 3) than for the dry weather events, suggesting a run-off contribution from these sources. Differences in the percentage of isolates classifying as wildlife between dry and wet weather events were less pronounced.

Results for unknown source isolates grouped by sampling station showed little effect of location. For the composite dataset, the human source contribution at each station was approximately the same. Slight differences were found amongst stations with respect to percentages of contribution from livestock and wildlife sources. Results from PFGE analysis confirmed this finding.

With the high rate of correct classification and corroboration between results from different analytical methods, known source isolates examined in this project are thought to adequately represent the potential contamination sources for the Mission and Aransas Rivers. Comparing findings of this study with the previously completed BST study of Copano Bay, it appears that sources are different in the bay compared with the rivers. However, following rainfall, the relative contributions of sources are more similar. For most bay stations the majority of isolates were collected following rainfall whereas, for the rivers numbers of *E. coli* were high enough to be collected during dry weather. The sources of fecal contamination into Mission and Aransas Rivers either differ from those of the main body of Copano Bay, or, sources may have changed since the completion of the previous project. The effect of rainfall on sources of contamination merits further study.

INTRODUCTION

Recreational waters contaminated with fecal wastes pose a serious health threat to the public. Fecal matter contains opportunistic and pathogenic bacteria and viruses that may infect humans that come into contact with the contaminated water body. Pathogens commonly transmitted by feces through ingestion of contaminated water include Hepatitis A virus, enteroviruses, *E. coli*, *Salmonella* sp., *Cryptosporidium parvum*, and *Giardia lamblia* (Moe, 2002). The sources of these pathogens include humans, livestock animals and wildlife.

Currently, due to low detection rates and cost limitations, pathogens themselves are not monitored in recreational waters. Fecal indicator organisms, such as fecal coliform bacteria and enterococci, are utilized to monitor quality of recreational surface waters. Standard bacteriological criteria are defined by levels of fecal indicator organisms and differ for water bodies based on their designated use. These criteria are set by U.S. Environmental Protection Agency (U.S. EPA) and, for Texas, the Texas Commission on Environmental Quality (TCEQ). Water bodies failing to meet these standards are classified as impaired. Further investigations are warranted to remedy the impairment when at all possible.

Mission River Tidal (Segment 2001 - from the confluence with Mission Bay in Refugio County to a point 7.4 kilometers [4.6 miles] downstream of U.S. 77 in Refugio County)) and Aransas River Tidal (Segment 2003 - from the confluence with Copano Bay in Aransas/Refugio County to a point 1.6 kilometers [1.0 mile] upstream of US 77 in Refugio/San Patricio County) are both listed on the 2004 Texas Water Quality Inventory and 303(d) list for bacteria, signifying that these water bodies do not meet bacteriological water quality standards. Both segments have been classified as impaired for bacteria with non-point source contributions. Both are listed as Category 5c with a rank of D, meaning the water bodies do not meet applicable water quality standards and additional data must be collected before a total maximum daily load (TMDL) is scheduled. Rankings are based on the current understanding of the causes of the non-support of the water quality standards and the sources of pollution, the importance of the resource, the severity of the impact, and the likelihood of TMDL success. As these segments are tidal, bacteriological water quality is assessed using the saltwater indicator, enterococci. However, both rivers flow into Copano Bay, a shellfish-harvesting bay, which is assessed using fecal coliforms (Texas Administrative Code, 2000, §307.7(b)(3)(B)).

Non-point source pollution originates from many disperse pollutants and is typically carried to water bodies through rainfall runoff (U.S. EPA, 1994). Common examples of non-point sources contributing to bacterial impairments include Concentrated Animal Feeding Operations (CAFOs), pet wastes, storm drain runoff and leaking septic systems. Non-point sources are more difficult to identify and control than their point-source counterparts such as waste water treatment plants. However, once identified, non-point sources may be remediated through Best Management Practices (BMPs). BMPs are considered to be the most effective and reasonable means of reducing or

preventing a pollutant from entering a water body (Boyer, accessed 2008). These types of methods have been implemented successfully in other watersheds to reduce bacterial impairments (U.S. EPA, 2005).

In order to remedy non-source pollution most effectively, contributing sources must be identified to narrow or broaden the scope of BMPs applied in the watershed. Bacterial source tracking (BST) is a grouping of analyses that may be used to determine the sources of fecal contamination.

Several different bacterial source tracking (BST) methods are currently in use. Some BST methods rely on detection of genetic markers specific for certain animal groups. These genotypic methods encompass a variety of techniques that look at specific genes. BST methods can also analyze phenotypic (physically expressed) characteristics that are indicative of specific groups of organisms. Phenotypic characteristics include metabolism of substrates and resistance to antimicrobial compounds. To provide the most accurate results, studies should combine two or more BST methods rather than relying on a single method (Jones et al., 2007). For this study two phenotypic methods were used and a subset of isolates was also analyzed using a genotypic method, Pulsed-field Gel Electrophoresis (PFGE), to provide confirmation of bacteria sources.

BST methods may also be classified as library-dependent or independent. A library is a collection of known source isolate profiles, with which to compare unknown source isolate profiles. The most accurate library is one that is representative of known sources that have potential impact on the watershed and is inclusive of a large number of non-clonal isolates (isolates that do not produce the same pattern as each other) (Johnson et al., 2004). All three methods used in this study are library-dependent.

Antibiotic resistance analysis (ARA) or antibiotic resistance profiling (ARP) is a widely applied library-dependent method that relies upon phenotypic characteristics. It is well established and has been used in a number of previous studies (Hagedorn et al., 1999; Harwood et al., 2000; Parveen et al., 2001; Webster et al., 2004, Whitlock et al., 2002, Wiggins et al., 1999). The advantage of ARA over other BST methods is cost-effectiveness and simplicity of the technical method (Sayah et al., 2005). The underlying assumption of ARA is that humans and different types of animals, both domesticated and wild, are subjected to different types of antibiotics. This exposure, which varies in frequency and concentration, produces varying resistance in the internal flora of animals to antibiotics via selective pressure mechanisms (Scott, 2002). These internal flora include bacteria normally found in the gut of warm-blooded animals such as *E. coli*, the indicator utilized in this study. The antibiotic resistance profiles of the normal gut flora produce patterns that act like a “fingerprint” that can identify different animal groups (Scott, 2002). The sources can then be differentiated using the statistical method, discriminant analysis (Huberty, 1994; Hagedorn et al., 1999; Harwood et al., 2000).

Carbon source utilization (CSU) is a phenotypic library-based technique that relies upon proprietary technology developed by Biolog, Inc. (Haywood, CA). A standardized suspension of each bacterial isolate is inoculated into a 96-well plate prepared by

Biolog, Inc. This plate contains 1 blank well (water) and 95 wells lined with different substrates to be used for metabolism by the bacteria. The wells are also coated with a color change reagent, tetrazolium violet. As bacteria metabolize the substrate within the well, the tetrazolium violet changes from a colorless, oxidized form to a purple reduced form (Biolog, 2006). This color change is measured in terms of intensity, approximately 22-24 hours after incubation of the plates, utilizing the Biolog MicroStation plate reader. The intensity readings from each well combine to form a profile for that isolate, which Biolog software compares with known profiles to provide an identification of the bacteria. The intensity readings from each well can also be used to compile a CSU profile for each isolate that can then be analyzed with discriminant analysis for classification.

Pulsed-Field Gel Electrophoresis (PFGE), a genotypic library-based technique, is considered one of the best techniques to discriminate between strains of bacteria in complex bacterial matrices (Hahm et al., 2003, Meays et al., 2004, McLellan et al., 2003, Zhechko et al., 2005). PFGE analysis separates genomic fragments in an electrophoresis gel by subjecting them to electrical fields that alternate in perpendicular orientations (Meays et al., 2004). The gels are stained and banding patterns emerge to produce "genetic fingerprints," which can be compared to known source profiles for classification. It is standardized, reliable, and reproducible which makes it useful in comparative genetic analysis (Cameron et al., 1994, Lu et al., 2004, Okwumabuna et al., 2005). PFGE, however, is expensive and requires a certain level of technical expertise.

As the three BST methods employed in this study (ARA, CSU, PFGE) are all library-dependent; library composition plays an important role in analysis. The ARA library utilized in previous studies was expanded under this project to include additional human, horse, cow and wildlife source isolates. New fecal samples were collected from wastewater treatment outflows (human) and ranches (horse, cow, feral hog), and isolates were identified, analyzed and added to the library. Additionally, known source isolates collected under previous projects lacking CSU data were revived from frozen storage at TAMU-CC and analyzed via CSU (Biolog system). This data was then added to the CSU library for known source isolates. The PFGE library utilized in previous studies was expanded to include additional human, feral hog, horse and duck isolates to create a larger and more representative library.

Objectives

The overall project objective was to identify and quantify the relative contributions of various sources of fecal contamination to the tidal portions of Aransas and Mission Rivers for development of a TMDL by TCEQ. This objective was accomplished via completion of the following tasks:

1. Expansion of the existing *E. coli* library at TAMU-CC using fecal samples from the Copano Bay area for comparison with water isolates of *E. coli*.

2. Isolation of *E. coli* from Mission and Aransas Rivers water samples and comparison of their antibiotic resistance profiles and pulsed-field gel electrophoresis "fingerprints" with those in the TAMU-CC database for source tracking purposes.
3. Submission of data report to the GMP and CBBEP Project Managers for use in determining sources of fecal contamination in the Mission and Aransas Rivers.

METHODS

All methods followed the approved Quality Assurance Project Plan (QAPP) for this study (Mott and Lehman, 2006).

Field Collections

Water samples were collected by TAMU-CC personnel from five stations currently monitored by the Nueces River Authority (NRA) Clean Rivers Program (CRP). Figure 1 shows the general location of the sampling area. Two stations are located along the Mission River, 12943 and 12944 (Figure 2) and three along the Aransas River, 12947, 12952, and 12945 (Figures 2, 3 and 4).

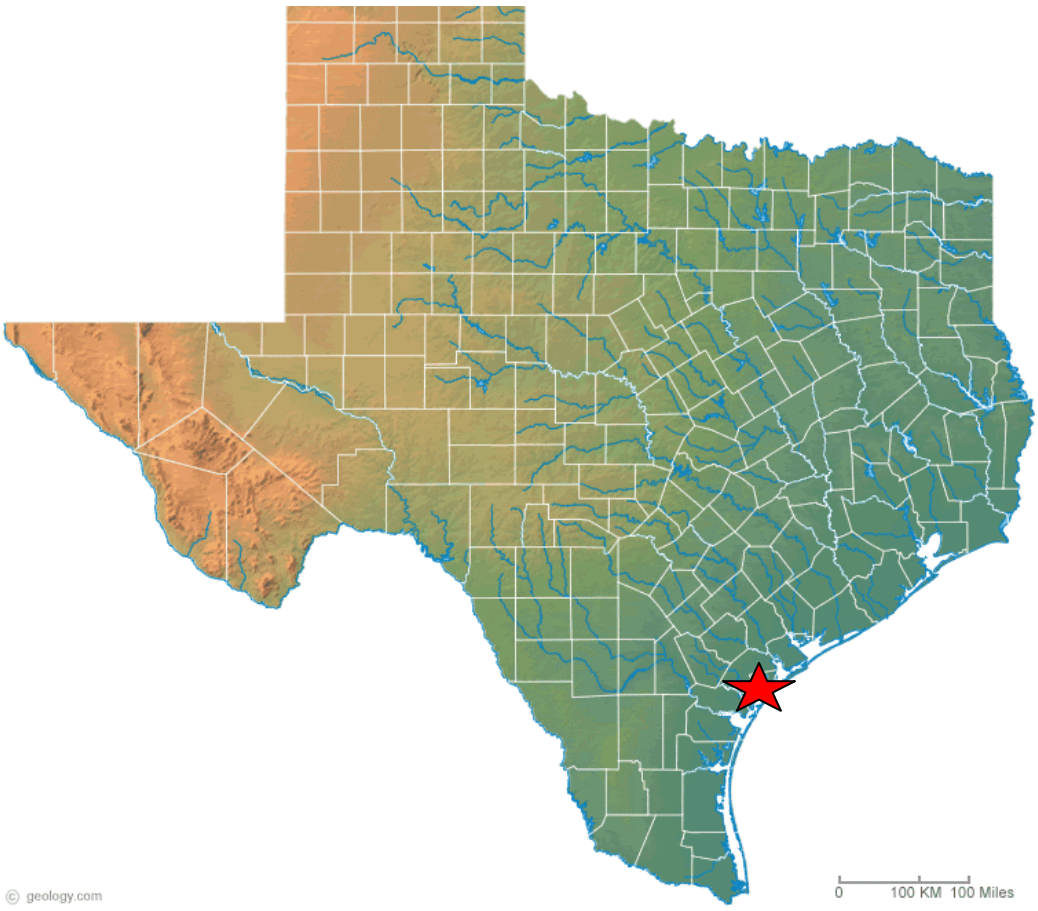
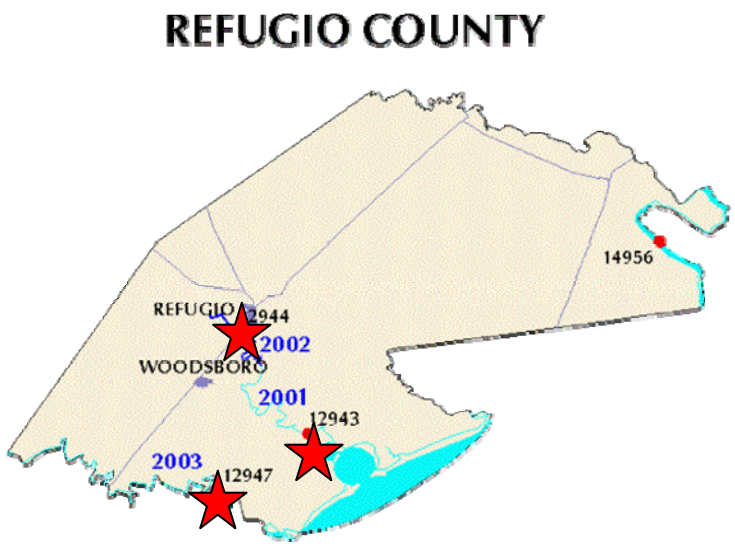


Figure 1. Map of Texas showing location of study area

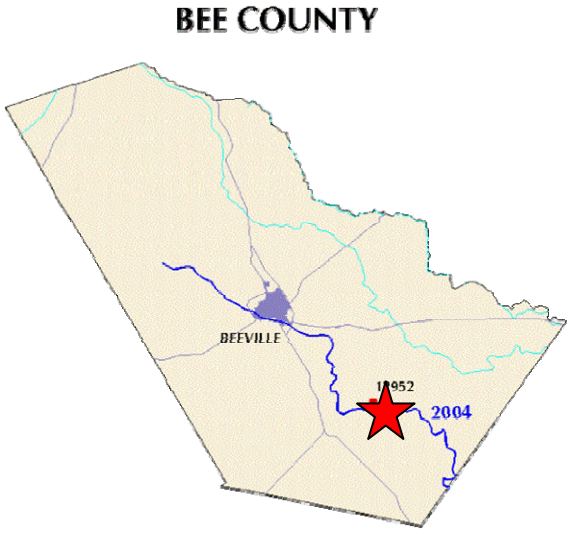


12947
28° 07'19"
-97° 18' 36"

12944
28° 17' 33"
-97° 16' 43"

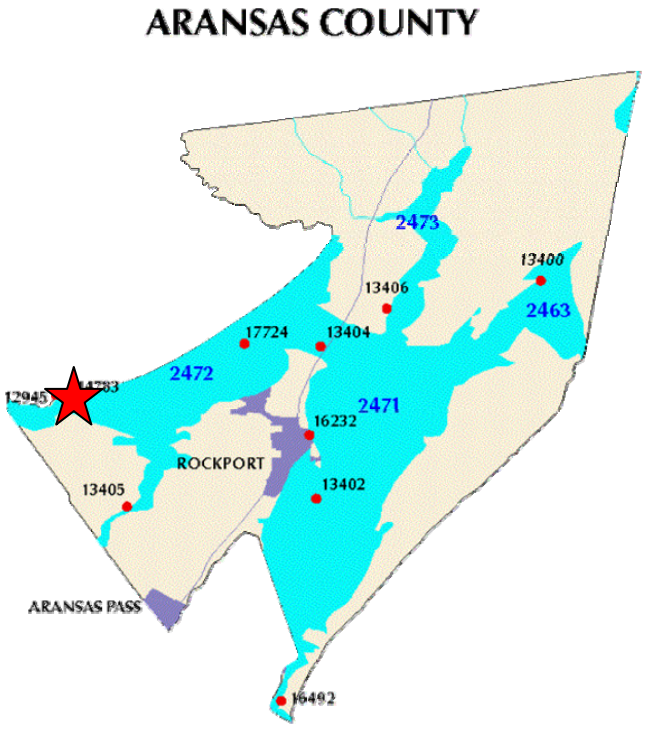
12943
28° 11' 02"
-97° 12' 49"

Figure 2. Mission River sampling stations, 12943 and 12944, and Aransas River sampling station, 12947. (Courtesy Nueces River Authority).
(Red dots indicate other stations monitored by the NRA, not included in this study).



12952
28° 16' 58"
-97° 37' 20"

Figure 3. Aransas River sampling station, 12952. (Courtesy Nueces River Authority).



12945
28° 04'41"
-97° 13' 16"

Figure 4. Aransas River sampling station, 12945. (Courtesy Nueces River Authority). (Red dots indicate other stations monitored by the NRA, not included in this study).

Field sampling procedures documented in the TCEQ *Surface Water Quality Monitoring Procedures Volume 1: Physical and Chemical Monitoring Methods for Water, Sediment and Tissue* (December 2003) were followed for collection of water samples and measurement of field parameters.

Water samples were collected in sterile, one liter polypropylene bottles from each station during three events: 3/26/07, 12/10/07, and 3/7/08. The final event, 3/7/08, was considered a rainfall event, with 1.16" of rainfall in the 24 hour period preceding the sampling event. The first two sampling events were considered dry events. All unknown source isolates analyzed for this project originated from water samples collected during these three events.

Field parameters for water collection events were documented on field data sheets for each station. These included water appearance, weather condition (clear, overcast, cloudy, drizzle or rain), wind intensity and direction, and air and water temperatures. Any human use (fishing, kayaking etc.) was also recorded. Precipitation data were obtained through the National Oceanic and Atmospheric Administration (NOAA) website (www.noaa.gov). Tables summarizing the field data collected during sampling events are located in Appendix A.

Fecal samples from nonhuman sources were collected from a variety of sites, including the Welder Wildlife Refuge, Fennessey Ranch, and several private ranches in the Mission and Aransas watershed. Samples from nonhuman sources (cow, horse, deer and hog) were collected using BD BBL EZ Culture swabs. Fecal sample collection data sheets with date, time, location of specimen collection and air temperature at site were completed for each nonhuman sample.

Fecal samples from human sources were collected from treated wastewater outflows at three facilities: City of Taft WWTP, St. Paul WCS, and Woodsboro WWTP. Fecal samples from humans were collected from the plant outfalls in polypropylene, screw-cap, sterile specimen containers. Chain-of-custody for all human specimens documented date, time and location of specimen collection. Table 1 summarizes the number of samples for each fecal source and the subsequent number of isolates analyzed from each source.

All samples were placed on ice after collection and transported directly to the TAMU-CC laboratory. All water samples (including wastewater treatment plant samples) were analyzed immediately after reception in the laboratory. Fecal samples were received into laboratory custody and refrigerated until analysis could commence (typically less than 24 hours after collection of sample).

Table 1. Numbers of known source *E. coli* isolates, verified and analyzed from Mission and Aransas River area fecal sample collections.

Animal Source	Collection Dates	Number of Samples	Number of Isolates
Cow	11/28/07-04/02/08	65	144
Deer	12/12/2007	2	3
Feral Hog	11/28/07-12/12/07	6	18
Horse	12/13/07-04/02/08	42	58
Human (treated wastewater)	10/02/08-10/04/08	8	125
Total	N/A	123	348

Laboratory Analysis

Isolation of *E. coli* from Water Samples

The water samples (including wastewater treatment plant samples) were filtered following USEPA Method 1103.1: the original *E. coli* method, developed by Alfred Dufour (1981) and adopted by the USEPA in 1986 (USEPA, 1986). Filtering procedures also followed the Improved Enumeration Methods for the Recreational Water Quality Indicators: Enterococci and *Escherichia coli* (2000) USEPA/821/R-97/004 and quality control methods outlined in Standard Methods for Examination of Water and Wastewater, 21st ed. (APHA, 2005).

Varying volumes of water were filtered (ranging from 50 to 200mL) for each sample onto 0.45 micrometer nitrocellulose filters with a total of approximately 1000mL filtered for each sample. For each event, any surplus water was stored at 4°C for up to 24 hr.

The membrane filters were placed onto m-TEC media plates and were incubated at 35°C for 2 hr and then at 44.5°C in a circulating water bath for 22 hr as required by the method. Following incubation, the isolates were verified by hydrolysis of urea as outlined in USEPA Method 1103.1. Initially, 25 *E. coli* isolates were to be obtained from each station; however, some stations yielded less than 25 isolates. Additional isolates were collected from other stations in these cases. The isolates were transferred from m-TEC to tryptic soy agar (TSA) slants and were stored at 4°C in the TAMU-CC Environmental Microbiology Laboratory. Each *E. coli* isolate was labeled according to the sampling date and station number.

To isolate pure colonies, each isolate was streaked onto Rainbow Agar O157® (Biolog, Inc.), a chromogenic media used for the isolation of *E. coli* strains, specifically O157:H7 (Biolog, 1994). The Rainbow Agar® plates inoculated with *E. coli* were incubated at 35°C for 18-24 hrs (Biolog, 1994). The plates were removed after incubation and colonies were evaluated. Cultures on Rainbow Agar plates containing a single color type indicative of *E. coli* (magenta, purple, pink, gray and black) were considered as pure cultures. Re-streaking on Rainbow Agar O157® was performed as necessary, until pure cultures were obtained.

Isolation of *E. coli* from Fecal Samples

Swabs of fecal samples were streaked onto mTEC agar plates and incubated at 35°C for 2 hr and then at 44.5°C for 22 hr. At least five yellow colonies were transferred from each mTEC plate (unless the plates contained < 5 colonies, in which case all colonies were transferred) onto Rainbow Agar O157® (Biolog, 1994) plates, and incubated at 35°C for 18-24 hr. After incubation, magenta, purple, pink, grey and black colonies, indicative of *E. coli*, were transferred from Rainbow Agar to trypic soy agar (TSA) slants and were stored at 4°C. Confirmation of isolates as *E. coli* using the MicroLog™ Microbial Identification System (Biolog, 1999), followed procedures described below.

Verification of Isolates as *E. coli* and CSU profiling

All isolates, known source and unknown source, were confirmed as *E. coli* using the MicroLog™ Microbial Identification System (MIS) (Biolog, Inc.) following the MicroLog™ System Release 4.0 User Guide (Biolog, 1999). Each isolate was transferred to a Biolog™ Universal Growth (BUG B) plate with 5% sheep's blood agar (Biolog, 1999). The plates were incubated at 35°C for 24 hrs. Growth from the plates was transferred to inoculating fluid (0.4% NaCl, 0.03% Pluronic F-68, 0.01% Phytagel™) to reach a transmittance level of 61% (±2%) at 600nm. This inoculum was pipetted into a 96 well GN2 Microplate™, and the plates were incubated for 16-24 hrs at 35°C. After the incubation period the plates were read using the MicroLog System™, Release 4.20.04 (Biolog, 2004). Only isolates that identified with a 90% or greater probability as *E. coli* were used for antibiotic resistance analysis. The carbon source utilization (CSU) profiles generated from the MicroLog System were also added to the data base for classification

of isolates. Each profile was composed of 95 data points corresponding to the individual wells of the microplate used in this analysis.

Isolates that were confirmed as another bacterium or could not be identified were discarded. Commonly identified non-*E. coli* isolates included *Buttiauxella izardii*, *Buttiauxella agrestis*, *Leclercia adecarboxylata*, and *Rahnella aquatilis*. These four organisms are typically found in aquatic habitats and share the common family, Enterobacteriaceae, with *E. coli* (Muller et al., 1996; Dorkin et al., 2006).

The *E. coli* isolates confirmed with the MicroLog™ Microbial Identification System were transferred to cryogenic storage vials. The *E. coli* isolates were transferred from TSA slants to 5 ml trypic soy broth (TSB) tubes and then incubated at 35°C for 16-18 hrs with shaking. After incubation, 600 µl of the bacteria culture was pipetted into a cryovial containing 400 µl of 80% glycerol (w/v). The sample was gently mixed and placed in a vial box holder, which was subsequently stored at – 70°C. Duplicate vials were made for each *E. coli* isolate.

Antibiotic Resistance Analysis

Known (fecal samples) and unknown (water samples) source isolates confirmed as *E. coli* with CSU were each analyzed to develop antibiotic resistance profiles (ARPs). ARA was performed via the Kirby-Bauer disk diffusion method following guidelines of the Clinical and Laboratory Standards Institute (formerly the National Committee for Clinical Laboratory Standards (NCCLS)) (NCCLS, 2002, CLSI, 2006a,b). Cultures of *E. coli* were transferred from TSA slants into 5 ml of TSB and placed on an orbital shaker. The cultures were incubated at 35°C for approximately four to six hours. Cultures were transferred into 12 x 75 mm cuvettes and turbidity levels were adjusted to an absorbency of 0.08 to 0.1 at 625 nm, using a calibrated spectrophotometer. Each culture was streaked three times onto two Mueller-Hinton agar (BBL) plates using a sterile swab to ensure a consistent lawn across the agar surface. Two BBL Sensi-Disk™ disk dispensers were used to place antibiotic-impregnated disks (BBL) onto MHA plates. Each dispenser contained a separate set of 10 antibiotics (Table 2), and each set of antibiotics was dispensed on a separate MHA plate inoculated with each isolate. The MHA plates were allowed to sit for five minutes before being placed (inverted) in a 35°C incubator for 16-18 hrs.

BIOMIC ®, an automated plate analyzer, was used to automatically measure zones of inhibition diameters and to calculate minimum inhibitory concentrations (MICs), using EXPERT software. Procedures were followed according to CLSI document M100-S16 (2006b). BIOMIC ® software was also used to classify the bacteria as Susceptible (S), Intermediate resistant (I), or Resistant (R) to a given antibiotic, as determined by the diameter measurements of inhibition zones (Table 3). The results of ARA were printed and also stored electronically. Discriminant analysis (Huberty, 1994) was performed on the zone diameters using SPSS software (Release 15.0, 2006).

Table 2. Antibiotics used to develop antibiotic resistance profiles for *E. coli* isolates from Mission and Aransas River Tidal Segments 2007-2008.

Antibiotic	Abbreviation	Concentration
Ampicillin	AMP	10 µg
Augmentin	AmC	30 µg
Cefazolin	CZ	30 µg
Cefotaxime	CTX	30 µg
Ceftazidime	CAZ	30 µg
Ceftriaxone	CRO	30 µg
Chloramphenicol	C	30 µg
Ciprofloxacin	CIP	5 µg
Doxycycline	D	30 mg
Enrofloxacin	ENO	5 µg
Gentamicin	GM	10 µg
Imipenem	IPM	10 µg
Kanamycin	K	30 µg
Nalidixic acid	NA	30 µg
Neomycin	N	30 µg
Spectinomycin	SPT	100 µg
Streptomycin	S	10 µg
Sulfamethoxazole Trimethoprim	SXT	23.75/1.25 µg
Sulfisoxazole	G	0.25 mg
Tetracycline	Te	30 µg

Table 3. Susceptible (S), Intermediate (I), and Resistant (R) ranges (mm) for *E. coli* using the BIOMIC® Microbiology Analyzer System (2007 Update).

Antibiotic	S	I	R
AMP	≥ 17	14-16	≤ 13
AMC	≥ 18	14-17	≤ 13
CZ	≥ 18	15-17	≤ 14
CTX	≥ 23	15-22	≤ 14
CAZ	≥ 18	15-17	≤ 14
CRO	≥ 21	14-20	≤ 13
C	≥ 18	13-17	≤ 12
CIP	≥ 21	16-20	≤ 15
D	≥ 14	13-11	≤ 10
ENO	≥ 21	16-20	≤ 15
GM	≥ 15	13-14	≤ 12
IPM	≥ 16	14-15	≤ 13
K	≥ 18	14-17	≤ 13
NA	≥ 19	14-18	≤ 13
N	≥ 17	13-16	≤ 12
SPT	≥ 18	15-17	≤ 14
S	≥ 15	12-14	≤ 11
SXT	≥ 16	11-15	≤ 10
G	≥ 7	NA	≤ 6
TE	≥ 15	14-12	≤ 11

Pulsed-Field Gel Electrophoresis
Known Source Isolates
Pulsed-field gel electrophoresis (PFGE) was used to obtain genetic 'fingerprints' to augment the existing known source library and was performed following Bio-Rad Methodology and Standards as described by the manufacturer (Bio-Rad Laboratories, 1995). Some of the isolates successfully analyzed were from an existing TAMU-CC library but had not been analyzed using PFGE (40 duck isolates) while others were from fecal samples collected during this study (13 cow, 28 horse, 4 hog, 26 human).

DNA was extracted, cut with the restriction enzyme *Not* I, embedded in agarose, and fingerprinted. After processing and running the DNA plugs for 20 hours in a CHEF-DR III Gel Electrophoresis Unit (Bio-Rad, Hercules, CA), the gels were stained with Ethidium Bromide, de-stained in double deionized distilled water with 1% TBE, and then photographed using the Gel-Doc System (Bio-Rad, Hercules, CA). A minimum of two photographs were printed and digital images for analysis with Quantity One (Bio-Rad, Hercules, CA) was created. For analysis, lanes were superimposed on the digital image from the bottom of each plug to the bottom of the gel. All samples that yielded distinct bands along with the standard had lane overlays traced on them and were adjusted for any curvatures. The lanes extended from the plug well to the bottom of the gel. A Gaussian curve was used to aid in establishing the banding patterns for each lane and background noise on each gel was subtracted (Bio-Rad Laboratories, 1999a).

A database of all the isolates (known and unknown) was created using Diversity Database Fingerprinting Software (Bio-Rad, Hercules, CA). The first band in the set was based on the first band of the standard, lambda, and the subsequent bands were based on the software's assignment. A band set was then assigned to each isolate to complete the PFGE profile. All gels were manually inspected and gels were adjusted to eliminate any software errors due to gel irregularities (Duck et al., 2003, McLellan et al., 2003).

Pulsed-Field Gel Electrophoresis
Unknown Source Isolates
Water (unknown) isolates were analyzed as described above, for known source isolates. A subset of the isolates analyzed by antibiotic resistance was analyzed (~10 per station per event). The unknown isolates were classified by comparing them with the known source isolates to determine their closest similarity using Diversity Database Fingerprinting Software.

Quality Control

Quality assurance and quality control protocols were followed according to the Quality Assurance Project Plan "Study to investigate sources of *E. coli* isolated from Mission and Aransas Rivers" (Mott and Lehman, 2006).

All data related to the sample (e.g. chain of custody, field data sheets) and log sheets related to equipment and reagents used during analysis were maintained according to

QA/QC protocols and stored in the TAMU-CC Environmental Microbiology Laboratory as per Table A.1 of the QAPP.

Accuracy was verified through the analysis of control standards and sterility checks. Positive (*E. coli*) and negative (*Enterobacter aerogenes*) culture controls were selected according to Table 9020:V (APHA, 2005). The American Type Culture Collection (ATCC) 25922 was selected for *E. coli* and ATCC 13048 was selected for *E. aerogenes*. These positive and negative controls were used for each batch of the selective media, m-TEC agar and Rainbow Agar O157®. Sterility checks were performed on all media before use. Positive controls were performed on all media, including MHA, TSA and TSB. One of three strains of *E. coli* (ATCC 25922, 8739, and 35218) was analyzed with each batch of GN2 Microplates™. Quality control for *E. coli* isolations followed that of the method (USEPA, 2000).

For quality assurance, ATCC strains *Staphylococcus aureus* ATCC 25923, *E. coli* ATCC 25922, and *E. coli* ATCC 35218 were included with each batch of samples analyzed with ARA to ensure that media, antibiotics, and computer software were within standards as indicated in the CLSI Performance Standards (NCCLS, 2002, CLSI 2006a, 2006b).

Control limits are specified in software associated with each technique to be used. MicroLog™ Microbial Identification System provides a % similarity of each isolate with known bacteria in the Biolog database and BIO-MIC® (for ARA analysis) follows NCCLS standards. The PFGE database was created with Diversity Database (Bio-Rad, Hercules, CA) and all samples (both known and unknown) were analyzed based on the standard lambda (Bio-Rad, Hercules, CA). Lambda ladders are frequently used as standards to normalize PFGE patterns for comparison between different gels (Duffy et al., 2005, Lu et al., 2004).

Statistical Analysis of the Known Source Library and Unknown Sources

For ARA, CSU and a composite dataset (ARA + CSU), the databases of profiles forming the library of known sources and unknown sources were analyzed using discriminant analysis with SPSS® Version 15.0 for Windows. Discriminant analysis was performed utilizing equal prior probabilities for each group. This allowed all groups of known source isolates to be represented equally regardless of the amount of representation in the overall library. With discriminant analysis each isolate was classified based on the similarities with groupings of known source profiles. Data for each method (ARA alone, CSU alone, and the composite data set) were analyzed by two-way (human vs. nonhuman), three-way (human vs. livestock vs. wildlife), and six-way (human vs. cow vs. horse vs. duck vs. gull vs. wildlife).

For PFGE analysis, each unknown isolate was run against the database using the Diversity Database Fingerprinting Software to determine the source of the unknown isolate. The known source with the highest percent similarity to the unknown isolate was

determined to be the source of the unknown. The similarity index was automatically calculated by the software as a function of the number of bands the compared isolates had in common divided by the number of bands in each isolate lane (Singer et al., 2004). The band types were compared as unweighted, as weighting the results compares the relative brightness of the band, which can be highly variable from gel to gel and even among lanes on the same gel. Numerous studies have analyzed their data using unweighted methods (Duffy et al., 2005, Singer et al., 2004). Unweighted band analysis resulted in only the position and number of bands in the lanes being compared to determine their percent similarity.

Band types were compared using the Jaccard matching coefficient. The Jaccard matching coefficient was used due to its high efficiency of distance projection in a two-dimensional space (Duarte et al., 1999). The Jaccard matching coefficient assigns a binary value of 0 or 1 with 0 representing the absence of a band at a particular location in the lane and 1 representing the presence of a band in a particular location (Soll et al., 2000). The Jaccard coefficient for two banding patterns is calculated using the formula $S_j=n_{AB}/(n_{AB}+a+b)$ (Soll et al., 2000). In this formula, S_j is the degree of commonness, nAB is the number of bands in common in lanes A and B (bands being compared), a is the total number of bands in lane A not present in lane B, b is the total number of bands in lane B not present in lane A, and $a+b$ is the total number of mismatches (Soll et al,. 2000). Degrees of similarity are based on increasing values between 0.0 and 1.0 (Soll et al., 2000).

RESULTS

Library development

The fecal sample collection results for this project are shown in Table 4. These isolates were added to an existing TAMU-CC *E. coli* database for the watershed, collected previously under approved QAPPs, to form a library for this study (Table 5). The total number of isolates for which ARPs were developed was 1176 and for CSU was 1033. However, some isolates from the existing library, which did not have CSU intensity profiles could not be re-grown for analysis; while other isolates collected in this study could not be confirmed to use for ARP development. Only known source profiles that had both ARA and CSU data were used in this library, yielding a combined library of 925 isolates. Adding the extra isolates to both ARA alone and CSU alone libraries did not significantly increase the accuracy of the results (Appendix B).

Table 4. Numbers of known source *E. coli* isolates, verified and analyzed for Mission and Aransas River area fecal sample collections through Antibiotic Resistance Profiles (ARP) and Carbon Source Utilization profiles.

Animal Source	Number of Isolates	Number CSU Completed	Number ARA Completed
Cow	144	64	23
Deer	3	3	2
Feral Hog	18	18	3
Horse	58	58	29
Human (treated wastewater)	125	79	61
Total	348	222	118

Table 5. Isolates composing the known source library for ARA and CSU, classified by animal source.

Animal Source	Number of Isolates
Cow	143
Ducks	98
Gulls	66
Horse	269
Human-untreated*	126
Human-treated effluent**	61
Wildlife (non-avian)	162
Total	925

*Note: Human untreated isolates originated from wastewater treatment inflows and were not taken directly from human sources

**Note: Human treated effluent isolates originated from wastewater treatment outflows

The composition of the additional PFGE known source isolates added for this project, and the earlier library, is outlined in Table 6.

Table 6. Isolates composing the known source library for PFGE, classified by animal source.

	Animal Source				
	Cow	Horse	Duck	Hog	Human
Number of Isolates collected and analyzed in previous Copano Bay Project	25	22	10	0	27
Number of Isolates collected in Copano project, analyzed in this project	0	0	40	0	0
Number of Isolates collected and analyzed in this project	13	28	0	4	26
Total	38	50	50	4	53

ARA and CSU Analysis

The percentages of known source isolates analyzed with ARA, CSU and the combination of the two (composite dataset) that were correctly classified using discriminant analysis are summarized in Table 7. Discriminant analysis was performed on each known source isolate using three different classification schemes with increasing specificity of known source categories. The most general classification scheme distinguished between human and nonhuman sources.

During the course of the statistical analysis for this project the two groups of known source isolates originating from humans were determined to be statistically different. The human isolates collected during a previous project were collected from untreated wastewater from two plants, Beeville WWTP and the Rockport Reclamation Plant. The human isolates added under this project were collected from treated wastewater from three facilities: City of Taft WWTP, St. Paul WCS, and Woodsboro WWTP. These two isolates formed distinct populations when analyzed with discriminant analysis and were considered to be two different human sources. All discriminant analyses were performed utilizing these two pools of humans, designated as human (untreated) and human (treated), respectively.

The most basic classification, between human and nonhuman sources, was designated a three-way classification, with the categories nonhuman, human (untreated) and human (treated). In addition, a four-way classification was performed that distinguished between human, livestock (cow and horse), and wildlife sources. Wildlife sources included feral hog, deer, skunk, raccoon, and other wildlife. The most complex classification, seven-way, classified isolates into the seven main source categories: cow, horse, duck, gull, wildlife, human (untreated) and human (treated).

Table 7. Percentage of Average Correct Classification (ARCC) of known source isolates (i.e. library) based on their Antibiotic Resistance Profiles (ARPs) and Carbon Source Utilization (CSU) profiles using three-, four- and seven-way classifications.

Human/Nonhuman (three-way)			Human/Livestock/Wildlife (four-way)			All categories (seven-way)		
ARP	CSU	ARP+CSU	ARP	CSU	ARP+CSU	ARP	CSU	ARP+CSU
59.1	89.9	93.2	58.6	83.4	88.4	55.2	80.9	86.4

Overall, the combination of ARA and CSU analysis provided the most powerful form of classification. CSU alone provided a higher level of correct classification than ARA alone. However, CSU utilized 95 variables (individual carbon sources) to classify each isolate; ARA used only 20 variables (individual antibiotics). The larger number of variables allowed for increased accuracy of classification of specific groups.

Average rates of correct classifications (ARCCs) decreased with increasing number of categories in the classification. This is expected, as more variation in classification of a single isolate may be introduced when there are a larger number of categories. The difference between the ARCC of the three-way composite data set (93.2%) and the seven-way set (86.4%) was 6.8%. This is a very minor reduction in correct classification rates in relation to the number of added categories and provided a high level of accuracy for a seven-way classification. The library was therefore deemed able to accurately classify isolates at all levels of discrimination of categories.

The unknown source isolates analyzed for this project originated from water samples taken during three sampling events - two dry weather events (3/26/07 and 12/10/07) and one event following 1.16 inches of rainfall (3/7/08). Table 8 summarizes the number of unknown source isolates analyzed by both ARA and CSU. Five stations were sampled, with similar numbers of isolates analyzed from each station for each event. Water samples from stations 12943 and 12945 did not provide the number of isolates (25), originally planned, from the second dry weather sampling event (12/10/07). Additional isolates from other stations were analyzed to compensate for this lower recovery of isolates from stations 12943 and 12945. The number of isolates analyzed from each station is shown in Table 9.

Table 8. Number of unknown source isolates analyzed using ARA and CSU, grouped by sampling event.

Event 1 (3/26/07)	Event 2 (12/10/07)	Event 3 (3/7/08)	Total
Dry Event	Dry Event	Wet Event	
147	96	121	364

Table 9. Number of unknown source isolates analyzed using ARA and CSU, grouped by sampling station.

Station					
12952	12947	12944	12943	12945	Total
82	70	78	67	67	364

Profiles (ARA and CSU) of unknown source isolates were compared with the known source library profiles to classify each isolates to a specific source using discriminant analysis. Table 10 summarizes the percentage of unknown source isolates from water samples attributed to specific sources using the composite database library. All unknown source isolates were classified against the known source libraries using profiles generated from ARA alone, CSU alone, and a composite dataset from ARA and CSU. The same classification schemes described in the above section on known source isolate analysis were utilized for unknown source isolates. Detailed tables from discriminant analysis for ARA alone, CSU alone, and the composite dataset are shown in Appendix C.

Table 10. Classifications of unknown source isolates into source categories (percentages) using the composite dataset (library).

Animal Source	Human/Nonhuman (three-way)	Human/Livestock/Wildlife (four-way)	All Categories (seven-way)
Cow	36.0	14.3	12.7
Horse			7.6
Duck		21.7	7.0
Gull			2.7
Wildlife			2.7
Human (untreated)	0.3	0.0	0.4
Human (treated)	63.7	64.0	66.9

As the composite data set (ARA+CSU), showed the highest ARCC (Table 7), the results of the discriminant analysis using the composite dataset (library) were used for source identifications. Using this data set 36% of the unknown source isolates were categorized as nonhuman sources of which 14.3% classified as livestock and 21.7% as wildlife. Further classification into seven sources attributed 12.7% of the unknown source isolates to cattle, with an additional 14.6% being divided almost equally between horse and duck categories. The remaining nonhuman pool, 5.4%, was classified

equally between gull and wildlife sources. Results of source categorizations of unknown sources using ARA alone and CSU alone, for the three levels of classification, are located in Appendix C (Table C23).

All three levels of classification attributed 63.7-66.9% of the unknown source isolates to human sources. The overwhelming majority of this human contribution was assigned to the category containing isolates from treated wastewater effluent.

The data was further analyzed by sampling event. The unknown source isolates from each event were grouped together for additional statistical analysis. Percentages of unknown source isolates profiles from the composite dataset classified as specific sources for the dry and wet events are shown in Table 11.

Table 11. Classifications of unknown source isolates into source categories (percentages) using the composite dataset (library), grouped by sampling event.

	Human/Nonhuman			Human/Livestock/Wildlife			All Categories		
	Event 1 (dry)	Event 2 (dry)	Event 3 (wet)	Event 1 (dry)	Event 2 (dry)	Event 3 (wet)	Event 1 (dry)	Event 2 (dry)	Event 3 (wet)
Cow	27.9	17.7	60.3	10.9	7.3	24.0	8.2	4.2	16.5
Horse							3.4	4.2	9.9
Duck				18.4	10.4	34.7	6.8	7.3	8.3
Gull							3.4	0.0	5.0
Wildlife							3.4	0.0	5.8
Human(un-treated)	0.0	0.0	0.8	0.0	0.0	0.0	0.0	0.0	0.8
Human (treated)	72.1	82.3	38.8	70.7	82.3	41.3	74.8	84.4	53.7

For both dry events (Events 1 and 2), more than 70% of the unknown source isolates classified as human (treated) with the three-way classification. In contrast, only 38.8% isolates from the sampling event following rainfall were attributed to treated human wastewater (three-way classification). This difference is consistent for both the four-way and seven-way classifications. With fewer unknown source isolates classifying as human, a greater proportion of Event 3 (wet event) isolates were categorized as livestock and wildlife sources. The number of isolates classifying as livestock (horse/cow) was more than double after rainfall (Event 3) than for the dry weather events. Differences in the percentage of isolates classifying as wildlife between dry and wet weather events were less pronounced.

Unknown source isolate profiles generated from the composite dataset were also grouped by station and analyzed with discriminant analysis. Percentages of unknown source isolates profiles from the composite dataset classifying as specific sources for each sampling station are summarized in Table 12. No major differences in source contribution were found between stations. All stations appeared to have approximately

the same contribution from human (treated) sources. Nonhuman contributions differed slightly between stations. For station 12947 21.4% of unknown source isolates were attributed to livestock (cow, horse) - slightly elevated in comparison with the other stations.

Table 12. Seven-way classification of unknown source isolates into source categories (percentages) using the composite dataset (library), grouped by station.

Animal Source	12952	12947	12944	12943	12945
Cow	14.6	10.0	7.7	6.0	10.4
Horse	2.4	11.4	5.1	7.5	3.0
Duck	7.3	5.7	9.0	4.5	10.4
Gull	4.9	0.0	2.6	4.5	3.0
Wildlife	2.4	1.4	2.6	7.5	3.0
Human (untreated)	0.0	0.0	1.3	0.0	0.0
Human (treated)	68.3	71.4	71.8	70.1	70.1

Pulsed-Field Gel Electrophoresis Analysis

A subset of unknown source isolates (~10 per water sample/50 per event) was analyzed with Pulsed-Field Gel Electrophoresis. The number of unknown source isolates analyzed and for which 'fingerprints' were obtained using PFGE is shown in Table 13. A total of 153 unknown source isolates were successfully analyzed with PFGE for this project.

Table 13. Number of unknown source isolates verified and analyzed using Pulsed-Field Gel Electrophoresis (PFGE).

Water Event	Number of Isolates Analyzed	Number of Isolates Successfully Fingerprinted
Event 1	142	51
Event 2	75	52
Event 3	80	50
TOTAL	297	153

For PFGE analysis, each unknown isolate was run against the known source database of 195 isolates (Table 6) using the Diversity Database Fingerprinting Software to determine the unknown isolate source. The source assigned to each unknown source

isolate was that of the known source isolate with the highest percent similarity to the unknown isolate. The classifications of unknown source isolates from PFGE analysis, both number of isolates and percentages, are summarized in Table 14. It must be noted that these isolates were a randomly selected subset of the original isolates obtained from each station, and therefore may not reflect the results of the ARA+CSU composite analysis of the entire set of samples. Of the 153 unknown source isolates analyzed with PFGE, 40.5% were classified as human. The remaining unknown source isolates largely classified as cow, duck and horse (58.1% total), with only a small percentage (1.3%) classified as feral hog.

Table 14. Classification of unknown source isolates using Pulsed-Field Gel Electrophoresis (PFGE) analysis .

	Cow	Horse	Duck	Hog	Human	Total
Number of Unknown Isolates Classified	27	28	34	2	62	153
Percentage of Unknown Isolates Classified	17.6	18.3	22.2	1.3	40.5	100.0

Classifications of unknown source isolate profiles generated with PFGE were also grouped by event (Table 15). Source contributions were similar for each event. As with the overall data (Table 14), human sources accounted for the largest proportion of isolates for each event.

Table 15. Classification of unknown source isolates (percentages) using Pulsed-Field Gel Electrophoresis (PFGE) analysis, grouped by sampling event.

Animal Source	Event 1 (Dry)	Event 2 (Dry)	Event 3 (Wet)
Cow	13.7	19.2	20.0
Horse	17.6	23.1	14.0
Duck	17.6	21.2	28.0
Feral Hog	2.0	1.9	0.0
Human	49.0	34.6	38.0

PFGE profiles for unknown source isolates were also grouped by sampling station. These results are summarized in Table 16. Contributing sources appeared to vary slightly with location; however, due to the small number of isolates per station this variation may not be significant. For each station, the largest percentage of unknown source isolates classified as human (33.3-53.3%). Isolates from only one station

(12944) included a small percentage that classified as hog. A fairly consistent proportion of isolates from all stations classified as duck. Numbers of isolates classifying as cow and horse varied slightly amongst stations, most notably stations 12947 and 12943.

Table 16. Classification of unknown source isolates (percentages) using Pulsed-Field Gel Electrophoresis (PFGE) analysis, grouped by station.

Station	Animal Source				
	Cow	Horse	Duck	Feral Hog	Human
12952	16.7	23.3	26.7	0.0	33.3
12947	20.0	6.7	20.0	0.0	53.3
12944	16.1	12.9	25.8	6.5	38.7
12943	9.4	28.1	21.9	0.0	40.6
12945	26.7	20.0	16.7	0.0	36.7

CONCLUSIONS AND RECOMMENDATIONS

The majority of unknown source isolates collected from water samples at the five sampling stations along the Mission and Aransas tidal segments classified as human utilizing the three library-dependent bacterial source tracking techniques: antibiotic resistance analysis (ARA), carbon source utilization (CSU) analysis, and pulsed-field gel electrophoresis (PFGE).

The known source libraries for this study were composed of isolates collected from various animal and human sources throughout the Mission and Aransas watershed over several years, spanning multiple projects. They were supplemented with isolates collected under this project to increase underrepresented sources in the existing library. Collection of samples for additional non human isolates was constrained by limited cooperation from land owners and, for feral hog and deer, availability of scat. Every effort was made to collect as many animal fecal samples from diverse areas within the watersheds as possible.

Over 1000 known source profiles were utilized in the statistical analysis. Only known source profiles that had both ARA and CSU data were used in the final ARA+CSU composite library of 925 isolates. Adding isolates to both the ARA alone and CSU alone libraries did not significantly increase the accuracy of the results. The known source ARA+CSU library utilized for this project was considered large enough to be representative of the possible sources contributing to elevated bacterial levels in the Mission and Aransas tidal segments. The PFGE known source library developed was a subset of the ARA+CSU library and was composed of 195 isolates.

The most accurate dataset (library) for ARA and CSU was the composite dataset combining both types of profiles with an average rate of correct classification (ARCC) of 93.2% with three-way classification. The addition of other source categories did not diminish the ARCC greatly. The seven-way classification produced an 86.4% ARCC (compared with a 14.3% chance of an isolate being randomly assigned correctly). The fact that the ARCC was significantly greater than would be expected from random classification instills confidence in the results produced from this study (Stoeckel and Harwood, 2007). This level of correct classification from a phenotypic method is higher than usually obtained from ARA alone, particularly with respect to the seven-way classification. ARCCs for published studies using ARA range from 62-84%, but as the number of categories increases, the ARCC decreases (Meays et al., 2004, Wiggins, 1996). Our seven-way ARCC was higher than the typical range for a two-way classification using ARA alone. Our results are in line with the only currently published study combining ARA and CSU profiles for classification (Moussa and Massengale, 2008).

Overall, 63.7-66.9% of unknown source isolate profiles from the composite data set were classified as treated (WWTP effluent) human sources. PFGE profile analysis of a subset of the unknown source isolates confirmed this finding with 40.5% of the isolates having the highest similarity to human sources. The remaining unknown source isolates

were classified as livestock and wildlife, with cow, horse and duck contributions accounting for the majority of the animal sources using both analyses. The discrepancy between the percentage of isolates classified as human using the composite dataset and PFGE is most likely due to the smaller library size, with a somewhat different composition, and number of isolates analyzed by the latter method.

When the data was grouped by sampling event, the composite database analysis. classified fewer isolates from sampling event following rainfall (Event 3) as human and a larger proportion as livestock when compared with the other two events. This suggests a runoff effect following rainfall compared with point sources of contamination (e.g. WWTP outflows) during dry weather. PFGE analysis did not reflect this difference; however, this may again be due to the reduced size of the database analyzed using this technique.

Results grouped by sampling station demonstrated no major differences with respect to location. Both the composite ARA+CSU analysis and the PFGE analysis showed similar human source contribution at each station. Slight differences were found amongst stations with respect to percentages of contribution of livestock and wildlife sources. PFGE analysis showed only Station 12944 on the Mission River to have a few unknown source isolates classifying as feral hog (less than 2% of all isolates). For the ARA+CSU analysis, hogs were included in the category 'wildlife'.

This study examined sources of contamination in two rivers which flow into Copano Bay. A previous BST study for Copano Bay found that, overall, of the unknown source *E. coli* isolated from water samples, 55.6% could be classified as livestock (cow or horse) 22.1% as human (untreated), and 21.3% as duck (Mott and Lehman, 2005). Only a small percentage (1.1%) classified as wildlife or gull (Mott and Lehman, 2005). However, it should be noted that these proportions varied at different stations and for different sampling events.

For this study the library was supplemented with isolates from 'treated' WWTP outflows. In order to determine whether this would have affected the outcome of the previous Copano Bay study the unknown source isolates from the Copano Bay project were re-analyzed with the supplemented library used for this study. Few differences were observed in the results using the two different libraries. This appears to confirm that there are real differences in sources, not attributable to changes in the library.

Differences between the findings of the two studies in terms of sources of *E. coli* may be accounted for in several ways:

- The majority of the isolates analyzed from the bay study were collected following rainfall as the numbers of *E. coli* present during dry weather were very low for most stations, while in this study the majority was collected during dry weather. The ARA+CSU analysis of the isolates collected in this study differed between dry and 'wet' events, with the sources of the isolates following rainfall being more similar to those from Copano Bay (higher proportions of livestock i.e. non-point source runoff).

- Source contributions do differ between the bay and the two rivers. Fecal contamination that impacts the Mission and Aransas Rivers may not survive downstream flow to Copano Bay. Additionally, flow from a narrow river will become diluted once reaching a large body of water such as Copano Bay. Thus there actually might be inherent differences in the types of contributing sources to the tributaries and the main body of Copano Bay.

- Contributing sources might have changed during the time between the two studies. Area residents and the business community (including ranching and agriculture) might have implemented Best Management Practices that have since reduced certain impacts into these impaired water bodies.

In summary, the majority of unknown source isolates collected from water samples at the five sampling stations along the Mission and Aransas Rivers classified as human (treated WWTP outflow) utilizing the three bacterial source tracking techniques antibiotic resistance analysis (ARA), carbon source utilization (CSU) analysis, and pulsed-field gel electrophoresis (PFGE). The remaining unknown source isolates were classified as livestock animals and wildlife, with cow, horse and duck contributions accounting for the majority of the animal sources in both the composite dataset and PFGE profiles. As the primary source identified in this study appears to be treated human effluent, wastewater facilities discharging into the Mission and Aransas Rivers there appears to be a controllable source of contamination which could be reduced with monitoring for effective disinfecting practices and permitting procedures.

Recommendations for future studies:

- Additional study of sources following rainfall compared to dry weather sources of contamination to further clarify the differences and importance of run-off contributions.
- Studies to determine the effects of improved water quality of WWTP outflows on levels of contamination in the rivers.

LITERATURE CITED

American Public Health Association. 2005. Standard methods for the examination of water and wastewater. 21st ed. American Public Health Association, Washington, D.C

Biolog. 1994. Rainbow Agar O157. Fast and easy isolation of *E. coli* O157:H7 and other *E. coli* strains. Technical Document, Biolog, Inc.

Biolog. 1999. MicroLog™ System Release 4.0 User Guide.

Biolog. 2004. MicroLog™ System Release 4.20.04 User Guide.

Biolog. 2006. Biolog Technical Guide.

Bio-Rad Laboratories. 1995. CHEF-DR III Pulsed Field Electrophoresis Systems: Instruction Manual and Applications Guide. Hercules, California.

Bio-Rad Laboratories. 1999. Quality One Quantitation Software: Instruction Manual and Applications Guide. Hercules, California.

Bio-Rad Laboratories. 1999. Diversity Database: Instruction Manual and Applications Guide. Hercules, California.

Boyer, A. Reducing bacteria with best management practices. www.wr.udel.edu/ChristinaTribTeam/Fact%20Sheets/ReducingBacteria_wBMPs.pdf. accessed 11/19/2008.

Cameron, D.N., F. M. Khambathy, I. K. Wachsmuth, R.V. Tauxe, and T.J. Barrett. 1994. Molecular characterization of *Vibrio cholerae* O1 strains by Pulsed Field Gel Electrophoresis. J. Clin. Microbiol. 32:1685-1690.

Clinical Laboratory Standards Insititute. 2006. Performance Standards for Antimicrobial Disc Susceptibility Tests; Approved Standard-Ninth Edition. CLSI document M2-A9.

Clinical Laboratory Standards Institute. 2006. Performance Standards for Antimicrobial Susceptibility Testing; Sixteenth Informational Supplement. CLSI document M100-S16.

Dorkin, M., S. Falkow, E. Rosenber, K.H. Schliefer, and E. Stackebrandt (Eds.). 2006. The Prokaryotes. 3rd Edition. Volume 6: Proteobacteria Gamma subclass. Springer, New York, NY.

Duarte, J.M., Bosco, J., and Melo, L.C. 1999. Comparison of similarity coefficients based on RAPD markers in the common bean. Genetics and Molecular Bio. 22, 427-432.

Duck, W.M., C.D. Steward, S.N. Banerjee, J.E. McGowan Jr., and F. C. Tenover. 2003. Optimization of computer software settings improves accuracy of Pulsed Field Gel Electrophoresis macrorestriction fragment pattern analysis. J. Clin. Microbiol. 41:3035-3042.

Duffy, G., S.B. O'Brien, E. Carney, J.J. Sheridan, D.A. McDowell, and I.S. Blair. 2005. Characterisation of *E. coli* O157 isolates from bovine hide and beef trimming in Irish abattoirs by pulsed field gel electrophoresis. J. Microbiol. Meth. 60: 375-382.

Dufour, A.P., E.R. Strickland and V.J. Cabelli. 1981. Membrane filter method for enumerating *Escherichia coli*. Appl. Environ. Microbiol. 41: 1152-1158.

Hagedorn, C., S.L. Robinson, J.R. Filtz, S.M. Grubbs, T.A. Angier, and R.B. Reneau, Jr. 1999. Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. Appl. Environ. Microbiol. 65: 5522-5531.

Hahm, B.-K., Y. Maldonado, E. Schreiber, A.K. Bhunia, and C.H. Nakatsu. 2003. Subtyping of foodbourne and environmental isolates of *Escherichia coli* by multiplex-PCR, rep-PCR, PFGE, ribotyping and AFLP. J. Microbiol. Meth. 53:387-399.

Harwood, V.J., J. Whitlock, and V. Withington. 2000. Classification of antibiotic resistance patterns of indicator bacteria by discriminant analysis: use in predicting the source of fecal contamination in subtropical waters. Appl. Environ. Microbiol. 66:3698-3704.

Huberty, C.J. 1994. Applied discriminant analysis. John Wiley and Sons, Inc. New York, N.Y.

Johnson, L.K., M.B. Brown, E.A. Carruthers, J.A. Ferguson, P.E. Dombek, and M.J. Sadowsky. 2004. Sample Size, Library Composition, and Genotypic Diversity among Natural Populations of *Escherichia coli* from Different Animals Influence Accuracy of Determining Sources of Fecal Pollution. Appl. Environ. Microbiol. 70: 4478-4485.

Jones, A., G. Di Giovanni, L. Hauck, J. Mott, H. Rifai, R. Srinivasan, and G. Ward. 2007. Bactera Total Maximum Daily Load Task Force Report. Submitted to and accepted by the Texas Commission on Environmental Quality and the Texas State Soil and Water Conservation Board.

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

McLellan, S.L., A.D. Daniels, and A.K. Salmore. 2003. Genetic characterization of *Escherichia coli* populations from host sources of fecal pollution by using DNA fingerprinting. Appl. Environ. Microbiol. 69:2587-2594.

Meays, C.L., K. Broersma, R. Nordin, and A. Mazumder. 2004. Source tracking fecal bacteria in water: a critical review of current methods. J. Environ. Manag. 73:71-79.

Moe, C.L. 2002. Waterborne Transmission of Infectious Agents p 184-204. In C. J. Hurst, R.L. Crawford, G.R. Knudsen, M.J. McInerney, and L.D. Stetzenbach (ed.), Manual of Environmental Microbiology, 2nd Ed. ASM Press, Washington, D.C.

Mott, J. and R. Lehman. 2005. Bacteria Source Tracking in Copano Bay Phase II, Final Report.

Mott, J. and R. Lehman. 2006. Study to investigate sources of *E. coli* isolated from Mission and Aransas Rivers. Quality Assurance Project Plan.

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

Müller, H.E. D.J. Brenner, G.F. Fanning, P.A. Grimont and P. Kämpfer. 1996. Emended description of *Buttiauxella agrestis* with recognition of six new species of *Buttiauxella* and two new species of *Kluyvera*: *Buttiauxella ferragutiae* sp. nov., *Buttiauxella gaviniae* sp. nov., *Buttiauxella brennerae* sp. nov., *Buttiauxella izardii* sp. nov., Buttiauxella *noackiae* sp. nov., *Buttiauxella warmboldiae* sp. nov., *Kluyvera cochleae* sp. nov., and *Kluyvera georgiana* sp. nov. Int. J. Syst. Bacteriol.. 46: 50-63.

National Committee on Clinical Laboratory Standards. 2002. Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals; approved standard-2nd ed. NCCLS document M31-A2.

Okwumabua, O., M. O'Connor, E. Shull, K. Strelow, M. Hamacher, T. Kurzynski, and D. Warshauer. 2005. Characterization of *Listeria monocytogenes* isolates from food animal clinical cases: PFGE pattern similarity to strains from human listeriosis cases. FEMS Microbiol. Letters. 249: 275-281.

Parveen, S., N.C. Hodge, R.E. Stall, S.R. Farrah, and M.L. Tamplin. 2001. Phenotypic and genotypic characterization of human and nonhuman *Escherichia coli*. Wat. Res. 35: 379-386.

Sayah, R.S., J.B. Kaneene, Y. Johnson, and R. Miller. 2005. Patterns of antimicrobial resistance observed in *Escherichia coli* isolates obtained from domestic and wild animal fecal samples, human septage, and surface water. Appl. Environ. Microbiol. 71:1394-1404.

Scott, T.M., J.B. Rose, T.M. Jenkins, S.R. Farrah, and J. Lakasik. 2002. Microbial source tracking: current methodology and future directions. Appl. Environ.Microbiol. 68: 5796-5803.

Singer, R.S., W.M. Sischo, and T.E. Carpenter. 2004. Exploration of biases that affect the interpretation of restriction fragment patterns produced by pulsed-field gel electrophoresis. J. Clin. Microbiol. 42:5502-5511.

Soll, D.R. 2000. The ins and outs of DNA fingerprinting the infectious fungi. Clin. Microbiol. Rev. 13, 332-370.

Stoeckel, D.M. and V.J. Harwood. 2007. Performance, design and analysis in microbial source tracking studies. Appl. Environ. Microbiol. 73:2405-2415.

Texas Commission on Environmental Quality. 2003. Surface water quality monitoring procedures Volume 1: Physical and chemical monitoring methods for water, sediment and tissue. Publication No. RG-415, December 2003, Austin, TX.

Texas Commission on Environmental Quality. 2004. 2004 Texas water quality inventory and 303(d) List. (August 1, 2006). http://www.tceq.state.tx.us/compliance/monitoring/water/quality/data/04twqi/twqi04.html

United States Environmental Protection Agency. 1986. Bacterial ambient water quality criteria. Federal Register 51(45): 8012-8016.

U.S. Environmental Protection Agency, 1994. Polluted brochure. EPA-841-F-94-005.

United States Environmental Protection Agency. 2000. Improved enumeration methods for the recreational water quality indicators: Enterococci and *E. coli*. EPA-821-R-97-004.

U.S. Environmental Protection Agency. 2005. Section 319 Nonpoint source program success story: Virginia. EPA 841-F-05-004I. www.epa.gov/owow/nps/Success319/state/pdf/va_holston.pdf. accessed 11/19/08.

Webster, L.F., B.C. Thompson, M.H. Fulton, D.E. Chestnut, R.F. Van Dolah, A.K. Leight, and G. I. Scott. 2004. Identification of sources of *Escherichia coli* in South Carolina estuaries using antibiotic resistance analysis. J. Exper. Mar. Biol. Ecol. 298: 179-195.

Whitlock, J.E., D.T. Jones, and V.J., Harwood. 2002. Identification of the sources of fecal coliforms in an urban watershed using antibiotic resistance analysis. Wat. Res. 36: 4273-4282.

Wiggins, B.A. 1996. Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. Appl. Environ. Microbiol. 62: 3997-4002.

Wiggins, B.A. R.W. Andrews, R.A. Conway, C.L. Corr, E.J. Dobratz, D.P. Dougherty, J.R Eppard, S.R. Knupp, M.C. Limjoco, J.M. Mettenburg, J.M. Rinehardt, J. Sonsino,

R.R. Torrijos, and M.E. Zimmerman. 1999. Use of antibiotic resistance analysis to identify non-point sources of fecal pollution. Appl. Environ. Microbiol. 65:3483-3486.

Zhechko, D., M. Michaylova, and S. Mincova. 2005. Characterization of *Lactobacillus helveticus* strains isolated from Bulgarian yoghurt, cheese, plants, and human faecal samples by sodium dodecilsulfate polyacrylamide gel electrophoresis of cell-wall proteins, ribotyping, and pulsed field gel fingerprinting. Inter. Dairy J. 15:998-1005.

Appendix A

Table A1. Field Parameters for Sampling Event 1 (03/26/07)

Parameters	12952	12947	12944	12943	12945
Depth Sample Collected (cm)	*NA	Bucket Sample	Bucket Sample	Bucket Sample	Bucket Sample
Air Temp (˚C)	22.4	22.9	24.5	23.1	22.9
Wind Intensity	Slight	Slight	Calm	Moderate	Moderate
Wind Direction	Slight	SE	SE	SE	SE
Present Weather	Overcast	Rain(light Sprinkle)	Rain(light Sprinkle)	Overcast	*NA
Water Temp (˚C)	22.8	23.0	22.9	23.0	22.5
Conductivity (µmhos/cm)	NA	NA	NA	NA	NA
Salinity (ppt)	NA	NA	NA	NA	NA
DO (mg/L)	NA	NA	NA	NA	NA
pH (s.u.)	NA	NA	NA	NA	NA
Secchi Disk (meters)	NA	NA	NA	NA	NA
Water Color	Brown (Dark)	Brown (Dark)	Brown (Medium)	Brown (Medium)	Brown (Medium)
Water Odor	None	None	None	None	None
Water Surface	Ripples	Ripples	Ripples	Ripples	Ripples
Tide Stage	Slack	Slack	Slack	Slack	Slack
Days Since Last Rainfall	0	0	0	0	0
Rainfall (Inches past 1 day)	Trace	Trace	Trace	Trace	Trace
Rainfall (inches past 7 days)	Trace	Trace	Trace	Trace	Trace
Human Use	NA	2 humans launching boat	NA	NA	NA
Other Comments	Moderate Trash,Waste water treatment plant downstream from station	Light trash	3 birds	Moderate trash on banks and in water	Light trash on Banks

*Information missing from original sheet; could not be retrieved from a reputable source

Table A2. Field Parameters for Sampling Event 2 (12/10/07)

Parameters	12952	12947	12944	12943	12945
Depth Sample Collected (cm)	6	12	60	60	60
Air Temp (˚C)	14.2	19.0	19.9	17.9	18.4
Wind Intensity	Slight	Slight	Slight	Slight	Moderate
Wind Direction	NE	NE	NE	NE	NE
Present Weather	Overcast	Overcast	Rain	Overcast	Overcast
Water Temp (˚C)	21.17	20.98	21.29	21.27	22.01
Conductivity (µmhos/cm)	1.663	2.078	1.602	3.059	10.20
Salinity (ppt)	0.84	1.06	0.81	1.60	5.77
DO (mg/L)	7.33	8.74	9.00	9.71	12.5
pH (s.u.)	7.7	8.24	7.90	8.08	8.06
Secchi Disk (meters)	NA	NA	NA	NA	NA
Water Color	Clear	Green	Green-Brown	Green-Brown	Green-Brown
Water Odor	None	None	None	None	None
Water Surface	Ripples	Ripples	Calm	Ripples	Waves
Tide Stage	*NA	*NA	*NA	*NA	*NA
Days Since Last Rainfall	2	2	2	2	2
Rainfall (Inches past 1 day)	0.0	0.0	0.0	0.0	0.0
Rainfall (inches past 7 days)	Trace	Trace	Trace	Trace	Trace
Human Use	None	None	None	None	None
Other Comments	Trash on banks	None	Bucket Sample	Trash on banks; bucket sample	None

Table A3. Field Parameters for Sampling Event 3 (03/07/08)

Parameters	12952	12947	12944	12943	12945
Depth Sample Collected (cm)	0.2	0.3	0.3	0.3	0.3
Air Temp (˚C)	13.7	13.7	14.0	6.2	8.8
Wind Intensity	Slight	Moderate	Slight	Slight	Slight
Wind Direction	NW	NW	SW	NW	NW
Present Weather	Clear	Clear	Clear	Clear	Clear
Water Temp (˚C)	14.03	14.59	14.13	16.05	9.42
Conductivity (µmhos/cm)	1521	2115	1469	2953	13648
Salinity (ppt)	NA	1.09	NA	1.55	7.87
DO (mg/L)	9.25	9.85	82.1	8.15	9.08
pH (s.u.)	8.05	8.52	7.76	8.09	8.2
Secchi Disk (meters)	0.25	0.14	0.21	0.21	0.07
Water Color	Clear	Brown	Brown	Brown	Brown
Water Odor	None	None	None	None	None
Water Surface	Ripples	Ripples	Calm	Ripples	Ripples
Tide Stage	NA	*NA	*NA	*NA	Low
Days Since Last Rainfall	1	1	1	1	1
Rainfall (Inches past 1 day)	1.16	1.16	1.16	1.16	1.16
Rainfall (inches past 7 days)	Trace	Trace	Trace	Trace	Trace
Human Use	None	None	None	None	None
Other Comments	None	Light Trash	Light Trash	43 birds, light trash	Light trash

*Information missing from original sheet; could not be retrieved from a reputable source

Appendix B

Table B1. Classification of known sources from library for this study; ARA analysis only

Classification Results[b,c]

			Predicted Group Membership			
		species	nonhuman	human-Copano	human-WWTP effluent	Total
Original	Count	nonhuman	402	199	137	738
		human-Copano	22	94	10	126
		human-WWTP effluent	6	4	51	61
	%	nonhuman	54.5	27.0	18.6	100.0
		human-Copano	17.5	74.6	7.9	100.0
		human-WWTP effluent	9.8	6.6	83.6	100.0
Cross-validated[a]	Count	nonhuman	392	204	142	738
		human-Copano	29	86	11	126
		human-WWTP effluent	9	6	46	61
	%	nonhuman	53.1	27.6	19.2	100.0
		human-Copano	23.0	68.3	8.7	100.0
		human-WWTP effluent	14.8	9.8	75.4	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 59.1% of original grouped cases correctly classified.

c. 56.6% of cross-validated grouped cases correctly classified.

Table B2. Classification of known sources from expanded library from this study (1176 known source isolates); ARA analysis only

Classification Results[b,c]

			Predicted Group Membership			
		species	nonhuman	human-copano	human-WWTP effluent	Total
Original	Count	nonhuman	553	243	174	970
		human-copano	25	109	11	145
		human-WWTP effluent	3	6	52	61
	%	nonhuman	57.0	25.1	17.9	100.0
		human-copano	17.2	75.2	7.6	100.0
		human-WWTP effluent	4.9	9.8	85.2	100.0
Cross-validated[a]	Count	nonhuman	547	247	176	970
		human-copano	28	104	13	145
		human-WWTP effluent	7	7	47	61
	%	nonhuman	56.4	25.5	18.1	100.0
		human-copano	19.3	71.7	9.0	100.0
		human-WWTP effluent	11.5	11.5	77.0	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 60.7% of original grouped cases correctly classified.

c. 59.4% of cross-validated grouped cases correctly classified.

Table B3. Classification of known sources from library for this study; CSU analysis only

Classification Results[b,c]

			Predicted Group Membership			
		species	domesticated	human-Copano	human-WWTP	Total
Original	Count	domesticated	673	33	32	738
		human-Copano	19	105	2	126
		human-WWTP	7	0	54	61
	%	domesticated	91.2	4.5	4.3	100.0
		human-Copano	15.1	83.3	1.6	100.0
		human-WWTP	11.5	.0	88.5	100.0
Cross-validated[a]	Count	domesticated	640	53	45	738
		human-Copano	26	97	3	126
		human-WWTP	10	0	51	61
	%	domesticated	86.7	7.2	6.1	100.0
		human-Copano	20.6	77.0	2.4	100.0
		human-WWTP	16.4	.0	83.6	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 89.9% of original grouped cases correctly classified.

c. 85.2% of cross-validated grouped cases correctly classified.

Table B4. Classification of known sources from expanded library from this study (1033 known source isolates); CSU analysis only

Classification Results[b,c]

			Predicted Group Membership			
		species	nonhuman	human-copano	human-WWTP effluent	Total
Original	Count	nonhuman	757	36	35	828
		human-copano	18	106	2	126
		human-WWTP effluent	9	0	70	79
	%	nonhuman	91.4	4.3	4.2	100.0
		human-copano	14.3	84.1	1.6	100.0
		human-WWTP effluent	11.4	.0	88.6	100.0
Cross-validated[a]	Count	nonhuman	733	52	43	828
		human-copano	25	98	3	126
		human-WWTP effluent	12	0	67	79
	%	nonhuman	88.5	6.3	5.2	100.0
		human-copano	19.8	77.8	2.4	100.0
		human-WWTP effluent	15.2	.0	84.8	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 90.3% of original grouped cases correctly classified.

c. 86.9% of cross-validated grouped cases correctly classified.

Appendix C

All of the following tables refer to discriminant analyses performed on the datasets utilizing the library of 925 isolates.

Table C1. Classification of unknown sources using ARA library alone, three-way classification (nonhuman v human (untreated) v human (treated));cross validation of known sources included.

Classification Results[b,c]

			Predicted Group Membership			
		species	nonhuman	human-Copano	human-WWTP	Total
Original	Count	nonhuman	402	199	137	738
		human-Copano	22	94	10	126
		human-WWTP	6	4	51	61
		Ungrouped cases	67	29	268	364
	%	nonhuman	54.5	27.0	18.6	100.0
		human-Copano	17.5	74.6	7.9	100.0
		human-WWTP	9.8	6.6	83.6	100.0
		Ungrouped cases	18.4	8.0	73.6	100.0
Cross-validated[a]	Count	nonhuman	392	204	142	738
		human-Copano	29	86	11	126
		human-WWTP	9	6	46	61
	%	nonhuman	53.1	27.6	19.2	100.0
		human-Copano	23.0	68.3	8.7	100.0
		human-WWTP	14.8	9.8	75.4	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 59.1% of original grouped cases correctly classified.

c. 56.6% of cross-validated grouped cases correctly classified.

Table C2. Classification of unknown sources using CSU library alone, three-way classification (nonhuman v human (untreated) v human (treated));cross validation of known sources included.

Classification Results[b,c]

			Predicted Group Membership			
		species	nonhuman	human-Copano	human-WWTP	Total
Original	Count	nonhuman	673	33	32	738
		human-Copano	19	105	2	126
		human-WWTP	7	0	54	61
		Ungrouped cases	198	0	166	364
	%	nonhuman	91.2	4.5	4.3	100.0
		human-Copano	15.1	83.3	1.6	100.0
		human-WWTP	11.5	.0	88.5	100.0
		Ungrouped cases	54.4	.0	45.6	100.0
Cross-validated[a]	Count	nonhuman	640	53	45	738
		human-Copano	26	97	3	126
		human-WWTP	10	0	51	61
	%	nonhuman	86.7	7.2	6.1	100.0
		human-Copano	20.6	77.0	2.4	100.0
		human-WWTP	16.4	.0	83.6	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 89.9% of original grouped cases correctly classified.

c. 85.2% of cross-validated grouped cases correctly classified.

Table C3. Classification of unknown sources using composite dataset, three-way classification (nonhuman v human (untreated) v human (treated));cross validation of known sources

Classification Results[b,c]

			Predicted Group Membership			
		species	nonhuman	human-Copano	human-WWTP	Total
Original	Count	nonhuman	694	19	25	738
		human-Copano	16	110	0	126
		human-WWTP	3	0	58	61
		Ungrouped cases	131	1	232	364
	%	nonhuman	94.0	2.6	3.4	100.0
		human-Copano	12.7	87.3	.0	100.0
		human-WWTP	4.9	.0	95.1	100.0
		Ungrouped cases	36.0	.3	63.7	100.0
Cross-validated[a]	Count	nonhuman	652	48	38	738
		human-Copano	29	96	1	126
		human-WWTP	4	0	57	61
	%	nonhuman	88.3	6.5	5.1	100.0
		human-Copano	23.0	76.2	.8	100.0
		human-WWTP	6.6	.0	93.4	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 93.2% of original grouped cases correctly classified.

c. 87.0% of cross-validated grouped cases correctly classified.

Table C4. Classification of unknown sources using ARA only, four-way classification (livestock v wildlife v human (untreated) v human (treated)); cross validation of known sources included.

Classification Results[b,c]

			Predicted Group Membership				
		species	domesticated	human-Copano	human-WWTP	wildlife	Total
Original	Count	domesticated	195	114	86	17	412
		human-Copano	26	88	9	3	126
		human-WWTP	5	4	50	2	61
		wildlife	48	36	33	209	326
		Ungrouped cases	54	25	269	16	364
	%	domesticated	47.3	27.7	20.9	4.1	100.0
		human-Copano	20.6	69.8	7.1	2.4	100.0
		human-WWTP	8.2	6.6	82.0	3.3	100.0
		wildlife	14.7	11.0	10.1	64.1	100.0
		Ungrouped cases	14.8	6.9	73.9	4.4	100.0
Cross-validated[a]	Count	domesticated	188	117	87	20	412
		human-Copano	28	85	10	3	126
		human-WWTP	7	6	43	5	61
		wildlife	48	39	34	205	326
	%	domesticated	45.6	28.4	21.1	4.9	100.0
		human-Copano	22.2	67.5	7.9	2.4	100.0
		human-WWTP	11.5	9.8	70.5	8.2	100.0
		wildlife	14.7	12.0	10.4	62.9	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 58.6% of original grouped cases correctly classified.

c. 56.3% of cross-validated grouped cases correctly classified.

Table C5. Classification of unknown sources using CSU only, four-way classification (livestock v wildlife v human (untreated) v human (treated)); cross validation of known sources included.

Classification Results[b,c]

			Predicted Group Membership				
		species	domesticated	human-Copano	human-WWTP	wildlife	Total
Original	Count	domesticated	332	22	3	55	412
		human-Copano	12	103	2	9	126
		human-WWTP	0	0	54	7	61
		wildlife	24	5	15	282	326
		Ungrouped cases	17	0	160	187	364
	%	domesticated	80.6	5.3	.7	13.3	100.0
		human-Copano	9.5	81.7	1.6	7.1	100.0
		human-WWTP	.0	.0	88.5	11.5	100.0
		wildlife	7.4	1.5	4.6	86.5	100.0
		Ungrouped cases	4.7	.0	44.0	51.4	100.0
Cross-validated[a]	Count	domesticated	309	30	8	65	412
		human-Copano	20	95	2	9	126
		human-WWTP	1	0	49	11	61
		wildlife	31	10	21	264	326
	%	domesticated	75.0	7.3	1.9	15.8	100.0
		human-Copano	15.9	75.4	1.6	7.1	100.0
		human-WWTP	1.6	.0	80.3	18.0	100.0
		wildlife	9.5	3.1	6.4	81.0	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 83.4% of original grouped cases correctly classified.

c. 77.5% of cross-validated grouped cases correctly classified.

Table C6. Classification of unknown sources using composite dataset, four-way classification (livestock v wildlife v human (untreated) v human (treated)); cross validation of known sources included.

Classification Results[b,c]

			Predicted Group Membership				
		species	domesticate	human-Copano	human-WWTP	wildlife	Total
Original	Count	domesticate	366	18	9	19	412
		human-Copano	16	104	1	5	126
		human-WWTP	0	0	58	3	61
		wildlife	25	3	8	290	326
		Ungrouped cases	52	0	233	79	364
	%	domesticate	88.8	4.4	2.2	4.6	100.0
		human-Copano	12.7	82.5	.8	4.0	100.0
		human-WWTP	.0	.0	95.1	4.9	100.0
		wildlife	7.7	.9	2.5	89.0	100.0
		Ungrouped cases	14.3	.0	64.0	21.7	100.0
Cross-validated[a]	Count	domesticate	332	27	17	36	412
		human-Copano	25	94	1	6	126
		human-WWTP	0	0	57	4	61
		wildlife	34	9	15	268	326
	%	domesticate	80.6	6.6	4.1	8.7	100.0
		human-Copano	19.8	74.6	.8	4.8	100.0
		human-WWTP	.0	.0	93.4	6.6	100.0
		wildlife	10.4	2.8	4.6	82.2	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 88.4% of original grouped cases correctly classified.

c. 81.2% of cross-validated grouped cases correctly classified.

Table C7. Classification of unknown sources using ARA alone, seven-way classification (cow v horse v wildlife v gull v duck v human (untreated) and human (treated)); cross validation of known sources included.

Classification Results[b,c]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gull	duck	Total
Original	Count	cow	43	33	26	17	2	5	17	143
		human-Copano	6	79	8	18	1	1	13	126
		human-WWTP	4	1	44	4	1	3	4	61
		horse	13	43	41	124	4	0	44	269
		wildlife	4	4	13	1	127	13	0	162
		gull	0	3	1	2	10	46	4	66
		duck	5	15	11	16	1	2	48	98
		Ungrouped cases	23	10	215	40	4	7	65	364
	%	cow	30.1	23.1	18.2	11.9	1.4	3.5	11.9	100.0
		human-Copano	4.8	62.7	6.3	14.3	.8	.8	10.3	100.0
		human-WWTP	6.6	1.6	72.1	6.6	1.6	4.9	6.6	100.0
		horse	4.8	16.0	15.2	46.1	1.5	.0	16.4	100.0
		wildlife	2.5	2.5	8.0	.6	78.4	8.0	.0	100.0
		gull	.0	4.5	1.5	3.0	15.2	69.7	6.1	100.0
		duck	5.1	15.3	11.2	16.3	1.0	2.0	49.0	100.0
		Ungrouped cases	6.3	2.7	59.1	11.0	1.1	1.9	17.9	100.0
Cross-validated[a]	Count	cow	42	33	26	17	2	5	18	143
		human-Copano	8	74	9	19	1	1	14	126
		human-WWTP	4	2	40	5	1	3	6	61
		horse	14	48	40	116	5	1	45	269
		wildlife	3	4	12	1	126	14	2	162
		gull	0	3	2	2	12	43	4	66
		duck	7	20	11	18	1	2	39	98
	%	cow	29.4	23.1	18.2	11.9	1.4	3.5	12.6	100.0
		human-Copano	6.3	58.7	7.1	15.1	.8	.8	11.1	100.0
		human-WWTP	6.6	3.3	65.6	8.2	1.6	4.9	9.8	100.0
		horse	5.2	17.8	14.9	43.1	1.9	.4	16.7	100.0
		wildlife	1.9	2.5	7.4	.6	77.8	8.6	1.2	100.0
		gull	.0	4.5	3.0	3.0	18.2	65.2	6.1	100.0
		duck	7.1	20.4	11.2	18.4	1.0	2.0	39.8	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 55.2% of original grouped cases correctly classified.

c. 51.9% of cross-validated grouped cases correctly classified.

Table C8. Classification of unknown sources using CSU alone, seven-way classification (cow v horse v wildlife v gull v duck v human (untreated) and human (treated)); cross validation of known sources included.

Classification Results[b,c]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gull	duck	Total
Original	Count	cow	104	5	2	7	8	13	4	143
		human-Copano	6	99	1	11	1	5	3	126
		human-WWTP	0	0	53	0	8	0	0	61
		horse	23	6	1	215	6	4	14	269
		wildlife	0	0	3	1	151	7	0	162
		gull	0	0	0	0	8	58	0	66
		duck	4	1	10	1	6	8	68	98
		Ungrouped cases	10	0	156	11	129	53	5	364
	%	cow	72.7	3.5	1.4	4.9	5.6	9.1	2.8	100.0
		human-Copano	4.8	78.6	.8	8.7	.8	4.0	2.4	100.0
		human-WWTP	.0	.0	86.9	.0	13.1	.0	.0	100.0
		horse	8.6	2.2	.4	79.9	2.2	1.5	5.2	100.0
		wildlife	.0	.0	1.9	.6	93.2	4.3	.0	100.0
		gull	.0	.0	.0	.0	12.1	87.9	.0	100.0
		duck	4.1	1.0	10.2	1.0	6.1	8.2	69.4	100.0
		Ungrouped cases	2.7	.0	42.9	3.0	35.4	14.6	1.4	100.0
Cross-validated[a]	Count	cow	83	9	2	19	10	15	5	143
		human-Copano	7	91	1	14	1	7	5	126
		human-WWTP	0	0	47	1	13	0	0	61
		horse	34	10	4	185	6	5	25	269
		wildlife	3	0	7	1	134	15	2	162
		gull	2	0	3	0	18	42	1	66
		duck	7	4	11	3	6	9	58	98
	%	cow	58.0	6.3	1.4	13.3	7.0	10.5	3.5	100.0
		human-Copano	5.6	72.2	.8	11.1	.8	5.6	4.0	100.0
		human-WWTP	.0	.0	77.0	1.6	21.3	.0	.0	100.0
		horse	12.6	3.7	1.5	68.8	2.2	1.9	9.3	100.0
		wildlife	1.9	.0	4.3	.6	82.7	9.3	1.2	100.0
		gull	3.0	.0	4.5	.0	27.3	63.6	1.5	100.0
		duck	7.1	4.1	11.2	3.1	6.1	9.2	59.2	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 80.9% of original grouped cases correctly classified.

c. 69.2% of cross-validated grouped cases correctly classified.

Table C9. Classification of unknown sources using composite dataset, seven-way classification (cow v horse v wildlife v gull v duck v human (untreated) and human (treated)); cross validation of known sources included.

Classification Results[b,c]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	60	2	317	36	13	13	33	474
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	12.7	.4	66.9	7.6	2.7	2.7	7.0	100.0
Cross-validated[a]	Count	cow	97	7	8	13	4	7	7	143
		human-Copano	13	91	1	14	0	0	7	126
		human-WWTP	0	0	55	0	1	1	4	61
		horse	37	9	5	191	1	1	25	269
		wildlife	0	0	10	0	145	5	2	162
		gulls	4	0	1	0	11	46	4	66
		ducks	6	2	4	10	0	4	72	98
	%	cow	67.8	4.9	5.6	9.1	2.8	4.9	4.9	100.0
		human-Copano	10.3	72.2	.8	11.1	.0	.0	5.6	100.0
		human-WWTP	.0	.0	90.2	.0	1.6	1.6	6.6	100.0
		horse	13.8	3.3	1.9	71.0	.4	.4	9.3	100.0
		wildlife	.0	.0	6.2	.0	89.5	3.1	1.2	100.0
		gulls	6.1	.0	1.5	.0	16.7	69.7	6.1	100.0
		ducks	6.1	2.0	4.1	10.2	.0	4.1	73.5	100.0

a. Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

b. 86.4% of original grouped cases correctly classified.

c. 75.4% of cross-validated grouped cases correctly classified.

Table C10. Classification of unknown source isolates using composite dataset, three-way classification; Event 1 only

Classification Results[a]

			Predicted Group Membership			
		species	nonhuman	human-Copano	human-WWTP	Total
Original	Count	nonhuman	694	19	25	738
		human-Copano	16	110	0	126
		human-WWTP	3	0	58	61
		Ungrouped cases	41	0	106	147
	%	nonhuman	94.0	2.6	3.4	100.0
		human-Copano	12.7	87.3	.0	100.0
		human-WWTP	4.9	.0	95.1	100.0
		Ungrouped cases	27.9	.0	72.1	100.0

a. 93.2% of original grouped cases correctly classified.

Table C11. Classification of unknown source isolates using composite dataset, three-way classification; Event 2 only

Classification Results[a]

			Predicted Group Membership			
		species	nonhuman	human-Copano	human-WWTP	Total
Original	Count	nonhuman	694	19	25	738
		human-Copano	16	110	0	126
		human-WWTP	3	0	58	61
		Ungrouped cases	17	0	79	96
	%	nonhuman	94.0	2.6	3.4	100.0
		human-Copano	12.7	87.3	.0	100.0
		human-WWTP	4.9	.0	95.1	100.0
		Ungrouped cases	17.7	.0	82.3	100.0

a. 93.2% of original grouped cases correctly classified.

Table C12. Classification of unknown source isolates using composite dataset, three-way classification; Event 3 only

Classification Results[a]

			Predicted Group Membership			
		species	nonhuman	human-Copano	human-WWTP	Total
Original	Count	nonhuman	694	19	25	738
		human-Copano	16	110	0	126
		human-WWTP	3	0	58	61
		Ungrouped cases	73	1	47	121
	%	nonhuman	94.0	2.6	3.4	100.0
		human-Copano	12.7	87.3	.0	100.0
		human-WWTP	4.9	.0	95.1	100.0
		Ungrouped cases	60.3	.8	38.8	100.0

a. 93.2% of original grouped cases correctly classified.

Table C13. Classification of unknown source isolates using composite dataset, four-way classification; Event 1 only

Classification Results[a]

			Predicted Group Membership				
		species	domesticated	human-Copano	human-WWTP	wildlife	Total
Original	Count	domesticated	366	18	9	19	412
		human-Copano	16	104	1	5	126
		human-WWTP	0	0	58	3	61
		wildlife	25	3	8	290	326
		Ungrouped cases	16	0	104	27	147
	%	domesticated	88.8	4.4	2.2	4.6	100.0
		human-Copano	12.7	82.5	.8	4.0	100.0
		human-WWTP	.0	.0	95.1	4.9	100.0
		wildlife	7.7	.9	2.5	89.0	100.0
		Ungrouped cases	10.9	.0	70.7	18.4	100.0

a. 88.4% of original grouped cases correctly classified.

Table C14. Classification of unknown source isolates using composite dataset, four-way classification; Event 2 only

Classification Results[a]

			Predicted Group Membership				
		species	domesticated	human-Copano	human-WWTP	wildlife	Total
Original	Count	domesticated	366	18	9	19	412
		human-Copano	16	104	1	5	126
		human-WWTP	0	0	58	3	61
		wildlife	25	3	8	290	326
		Ungrouped cases	7	0	79	10	96
	%	domesticated	88.8	4.4	2.2	4.6	100.0
		human-Copano	12.7	82.5	.8	4.0	100.0
		human-WWTP	.0	.0	95.1	4.9	100.0
		wildlife	7.7	.9	2.5	89.0	100.0
		Ungrouped cases	7.3	.0	82.3	10.4	100.0

a. 88.4% of original grouped cases correctly classified.

Table C15. Classification of unknown source isolates using composite dataset, four-way classification; Event 3 only

Classification Results[a]

			Predicted Group Membership				
		species	domesticated	human-Copano	human-WWTP	wildlife	Total
Original	Count	domesticated	366	18	9	19	412
		human-Copano	16	104	1	5	126
		human-WWTP	0	0	58	3	61
		wildlife	25	3	8	290	326
		Ungrouped cases	29	0	50	42	121
	%	domesticated	88.8	4.4	2.2	4.6	100.0
		human-Copano	12.7	82.5	.8	4.0	100.0
		human-WWTP	.0	.0	95.1	4.9	100.0
		wildlife	7.7	.9	2.5	89.0	100.0
		Ungrouped cases	24.0	.0	41.3	34.7	100.0

a. 88.4% of original grouped cases correctly classified.

Table C16. Classification of unknown source isolates using composite dataset, seven-way classification; Event 1 only

Classification Results[a]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	12	0	110	5	5	5	10	147
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	8.2	.0	74.8	3.4	3.4	3.4	6.8	100.0

a. 86.4% of original grouped cases correctly classified.

Table C17. Classification of unknown source isolates using composite dataset, seven-way classification; Event 2 only

Classification Results[a]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	4	0	81	4	0	0	7	96
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	4.2	.0	84.4	4.2	.0	.0	7.3	100.0

a. 86.4% of original grouped cases correctly classified.

Table C18. Classification of unknown source isolates using composite dataset, seven-way classification; Event 3 only

Classification Results[a]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	20	1	65	12	7	6	10	121
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	16.5	.8	53.7	9.9	5.8	5.0	8.3	100.0

a. 86.4% of original grouped cases correctly classified.

Table C19. Classification of unknown source isolates using composite dataset, seven-way classification; Site 12952 only

Classification Results[a]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	12	0	56	2	2	4	6	82
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	14.6	.0	68.3	2.4	2.4	4.9	7.3	100.0

a. 86.4% of original grouped cases correctly classified.

Table C20. Classification of unknown source isolates using composite dataset, seven-way classification; Site 12947 only

Classification Results[a]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	7	0	50	8	1	0	4	70
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	10.0	.0	71.4	11.4	1.4	.0	5.7	100.0

a. 86.4% of original grouped cases correctly classified.

Table C21. Classification of unknown source isolates using composite dataset, seven-way classification; Site 12944 only

Classification Results[a]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	6	1	56	4	2	2	7	78
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	7.7	1.3	71.8	5.1	2.6	2.6	9.0	100.0

a. 86.4% of original grouped cases correctly classified.

Table C22. Classification of unknown source isolates using composite dataset, seven-way classification; Site 12943 only

Classification Results[a]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	4	0	47	5	5	3	3	67
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	6.0	.0	70.1	7.5	7.5	4.5	4.5	100.0

a. 86.4% of original grouped cases correctly classified.

Table C22. Classification of unknown source isolates using composite dataset, seven-way classification; Site 12945 only

Classification Results[a]

			Predicted Group Membership							
		species	cow	human-Copano	human-WWTP	horse	wildlife	gulls	ducks	Total
Original	Count	cow	119	2	6	5	2	4	5	143
		human-Copano	7	102	1	10	0	0	6	126
		human-WWTP	0	0	60	0	0	0	1	61
		horse	22	5	3	227	0	0	12	269
		wildlife	0	0	6	0	149	5	2	162
		gulls	3	0	1	0	5	55	2	66
		ducks	5	0	2	3	0	1	87	98
		Ungrouped cases	7	0	47	2	2	2	7	67
	%	cow	83.2	1.4	4.2	3.5	1.4	2.8	3.5	100.0
		human-Copano	5.6	81.0	.8	7.9	.0	.0	4.8	100.0
		human-WWTP	.0	.0	98.4	.0	.0	.0	1.6	100.0
		horse	8.2	1.9	1.1	84.4	.0	.0	4.5	100.0
		wildlife	.0	.0	3.7	.0	92.0	3.1	1.2	100.0
		gulls	4.5	.0	1.5	.0	7.6	83.3	3.0	100.0
		ducks	5.1	.0	2.0	3.1	.0	1.0	88.8	100.0
		Ungrouped cases	10.4	.0	70.1	3.0	3.0	3.0	10.4	100.0

a. 86.4% of original grouped cases correctly classified.

Table C23. Percentage of classification of unknown source isolates using ARA alone and CSU alone

	Human/Nonhuman (three-way)		Human/Livestock/Wildlife (four-way)		All Categories (seven-way)	
	ARA	CSU	ARA	CSU	ARA	CSU
cow	18.4	54.4	14.8	4.7	6.3	2.7
horse					11.0	3.0
duck			4.4	51.4	17.9	1.4
gull					1.9	14.6
wildlife					1.1	35.4
human (untreated)	8.0	0.0	6.9	0.0	2.7	0.0
human (treated)	73.6	44.6	73.9	44.0	59.1	42.9