FINAL REPORT

JUL 2 9 2004

2001 Project Abstract
For the Period Ending June 30, 2004 (extended from June 30, 2003)

TITLE: Determination of Fecal Pollution Sources in Minnesota Watersheds
PROJECT MANAGER: Dr. Michael J. Sadowsky
ORGANIZATION: University of Minnesota
ADDRESS: Department of Soil, Water & Climate, 1991 Upper Buford Circle, 439 Borlaug Hall St. Paul, MN 55108
WEB SITE ADDRESS: www.ecolirep.umn.edu
FUND: Future Resources Fund
LEGAL CITATION: ML 2001, 1st Special Session, Ch. 2, Sec.14, Subd. 6 (d)

APPROPRIATION AMOUNT: $275,000

Overall Project Outcome and Results

We used a library of DNA fingerprints, created using the rep-PCR and HFERP techniques, in an attempt to define sources of fecal bacterial pollution, *E. coli*, in three Minnesota watersheds, Minneopa Creek (Blue Earth County), High Island Creek (Sibley County), and Vermillion River (Dakota County). Sampling from 10 sites per watershed took place in 2001 and 2002. Approximately 25 *E. coli* isolates were obtained from each site per sampling date. About 1,776, 1,651, and 1,762 E. coli were DNA fingerprinted from the Vermillion River, High Island Creek, Minneopa Creek Watersheds, respectively. The most reliable results from data came from bootstrap analyses of fecal bacteria segregated into Human vs. Non-human categories, or into groupings consisting of Humans, Pets (dogs and cats), Waterfowl (geese, ducks), Wildlife (deer), and Domesticated animals (chickens, cows, goats, horses, pigs, sheep, turkeys). Analysis of the Vermillion River showed that 93 and 6.1 % of the isolates identified were of Non-Human and Human origin, respectively. The greatest potential contributors to fecal pollution in this watershed were domesticated animals (23 %), pets (45%), and deer (19%). Similar results were found with the Minneopa Creek isolates, where 90 and 10% of the isolates were from non-human and human origin, respectively. Of these 23% were from Domesticated animals, 36% from Pets, and 21% from deer. In contrast, while 84 and 16% of High Island Creek isolates were Non-Human and Human sources, respectively, the majority came from domesticated animals (42%, mostly from cows), with the remainder contributed by geese, 14%, and humans 16%. It should be noted however, that our research showed that much larger database of DNA fingerprints is needed for more accurate assignments to the animal level. A reliable bacterial source tracking method would aid watershed managers tremendously, giving them another tool to efficiently direct efforts clean watersheds of bacterial pollutants.

Project Results Use and Dissemination

Results from this project have been disseminated in reports made to the LCMR, in periodic update reports made to cooperators, in seminars given throughout the state, nationally and internationally, and in scientific publications in peer-reviewed journals. In addition, results from our studies are posted and will be updated on the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) which is housed on computers at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our previous LCMR projects. Data obtained from our studies will be utilized by cooperating agencies and the U.S. EPA to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas.

2001 Project Abstract
For the Period Ending June 30, 2004 (extended from June 30, 2003)

TITLE: Determination of Fecal Pollution Sources in Minnesota Watersheds
PROJECT MANAGER: Dr. Michael J. Sadowsky
ORGANIZATION: University of Minnesota
ADDRESS: Department of Soil, Water & Climate, 1991 Upper Buford Circle, 439 Borlaug Hall St. Paul, MN 55108
WEB SITE ADDRESS: www.ecolirep.umn.edu
FUND: Future Resources Fund
LEGAL CITATION: ML 2001, 1st Special Session, Ch. 2, Sec.14, Subd. 6 (d)

APPROPRIATION AMOUNT: $275,000

Overall Project Outcome and Results

We used a library of DNA fingerprints, created using the rep-PCR and HFERP techniques, in an attempt to define sources of fecal bacterial pollution, *E. coli*, in three Minnesota watersheds, Minneopa Creek (Blue Earth County), High Island Creek (Sibley County), and Vermillion River (Dakota County). Sampling from 10 sites per watershed took place in 2001 and 2002. Approximately 25 *E. coli* isolates were obtained from each site per sampling date. About 1,776, 1,651, and 1,762 E. coli were DNA fingerprinted from the Vermillion River, High Island Creek, Minneopa Creek Watersheds, respectively. The most reliable results from data came from bootstrap analyses of fecal bacteria segregated into Human vs. Non-human categories, or into groupings consisting of Humans, Pets (dogs and cats), Waterfowl (geese, ducks), Wildlife (deer), and Domesticated animals (chickens, cows, goats, horses, pigs, sheep, turkeys). Analysis of the Vermillion River showed that 93 and 6.1 % of the isolates identified were of Non-Human and Human origin, respectively. The greatest potential contributors to fecal pollution in this watershed were domesticated animals (23 %), pets (45%), and deer (19%). Similar results were found with the Minneopa Creek isolates, where 90 and 10% of the isolates were from non-human and human origin, respectively. Of these 23% were from Domesticated animals, 36% from Pets, and 21% from deer. In contrast, while 84 and 16% of High Island Creek isolates were Non-Human and Human sources, respectively, the majority came from domesticated animals (42%, mostly from cows), with the remainder contributed by geese, 14%, and humans 16%. It should be noted however, that our research showed that much larger database of DNA fingerprints is needed for more accurate assignments to the animal level. A reliable bacterial source tracking method would aid watershed managers tremendously, giving them another tool to efficiently direct efforts clean watersheds of bacterial pollutants.

Project Results Use and Dissemination

Results from this project have been disseminated in reports made to the LCMR, in periodic update reports made to cooperators, in seminars given throughout the state, nationally and internationally, and in scientific publications in peer-reviewed journals. In addition, results from our studies are posted and will be updated on the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) which is housed on computers at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our previous LCMR projects. Data obtained from our studies will be utilized by cooperating agencies and the U.S. EPA to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas.

Date of Report: July 30, 2004

Date of Workprogram Approval: June 29, 2001

Project Completion Date: June 30, 2004

LCMR Final Work Program Report

I. PROJECT TITLE: Determination of Fecal Pollution Sources in Minnesota Watersheds

Project Manager: Dr. Michael J. Sadowsky

Affiliation: University of Minnesota
Mailing Address: Department of Soil, Water & Climate
1991 Upper Buford Circle
439 Borlaug Hall
St. Paul, MN 55108

Telephone Number: (612) 624-2706
E-Mail: sadowsky@umn.edu
Fax: (612) 625-6725
Web Page address: http://www.ecolirep.umn.edu

Total Biennial Project Budget:

$ LCMR Appropriation	- $ Amount Spent	= $ Balance:
$275,000	$273,494	$1,506

Legal Citation: ML 2001, 1st Special Session, Ch. 2, Sec.14, Subd. 6 (d).

Appropriation Language:

Determination of Fecal Pollution Sources in Minnesota Watersheds $275,000 is from the future resources fund to the University of Minnesota for the second biennium to determine sources of fecal pollution in three impacted watersheds utilizing DNA fingerprinting techniques, and evaluate the efficacy of implemented and proposed abatement procedures to remediate fecal contamination.

The availability of the appropriation for the following project is extended to June 30, 2004, unless an earlier date is specified in the work program: ML 2003, Art. 1,Ch.128, Sec. 9, Subd. 20(a): 6 (d) Determination of fecal pollution sources in Minnesota.

II. and III. FINAL PROJECT SUMMARY

We used a library of DNA fingerprints, created using the rep-PCR and HFERP techniques, in an attempt to define sources of fecal bacterial pollution, *E. coli*, in three Minnesota watersheds, Minneopa Creek (Blue Earth County), High Island Creek (Sibley County), and Vermillion River (Dakota County). Sampling from 10 sites per watershed took place in 2001 and 2002. Approximately 25 *E. coli* isolates were obtained from each site per sampling date. About 1,776, 1,651, and 1,762 *E. coli* were DNA fingerprinted from the Vermillion River, High Island Creek, Minneopa Creek Watersheds, respectively. The most reliable results from data came from bootstrap analyses of fecal bacteria segregated into Human vs. Non-human categories, or into groupings consisting of Humans, Pets (dogs and cats), Waterfowl (geese, ducks), Wildlife (deer), and Domesticated animals (chickens, cows, goats, horses, pigs, sheep, turkeys). Analysis of the Vermillion River showed that 93 and 6.1 % of the isolates identified were of Non-Human and Human origin, respectively. The greatest potential contributors to fecal pollution in this watershed were domesticated animals (23 %), pets (45%), and deer (19%). Similar results were found with the Minneopa Creek isolates, where 90 and 10% of the isolates were from non-human and human origin, respectively. Of these 23% were from Domesticated animals, 36% from Pets, and 21% from deer. In contrast, while 84 and 16% of High Island Creek isolates were Non-Human and Human sources, respectively, the majority came from domesticated animals (42%, mostly from cows), with the remainder contributed by geese, 14%, and humans 16%. It should be note however, that our research showed that much larger database of DNA fingerprints is needed for more accurate assignments. A reliable bacterial source tracking method would aid watershed managers tremendously, giving them another tool to efficiently direct efforts clean watersheds of bacterial pollutants.

IV. OUTLINE OF PROJECT RESULTS:

Result 1: Acquisition of fecal coliform bacteria from watershed areas.

Fecal coliform bacteria were isolated from water samples collected in each of the 3 watershed areas in the Spring, Summer, and Fall months (during baseline and critical run-off periods). Approximately 80 water samples were collected from each watershed (10 samples collected per watershed on at least eight separate sampling occasions). Samples were analyzed for fecal coliform bacteria by an EPA-certified contract laboratory (see below). *E.coli* bacteria from fecal coliform plates were isolated (about 1600 isolates per watershed for a total of approximately 4800 isolates) and the identity of bacteria was confirmed by using selective and differential microbiological media and biochemical tests. The confirmed *E. coli* isolates were cataloged, stored, and preserved in glycerol at -80°C until DNA fingerprinting was completed.

LCMR Budget:	~~**$101,200**~~	**$100,989**
Balance:	~~**$846**~~	**$635**

Personnel:	$80,000	(Sr. Scientist /Assistant Scientist [40%], Jr. Scientist [40%], and Student lab Techs [70%])
Supplies:	$14,900	(Consumables: $2.38/isolate x 5232 isolates = $12,400,Pipetors: $1000, Miscellaneous lab supplies: $1000, Sampling Supplies $500)
Local Travel:	$1,500	Approx. 5 trips @ $180/trip mileage, lodging and meals; 10 trips @ $50/trip mileage)
Contracted Services:	~~$4,800~~ $4,589	(Fecal coliform counts, 80 samples/ watershed x 3 watersheds @ $20/sample)

Total $100,989

Water Sample Collection - Collection of water samples from the three watersheds was completed in July 2002. Each watershed was sampled on at least eight sampling dates according to our original plan. Approximately 18 ml of water was collected from the center of each sample site in sterile Whirl-Pak® bags (Nasco, Fort Atkinson, WI) attached to a homemade pole with an end clamp. Samples were labeled and stored on ice at 4°C until delivered to the laboratory for analysis.

a. The isolation of bacteria from the three watersheds generally proceeded as originally planned. However, it should be noted that due to drought conditions during the summer of 2001, there was insufficient water flow at the High Island Creek watershed to obtain samples once per month. Nevertheless, we did collect snowmelt water samples from High Island on 4/9/01 (before this current project started) and a set of samples on 7/18/01. Wet weather conditions and substantial water flows in the late Spring and early Summer in 2002 allowed us to obtain additional water samples from the High Island Creek watershed. Likewise, we collected a set of samples from Minneopa Creek during Spring flooding on 4/3/01. In the other watersheds, Vermillion and Minneopa, water samples were collected once per month as originally proposed (see Table 1 below).

All water samples obtained were processed according to our original plan. Water samples were analyzed for fecal coliform bacteria by an EPA-certified laboratory at Metropolitan Council Environmental Services laboratory in St. Paul MN, using the membrane filtration method and mFC Agar plates (*Standard Methods for the Examination of Water and Wastewater,* 1995). The mFC agar plates used to enumerate fecal coliform bacteria were used by project staff for the isolation, identification and confirmation of *E. coli* strains. Typical fecal coliform colonies (blue in color) were picked and restreaked for purity onto the same medium. The purified presumptive coliform bacteria were subjected to confirmatory tests using EC broth with MUG (with Durham tubes). The EC-MUG Broth was incubated at 44.5 ± 0.2 °C. The EC medium differentiates between coliform bacteria of fecal or other origin. Presumptive *E. coli* colonies were streaked for isolation on MacConkey agar and plated onto ChromAgar ECC to differentiate between *E. coli* and *Klebsiella*. Confirmed coliform bacteria are indicated by the production of gas in Durham tubes, strong fluorescence

when EC-MUG tubes are examined by using ultraviolet light, and the production of blue colonies on ChromAgar. Isolates giving atypical responses with any test were further screened using API 20E test kits (bioMerieux, Inc., St. Louis, MO). Isolates yielding a "good" to "excellent" *E. coli* identification by the API 20E kit were used for DNA fingerprinting. Three *E. coli* colonies from each individual fecal sample were used for DNA fingerprinting and were stored at –80°C in 50% glycerol.

b. The isolation of bacteria from the three watersheds was completed in August 2002. Fecal *E. coli* bacteria were isolated from each water sample (see Table 1) using selective and differential microbiological plating media (as described above). The final number of *E. coli* isolates from each watershed exceeded the original goal of 1600 (see Table 1). In total, 5,232 *E. coli* isolates were obtained from all three watersheds. Of these, 5,189 were confirmed as *E. coli* and subjected to DNA fingerprinting.

Table 1. Water samples obtained and bacteria isolated.

Watershed	Date Sampled	*E. coli* Isolates Obtained	Total Isolates (% of Goal)
Minneopa Creek	4/3/01	100	
	7/30/01	204	
	8/15/01	216	1783
	9/12/01	185	(111)
	10/10/01	217	
	4/18/02	177	
	5/15/02	224	
	6/12/02	239	
	7/17/02	221	
High Island Creek	4/9/01	208	
	7/18/01	214	
	4/10/02	178	1651
	4/24/02	196	(103)
	5/8/02	225	
	5/29/02	180	
	6/19/02	233	
	7/10/02	217	
Vermillion River	7/11/01	237	
	8/8/01	229	
	9/5/01	238	1798
	10/3/01	238	(112)
	3/27/02	215	
	5/1/02	200	
	6/5/02	226	
	7/2/02	215	

Result 2: Generate DNA fingerprints from *E. coli* isolates obtained from watersheds.

DNAs from confirmed *E. coli* isolates were subjected to rep-PCR and HFERP DNA fingerprinting using BOXA1R primers. The resulting 5,189 DNA fingerprints were captured as digital images, band migration on gels was normalized to internal molecular weight standards, and compared and analyzed using BioNumerics pattern recognition and statistical analysis software. The animal(s) and animal groups contributing to *E. coli* in watersheds were determined by cluster and discriminant statistical analyses, by comparison to our known source DNA fingerprint library.

	LCMR Budget:	~~**$143,553**~~	**$143,764**
	Balance:	~~**$1,228**~~	**($3)**
Personnel:	~~$121,053~~	(Sr. Scientist /Assistant Scientist [40%],	
	$121,833	Jr. Scientist [60%], and Student Lab Techs [30%])	
Supplies:	~~$15,200~~	(Consumables; $2.92 /isolate x 5211 isolates)	
	$14,631		
Equipment:	$7,300	(Computer, $2500; Gel Analysis Software; $4,800)	

Total: $143,764

Completion Date: December 31, 2003

Results Status:

We previously used the rep-PCR DNA fingerprinting technique, with BOXA1R primers, to generate a DNA fingerprint library from *E. coli* bacteria obtained from 12 known animal sources and humans. This known source library was used in this current LCMR project to determine the sources of fecal bacteria in the three watersheds. Before we analyzed for the potential sources of *E. coli* isolates from the three watersheds, we performed cluster analysis of 29 DNA fingerprints generated from the same control *E. coli* strain. This was used to determine the reproducibility of rep-PCR DNA technique over a large number of gels. We found an average similarity of about 88 % between the 29 fingerprints (Fig. 1). While this level of reproducibility is sufficient to examine genetic diversity among the known source bacteria and place these bacteria into their respective source groups, it may not be adequate to assign unknown *E. coli* to the correct animal source group with a high degree of certainty. That is because the unknown isolates may be very closely related to several animal source groups at levels greater than 90%. As such, only small differences separate some isolates from different source groups, and this may reduce the statistical certainty of source group assignment. Consequently, in order assign source groups to the unknown *E. coli* bacteria from the watersheds with a greater degree of statistical certainty, and to reduced within gel grouping of DNA fingerprints and improved alignment of DNA fingerprints between gels, we modified the rep-PCR fingerprinting technique to improve

precision. The modifications used fluorescently labeled BOX primers in the rep-PCR reaction, to generate labeled PCR products, and included the use of molecular weight standards in each fingerprint lane that are labeled with a second fluorophore. This allowed accurate normalization of DNA bands in each fingerprint lane, and allows for more precise assignment of DNA bands within and across several fingerprint gels. The revised HFERP DNA fingerprinting protocol is listed below:

***E. coli* preparation and PCR conditions for HFERP Fingerprinting**. *E. coli* isolates were streaked onto Plate Count Agar (Difco, BD Diagnostic Systems, Sparks, MD) and grown overnight at 37°C. Colonies were picked with a 1 µl sterile inoculating loop (Fisher Scientific, Pittsburgh, PA), suspended in 100 µl of 0.05 M NaOH in 96-well, low profile, PCR plates (MJ Research, Waltham, MA), heated to 95 °C for 15 min, and centrifuged at 640 RPM for 10 min in a Hermle/Labnet Z383K centrifuge. A 2µl aliquot of the supernatant in each well was used as template for PCR. The PCR master mix (described here in µl per reaction), consisted of 12.65µl ddi H_2O, 5µl 5x Gitscher buffer, 2.5µl DMSO, 1.0µl 6FAM-BOX primers, 1.25µl dNTP's, 0.2µl, and 0.4µl TAQ polymerase, The primer consisted of a mixture of 0.09 µg of unlabeled Box A1R primer per µl and 0.03 µg of 6-FAM fluorescently labeled Box A1R primer per µl (Integrated DNA Technologies, Coralville, IA). The PCR was performed using an MJ Research PTC 100 (MJ Research, Waltham, MA) using the protocol specific for this thermocyclers and the Box A1R primer. PCR was initiated with an incubation at 95°C for 2 minutes, followed by 30 cycles, consisting of 94°C for 3 seconds, 92°C for 30 seconds, 50°C for 1 minute, and 65°C for 8 minutes. PCR reactions were terminated after an extension at 65°C for 8 min, and stored at 4°C. Reactions that were not used immediately for gel electrophoresis analysis were stored at -20°C. A 6.6 µl aliquot of a mixture of 50 µl Genescan-2500 ROX internal lane standard (Applied Biosystems, Foster City, CA) and 200 µl non-migrating loading dye (150 mg Ficoll 400 per ml, and 25 mg blue dextran per ml) was added to each 25 µl PCR reaction prior to loading the PCR reaction into agarose gels, 12 µl of the resulting mixture was loaded per gel lane. DNA fragments were separated by electrophoresis, which was done at 4°C for 17-18 hours at 70V with constant buffer recirculation. Gel Images were captured as TIF files, using a Typhoon 8600 Variable Mode Imager (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) operating in the fluorescence acquisition mode using the following settings: green (532 nm) excitation laser; 610 BP 30 and 526 SP emission filters in the autolink mode with 580 nm beam splitter; normal sensitivity; 200 micron/pixel scan resolution; + 3 mm focal plane; and 800 V power.

Using this modification we now show an average similarity of about 92% between fingerprints from the same control *E. coli* strain (Fig. 2). This improvement will also reduce between gel variability and increase the overall precision of our results. While this took some time to do, we think it was well worth the effort. Efforts to further refine this modification continued into January 2002, at which time we began to DNA fingerprint the watershed isolates obtained in

Result 1.

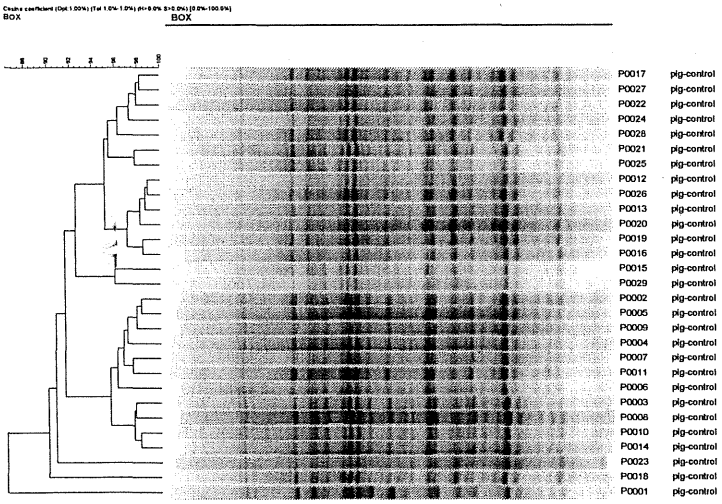


Figure 1. Similarity of DNA fingerprints from 29 control strains using old rep-PCR method.

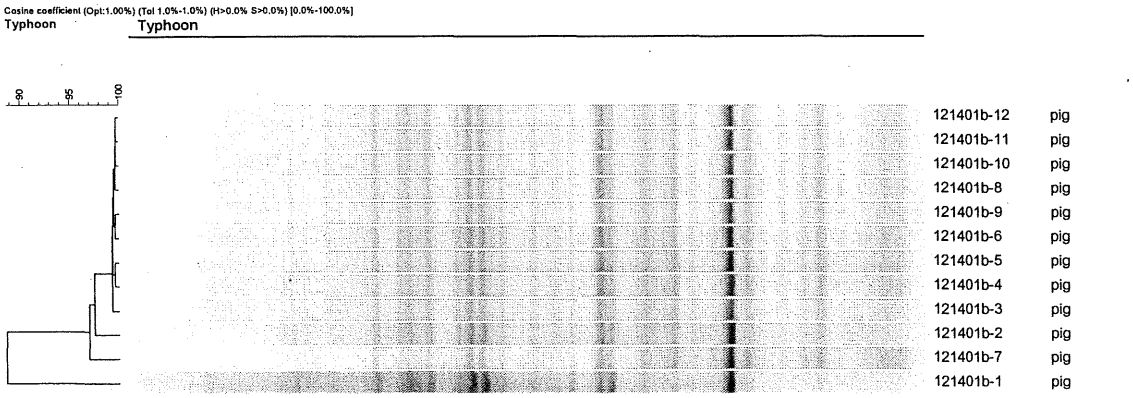


Figure 2. Similarity of DNA fingerprints from 12 control strains using New Modified rep-PCR method.

To test whether HFERP reduced within-gel groupings of DNA fingerprints, we analyzed DNA fingerprints from 40 *E. coli* strains obtained from dogs on 2 different gels using Pearson's product-moment coefficient. Results of these studies indicated that rep-PCR DNA fingerprints from strains run on the same gel were, on average, 50% (range 29 – 57%) more likely to be grouped together as the same strains analyzed by using the HFERP technique. This indicates that HFERP method considerably reduces within gel grouping of DNA fingerprints. In addition, the

HFERP method reduced alignment difficulties due to within- and between-gel variation in band migration found with rep-PCR gels.

Due to increased method precision, our data obtained using the fluorophore-enhanced technique reduced inter- and intra-gel variability contributing to error in the correct classification of known (and presumably unknown) isolates. The precision of fluorophore-enhanced DNA fingerprinting method was determined by repeated fingerprinting of a reference *E. coli* strain (pig isolate number 294). Fifty-eight (58) positive controls were generated using this method. Cluster analysis of the control strains revealed the level at which the DNA fingerprints could be repeated between experimental runs. When analyzed using the curve-based Pearson correlation coefficient and 1% optimization, the DNA fingerprints obtained using the fluorophore-enhanced technique had a 91.2% average similarity.

The result of our new normalization process is that fingerprint patterns from different gels can be accurately compared. It should be noted, however, that the intensity of HFERP bands are more variable than those generated by rep-PCR, and that some of the gains achieved by more precise alignment of bands may be offset by more variation in band intensity. We found that this variation in intensity can be overcome by the careful mixing of all reagents in the PCR master mix and greater pipetting precision when loading gels (data not presented). Further improvements to increasing the intensity of HFERP-generated DNA fingerprints may also be obtained by varying the ratio of labeled to unlabeled primer and the final concentration of the primer mixture in PCR reactions. Nevertheless, our results clearly show that HFERP-derived DNA fingerprint bands are more precisely aligned than the rep-PCR bands. In addition, we show that HFERP DNA fingerprints generated by our method reduce within gel groupings of fingerprints, which can have major ramifications for the assembly of libraries and the analysis of unknown environmental isolates.

While we previously described the use of rep-PCR DNA fingerprinting to determine sources of fecal bacteria (Dombek et al. 2000), our initial studies, and many others by most researchers, used libraries consisting of a relatively small number of samples, some of which were obtained from the same individual animal. To test the influence of library size and duplication of samples in libraries, 2,466 high-quality rep-PCR DNA fingerprints were generated using the Box A1R primer and template DNA from *E. coli* strains obtained from the 13 human and animal sources (Table 1). Of the 2,466 DNA fingerprints analyzed, 1,535 (62%) remained in the "unique" DNA fingerprint library (Table 2). The influence of duplicate DNA fingerprints on the correct classification of library strains is shown in Table 3.

Table 2. Animal source groups and rep-PCR DNA fingerprints generated from *E. coli* isolates.

Animal Source Group	Individuals Sampled	Total Fingerprints	Unique Fingerprints[a]
Cat	37	108	48
Chicken	86	231	144
Cow	115	299	191
Deer	64	179	96
Dog	71	196	106
Duck	42	122	81
Goat	36	104	42
Goose	73	200	135
Horse	44	114	79
Human	197	307	211
Pig	111	303	215
Sheep	37	101	61
Turkey	69	202	126
Total	**982**	**2,466**	**1,535**

[a]Identical *E. coli* genotypes from each individual animal were removed.

Jackknife analysis performed on the 2,466 DNA fingerprints from the entire known-source rep-PCR DNA fingerprint database, using Pearson's product-moment correlation coefficient, indicated that 69-97% of animal and human *E. coli* isolates were assigned into correct source groups (Table Y). This corresponds to an 82.2% average rate of correct classification for the 2,466 rep-PCR DNA fingerprints. However, since identical DNA fingerprints from *E. coli* strains obtained from the same individual most likely represent isolates of clonal origin, and can artificially bias subsequent analyses of strain groupings (e.g. increase the average rate of correct classification) and the fidelity of the database, we eliminated duplicate DNA fingerprints originating from *E. coli* strains obtained from the same individual animal or human. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%.

Of the 2,466 DNA fingerprints analyzed, 1,535 (62%) remained in the "unique" DNA fingerprint library (Table 2). The influence of duplicate DNA fingerprints on the correct classification of library strains is shown in Table 2. When the 1,535 DNA fingerprints from the unique *E. coli* isolates were examined, Jackknife analyses indicated that only 44-74% of the isolates were assigned to the correct source group (Table 3). The average rate of correct classification for these 1,535 unique rep-PCR DNA fingerprints was 60.5%. Taken together, these results indicate that

inclusion of duplicate DNA fingerprints in the library can artificially influence strain groupings and increase percentages of strains correctly assigned to source groups.

Table 3. Percentage of known-source rep-PCR DNA fingerprints assigned to the correct source group by Jackknife analysis[a].

Animal Source	**All Fingerprints (n=2,466)**	**Unique Fingerprints (n=1,535)**
	Percent Correctly Classified Isolates	
Pets[b]	91.8 (279)[d]	61.7 (95)
Chicken	81.4 (188)	59.7 (86)
Cow	79.6 (238)	55.0 (105)
Deer	85.5 (145)	55.2 (53)
Waterfowl[c]	81.4 (262)	66.2 (143)
Goat	97.1 (101)	66.7 (28)
Horse	69.3 (79)	44.3 (35)
Human	78.3 (240)	59.2 (125)
Pig	77.9 (236)	63.7 (137)
Sheep	79.0 (80)	47.5 (29)
Turkey	88.6 (179)	73.8 (93)
Overall	82.2 (2,027)	60.5 (929)

In addition, our studies reported here we show that increasing the size of the known source library to 2,466 isolates does not necessarily lead to an increase in the ability to correctly assign strains to the correct source group. In fact, the average rate of correct classification decreased 4.2% using the larger library reported here, relative to what was seen using a smaller library in our previous studies. This may in part be due to the uncovering of increased genetic diversity among isolates, increased accumulation of errors due to gel-to-gel variation, or the presence of duplicate genotypes (DNA fingerprints) from the same individual within our original library. Reduction in the percentage of known-source *E. coli* isolates that were correctly classified was especially apparent when our unique library of 1535 *E. coli* isolates was examined. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%. Since DNA fingerprints from *E. coli* strains obtained from the same individual represent isolates of clonal origin, these duplicate strains (or fingerprints) can artificially bias the average rate of correct classification and the fidelity of the database. Results in Table 3 show that there was a 21.7% reduction in the average rate of correct classification by using the unique DNA fingerprint library, relative to that seen with the complete library. More importantly, our results show that failure to remove identical fingerprints from analyses resulted in an overestimation of the ability of the database to assign isolates to their correct source group, perhaps in part due to the clonal composition of *E. coli*

populations. Taken together, our results indicate that inclusion of duplicate DNA fingerprints in the library can artificially influence strain groupings and incorrectly increases percentages of strains correctly assigned to source groups.

The fluorophore-enhanced DNA fingerprinting method was then applied to 1,531 *E. coli* isolates included in the "unique" isolate subset. Table 4 summarizes the ability of this method to correctly classify *E. coli* isolates into each of the thirteen human and animal source groups based on Pearson correlation calculations. The ability to correctly classify isolates into their source groups was 57.6% overall correct classification for the fluorophore-enhanced technique.

Table 4. *E. coli* isolates correctly classified into thirteen source groups using the fluorophore-enhanced DNA fingerprinting (HFERP) method.

Source group	Number of DNA Fingerprints	Percent correctly classified[1]
Cat	48	37.5 (18)
Chicken	144	63.2 (91)
Cow	189	62.4 (118)
Deer	96	49.0 (47)
Dog	106	58.5 (62)
Duck	81	60.5 (49)
Goat	42	50.0 (21)
Goose	135	65.2 (88)
Horse	78	52.6 (41)
Human	210	55.7 (117)
Pig	215	54.0 (116)
Sheep	61	39.3 (24)
Turkey	126	71.4 (90)
Overall	1531	57.6 (882)

[1]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) calculation.
[2]Values in parentheses are number of isolates (n) correctly classified.

Further refinements to the jackknife analysis, including pooling of animal source groups (Tables 5 and 6) and limiting the analysis to relevant source groups, were found to improve the ability of the DNA fingerprint library to correctly classify isolate sources by both DNA fingerprinting methods. When all animal sources were pooled into one group, the overall correct classification rate for humans and animals by the fluorophore-enhanced technique was 88.1% (Table 5). Accordingly, these results indicated that (1) broader classifications of source groups should be used when appropriate, or (2) a targeted subset of the DNA fingerprint database should be used to more precisely determine sources of fecal pollutants in watersheds where specific source groups are known to be present or absent. When isolates were separated into domesticated animals, wildlife, and humans, the average rate of correct classification was 78.5%, with domesticated animals being most successfully classified (Table 6).

Table 5. Percentages of *E. coli* isolates correctly classified into human and animal source groups by the fluorophore-enhanced DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent correctly classified[1]
Animal	1321	93.3 (1232)[2]
Human	210	55.7 (117)
Overall	1531	88.1 (1349)

[1]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) similarity calculation.
[2]Values in parentheses are number of isolates (n) correctly classified.

Table 6. Percentages of *E. coli* isolates correctly classified into domesticated, human and wildlife source groups by the fluorophore-enhanced DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent correctly classified[1]
Domesticated[3]	1009	86.2 (870)
Human	210	55.7 (117)
Wildlife[4]	312	68.9 (215)
Overall	1531	78.5 (1202)

[1]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) similarity calculation.
[2]Values in parentheses are number of isolates (n) correctly classified.
[3]The domesticated group includes cat, chicken, cow, dog, goat, horse, pig, sheep and turkey.
[4]The wildlife group includes deer, duck and goose.

To estimate genetic diversity of the *E. coli* comprising the known source database, an accumulation curve was constructed using fluorophore-enhanced fingerprints from the known-source DNA fingerprint library. To do this, each *E. coli* isolate in the library was assigned to a genotype. A genotype was defined as a cluster of DNA fingerprints with similarity of 90% or greater (based on Pearson correlation, 1% optimization and UPMGA). Using this definition, 657 genotypes were identified from the 1,531 unique *E. coli* isolates in the known-source database. The isolates were randomized, and an accumulation curve was constructed by summing the number of genotypes represented by the isolates. The resulting accumulation curve is shown in Figure 3, and is essentially linear over its entire range. The linear, non-asymptotic nature of the curve is indicative of high genetic diversity. Based on these results it is not possible to predict how many new isolates would be required before *E. coli* no new genotypes are acquired. It is clear, however, that the 1531 unique DNA fingerprints contained in the known-source database do not adequately capture the genotypic diversity that exists among naturally-occurring *E.coli*, and that significantly more samples would be required to obtain a library that is adequate enough

to identify all naturally-occurring unknown isolates in watersheds. This problem is not unique to our technique, as all library-based source tracking methods using *E. coli* have the same inherent problem.

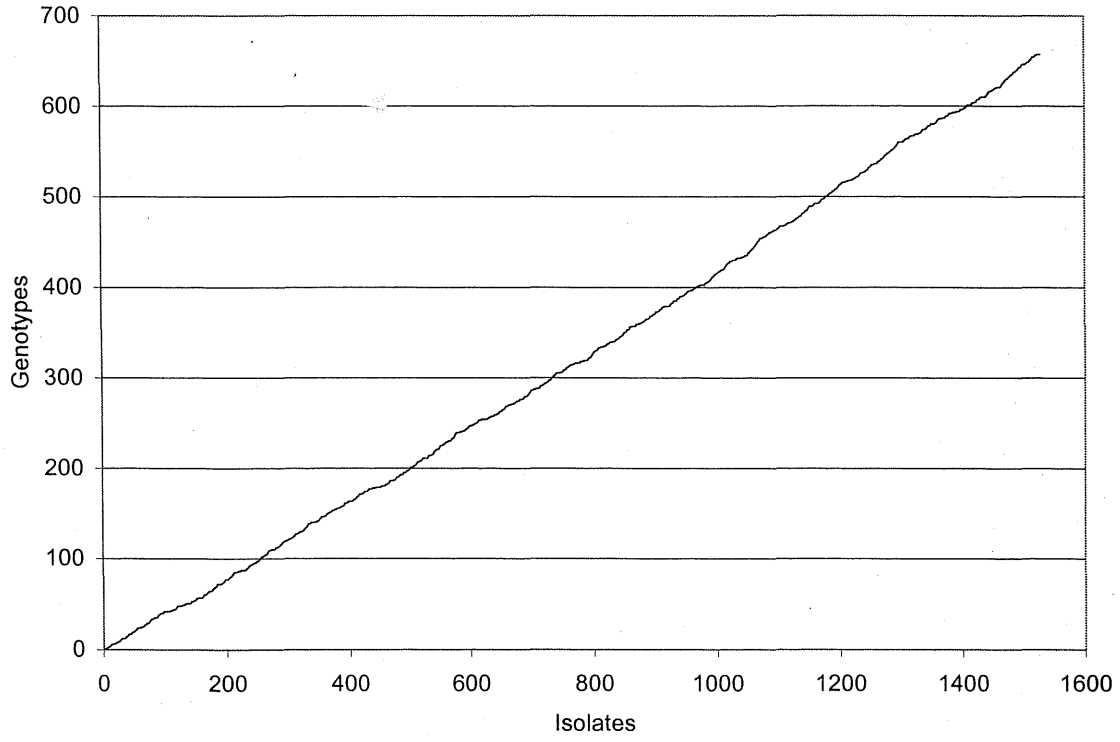


Figure 3. Diversity among *E. coli* genotypes in the unique isolate subset of the known-source library (n=1531). A genotype was defined as a cluster of DNA fingerprints with similarity of 90% or greater (based on Pearson correlation coefficient, 1% optimization and UPMGA).

Results presented here also show that despite having a known source library or over 1,500 unique isolates, the number of genotypes uncovered by DNA fingerprinting continued to increase at a constant rate. Moreover, across all animal hosts, the majority of these fingerprints occurred only once. For a library to be truly representative it needs to be large enough to capture all the unknowns present in an environmental sample, otherwise strain assignment will most likely be incorrect, or a large number of isolates will be characterized as being unknowns or cosmopolitan. Since the rarefaction curve in Figure 1 has not become asymptotic, our data cannot be used to predict the ultimate size that this library needs to be. However, data presented in Figure 3 indicates that with our current library size, each new isolate added to the library only has a greater than 50% chance of being new. It has been suggested that a library size of 20,000 to 40,000 isolates may be needed to capture all the genetic diversity present in *E. coli* (Mansour Samadpour, personal communication). One suggested strategy to avoid this under-representation problem in large regional or national libraries, is to develop moderate sized libraries for a highly confined geographical region, wherein isolates are only obtained from the animals in the study

area. In this way only animals pertinent to the study site, and those likely to have an impact on the targeted watershed, need to be examined in detail

In October 2002, work on DNA fingerprinting the remaining *E. coli* isolates from the three watersheds was temporarily suspended to allow us to participate in a "round-robin" study of source tracking methods (see below).

The Southern California Coastal Water Research Project (SCCWRP) and the U.S. EPA asked us to participate in a "round-robin" study to evaluate methods for determining sources of fecal pollution in waterways. SCCWRP and U.S. EPA provided funds for personnel and supplies for us to evaluate the use of our LCMR-funded rep-PCR DNA fingerprinting method to determine sources of fecal pollutants in water samples. About 22 laboratories are participating in this study, and are examining the usefulness of 8 different methods to determine sources of fecal bacteria in replicated and identical samples. All participants in the study received samples on October 8, 2002 and have about 4 months to complete their analyses. The study offers a unique opportunity to verify and validate our DNA fingerprinting methods that were developed from our LCMR-funded project. We will complete work on the SCCWRP-EPA project by the end of February 2003, and at that time we will resume work on our LCMR-sponsored project. We requested and were granted a no-cost extension of our LCMR appropriation, so that we could participate in the SCCWRP-EPA round-robin study.

Results from this round robin study were published in the November issue of the Journal of Water and Health. We contributed to two manuscripts in this publication: one dealing with genotypic methods for source tracking (Comparison of genotypic-based microbial source tracking methods requiring a host origin database by Samuel P. Myoda, C. Andrew Carson, Jeffry J. Fuhrmann, Byoung-Kwon Hahm, Peter G. Hartel, Helen Yampara-Iquise, LeeAnn Johnson, Robin L. Kuntz, Cindy H. Nakatsu, Michael J. Sadowsky and Mansour Samadpour, pp. 167-180), and the other a paper (Assessment of statistical methods used in library-based approaches to microbial source tracking by Kerry J. Ritter, Ethan Carruthers, C. Andrew Carson, R. D. Ellender, Valerie J. Harwood, Kyle Kingsley, Cindy Nakatsu, Michael Sadowsky, Brian Shear, Brian West, John E. Whitlock, Bruce A. Wiggins and Jayson D. Wilbur, pp. 209-223) on the proper statistical analyses to use to analyze DNA fingerprint data generated by rep-PCR. When our data was re-analyzed with stringent assignments to source groups, our rep-PCR based fingerprinting method was found to be far superior to almost all other methods used by participants of the study.

Our data analysis was improved, due to our collaboration with the SCCWRP statistician, Ms. Kerry Ritter, the development of a new software module, ID bootstrap, produced by Bionumerics, and the use of quality factors in our data analysis. Briefly, the ID bootstrap software addition applies reiterative analysis of the integrity of known source groups and applies the resulting correlation statistic to assign identities to unknown isolates. This determines the fidelity of our assignments of unknown isolates, and reduces the number of false positive results from 48% to 6% and false negative results from 4% to 2%.

In February 2003, following completion of the SCCWRP project, our staff resumed DNA fingerprinting of environmental *E. coli* isolates from the three watersheds using the newly developed fluorophore-enhanced modification of the rep-PCR DNA fingerprinting technique (FERP). Fingerprinting and analysis proceeded until December 2003.

Reanalysis of known source isolates - Due to improved statistical analysis techniques that we learned working on the SCCWRP project and with our consultation with Ms. Kerry Ritter, a statistician at SCCWRP, the inclusion of an amended grouping system, and the accumulation of a larger number of positive control strains, in the Fall of 2003 we reanalyzed the isolates and groups previously identified in Tables 4, 5 and 6 (see new analysis results in Tables 7, 8 and 9).

To put this analysis in perspective, it is important to note that a variety of similarity measures exist. Binary similarity coefficients are mostly used to analyze presence/absence data and band-matching data obtained from DNA fingerprints can be analyzed using binary coefficients. However, quantitative similarity coefficients require a measure of relative abundance. Quantitative coefficients can be applied to DNA fingerprints when the fingerprints are analyzed as densitometric curves that take into account both peak position and intensity (peak height). For complex DNA fingerprints, such as those produced with the techniques we used here, a curve-based method such as Pearson's product-moment correlation coefficient more reliably identified similar or identical DNA fingerprints than band matching formulas, such as simple matching, Dice, or Jaccard. Results presented here confirm that the curve-based Pearson's product-moment correlation coefficient was superior to the band-based Jaccard algorithm is correctly assigning isolates to the correct source group. The influence of analysis method on the classification of source group isolates is shown in Table 7.

Table 7. Unique *E. coli* isolates correctly classified into source groups by rep-PCR and HFERP DNA fingerprinting methods.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]			
		rep-PCR		HFERP	
		Pearson	Jaccard	Pearson	Jaccard
Pets[b]	154	61.7 (95)[d]	45.5 (70)	59.1 (91)	44.8 (69)
Chicken	144	59.7 (86)	38.9 (56)	63.2 (91)	31.9 (46)
Cow	189	55.0 (104)	47.6 (90)	62.0 (117)	48.2 (91)
Deer	96	55.2 (53)	36.5 (35)	62.2 (60)	42.6 (41)
Waterfowl[c]	216	66.2 (150)	52.8 (114)	70.4 (152)	56.5 (122)
Goat	42	66.7 (27)	59.5 (25)	47.6 (20)	42.9 (18)
Horse	78	44.3 (35)	34.2(27)	52.6 (41)	32.1 (25)
Human	210	59.2 (124)	47.4(100)	53.8 (113)	45.2 (95)
Pig	215	63.7 (137)	43.7 (94)	54.4 (117)	36.3 (78)
Sheep	61	7.5 (29)	39.3 (24)	37.7 (23)	8.2 (5)
Turkey	126	73.8 (93)	52.4 (66)	73.0 (92)	54.8 (69)
Overall	1,531	60.9 (933)	45.8 (701)	59.9 (917)	43.0 (659)

[a]Based on Jackknife analysis with 1% optimization and maximum similarities using curve-based (Pearson's product moment correlation coefficient) or band-based (Jaccard's coefficient) similarity calculations.
[b]Pet group consists of cats and dogs.
[c]Waterfowl group consists of ducks and geese.
[d]Values in parentheses are number of isolates correctly classified.

The 1,535 previously selected unique *E. coli* isolates from animals and humans were subjected to HFERP DNA fingerprinting using a combination of fluorescently labeled and unlabeled Box A1R PCR primers. Jackknife analyses of HFERP gels done using the curve-based Pearson's correlation coefficient indicated that 38-73% of the isolates were assigned to the correct source group using this technique (Table 7). For the curve-based analysis, the HFERP technique had the lowest percent of correctly classified strain in cases where the numbers of analyzed fingerprints were relatively small (for sheep, horses, and goats). The average rate of correct classification for the unique HFERP-generated DNA fingerprints was 59.9%. In contrast, Jacknife analyses of HFERP-generated DNA fingerprints done using the band-based Jaccard analysis showed that only 8-56% of the *E. coli* isolates were assigned to the correct source group, with a 43.0% average rate of correct classification. This indicates that for this type of data, the Pearson's product-moment correlation coefficient was superior to Jaccard's band matching algorithm for assigning known isolates to the correct source groups. Interestingly, results in Table 3 also show that despite problems associated with within- and between-gel variation, within-gel grouping of isolates, and repeatability issues, Jacknife analysis of rep-PCR DNA fingerprints, analyzed using Pearson's correlation coefficient, indicated that 48-74% of the isolates were assigned to the correct source group, a 60.9% average rate of correct classification. Analysis of rep-PCR DNA fingerprint data using the Jaccard band-based method was not as useful in separating *E. coli* isolates into their correct source group as was the curve-based method.

In some instances, it may be sufficient to identify unknown watershed *E. coli* isolates to larger groupings, rather than to individual animal types. To determine if the HFERP-generated DNA fingerprint data from our library of unique *E. coli* isolates grouped well into larger categories, we assembled DNA fingerprints from pets (dogs and cats), domesticated animals (chickens, cows, goats, horses, pigs, sheep, and turkeys), wild-life (deer, ducks, and geese), and humans, and used Jacknife analysis to assess the percent of correctly classified strains. Results in Table 8 show that the HFERP DNA fingerprints, analyzed using Pearson's product-moment correlation coefficient, correctly classified 83.2, 53.8, 71.4, and 59.1% of the isolates into the domesticated, human, wildlife, and pet categories, respectively. The average rate of correct classification for these groups was 74.3%. However, when DNA fingerprints were analyzed using Jaccard's coefficient, the average rate of correct classification was 66.2%. As before, the least precision was found in categories having the smallest number of fingerprints, pets and humans, suggesting that there is an apparent relationship between the number of fingerprints analyzed and the percentage of correctly classified isolates.

In microbial source tracking studies it may often be useful to determine if unknown isolates belong to either animal or human source groups, rather than to more specific categories. Results in Table 9 show that about 94% and 54% of *E. coli* from animals and humans, respectively, were

assigned to the correct source groups using HFERP-generated DNA fingerprints and Pearson's correlation coefficient. The average rate of correct classification was 88.2 and 86.1% for analyses done using Pearson's and Jaccard's algorithms, respectively. The lower percentage of correctly classified human isolates may, in part, be due to the smaller size of fingerprints analyzed for this category.

Table 8. Percentages of *E. coli* isolates correctly reclassified into domesticated, human, pets and wildlife source groups by the fluorophore-enhanced DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]
Domesticated[b]	855	83.2 (711)[d]
Human	210	53.8 (113)
Wildlife[c]	312	71.4 (223)
Pets	154	59.1 (91)
Overall	1531	78.5 (1202)

[a]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient)
[b]The domesticated group includes cat, chicken, cow, dog, goat, horse, pig, sheep and turkey.
[c]The wildlife group includes deer, duck and goose.
[d]Values in parentheses are number of isolates (n) correctly classified.

Table 9. Percentages of *E. coli* isolates correctly reclassified into human and animal source groups by the fluorophore-enhanced DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]
Animal	1321	93.7 (1237)[b]
Human	210	53.8 (113)
Overall	1531	88.2 (1350)

[a]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) similarity calculation.
[b]Values in parentheses are number of isolates (n) correctly classified.

Results of our studies indicated that further refinements to the Jackknife analysis, including the pooling of source groups into domesticated , human, and wild-life categories, were found to improve the ability to correctly classify isolate to their respective source groups. Over 83, 53, and 71% of domesticated animals, humans, and wild-life animals, respectively, were correctly classified using this approach with the unique DNA fingerprint library analyzed by HFERP.

When all animal sources were pooled into one group, the overall correct classification rate for humans and animals by HFERP was improved to about 94 and 54%, respectively, when analyzed using the curve-based Pearson's correlation coefficient. Accordingly, these results indicated that (1) broader classifications of source groups should be used when appropriate, or (2) a targeted subset of the DNA fingerprint database should be used to more precisely determine sources of fecal pollutants in watersheds where specific source groups are known to be present. The pooling of source groups into a more limited number of categories has previously been shown to increase the average rate of correct classification following discriminant analysis of antibiotic resistance, ribotype, and rep-PCR DNA fingerprint analyses.

In summary, our results suggest that HFERP-generated Box A1R DNA fingerprints of *E. coli* are useful to differentiate between different *E. coli* subtypes of human and animal origin and that this method reduces within gel groupings of DNA fingerprints, and ensures more proper alignment and normalization of fingerprint data. However, our results further indicate that other important issues must also be resolved to more fully understand the potential applications and limitations of this and other library-based microbial source tracking methodologies. Among these are questions concerning the inclusion of identical DNA fingerprints from the same animal in the library, the number of fingerprints that must be included in an *E. coli* known source library to adequately capture the diversity of *E. coli* genotypes that exist among potential host animals, and ultimately, whether *E. coli* exhibits a sufficient level of host specificity to allow unambiguous assignment of unknown environmental *E. coli* to specific host animals.

Results of these studies will be published in the August issue of Applied and Environmental Microbiology for (see Appendix B). The manuscript is entitled: "Sample Size, Library Composition, and Genotypic Diversity Influence Accuracy of Determining Sources of Fecal Pollution Among Natural Populations of *Escherichia coli* from Different Animals" by LeeAnn K. Johnson, Mary B. Brown, Ethan A. Carruthers, John A. Ferguson, Priscilla E. Dombek and Michael J. Sadowsky.

WATERSHED ANALYSES

Computer-assisted HFERP DNA fingerprint analysis.

DNA fingerprinting of environmental *E. coli* isolates, using the newly developed HFERP fluorophore-enhanced modification of the rep-PCR DNA fingerprinting technique, was initiated in September 2002. The watershed *E. coli* isolates, or "unknowns", were compared with a previously described existing known source DNA fingerprint library, consisting of 12 animal species and humans, combined into either 11, 4 or 2 groups. Results in Table 10 report the total number of isolates obtained from each watershed, and the number of bonefide *E. coli* isolated fingerprinted. The final number of *E. coli* isolates from each watershed exceeded the original goal of 1600 by 3 – 12%. In total, 5,232 *E. coli* isolates were obtained from all three watersheds. Of these, 5,189 (99%) were confirmed as *E. coli* and subjected to DNA fingerprinting.

Below are results our analysis of the identity of *E. coli* isolates obtained from each watershed. At the end of each watershed is a summary of the analysis.

Table 10. *E.coli* isolates from watersheds that were DNA fingerprinted.

Watershed	Number *E. coli* Isolated	Number *E. coli* Fingerprinted
Vermillion River	1,798	1,776
High Island Creek	1,651	1,651
Minneopa Creek	1,783	1,762

SUMMARY OF HIGH ISLAND CREEK ANALYSES

We analyzed 1,651 *E.coli* isolates obtained from the High Island Creek Watershed. Sampling took place on 4/9/2001, 7/18/2001, 4/10/2002, 4/24/2002, 5/8/2002, 5/29/2002, 6/19/2002, and 7/10/2002. Approximately 25 *E.coli* isolates were obtained from each sampling site on each date. Sites sampled were: 1s, 2p, 3p, T (tile-line), 5p, 6s, 7s, 8s, 9p, 10p, and 11. (Site 11 was sampled only on 4/9/2001 (see Appendix B for sample Map).

As reported in the July 2003 status report, we did a preliminary study in which we applied statistical analysis to unknown environmental *E. coli* isolates from High Island Creek. The Pearson cosine coefficient analysis and ID Bootstrap analysis was used on DNA fingerprint fragments sizes of 287 to 14,051 basepairs. Please note that the following figures and tables are a result of the use of improved settings and parameters, as described in the method section above. While we initially reported in the July 2003 progress report that it was our plan to exclude cat isolates from analyses due to difficulties they present, upon further examination we decided to include these isolates in the combined pet category.

Results presented in Table 11 show our analysis of the probable identity of *E. coli* isolates in High Island Creek, using Pearson's Correlation Coefficient with 1% optimization. Similar tables and figures are shown on a watershed-by-watershed basis, and discussed at the end.

Table 11. Percent Identity of *E. coli* isolates from Separate Animals, by Date, obtained from High Island Creek.

	Sample Date							
Animal	**04/09/01**	**07/18/01**	**04/10/02**	**04/24/02**	**05/08/02**	**05/29/02**	**06/19/02**	**07/10/02**
	Percent Isolates							
Cat	2.9	1.8	16.2	4.8	1.7	3.7	2.7	1.4
Chicken	4.6	1.8	0.7	3.4	6.3	8.1	12.4	10.3
Cow	13.9	18.8	4.2	15.0	16.5	12.5	9.2	12.3
Deer	28.3	29.4	23.9	25.2	6.3	16.9	18.9	13.7
Dog	2.9	5.9	14.8	6.1	6.8	8.1	2.7	3.4
Duck	3.5	2.4	0.7	3.4	5.7	1.5	3.8	2.1
Goat	0.6	1.2	0.0	1.4	2.3	5.2	1.6	5.5
Goose	8.1	11.8	2.8	9.5	10.8	14.0	13.5	17.1
Horse	0.0	0.0	0.0	1.4	6.8	2.9	0.5	0.7
Human	6.4	5.3	19.0	6.8	8.5	6.6	10.3	6.2
Pig	18.5	15.3	10.6	12.9	19.9	11.8	10.3	8.2
Sheep	7.5	4.1	3.5	2.7	2.3	1.5	2.2	4.1
Turkey	2.9	2.4	3.5	7.5	6.3	7.4	11.9	15.1

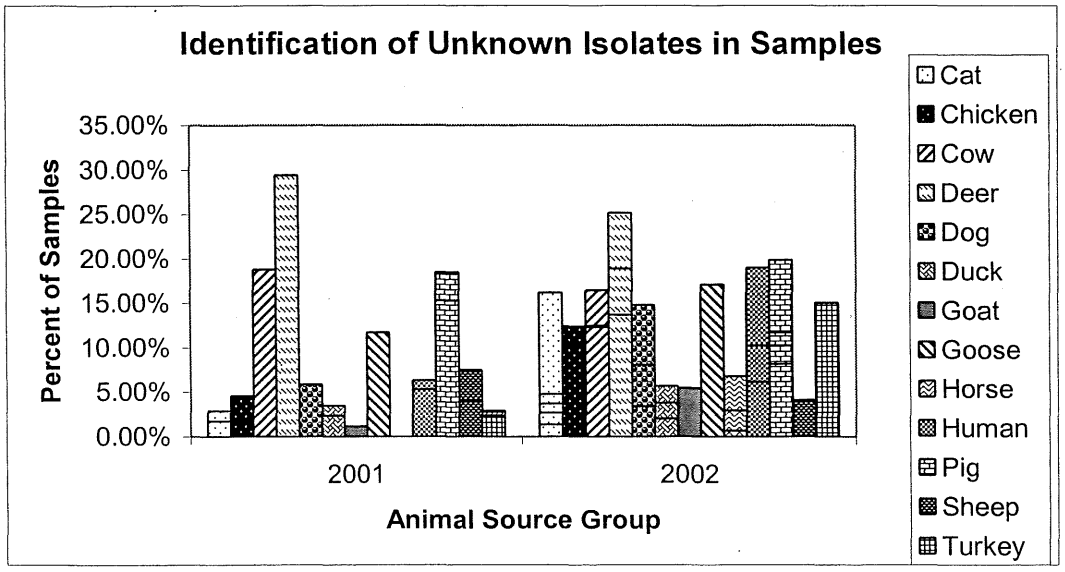


Figure 4. Percent of samples from each animal source group recovered in 2001 and 2002.

Table 12. Probable Identity of *E. coli* isolates from Separate Animals, by Site, obtained from High Island Creek.

Animal Type	Site										
	1s	2p	3p	T	5p	6s	7s	8s	9p	10p	11
	Percent of Samples										
Cat	1.4	2.6	4.4	6.3	3.7	2.2	5.0	10.4	5.9	2.7	0.0
Chicken	3.6	7.1	6.6	4.5	5.9	6.6	6.7	8.5	6.9	5.3	0.0
Cow	7.9	16.2	21.2	13.4	10.3	19.0	15.1	9.4	5.9	8.9	5.0
Deer	25.2	18.2	15.3	16.1	16.2	24.8	15.1	11.3	24.5	35.4	30.0
Dog	7.9	7.1	4.4	7.1	8.8	2.2	10.1	1.9	6.9	4.4	5.0
Duck	4.3	1.3	2.9	1.8	2.9	0.7	1.7	3.8	8.8	3.5	0.0
Goat	3.6	2.6	0.7	5.4	5.9	1.5	0.0	0.0	0.0	0.9	0.0
Goose	13.7	9.1	8.0	10.7	12.5	8.8	10.1	15.1	9.8	12.4	15.0
Horse	2.2	0.7	1.5	1.8	0.7	1.5	2.5	2.8	1.0	1.8	0.0
Human	10.8	17.5	2.9	8.0	4.4	5.1	10.1	10.4	11.8	4.4	5.0
Pig	15.1	9.1	17.5	16.1	15.4	13.9	11.8	17.9	7.8	10.6	20.0
Sheep	1.4	2.6	6.6	3.6	4.4	3.7	5.9	2.8	2.0	0.9	10.0
Turkey	2.9	5.8	8.0	5.4	8.8	10.2	5.9	5.7	8.8	8.9	10.0

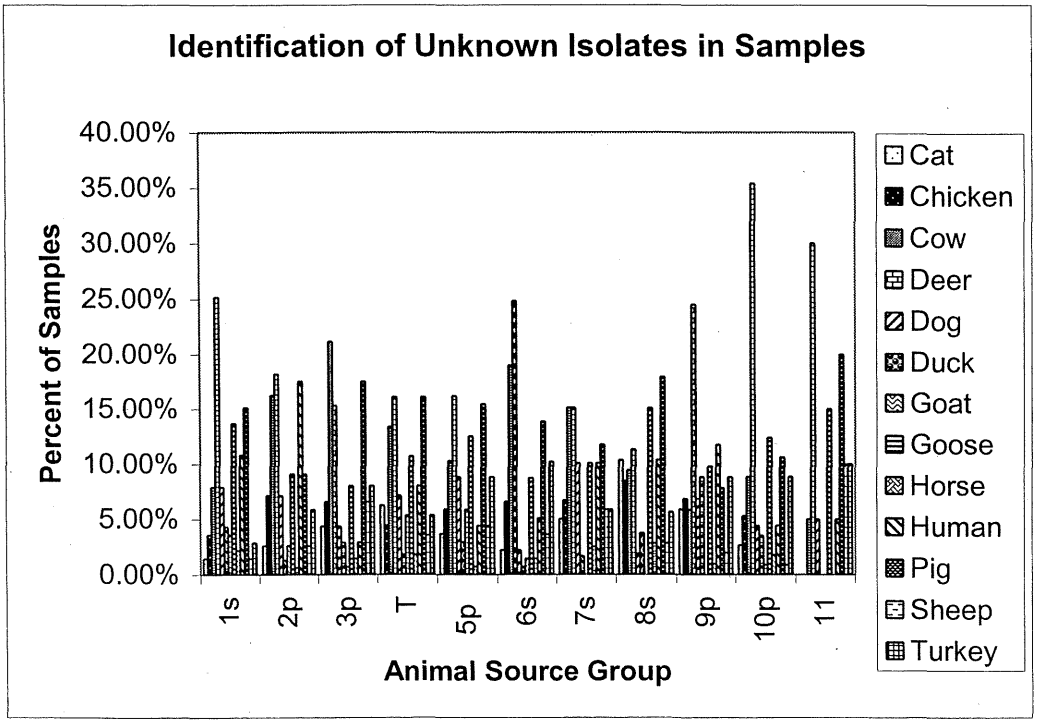


Figure 5. Percent of samples having isolates in each animal source group recovered by Site, obtained from High Island Creek.

Table 13. Probable Identity of *E. coli* isolates from Domestic, Human, Pets, Waterfowl, Wild source groups, by Sample Date, obtained from High Island Creek.

Animal Type	Date							
	04/09/01	07/18/01	04/10/02	04/24/02	05/08/02	05/29/02	06/19/02	07/10/02
	Percent of Isolates							
Domesticated	48.0	43.5	22.5	47.8	60.2	49.3	48.1	56.2
Human	6.4	5.3	19.0	6.4	8.5	6.6	10.3	6.2
Pets	5.8	7.7	31.0	10.2	8.5	11.8	5.4	4.8
Waterfowl	11.6	14.1	3.5	12.1	16.5	15.4	17.3	19.2
Wild Animal (Deer)	28.3	29.4	23.9	23.6	6.3	16.9	18.9	13.7

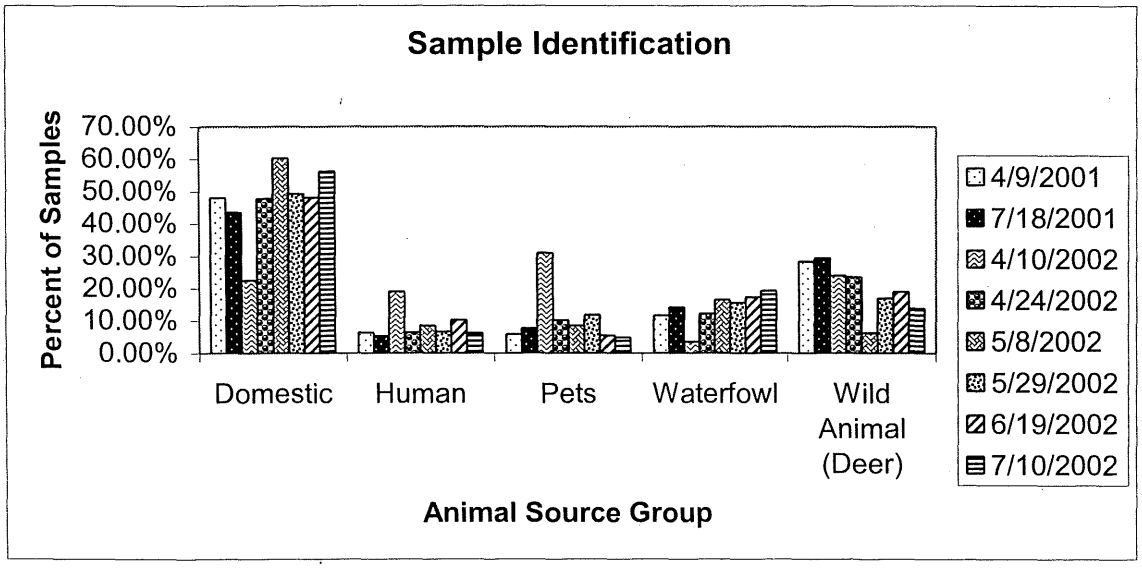


Figure 6. Percent of samples having isolates in animal source groups by sample date.

Table 14. Probable Identity of *E. coli* isolates from Domestic, Human, Pets, Waterfowl, Wild source groups, by Site, obtained from High Island Creek.

Animal Type	1s	2p	3p	T	5p	6s	7s	8s	9p	10p	11
	Percent of Samples										
Domesticated	36.7	44.2	62.0	50.0	51.5	56.2	47.9	47.2	32.4	37.2	45.0
Human	10.8	17.5	2.9	8.0	4.4	5.1	10.1	10.4	11.8	4.4	5.0
Pets	9.4	9.7	8.8	13.4	12.5	4.4	15.1	12.3	12.8	7.1	5.0
Waterfowl	18.0	10.4	11.0	12.5	15.4	9.5	11.8	18.9	18.6	15.9	15.0
Wild Animal (Deer)	25.2	18.2	15.3	16.1	16.2	24.8	15.1	11.3	24.5	35.4	30.0

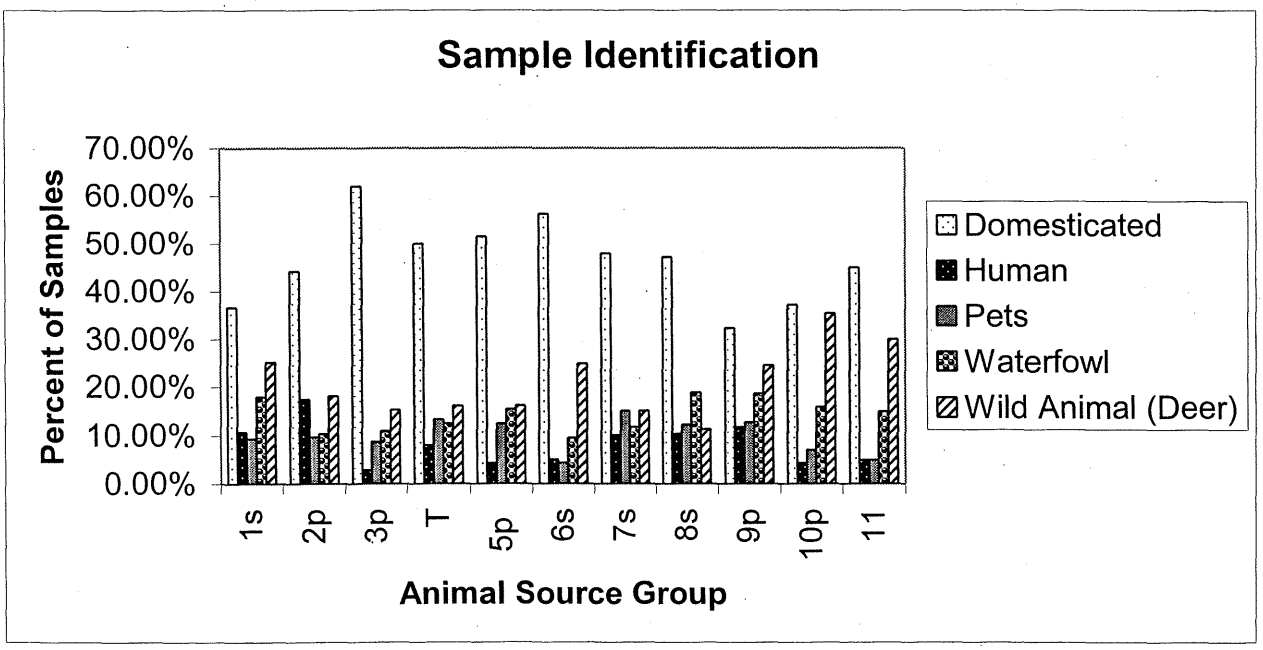


Figure 7. Percent of samples having isolates in Domestic, Human, Pets, Waterfowl, and Wild animal source groups by site.

Table 15. Probable Identity of *E. coli* isolates from Household (human) and Non-Households, by Sample Date, obtained from High Island Creek.

Animal Type	Date							
	04/09/01	07/18/01	04/10/02	04/24/02	05/08/02	05/29/02	06/19/02	07/10/02
	Percent of Samples							
Household	12.1	12.9	50.0	17.7	17.1	18.4	15.7	11.0
Non-Household	87.9	87.1	50.0	82.3	83.0	81.6	84.3	89.0

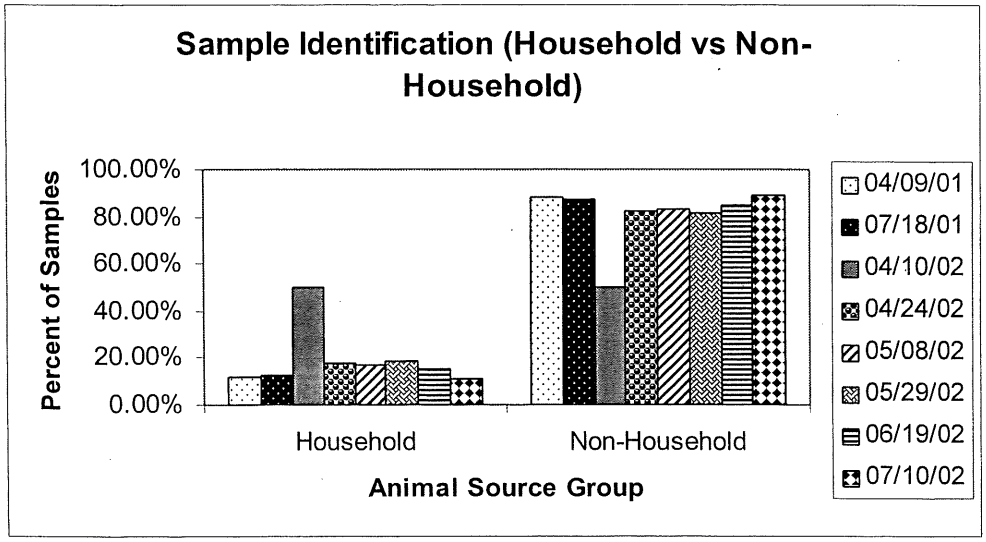


Figure 8. Percent of samples having isolates from Household (human) and Non-Households, by Sample Date, obtained from High Island Creek.

Table 16. Probable Identity of *E. coli* isolates from Household and Non-Households, by Site, obtained from High Island Creek.

Animal Type	Site										
	1s	2p	3p	T	5p	6s	7s	8s	9p	10p	11
	Percent of Samples										
Household	20.1	27.3	11.7	21.4	16.9	9.5	25.2	22.6	24.5	11.5	10.0
Non-Household	79.9	72.7	88.3	78.6	83.1	90.5	74.8	77.4	75.5	88.5	90.0

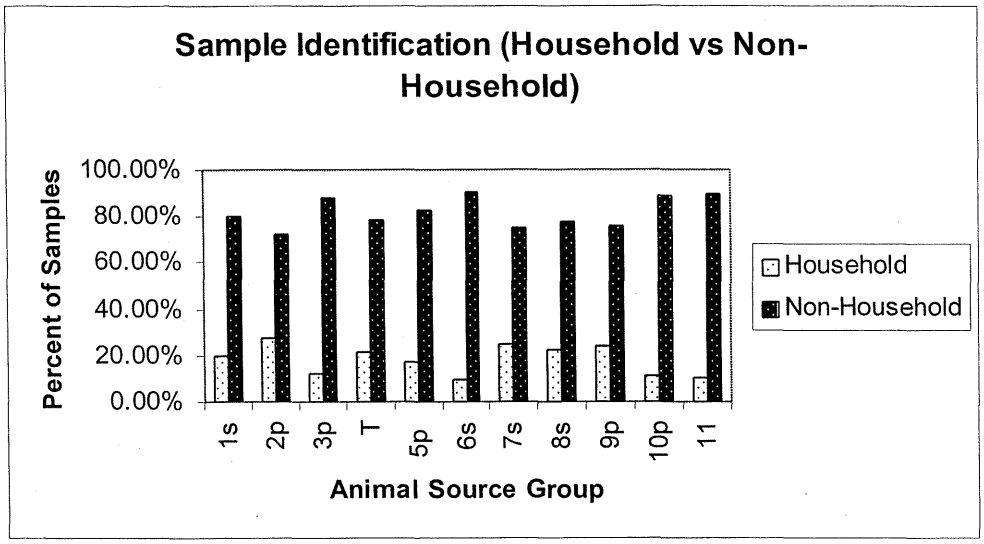


Figure 9. Percent of samples having isolates from Household and Non-Households, by Sample Site, obtained from High Island Creek.

Table 17. Probable Identity of *E. coli* isolates from Humans and Non-Humans, by Sample Date, obtained from High Island Creek.

Animal Type	Date							
	04/09/01	07/18/01	04/10/02	04/24/02	05/08/02	05/29/02	06/19/02	07/10/02
	Percent of Samples							
Human	6.4	5.3	19.0	6.8	8.5	6.6	10.3	6.2
Non-Human	93.6	94.7	81.0	93.2	91.5	93.4	89.7	93.8

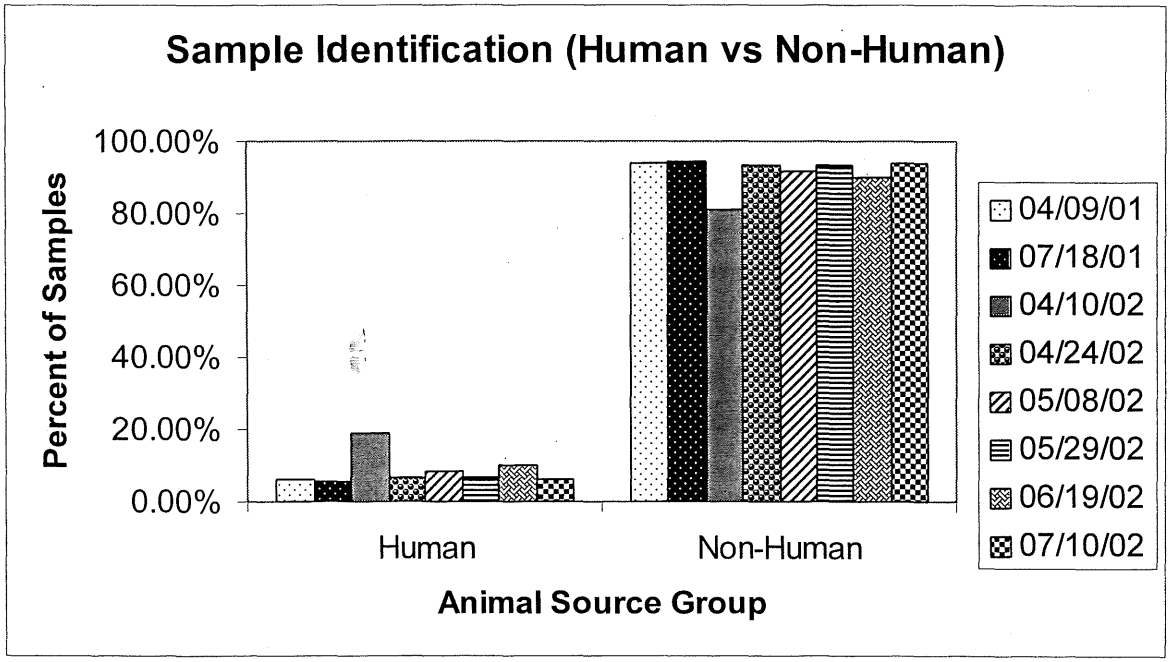


Figure 10. Percent of samples having isolates from Humans and Non-Humans, by Sample Date, obtained from High Island Creek.

Table 18. Probable Identity of *E. coli* isolates from Humans and Non-Humans, by Site, obtained from High Island Creek.

Animal Type	Site										
	1s	**2p**	**3p**	**T**	**5p**	**6s**	**7s**	**8s**	**9p**	**10p**	**11**
	Percent of Samples										
Human	10.8	17.5	3.0	8.0	4.4	5.1	10.1	10.4	11.8	4.4	5.0
Non-Human	89.2	82.5	97.1	92.0	95.6	94.9	89.9	89.6	88.2	95.6	95.0

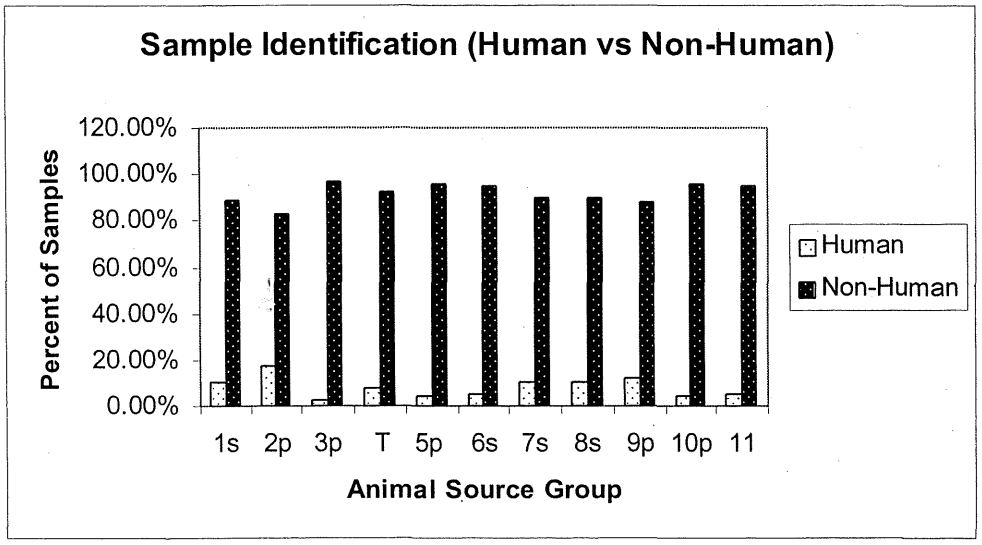


Figure 11. Percent of samples having isolates from Humans and Non-Humans, by Site, obtained from High Island Creek.

ID Bootstrap Analysis of High Island Creek Isolates

While the above analyses show the probable identities of various *E. coli* isolates from the High Island Creek watershed, our initial studies and subsequent analyses done during the SCCWRP project indicated that despite the use of quality factors in our analyses, our library size limitations and the presence of isolates with similar ID values makes definitive classification of isolates somewhat tenuous. To overcome this limitation, we used ID bootstrap software provided by Bionumerics to make more definitive assignment of the unknown isolates. ID boostrapping applies reiterative analysis of the integrity of known source groups and applies the resulting correlation statistic to assign identities to unknown isolates. This analysis reduces the number of false positive results by 42% and false negative results by 2%. Moreover, while the analysis results in us discarding isolates from our final assignments, the isolates that remain are assured (at a $\geq$90% confidence level) more correct classification. Below is the ID bootstrap analysis for High Island Creek Isolates. Given that isolates had to be discarded from consideration due to non-statistical assignment reasons, only analyses by animal group (rather than by date or site) is given. Nevertheless, this gives a better picture of the probable source of isolates present in the watershed.

Table 19. Probable Identity of *E. coli* isolates obtained from High Island Creek, by Animals Group, using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Cat	9	3.9
Chicken	8	3.5
Cow	49	21.3
Deer	24	10.4
Dog	21	9.1
Duck	9	3.9
Goat	4	1.7
Goose	33	14.4
Horse	3	1.3
Human	37	16.1
Pig	3	1.3
Sheep	15	6.5
Turkey	15	6.5
TOTAL	**230**	

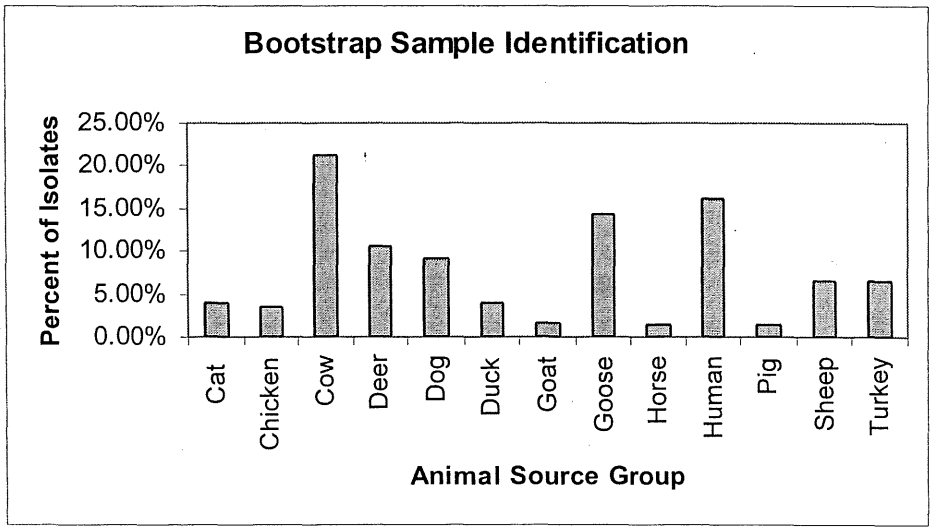


Figure 12. Percent of Isolates from obtained from High Island Creek in Animals Group by using ID Bootstrap Analysis.

Table 20. Probable Identity of *E. coli* Isolates Obtained from High Island Creek, in Domestic, Human, Pet, Waterfowl, and Wild Animal Groups, Using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Domesticated	97	42.2
Human	37	16.1
Pets	30	13.0
Waterfowl	42	18.3
Wild Animal (Deer)	24	10.4
TOTAL	230	

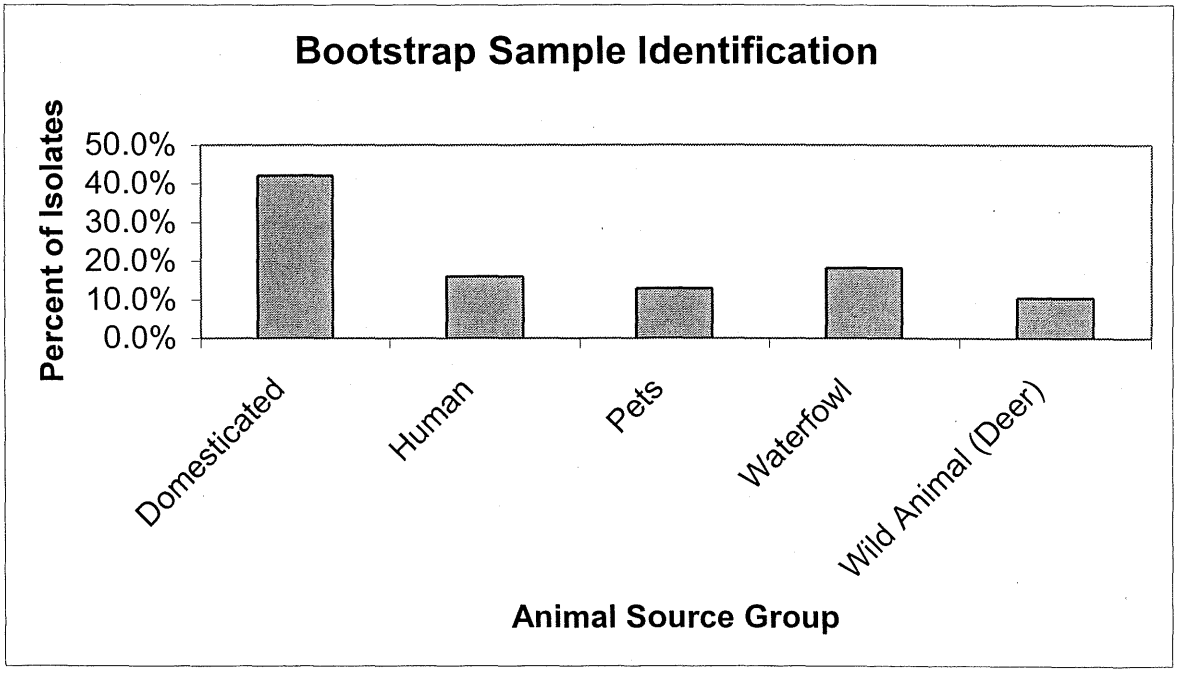


Figure 12. Percent of Isolates from obtained from High Island Creek in Domestic, Human, Pet, Waterfowl, and Wild Animal Groups, using ID Bootstrap Analysis.

Table 21. Probable Identity of *E. coli* isolates from Household and Non-Households obtained from High Island Creek, Using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Household	67	29.1
Non-Household	163	70.9
TOTAL	230	

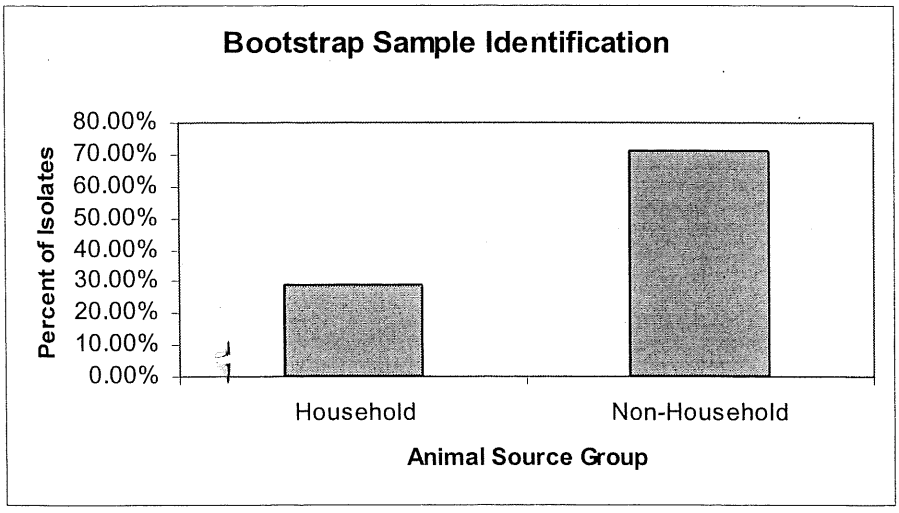


Figure 13. Percent of Isolates Obtained from High Island Creek in Household and Non-Household groups Using ID Bootstrap Analysis.

Table 22. Probable Identity of *E. coli* isolates obtained from High Island Creek in Humans and Non-Humans groups Using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Human	37	16.1
Non-Human	193	83.9
TOTAL	230	

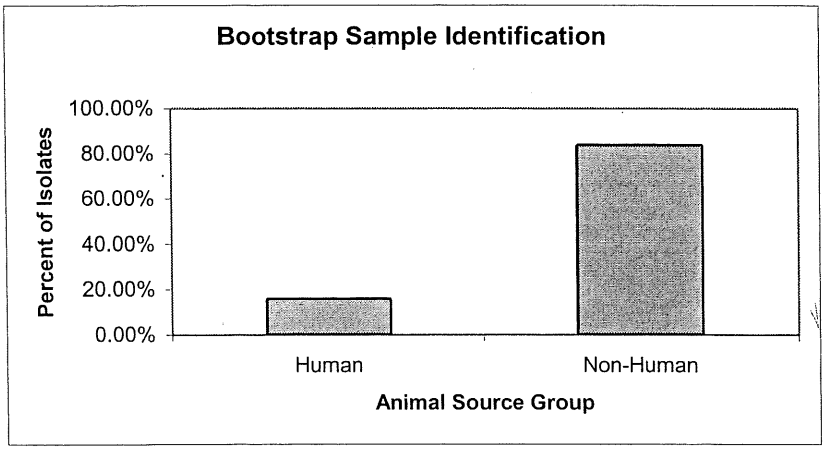


Figure 14. Percent of Isolates Obtained from High Island Creek in Human and Non-Human groups Using ID Bootstrap Analysis.

DISCUSSION OF HIGH ISALND CREEK ANALYSES

Results presented in Tables 11-18 and Figures 4-11 report on our analyses of the identities of watershed isolates obtained the High Island Creek Watershed. In total 1,651 *E. coli* isolates were analyzed. As we indicated above, the most stringent analysis for the potential identification of these isolates comes from ID bootstrap studies (Tables 19-22 and Figures 12-14). Thus while the trends for Pearson's analyses Tables 11-18 are similar, we believe that table data from ID bootstrap analyses more accurately reflect the identity of isolates to a 90% certainty. Consequently, we will discuss these results in more detail. Results in Table 19 show that the majority of isolates were removed from the study using ID bootstrap analysis. However, of the remaining isolates, the majority have cow as their source. Our data however, points to the fact that there is some input into this watershed from Geese (14% of isolates) and humans (16%). Results in Table 20 and Figure 13 show that when we break down the isolates into larger groups, that domesticated animals (chickens, cows, goats, horses, pigs, sheep, and turkeys) contribute 42% of the *E. coli* isolates to this watershed, the remainder mostly being contributed by humans and waterfowl (ducks and geese). Thus, the majority of isolates in the watershed come from non-household sources (Table 21 and Figure 14). On an even larger scale, this is further reflected in Table 22 and Figure 14 which show that over 80% of the *E. coli* isolates in the High Island Creek watershed come from non-human sources.

SEE ATTACHEMNT C FOR ADDITIONAL RAW DATA BY SITE AND DATE

SUMMARY OF VERMILLION RIVER ANALYSIS – Analyses done using Pearson's cosine coefficient analysis with 1% optimization.

We analyzed 1,776 *E. coli* isolates obtained from the Vermillion River Watershed. Sampling took place on 7/11/01, 08/08/01, 09/05/01, 10/03/01, 03/27/02, 05/01/02, 06/05/02, and 07/02/02. Approximately 25 *E.coli* isolates were obtained from each site at each date. The sites sampled were: VMCwest, VMCeast, VMC, VNC175, VNC, VSBtrib, VSB, V31, Vverm, and V47.

Table 23. **Percent Identity of *E. coli* isolates from Separate Animals, by Date, obtained from the Vermillion River watershed.**

Animal Type	Date							
	07/11/01	08/08/01	09/05/01	10/03/01	03/27/02	05/01/02	06/05/02	07/02/02
	Percent of Isolates							
Cat	15.2	11.6	12.4	16.8	3.3	10.8	13.0	11.2
Chicken	0.5	0.0	2.3	3.2	3.3	8.5	1.6	0.6
Cow	7.3	2.9	4.5	7.9	16.7	18.2	10.9	11.8
Deer	14.7	11.6	16.9	8.4	3.9	5.7	5.2	4.1
Dog	5.2	12.1	9.0	6.3	17.2	12.5	4.7	8.3
Duck	5.8	1.7	5.6	2.1	2.2	2.8	1.6	6.5
Goat	0.5	0.0	0.0	1.1	0.6	0.6	1.0	0.0
Goose	14.1	17.3	9.0	17.4	8.3	11.9	22.4	13.6
Horse	1.6	0.0	0.0	0.0	0.0	0.0	2.6	0.6
Human	3.1	13.9	6.2	4.2	21.7	8.0	6.3	4.7
Pig	16.2	13.9	13.5	12.1	11.1	8.0	15.1	7.7
Sheep	5.8	6.4	6.7	14.2	8.3	6.3	9.9	13.6
Turkey	10.0	8.7	14.0	6.3	3.3	6.8	5.7	17.2

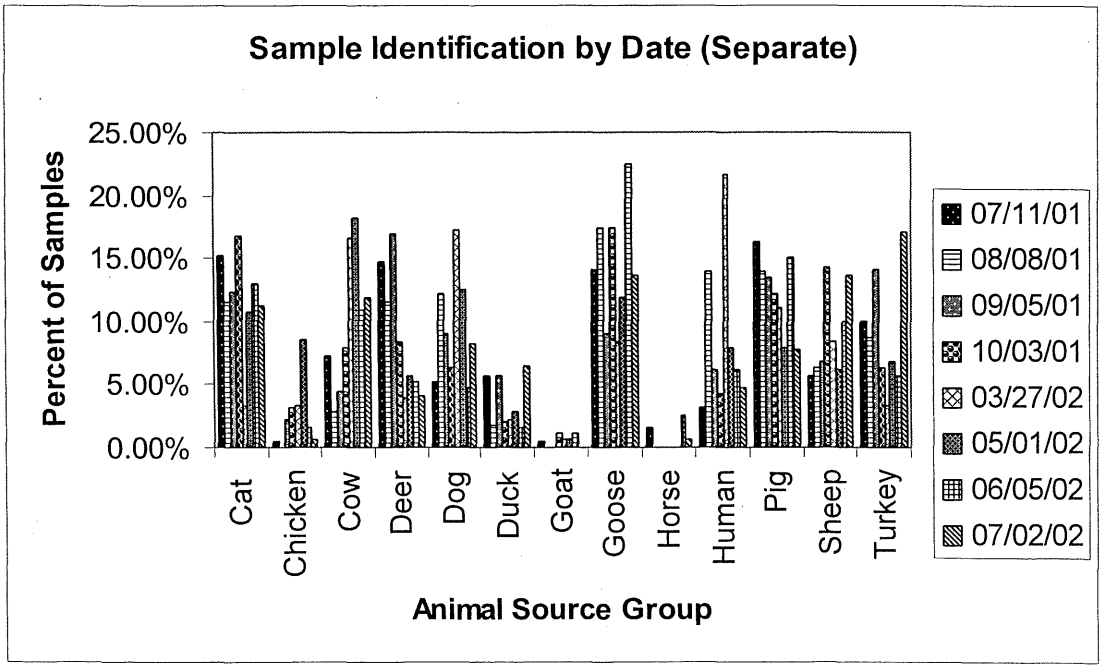


Figure 15. Percent of samples in each animal source group recovered, by Sample date, in the Vermillion River Watershed.

Table 24. Percent Identity of *E. coli* isolates from Separate Animals, by Site, obtained from the Vermillion River watershed.

Animal Type	Site									
	VMC west	VMCeast	VMC	VNC175	VNC	VSBtrib	VSB	V31	Vverm	VMC west
	Percent of Samples									
Cat	15.3	12.1	13.0	10.7	8.2	16.8	8.4	9.0	12.3	13.3
Chicken	0.8	0.0	0.7	0.7	2.2	7.7	5.4	3.8	2.1	1.3
Cow	3.1	4.7	18.8	8.1	4.4	6.3	16.2	7.5	15.8	13.3
Deer	6.1	9.4	16.0	10.7	9.6	9.8	6.6	5.3	6.9	8.2
Dog	9.9	4.0	4.4	13.4	11.9	9.1	9.0	15.8	6.9	9.5
Duck	4.6	2.7	5.1	11.4	0.7	0.7	1.8	2.3	3.4	2.5
Goat	0.0	0.7	0.0	0.7	0.7	0.0	0.6	2.3	0.0	0.0
Goose	9.2	16.1	8.7	10.7	17.0	11.9	16.8	18.8	16.4	17.1
Horse	0.0	0.0	0.7	0.0	0.7	0.7	0.6	3.0	0.7	0.0
Human	11.5	6.0	6.5	9.4	6.7	13.3	6.0	9.0	6.2	10.1
Pig	16.0	16.8	13.8	10.1	17.8	4.9	14.4	7.5	9.6	12.0
Sheep	6.1	7.4	7.3	4.7	11.9	14.0	7.8	9.8	11.0	9.5
Turkey	17.6	20.1	5.1	9.4	8.2	4.9	6.6	6.0	8.9	3.2

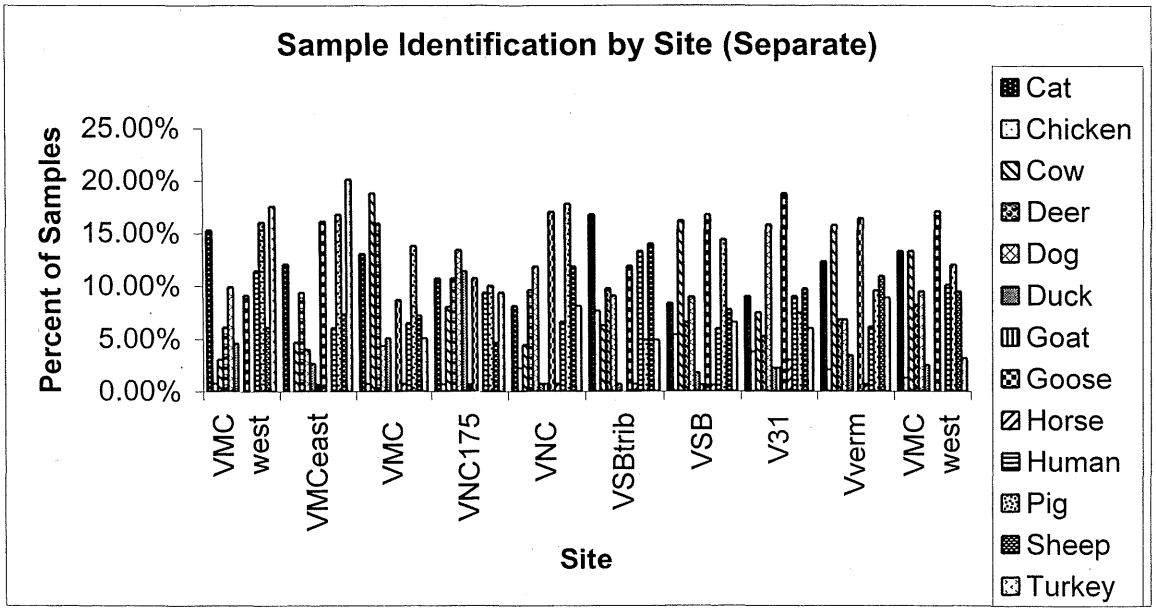


Figure 16. Percent of samples in each animal source group recovered by Site, in the Vermillion River Watershed.

Table 25. Percent of *E. coli* samples obtained from the Vermillion River watershed in Domesticated, Human, Pets, Waterfowl, and Wild animal Source Groups, by Date.

Animal Type	Date							
	07/11/01	08/08/01	09/05/01	10/03/01	03/27/02	05/01/02	06/05/02	07/02/02
	Percent of Samples							
Domesticated	41.9	31.8	41.0	44.7	43.3	48.3	46.9	51.5
Human	3.1	13.9	6.2	4.2	21.7	8.0	6.3	4.7
Pets	20.4	23.7	21.4	23.2	20.6	23.3	17.7	19.5
Waterfowl	19.9	19.1	14.6	19.5	10.6	14.8	24.0	20.1
Wild Animal (Deer)	14.7	11.6	16.9	8.4	3.9	5.7	5.2	4.1

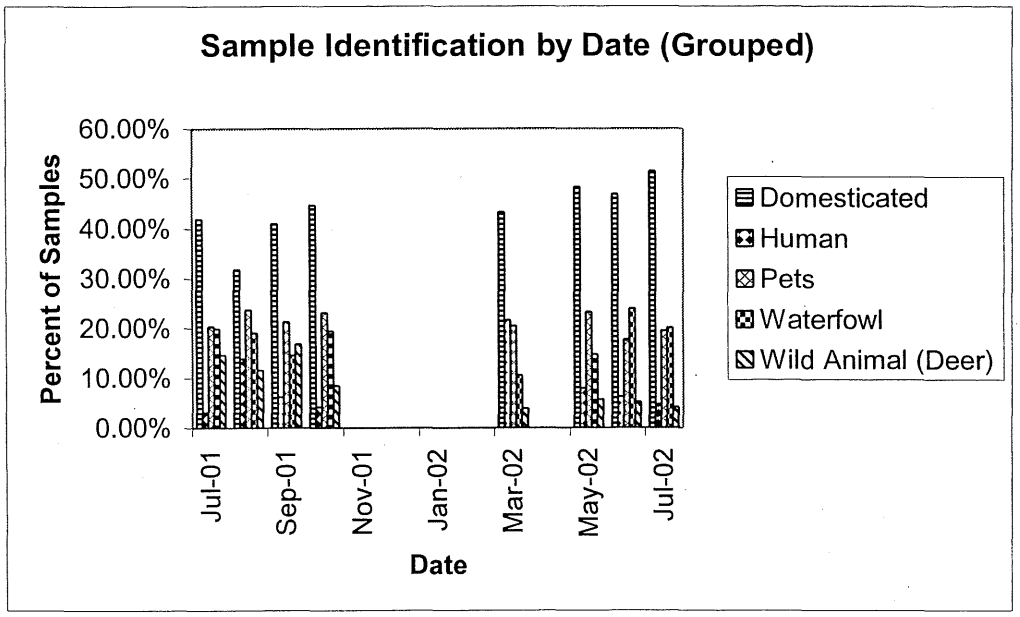


Figure 17. Percent of samples in Domesticated, Human, Pets, Waterfowl, and Wild animal Source Groups recovered by Date, in the Vermillion River Watershed.

Table 26. Percent of samples obtained from the Vermillion River watershed in Domesticated, Human, Pets, Waterfowl, and Wild animal Source Groups, by Site.

	Site									
Animal Type	**VMC west**	**VMCeast**	**VMC**	**VNC175**	**VNC**	**VSBtrib**	**VSB**	**V31**	**Vverm**	**VMC west0**
	Percent of Samples									
Domesticated	43.5	49.7	46.4	33.6	45.9	44.4	51.5	39.9	48.0	39.2
Human	11.5	6.0	6.5	9.4	6.7	15.3	6.0	9.0	6.2	10.1
Pets	25.2	16.1	17.4	24.2	20.0	14.5	17.4	24.8	19.2	22.8
Waterfowl	13.7	18.8	13.8	22.2	17.8	14.5	18.6	21.1	19.9	19.6
Wild Animal (Deer)	6.1	9.4	15.9	10.7	9.6	11.3	6.6	5.3	6.9	8.2

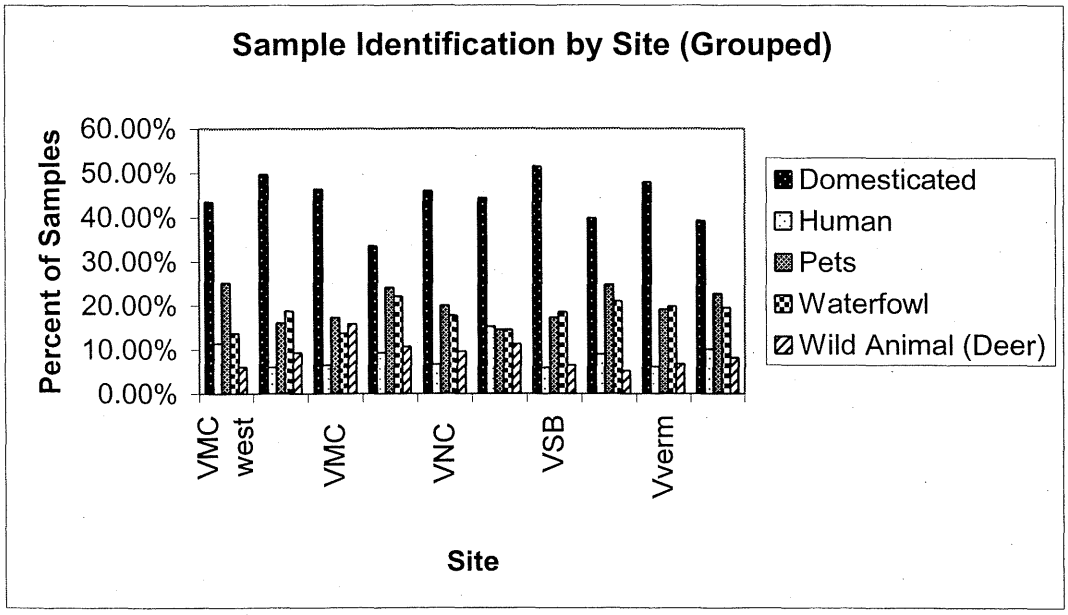


Figure 18. Percent of samples in Domesticated, Human, Pets, Waterfowl, and Wild animal Source Groups recovered by Site, in the Vermillion River Watershed.

Table 23. Percent of samples obtained from the Vermillion River watershed in Household and non-household Source Groups, by Date.

	Date							
Animal Type	**07/11/01**	**08/08/01**	**09/05/01**	**10/03/01**	**03/27/02**	**05/01/02**	**06/05/02**	**07/02/02**
	Percent of Samples							
Household	23.6	37.6	27.5	27.4	42.2	31.3	24.0	24.3
Non-Household	76.4	62.4	72.5	72.6	57.8	68.8	76.0	75.7

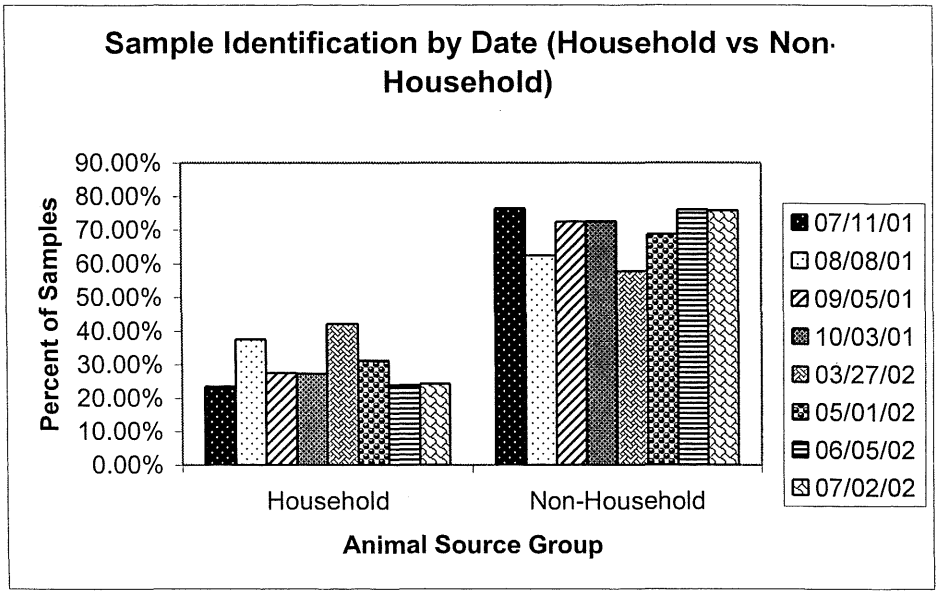


Figure 19. Percent of samples in Household and non-Household Source Groups recovered by Date, in the Vermillion River Watershed.

Table 24. Percent of samples obtained from the Vermillion River watershed in Household and non-household Source Groups, by Site.

	Site									
Animal Type	**VMC west**	**VMCeast**	**VMC**	**VNC175**	**VNC**	**VSBtrib**	**VSB**	**V31**	**Vverm**	**VMC west**
	Percent of Samples									
Household	36.6	22.2	23.9	33.6	26.7	39.2	23.4	33.8	25.3	32.9
Non-Household	63.4	77.9	76.1	66.4	73.3	60.8	76.7	66.2	74.7	67.1

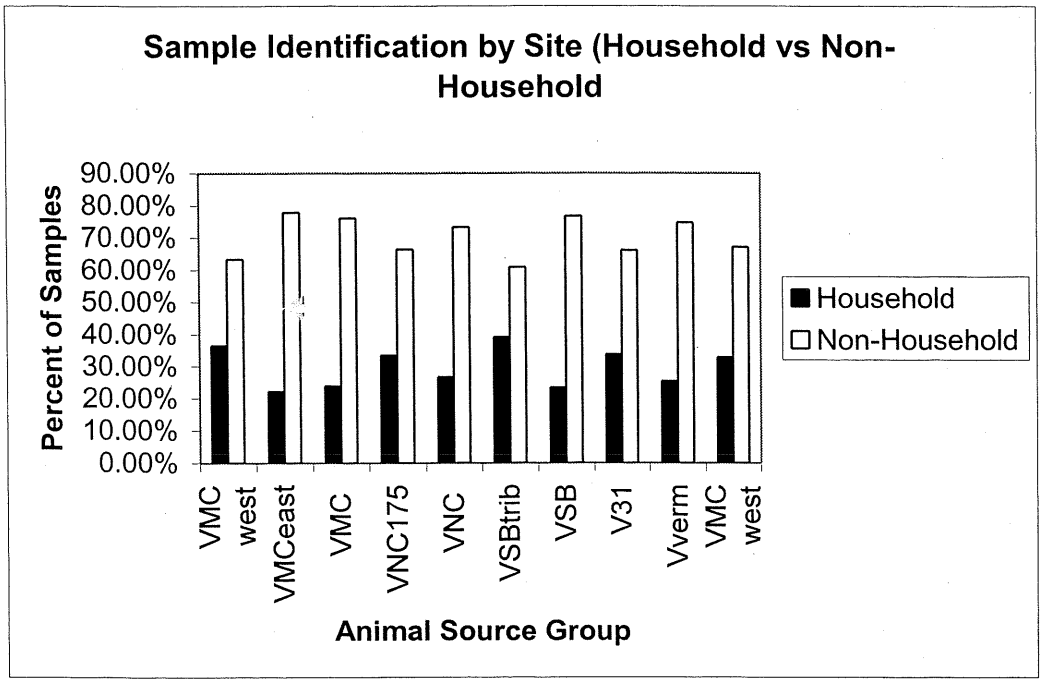


Figure 20. Percent of samples in Household and non-Household Source Groups recovered by Site, in the Vermillion River Watershed

Table 25. Percent of samples obtained from the Vermillion River watershed in Human and Non-Human Source Groups, by Date.

Animal Type	Date							
	07/11/01	08/08/01	09/05/01	10/03/01	03/27/02	05/01/02	06/05/02	07/02/02
	Percent of Samples							
Human	3.1	16.8	6.7	8.1	22.5	11.5	9.6	14.4
Non-Human	96.9	83.2	93.3	91.9	77.5	88.5	90.5	85.6

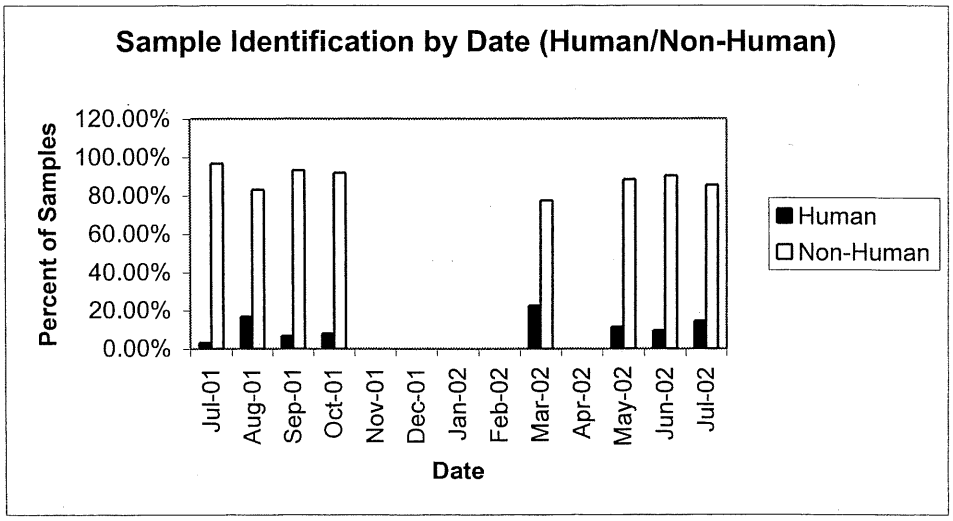


Figure 21. Percent of samples in Human and non-Human Source Groups recovered by Date, in the Vermillion River Watershed

Table 26. Percent of samples obtained from the Vermillion River watershed in Human and Non-Human Source Groups, by Site.

Animal Type	Site									
	VMCwest	VMCeast	VMC	VNC175	VNC	VSBtrib	VSB	V31	Vverm	V47
	Percent of Samples									
Human	20.6	9.7	9.2	14.6	8.7	16.2	7.1	14.8	7.4	11.8
Non-Human	79.5	90.3	90.9	85.4	91.3	83.8	92.9	85.2	92.6	88.2

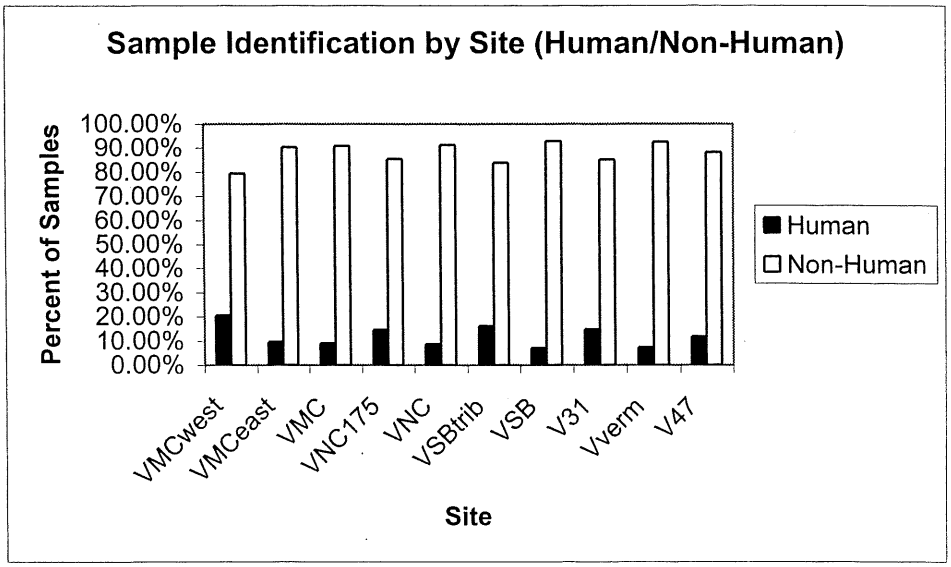


Figure 22. Percent of samples in Human and non-Human Source Groups recovered by Site in the Vermillion River Watershed

BOOTSTRAP ANALYSIS OF VERMILLION RIVER WATERSHED

Table 31. Bootstrap analysis of isolates from the Vermillion River Watershed, percentage and frequency by animals.

Animal Type	Frequency	Percentage
Cat	49	23.0
Chicken	6	2.8
Cow	28	13.2
Deer	41	19.3
Dog	47	22.2
Duck	3	1.4
Goat	4	1.9
Goose	12	5.6
Horse	0	0.0
Human	13	6.1
Pig	0	0.0
Sheep	6	2.8
Turkey	4	1.9
TOTAL	213	

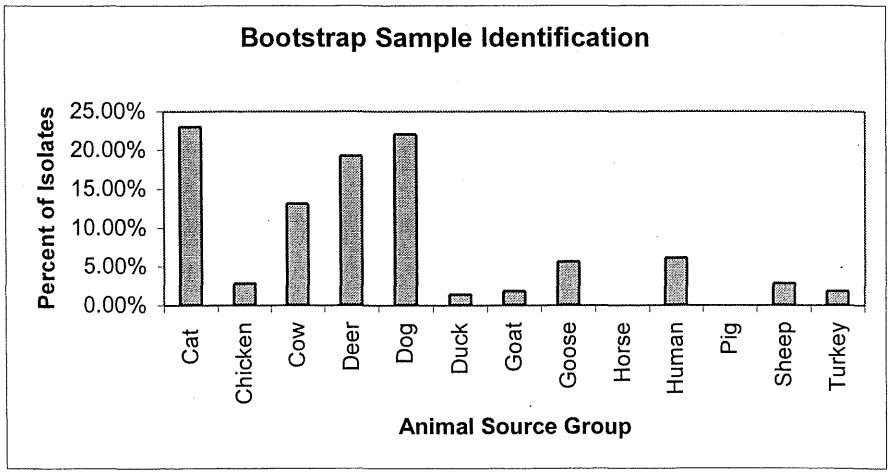


Figure 23. Percent of isolates in Animal Source Groups recovered in the Vermillion River Watershed, ID Bootstrap Analysis.

Table 32. Bootstrap analysis of isolates from the Vermillion River Watershed, percentage and frequency of isolates in domesticated, human, pet, waterfowl, and wild animal source groups.

Animal Type	Frequency	Percentage
Domesticated	48	22.5
Human	13	6.1
Pets	96	45.1
Waterfowl	15	7.0
Wild Animal (Deer)	41	19.3
TOTAL	213	

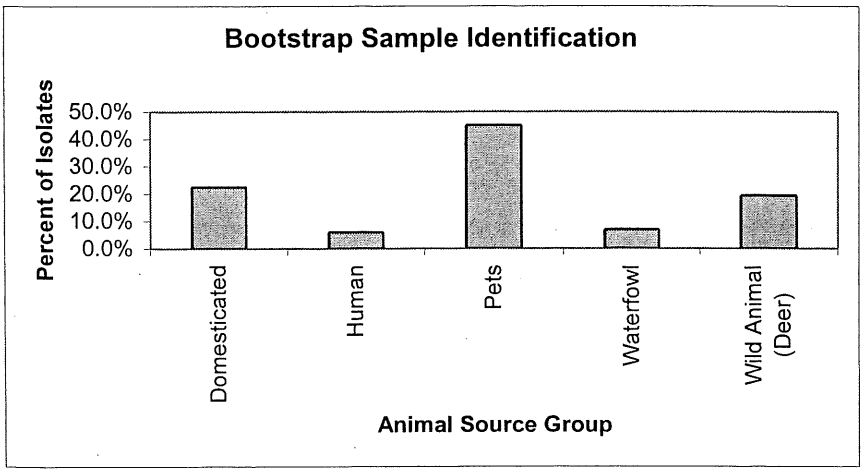


Figure 24. Percent of isolates in domesticated, human, pet, waterfowl, and wild animal source groups in the Vermillion River Watershed, ID Bootstrap Analysis

Table 33. Bootstrap analysis of isolates from the Vermillion River Watershed, percentage and frequency of isolates in Household and Non-household source groups.

Animal Type	Frequency	Percentage
Household	109	51.2
Non-Household	104	48.8
TOTAL	213	

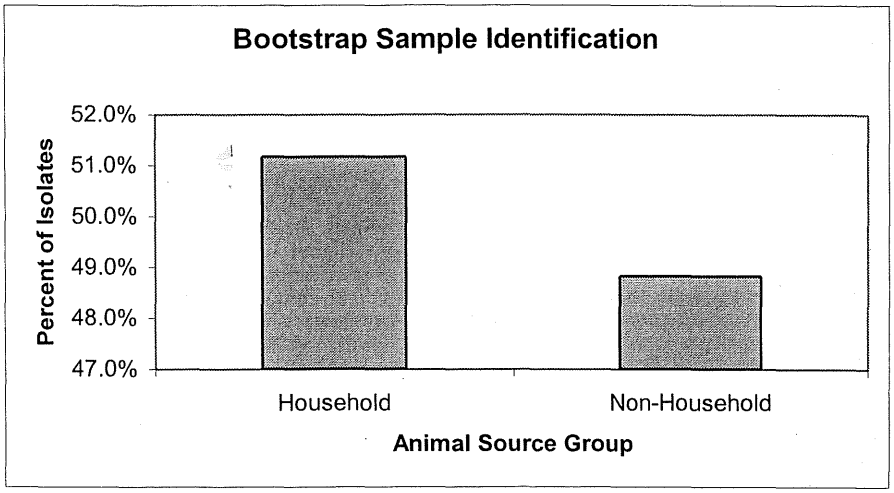


Figure 25. Percent of isolates in Household and Non-Household source groups in the Vermillion River Watershed, ID Bootstrap Analysis

Table 34. Bootstrap analysis of isolates from the Vermillion River Watershed, percentage and frequency of isolates in Human and Non-Human source groups.

Animal Type	Frequency	Percentage
Human	13	6.1
Non-Human	200	93.9
TOTAL	213	

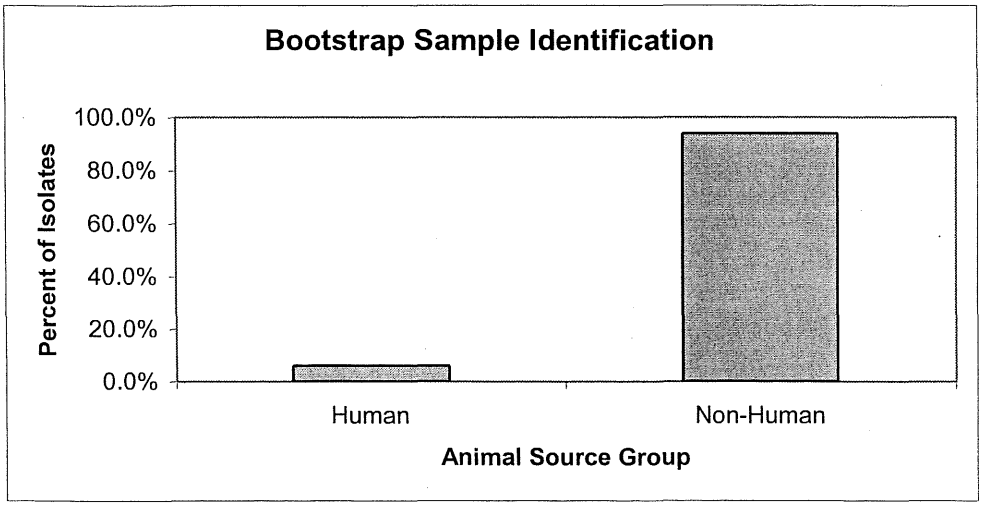


Figure 26. Percent of isolates in Human and Non-Human source groups in the Vermillion River Watershed, ID Bootstrap Analysis.

SEE ATTACHMENT C FOR ADDITIONAL RAW DATA BY SITE AND DATE

DISCUSSION OF VERMILLION RIVER ANALYSES

Results presented in Tables 23-30 and Figures 15-22 report on our analyses of the identities of watershed isolates obtained from the Vermillion River Watershed. In total 1,798 *E. coli* isolates were analyzed. As we indicated above, the most stringent analysis for the potential identification of these isolates comes from ID bootstrap studies (Tables 31-34 and Figures 23-26). Thus while the trends for Pearson's analyses Tables 23-30 are similar, we believe that table data from ID bootstrap analyses more accurately reflect the identity of isolates to a 90% certainty. There was a distinct clustering of data by sample date and site, suggesting that climatic and land use factors affected the origin of isolates present (Figures 17 and 18). However, generally speaking, the sites were dominated by bacteria from domesticated, non-household, non-human animals, regardless of the date or site. Since ID bootstrap gives more reliable interpretation, we will discuss these results in more detail. Results in Table 31 show that the majority of isolates were removed from the study using ID bootstrap analysis. However, of the remaining isolates, the majority have cow and pets as their source. Our data however, points to the fact that there is some input into this watershed from Deer (19% of isolates), but very few humans. As we previously discussed, we are unsure why pets contribute so many isolates in this watershed, but we and others have noted similarities between human and pet bacteria in the past. It may also however be due to the fact that we have a limited number of these isolates to analyze and thus their contribution to bacterial load in the watershed may be overestimated. Nevertheless, results in Table 32 show that 45, 22, and 19% of the isolates come from pets, domesticated animals, and wild animals (deer), respectively. This result is also reflected in Table 33 showing a roughly equal contribution of household and non-household bacteria to this watershed. However, as was presented for the High-Island Creek watershed, the Vermillion River is impacted mostly by non-human sources. Thus, on a larger scale, results in Tables 34 and Figure 26 show that over 93% of the *E. coli*

isolates in the Vermillion River watershed come from non-human sources. This will have surely guide remediation efforts.

SUMMARY OF MINNEOPA CREEK ANALYSIS - Analyses done using Pearson's cosine coefficient analysis with 1% optimization.

We analyzed 1,762 *E. coli* isolates obtained from the Minneopa Creek Watershed. Sampling was done on 4/03/2001, 7/30/2001, 8/15/2001, 9/12/2001, 10/10/2001, 4/18/2002, 5/15/2002, 6/12/2002, and 7/17/2002. Approximately 25 *E.coli* isolates were obtained from each site on each date. The sites sampled were labeled 1-10. Site 1.5 was sampled once only (this was not an official project site).

Table 35. Percent Identity of *E. coli* isolates from Separate Animals, by Date, obtained from the Minneaopa Creek Watershed.

Animal Type	Date								
	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
	Percent Isolates								
Cat	8.5	7.2	10.0	12.7	10.2	5.1	4.6	5.8	2.4
Chicken	4.2	0.6	2.4	2.7	10.2	5.7	6.3	4.2	0.6
Cow	2.8	10.2	5.9	10.7	3.7	5.1	13.1	7.4	28.3
Deer	53.5	30.7	34.1	30.7	22.5	26.8	21.7	31.1	18.7
Dog	2.8	6.6	3.5	4.0	6.4	8.9	13.7	3.7	4.8
Duck	0.0	2.4	1.8	2.7	10.7	1.3	1.7	4.7	2.4
Goat	0.0	2.4	0.6	4.7	0.0	1.9	0.0	0.0	4.8
Goose	1.4	14.5	15.9	7.3	10.2	20.4	12.0	10.0	9.6
Horse	0.0	0.0	0.6	2.0	0.0	0.0	0.0	2.6	0.6
Human	7.0	3.0	8.2	4.7	8.0	7.0	11.4	4.7	3.0
Pig	4.2	4.2	9.4	13.3	67.0	15.3	6.9	15.8	12.7
Sheep	5.6	12.1	2.4	0.7	0.5	0.6	8.0	3.7	7.8
Turkey	9.9	6.0	5.3	4.0	10.7	1.9	0.6	6.3	4.2

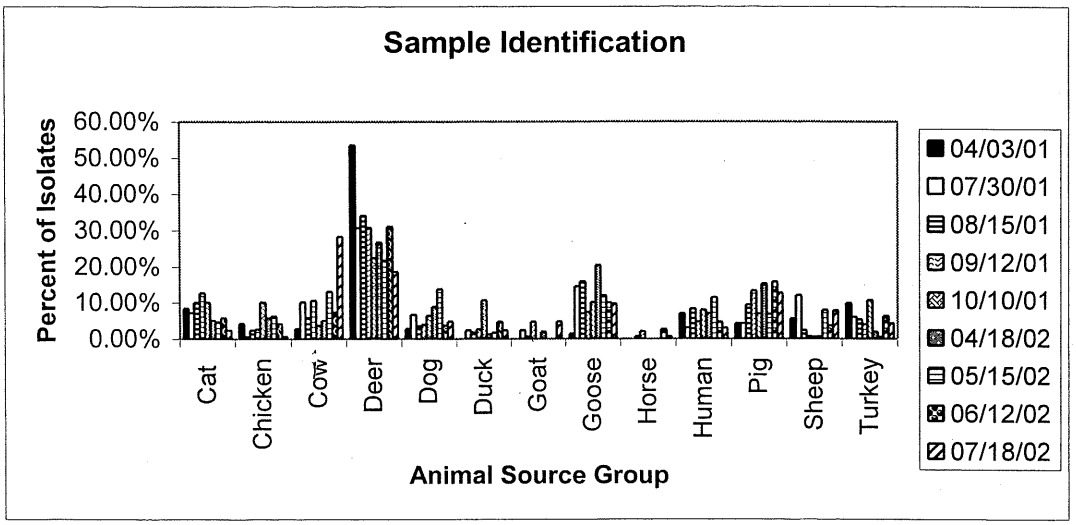


Figure 27. Percent of isolates in All Animal source groups in the Minneopa Creek Watershed, by Date.

Table 36. Percent Identity of *E. coli* isolates from Separate Animals, by Site, obtained from the Minneaopa Creek Watershed.

	Site										
Animal Type	**Site 1**	**Site 2**	**Site 3**	**Site 4**	**Site 5**	**Site 6**	**Site 7**	**Site 8**	**Site 9**	**Site 10**	**Site 1.5**
	Percent Isolates										
Cat	8.9	5.6	10.5	7.6	8.4	1.9	4.7	4.6	9.0	13.5	0.0
Chicken	4.8	4.9	3.0	8.5	2.4	0.0	5.4	0.0	3.0	10.3	16.7
Cow	11.3	7.6	5.3	10.4	10.8	11.5	8.1	12.9	10.5	13.5	0.0
Deer	33.3	32.6	24.1	23.6	27.1	26.1	36.9	28.0	21.1	23.8	50.0
Dog	4.2	5.6	9.0	9.4	12.6	7.6	3.4	3.8	3.0	4.8	0.0
Duck	1.8	2.1	5.3	0.9	1.2	12.1	4.7	1.5	0.8	3.2	0.0
Goat	0.6	2.1	3.0	0.0	7.2	0.0	0.7	0.8	0.8	0.0	0.0
Goose	13.7	9.7	13.5	7.6	7.2	10.2	13.4	18.9	16.5	9.5	0.0
Horse	0.6	0.7	0.8	0.0	1.2	0.0	2.0	0.8	0.8	0.0	0.0
Human	2.4	11.8	6.8	12.3	4.2	8.9	4.7	3.8	9.0	0.8	11.1
Pig	6.0	6.9	7.5	7.6	10.2	14.7	6.7	14.4	18.1	11.1	5.6
Sheep	4.8	5.6	4.5	4.7	4.2	3.2	4.0	6.8	1.5	4.8	16.7
Turkey	7.7	4.9	6.8	7.6	3.0	3.8	5.4	3.8	6.0	4.8	0.0

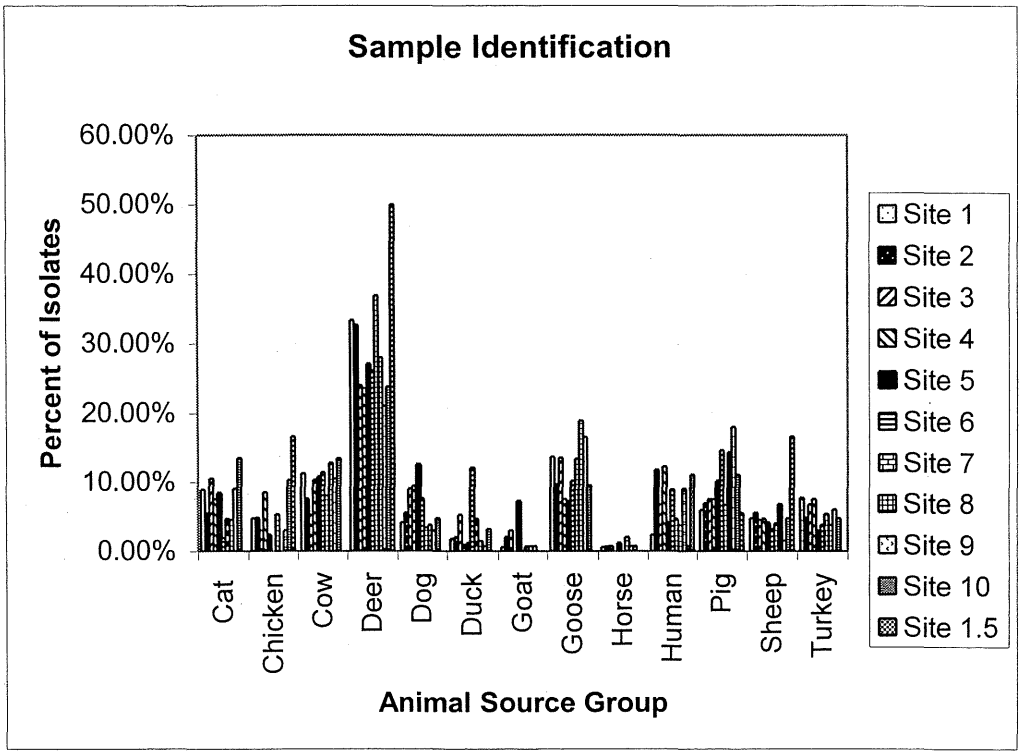


Figure 28. Percent of isolates in All Animal source groups in the Minneopa Creek Watershed, by Site.

Table 37. Percent Identity of *E. coli* isolates from Domesticated, Human, Pets, Waterfowl, and Wild-Animal Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Animal Type	Date								
	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
	Percent of Isolates								
Domesticated	26.8	35.5	26.5	41.1	32.1	30.6	34.9	40.0	59.0
Human	7.0	3.0	8.2	4.4	8.0	7.0	11.4	4.7	3.0
Pets	11.3	13.9	13.5	15.8	16.6	14.0	18.3	9.5	7.2
Waterfowl	1.4	16.9	17.7	9.5	20.9	21.7	13.7	14.7	12.1
Wild Animal (Deer)	53.5	30.7	34.1	29.1	22.5	26.8	21.7	31.1	18.7

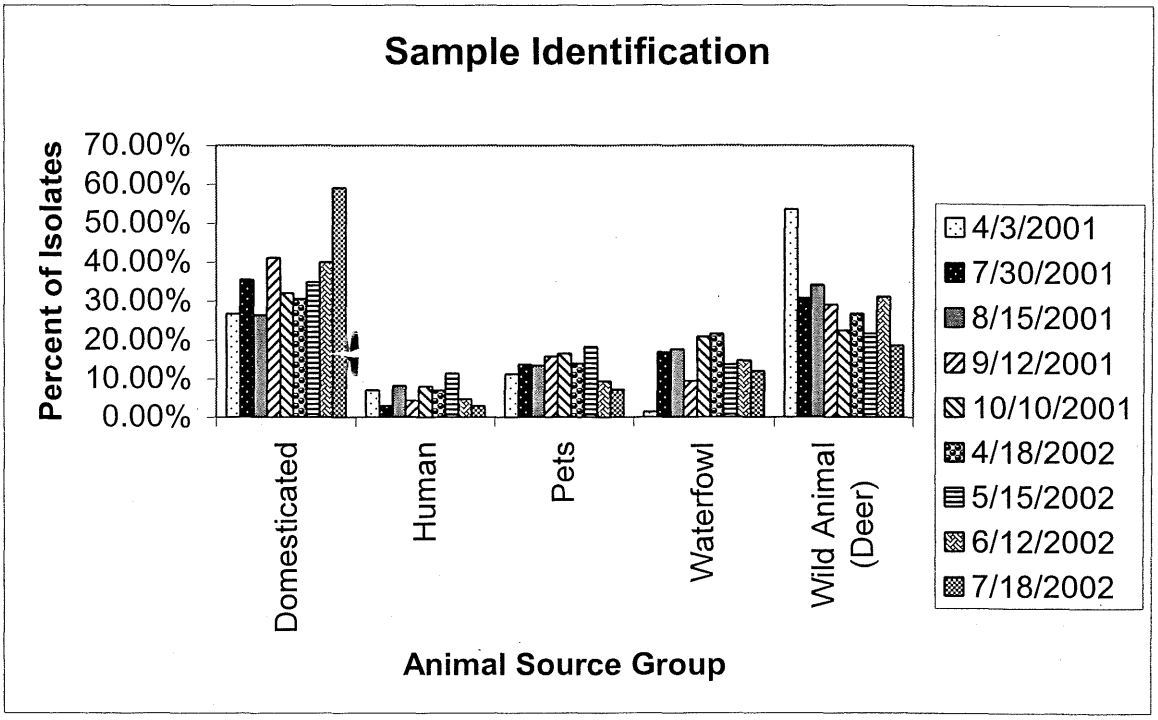


Figure 29. Percent of *E. coli* isolates in Domesticated, Human, Pets, Waterfowl, and Wild-Animal Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Table 38. Percent Identity of *E. coli* isolates in Domesticated, Human, Pets, Waterfowl, and Wild-Animal Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

Animal Type	Site										
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 1.5
	Percent Isolates										
Domesticated	35.7	32.6	30.8	38.7	39.2	33.1	32.2	39.4	40.6	44.4	38.9
Human	2.4	11.8	6.8	12.3	4.2	8.9	4.7	3.8	9.0	0.8	11.1
Pets	13.1	11.1	19.6	17.0	21.1	9.6	8.1	8.3	12.0	18.3	0.0
Waterfowl	15.5	11.8	18.8	8.5	8.4	22.3	18.1	20.5	17.3	12.7	0.0
Wild Animal (Deer)	33.3	32.6	24.1	23.6	27.1	26.1	36.9	28.0	21.1	23.8	50.0

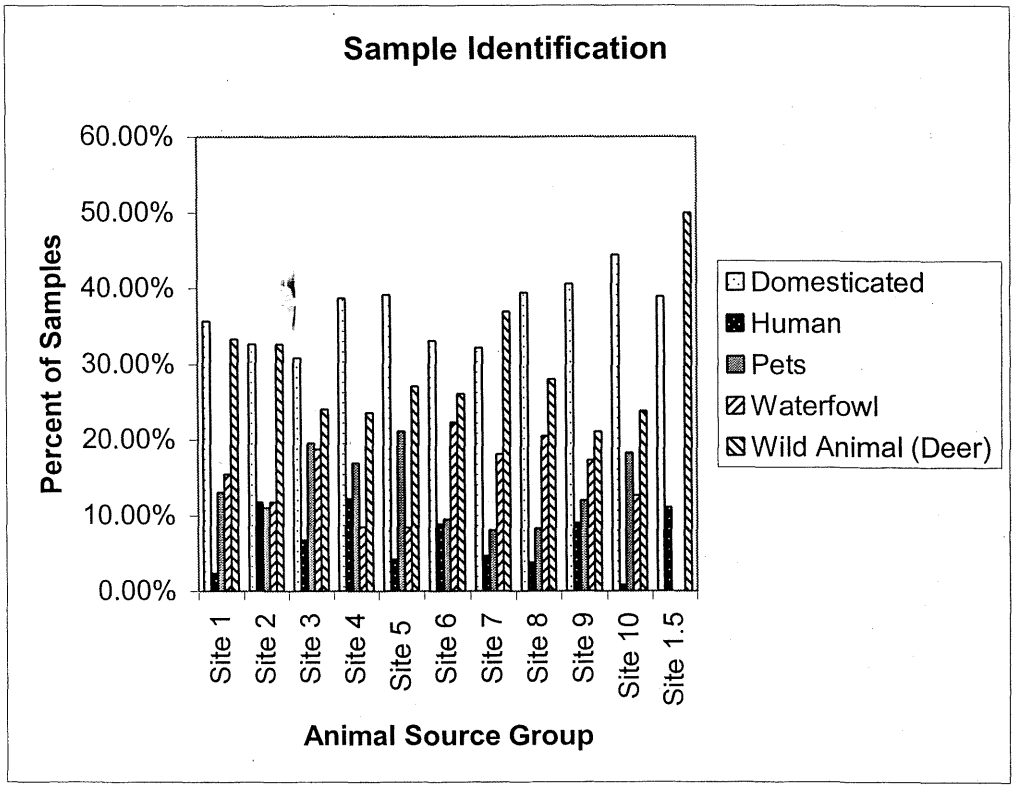


Figure 30. Percent of samples in Domesticated, Human, Pets, Waterfowl, and Wild-Animal Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

Table 39. Percent Identity of *E. coli* isolates in Household and Non-Household Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Animal Type	Date								
	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
	Percent Isolates								
Household	18.3	16.9	21.8	21.3	24.6	21.0	29.7	14.2	10.2
Non-Household	81.7	83.1	78.2	78.7	75.4	79.0	70.3	86.0	90.0

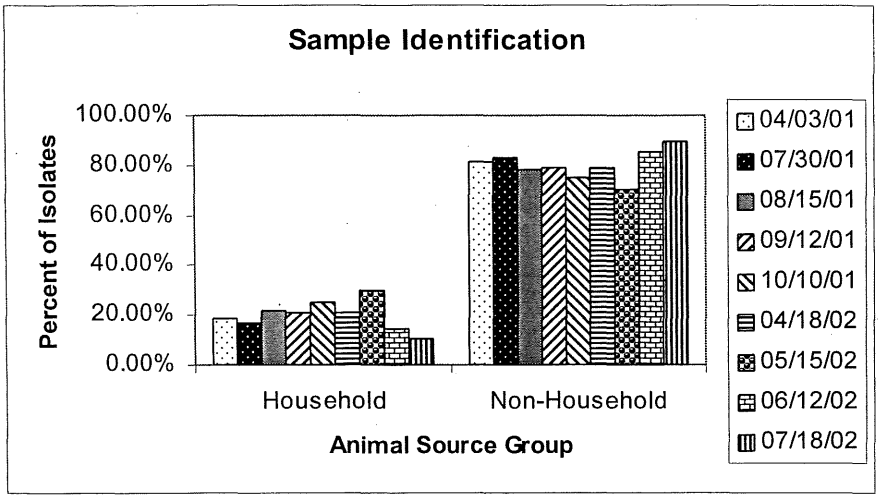


Figure 31. Percent of *E. coli* isolates in Household and Non-Household Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Table 40. Percent Identity of *E. coli* isolates in Household and Non-Household Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

Animal Type	Site										
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 1.5
	Percent Isolates										
Household	15.5	22.9	26.3	29.3	25.3	18.5	12.8	12.1	21.1	19.1	11.1
Non-Household	84.5	77.1	73.7	70.8	74.7	81.5	87.3	87.9	79.0	81.0	89.0

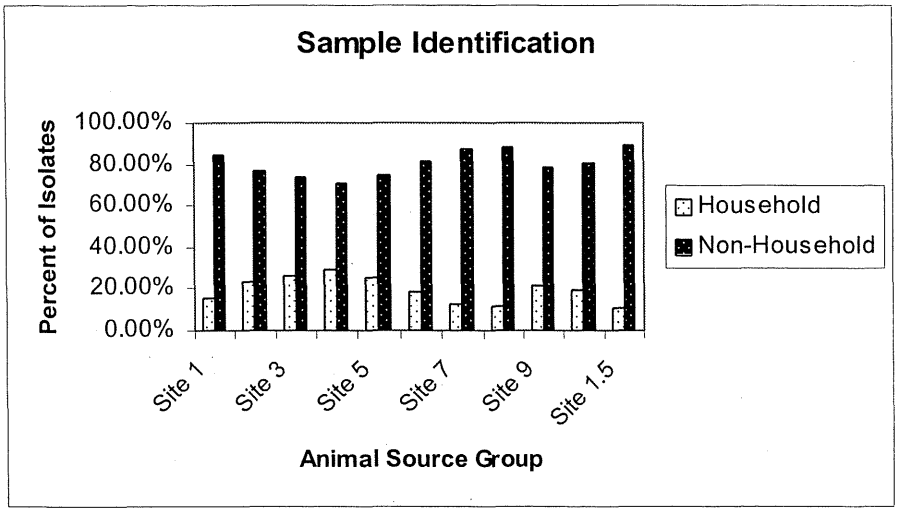


Figure 32. Percent of *E. coli* isolates in Household and Non-Household Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

Table 41. Percent Identity of *E. coli* isolates in Human and Non-Human Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Animal Type	Date								
	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
	Percent Isolates								
Human	7.0	3.0	8.2	4.7	8.0	7.0	11.4	4.7	3.0
Non-Human	93.0	97.0	92.0	95.3	92.0	93.0	88.6	95.3	97.0

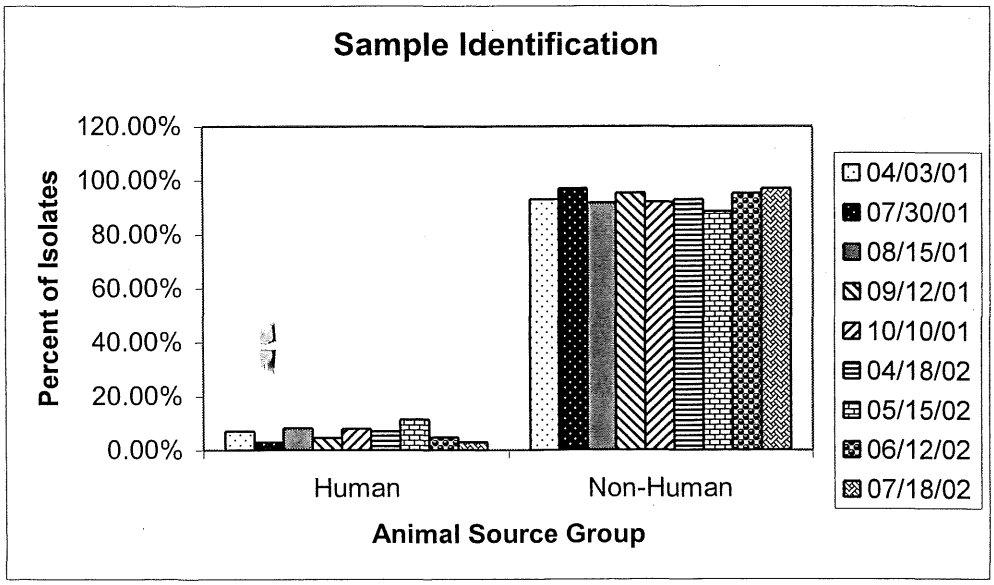


Figure 33. Percent of *E. coli* isolates in Human and Non-Human Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Table 42. Percent Identity of *E. coli* isolates in Human and Non-Human Source Groups, by Site, obtained from the Minneaopa Creek Watershed

Animal Type	Site										
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 1.5
	Percent Isolates										
Human	2.4	11.8	6.8	12.3	4.2	8.9	4.7	3.8	9.0	0.8	11.1
Non-Human	97.6	88.2	93.2	87.7	95.8	91.1	95.3	96.2	91.0	99.2	88.9

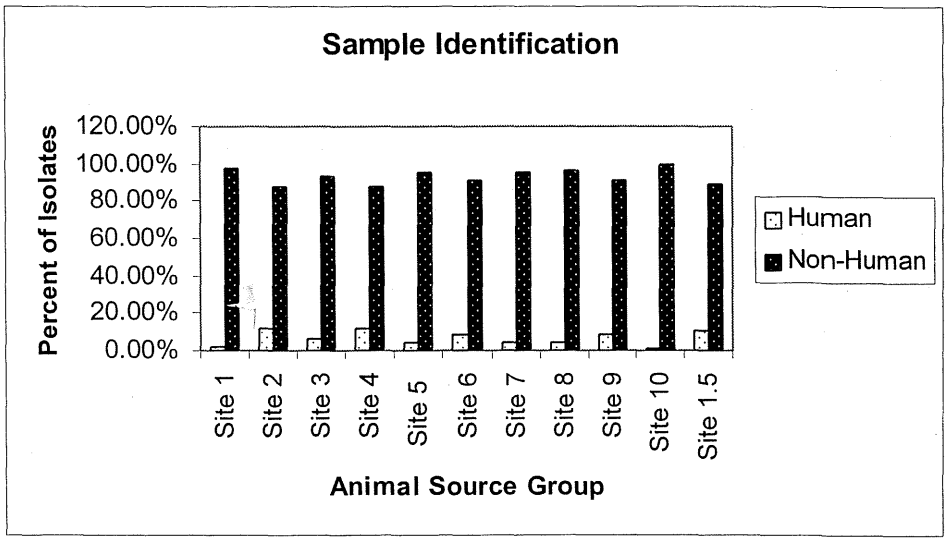


Figure 34. Percent of *E. coli* isolates in Human and Non-Human Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

SEE ATTACHMENT C FOR ADDITIONAL RAW DATA BY SITE AND DATE

BOOTSTRAP ANALYSIS OF MINNEOPA CREEK WATERSHED

Table 43. Bootstrap analysis of isolates from the Minneopa Creek Watershed, percentage and frequency by animals.

Animal Type	Frequency	Percentage
Cat	36	19.4
Chicken	4	2.2
Cow	7	3.8
Deer	40	21.5
Dog	31	16.7
Duck	1	0.5
Goat	14	7.5
Goose	17	9.1
Horse	1	0.5
Human	18	9.7
Pig	2	1.1
Sheep	11	5.9
Turkey	4	2.2
TOTAL	186	

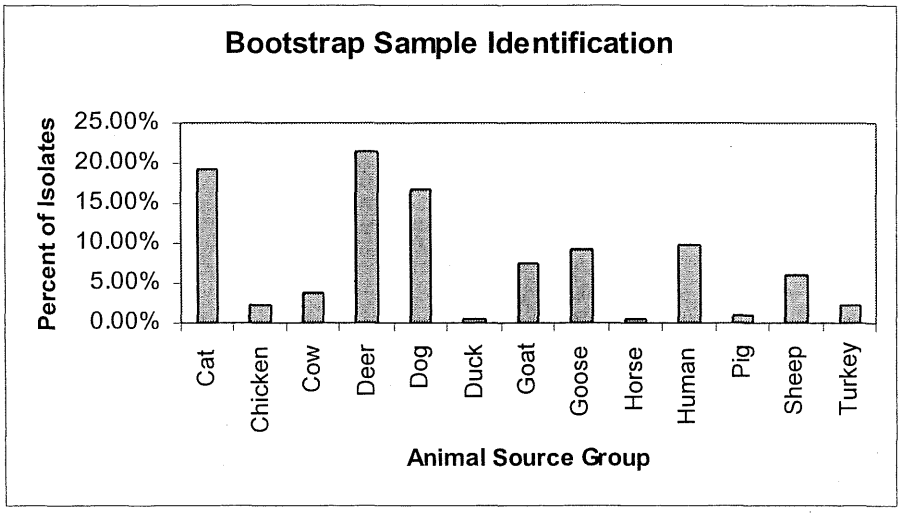


Figure 35. Percent of isolates in all Animal Source Groups recovered in the Minneopa Creek Watershed by ID Bootstrap Analysis.

Table 44. Probable Identity of *E. coli* isolates from Domestic, Human, Pets, Waterfowl, and Wild Animal Groups in the Minneopa Creek Watershed using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Domestic	43	23.1
Human	18	9.7
Pets	67	36.0
Waterfowl	18	9.7
Wild Animal (Deer)	40	21.5
TOTAL	186	

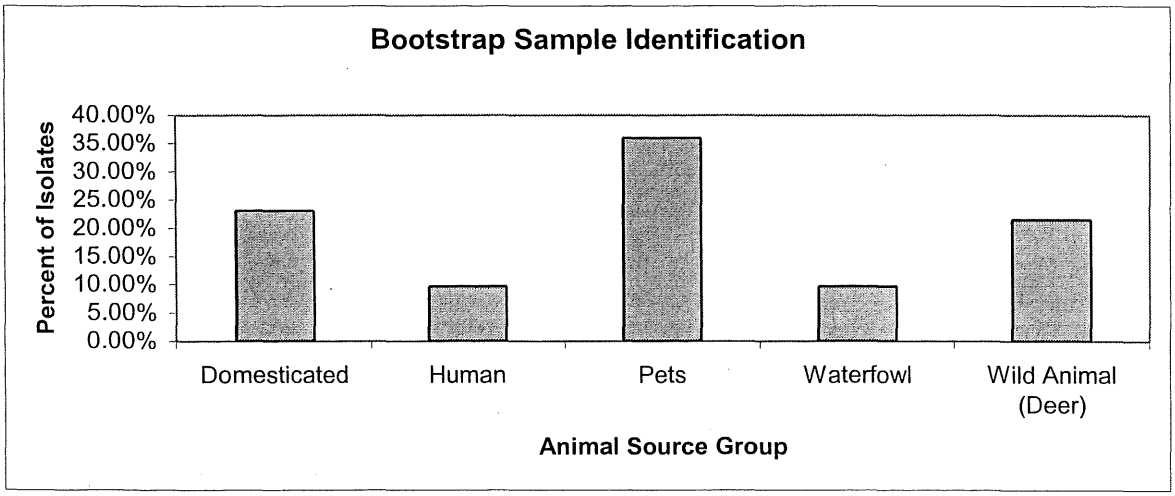


Figure 36. Percent of isolates in domesticated, human, pet, waterfowl, and wild animal source groups in the Minneopa Creek Watershed using ID Bootstrap Analysis

Table 45. Frequency and percentage of *E. coli* isolates from Household and Non-Household Groups in the Minneopa Creek Watershed using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Household	85	45.7
Non-Household	101	54.3
TOTAL	186	

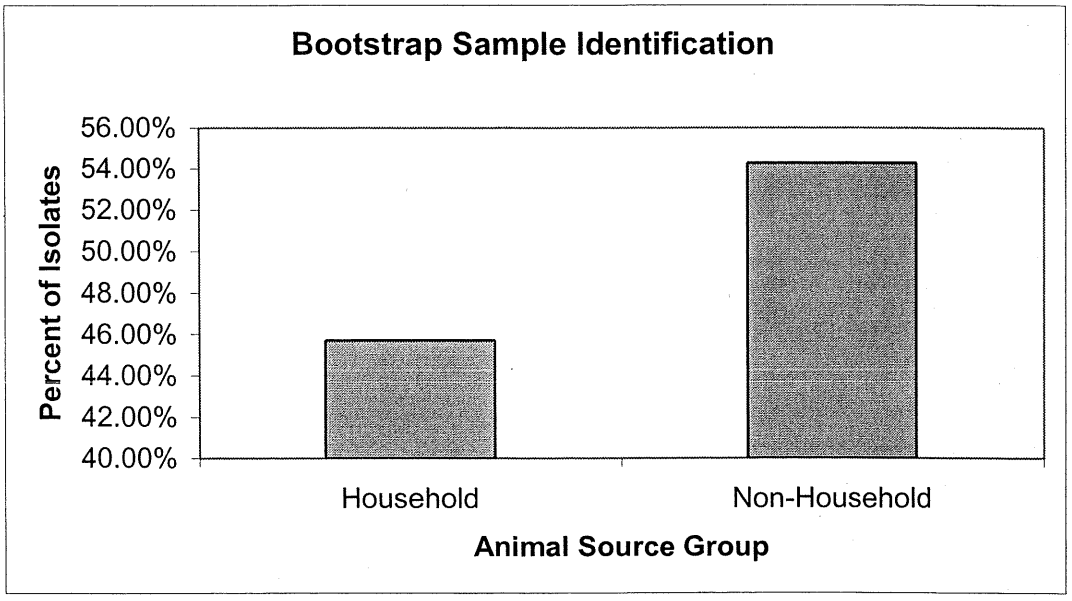


Figure 37. Percent of isolates in Household and Non-Household source groups in the Minneopa Creek Watershed using ID Bootstrap Analysis.

Table 46. Frequency and percentage of *E. coli* isolates from Human and Non-Human source groups in the Minneopa Creek Watershed using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Human	18	9.7
Non-Human	168	90.3
TOTAL	186	

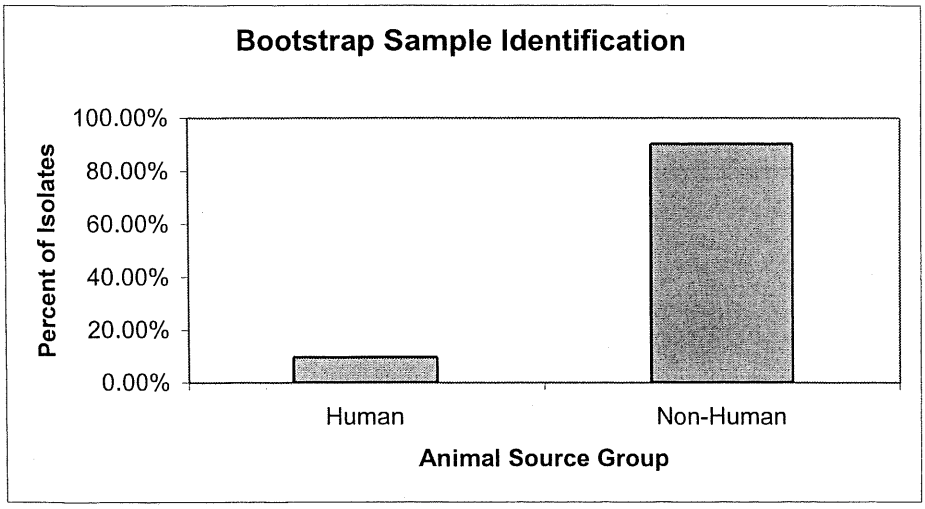


Figure 38. Percent of isolates in Human and Non-Human source groups in the Minneopa Creek Watershed using ID Bootstrap Analysis

SEE ATTACHMENT C FOR ADDITIONAL RAW DATA BY SITE AND DATE

DISCUSSION OF MINNEOPA CREEK WATERSHED ANALYSES

Results presented in Tables 35-42 and Figures 27-34 report on our analyses of the identities of watershed isolates obtained from the Minneopa Creek Watershed. In total 1,762 *E. coli* isolates were analyzed. As we indicated above, the most stringent analysis for the potential identification of these isolates comes from ID bootstrap studies (Tables 44-46 and Figures 35-38). Thus while the trends for Pearson's analyses Tables 35-42 are similar, we believe that table data from ID bootstrap analyses more accurately reflect the identity of isolates to a 90% certainty. As before, there was a distinct clustering of data by sample date and site, suggesting that climatic and land use factors affected the origin of isolates present. However, generally there was a dominance of deer and perhaps geese *E. coli* in the watershed across all sites and dates (Tables 35 and 36, Figures 27 and 28). However, generally speaking, the sites were dominated by bacteria from domesticated, non-household, non-human animals, regardless of the date or site. Since ID bootstrap gives more reliable interpretation, we will discuss these results in more detail. Results in Table 43 show that the majority of isolates were removed from the study using ID bootstrap analysis. However, of the remaining isolates, the majority have *E. coli* from deer and pets as their source. Our data however, points to the fact that there is some lower input into this watershed from human and geese (about 9% of isolates), but very few from other animals. As we previously discussed, we are unsure why pets contribute so many isolates in this watershed, but we, and others, have noted similarities between human and pet bacteria in the past. It may also however be due to the fact that we have a limited number of these isolates to analyze and thus their contribution to bacterial load in the watershed may be overestimated. Nevertheless, results in Table 44 and Figure 36 show that 36, 23, and 21% of the isolates come from pets, domesticated animals, and wild animals (deer), respectively. This is similar to what was seen in the Vermillion Rover watershed. Our overall results are also reflected in Table 45 showing a roughly equal contribution of household and non-household bacteria to this watershed. However,

as was presented for the High-Island Creek and Vermillion River Watersheds, the Minneopa Creek watershed is impacted mostly by non-human and non-household sources. Results in Tables 45 and 46 and Figure 37 and 38 show that over 90% of the *E. coli* isolates in the Minneopa Creek Watershed come from non-human sources.

Result 3: Dissemination and Implementation of Results.

Results from this project have been disseminated in reports made to the LCMR, in periodic update reports made to cooperators, in seminars given throughout the state, nationally and internationally, and in scientific publications in peer-reviewed journals. In addition, results from our studies will be posted and updated on the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) which is housed on computers at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our previous LCMR projects. Data obtained from our studies will be utilized by cooperating agencies to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas.

Our 1999 LCMR project generated a great deal of interest among water resources management and pollution control professionals, farm organizations, scientific researchers, and citizen groups throughout Minnesota and elsewhere. As a result we were invited to present information and findings of our research project at several State and local government-sponsored conferences and at national meetings. In all cases, we were able to accommodate the requests for presentations. The 2001 LCMR project generated a similar level of interest, and we were invited to give several presentations of our research findings.

	LCMR Budget:	**$30,247**	
	Balance:	~~**$1,851**~~	**$875**
Personnel:	~~$27, 547~~ $28,288	(Sr. Scientist /Assistant Scientist [20%])	
Software:	~~$1,000~~ $495		
Publication Costs:	$1,000		
Local Travel:	~~$700~~ $464	(2 trips @ $200 mileage, food & lodging; 3 trips @ $100 mileage)	

Total: $30,247

Completion Date: June 30, 2004

Dissemination activities include:

- 8/16/01 - Newspaper article in Farmington Independent, entitled "Source of Pollution in Vermillion River Sought by University".

- 9/5/01 - "Environmental Journal" television series segment recorded, with subsequent airing of "Bacteria Busters" on Cable TV.

- 9/13/01 - Presentation to Metropolitan Council Environmental Services Environmental Planning and Evaluation managers, St. Paul, MN.

- 9/27/01 - Presentation to Sibley County Commissioners, Soil and Water Conservation District staff, and Environmental Services staff, Gaylord, MN.

- 12/18/01 - Presentation to Rice County Extension Service staff, DNR staff, Rice County SWCD staff, and Rice County Commissioners and township officials, Faribault, MN.

- Maintain the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) which is housed on computers at the University of Minnesota, Department of Soil, Water, and Climate. The Website specific for this project was developed as part of our previous LCMR project and will be updated through this project period.

- 2/5-7/02 – Invited speaker and participant at US EPA and Southern California Coastal Water Research Project (SCCWRP)-sponsored workshop: "Microbiological Source Tracking Workshop," Irvine, CA

- 4/17/02 – Presentation at Minnesota Water 2002 Conference, St. Cloud, MN

- 5/3/02 – Presentation at Minnesota Environmental Health Association Annual Spring Conference, Nisswa, MN

- 5/8/02 – Presentation to University of Minnesota Extension Natural Resources Planning Group, St. Paul, MN

- 5/21/02 – Invited speaker at American Society for Microbiology General Meeting symposium entitled "Development and Application of Methods to Identify Sources of Fecal Pollution in Water," Salt Lake City, UT

- 5/22/02 - Convened Colloquium at American Society for Microbiology General Meeting entitled "Tracking Sources and Sinks of Microorganisms", Salt Lake City, UT

- 6/20/02 – Presentation at American Farm Bureau Federation Watershed Heroes Conference, St. Peter, MN

- 7/267/02 – MPH thesis seminar and defense, University of Minnesota School of Public Health, Minneapolis, MN

- 10/14/02 – Presentation at Minnesota Department of Agriculture Water Quality Seminar, St. Paul, MN

- 11/5/02 – Invited speaker at Seoul National University entitled "The Use of Molecular Methods to Track Sources and Sinks of Microorganisms in the Environment", Seoul Korea.

- 11/6/02 – Invited speaker at Kwangju Institute of Science and Technology entitled "The Use of Molecular Methods to Track Sources and Sinks of Microorganisms in the Environment", Gwangju, Korea

- 11/7/02 – Invited speaker at Yunsei University entitled "The Use of Molecular Methods to Track Sources and Sinks of Microorganisms in the Environment", Seoul, Korea

- 11/16/02 – Invited speaker at the Water Environment Federation 2002 National TMDL Science and Policy Conference, Phoenix, AZ

- 2/19-2/21/03 – Invited speaker and participant at US EPA and Southern California Coastal Water Research Project (SCCWRP)-sponsored workshop: "Microbiological Source Tracking Workshop," Irvine, CA

- 6/19-6/20/03 - Invited participant in Health Canada workshop on Microbial Source Tracking, Ottawa, Ontario, Canada

- 8/7/03 – Presentation and planning meeting with Vermillion River collaborator, St. Paul, MN.

- 8/8/03 – Presentation and planning meeting with High Island collaborators and MNPCA, St. Paul, MN.

- 9/19/03- Presentation and meeting with Vermillion River collaborator, St. Paul, MN.

- 3/23/04 – Invited participant and speaker at Minnesota Water 2004, Minneapolis, MN.

- 6/15/04 – Invited speaker at Nara Institute of Advanced Technology, "The Use of DNA Fingerprinting Technologies to Determine Sources and Sinks of Bacteria in the Environment", Nara, Japan.

- 08/11/04 – Presentation at Minnesota Department of Agriculture Water Quality Workshop, Rochester, Minnesota.

Formal Review of Project

The methods, approach, and overall impact of this project have been formally peer reviewed as part of documentation provided to Sea Grant for three research proposals and to U.S. EPA for a research proposal/contract. These projects are:

Title: Identifying Sources of Fecal Coliform Bacteria in Coastal Ecosystems and Their Relationship to Land Use", Sea Grant, 01/01/01 - 01/31/04.

Title: Sources and Impacts of "Naturalized" Escherichia coli in Coastal Environments", Sea Grant, 02/01/03 - 01/31/05.

Title: "Comparative Evaluation of Microbiological Source Tracking Techniques: rep-per DNA Fingerprinting", U.S. EPA, 03/03/03 - 03/02/04.

Title: "Seasonal Variation in Sources of Escherichia coli Fecal Bacteria Contributing to Beach Closures", Submitted to Sea Grant, 4/2004 for consideration of funding.

V. TOTAL PROJECT BUDGET:

All Results: Personnel:	~~**$228,600**~~ **$230,121**
All Results: Supplies:	~~**$30,100**~~ **$29,531**
All Results: Computer & Software:	~~**$8,300**~~ **$7,795**
All Results: Publication Costs	**$1,000**
All Results: Contracted Services:	~~**$4,800**~~ **$4,589**
All Results: Local Travel:	~~**$2,200**~~ **$1,964**
TOTAL BUDGET:	**$275,000**

Budget Details: See Attachment A

1. A requested funds shifts between results was approved on July 31, 2003. Those changes included the following:

- Transfer of $17,000 from personnel in result 1 to personnel in result 2.
- Transfer of $500 office supplies, $600 mileage, and $2,000 travel expenses, from result 1, to personnel in result 2.
- Transfer of $800 from lab supplies in result 1 to lab supplies in result 2.
- Transfer of $4,747 from personnel in result 2 to personnel in result 3.

2. On January 12, 2004 we requested permission to shift funds within results to pay for severance and vacation buy-out of personnel. An explanation of this request for shifting funds in indicated below:

As the majority of work on the current project was terminated on December 31, 2003, I ended the employment of two employees, Mary Brown and Ethan Carruthers. As mandated by University policy, I was required to pay severance benefits to Mary Brown ($2,307, $1,384 from Result #2 and $923 from Result 3) and vacation buy-out to Ethan Carruthers ($550 from Result 2). Accordingly, since the personnel categories did not have sufficient funds to cover these cost, we requested that funds were transferred, within a result, from the following categories to cover these required salary payments:

$568 from Laboratory Supplies from Result 2 to personnel Result 2
$236 from Local mileage from Result 3 to personnel Result 3
$505 from Software Result 3 software to personnel Result 3
$270 from Salary Student Worker Result 2 to Salary Junior Scientist Result 2

3. **On June 7, 2004 we requested permission to shift funds between results to cover personnel costs for Result 2.**

The discrepancy in salary and fringe funds (total personnel costs) was most likely due to the fact that I originally needed to estimate the payoffs for Ethan's and Mary's vacation time and the U's financial system was not up to date at that time. In addition, salary and fringe costs increased during the granting period. Accordingly, since the personnel categories did not have sufficient funds to cover these cost, we requested that funds were transferred, between results, from the following category to cover these required salary payments:

$211 from Professional/Technical Service Result 1 to Senior Scientist Result 2.

4. **Final Budget Analysis**

As indicated below, a project balance of $1,506 will be returned to the LCMR. All result sections had a positive balance at the end of the granting period, except for result 2. In addition, no individual categories in each result section were overspent, except for personnel costs for result 2. However, the -$3 overage in the results 2 budget was actually due to rounding errors for the cents values from each number, and our records which use a cents column indicate that the column total for result 2 is actually $0. Appendix A also shows a deficit in personnel costs for result 2 amounting to -$1,143 (about 2% of result 2 budget and 0.4% of total budget). This is primarily due to increased personnel costs from University mandated raises and large increases in fringe benefit costs that occurred during the granting period. We regret that this overage occurred and apologize for this.

Additional End of Project Costs

None

Anticipated Final Project Balance

The current project balance is estimated as of 7/27/04 to be $1,506. These funds will be returned to the LCMR.

VI. PAST, PRESENT AND FUTURE SPENDING:

A. Past Spending:

The LCMR funded project W13 in the 1999-2001 biennium (for $300,000) to develop DNA fingerprinting tools for tracking human and animal sources of fecal pollution. The proposed project will leverage the results and resources of that project by utilizing a library of DNA fingerprints generated from known human and animal sources for the identification of unknown environmental isolates. Laboratory equipment purchased under the current LCMR-sponsored project also will be fully utilized for the proposed project.

In March 2000 the Metropolitan Council Environmental Services provided us with about $38,000 in funds and $22,000 in equipment and supplies, in lieu of initially promised in-kind services, to support our LCMR–funded project. These funds were used to provide personnel to aid in DNA fingerprinting efforts and to create antibiotic and metabolic profiles of the isolated *E. coli* bacteria.

B. Current and Future Spending:

We entered into an agreement with Bacterial Bar Codes, Inc. (Houston, Texas) to sell them up to 1500 *E. coli* bacteria that were isolated during 1999 W13 LCMR-sponsored project. The sale will generate up to $15,000 in program income. These funds were deposited in an auditable account managed by the Sponsored Projects Administration at the University of Minnesota. As per our discussions with LCMR staff, these funds were only used to offset projected increases in fringe benefit rates for project personnel and to conduct additional biochemical testing of atypical *E. coli* bacteria. A total of 1010 isolates were sold to Bacterial Barcodes, Inc., which generated $10,100 in programmatic income. As approved by the LCMR, these funds were used to cover fringe benefit shortfalls (due to increases and the UM) and salary for personnel.

C. Project Partners: Salary costs for all project partners is at no cost to the project

High Island Creek

Lauren Klement – Water Plan Coordinator, Sibley County Soil and Water Conservation District, Gaylord, MN

Scott Matteson – Project Coordinator High Island Watershed Assessment Project, Sibley County Soil and Water Conservation District, Gaylord, MN

Minneopa Creek

Dr. Beth Proctor – Professor and Associate Director, Minnesota State University-Mankato Water Resources Center, Mankato, MN

Julie Conrad – Water Plan Coordinator, Blue Earth County Environmental Services Department, Mankato, MN

Vermillion River

Laura Jester – Watershed Conservationist, Dakota County Soil and Water Conservation District, Farmington, MN

D. Time: July 1, 2001 – June 30, 2004

VII. DISSEMINATION: Results from this project were disseminated in reports made to the LCMR, in periodic update reports made to cooperators, in scientific publications in peer-reviewed journals, and in scientific presentations. In addition, results from our studies will be posted on the *E. coli* rep-PCR web page (see http://www.ecolirep. umn.edu/) at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our 1999 LCMR project. Data obtained from our studies will be utilized by the cooperating agencies to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas. We anticipate that we will continue to receive several invitations to present our research results at local and regional conferences and meetings. We also recently published a paper in Applied and Environmental Microbiology, entitled: "Frequency and Distribution of Tetracycline Resistance Genes in Genetically Diverse, Nonselected, and Nonclinical *Escherichia coli* Strains Isolated from Diverse Human and Animal Sources", by Andrew Bryan, Nir Shapir, and Michael J. Sadowsky (2004, Vol. 70, pages 2503-2507, See Appendix B). This work, while not directly related to source tracking bacteria, arose from our work on *E. coli* from diverse animal hosts and shows that fecal bacteria in the environment and their resistance to antibiotics is directly related to human activity of feeding antibiotics to animals. We are currently are writing up two additional research publications concerning host origin of *E. coli* bacteria in the watersheds that we examined. We will send these to the LCMR when these are accepted for publication.

VIII. LOCATION: The project was conducted in the following areas (see maps in Attachment B):

Minneopa Creek Watershed in Blue Earth County
High Island Creek Watershed in Sibley County
Vermillion River Watershed in Dakota County

IX. REPORTING REQUIREMENTS: Periodic workprogram progress reports will be submitted not later than December 31, 2001, June 30, 2002, December 31, 2002, and July 25, 2003. A final workprogram report and associated products will be submitted by December 31, 2003. We have requested, and received a no-cost extension for this project, and specified a June 30, 2004 completion date.

X. RESEARCH PROJECTS: See Attachment B

ATTACHMENT A

Project Title: Determination of Fecal Pollution Sources in Minnesota Watersheds

Project Number: IR13

LCMR Recommended Funding: $275,000

Attachment A Deliverable Products and Related Budget												
2001 LCMR Project Biennial Budget[A]	Objective/Result											
	Result 1 Budget:	Result 1 Current invoice:	Result 1 Balance:	Result 2 Budget:	Result 2 Current Invoice:	Result 2 Balance:	Result 3 Budget:	Result 3 Current Invoice:	Result 3 Balance:	PROJECT TOTAL:		
Budget Item	Acquire E. coli Bacteria			Generate DNA Fingerprints			Dissemination Activities			BUDGET TOTAL:	CURRENT INVOICE TOTAL:	BALANCE TOTAL:
Wages, salaries & benefits												
Senior Scientist	40,100	40,064	36	57,380	58,523	-1,143	28,288	27,759	529	125,768	126,346	-578
Junior Scientist	28,000	27,899	101	60,470	60,462	8				88,470	88,361	109
Student Workers (2@15 hours/week)	11,900	11,879	21	3,983	3,754	229				15,883	15,633	250
Contracts										0		
Professional/technical (Metropolitan Council Environmental Services for fecal coliform analysis)	4,589	4,196	393							4,589	4,196	393
Printing and Publication Costs							1,000	1,000	0	1,000	1,000	0
Laboratory Supplies	14,400	14,378	22	13,631	13,502	129				28,031	27,880	151
Office Supplies	500	457	43	1,000	416	584				1,500	873	627
Local automobile mileage paid	1,500	1,480	20				264	264	0	1,764	1,744	20
Other travel expenses in Minnesota (lodging and Meals)	0	0	0				200	0	200	200	0	200
Office equipment & computers				2,500	2,411	89				2,500	2,411	89
Software				4,800	4,699	101	495	349	146	5,295	5,048	247
COLUMN TOTAL	$100,989	$100,354	$635	$143,764	$143,767	-$3	$30,247	$29,372	$875	$275,000	$273,494	$1,506

[A] Dollar amounts are estimates of balances on 12/31/03

ATTACHMENT B: Research Format

I. Abstract

Many of Minnesota's rivers and streams do not achieve the Clean Water Act "swimmable" goal due to elevated numbers of fecal coliform bacteria. Sources of fecal coliform bacteria include runoff from feedlots and manure-amended agricultural land, wildlife, inadequate septic systems, urban runoff, and sewage discharges. In this project we propose to define sources of fecal pollution in waters with excessive levels of fecal coliform bacteria. To achieve our goals, we will use a DNA fingerprinting technique to differentiate between strains of fecal coliform bacteria originating from animal and human sources. In our research studies we will use the polymerase chain reaction (PCR) technique, coupled with the use of a specific nucleic acid primer, BOXA1R, to characterize *E. coli* strains in three Minnesota watersheds. The DNA fingerprints that are generated will be compared to a library of DNA fingerprints from known human and animals sources. The library was created with funding from the Legislative Commission on Minnesota Resources during the 1999-2001 biennium. Fingerprints from the watershed isolates will be compared to those in the DNA fingerprint library using a sophisticated pattern recognition and statistical analysis software, BioNumerics. Our previous studies have demonstrated that the rep-PCR technique has the necessary sensitivity, specificity, and resolving power to differentiate between strains of fecal coliform bacteria originating from different human and animal sources. Results of these studies will be used to identify the sources of fecal pollution and target appropriate water pollution abatement efforts.

II. Background and Hypothesis

Currently, many of Minnesota's rivers and streams do not meet the water quality standard for fecal coliform bacteria. Sources of fecal coliform bacteria include runoff from feedlots and manure-amended agricultural land, wildlife, inadequate septic systems, urban runoff, and sewage discharges. High levels of fecal bacteria in Minnesota's rivers, lakes, streams, and aquifers threaten the use of these water resources for swimming and drinking. The State's water quality standard for fecal coliform bacteria is 200 microorganisms per 100 milliliters of water (as a monthly average). This number is used as an indicator of the possible presence of human pathogenic microorganisms. According to the 1996 report to the U.S. Congress on the condition of Minnesota's rivers, lakes and streams (as reported by the Minnesota Pollution Control Agency), 47% of the river miles assessed could not support swimming due to high levels of fecal bacteria.

In our 1999 LCMR-supported project, we generated a database of DNA fingerprints from a large number (greater than 2400) of fecal coliform bacteria. These DNA fingerprints have proven useful in differentiating between fecal coliform bacteria of animal and human origin. **In this project we will leverage the existing database and**

resources to determine and track sources of fecal contamination in three watershed areas. We will systematically sample multiple locations in three watershed areas, during baseline and critical run-off periods, to determine likely sources and locations of fecal contamination. Many of these watersheds have been intensively characterized with respect to potential sources of fecal pollution, feedlot locations and size, water quality, and status of septic systems. In an effort to validate local, state, and federal-supported abatement efforts, we will coordinate our monitoring program with cooperators prior to and following the implementation of best management practices in the watershed areas.

A better understanding of the source of fecal contamination will be a valuable tool in efforts to minimize the deleterious environmental consequences of fecal pollution. The human health risks associated with the ingestion of water contaminated with human fecal materials is well documented. Also, there is increasing concern about possible pathogens associated with fecal material from animal sources (e.g. the *Cryptosporidium* outbreak in Milwaukee and the *E. coli* outbreak in Canada).

The ability to distinguish between human and animal sources of fecal contamination is an important assessment tool. From a public health perspective, fecal contamination originating from human sources poses a greater human health risk than that originating from animal sources. Armed with knowledge about contamination sources, agencies could respond more quickly and more directly to inform that segment of the population at the greatest risk, without unnecessarily alarming people at low or insignificant risk. From a water quality perspective the ability to narrow the source of fecal contamination among the many potential sources will facilitate more tailored and cost effective pollution abatement efforts.

Conventional microbiological methods cannot differentiate between sources of fecal pollution giving rise to elevated coliform counts. While various methods have been proposed to determine the source of water-borne fecal contamination, many problems with these procedures are yet to be satisfactorily resolved. However, recent developments in molecular biology have provided some of the answers. Modern molecular biological approaches have been used to detect and track coliform bacteria and specific microbial pathogens in water. The polymerase chain reaction (PCR) technique, coupled with the use of specific nucleic acid primers and gene probes, has been used successfully to detect *E. coli* and the enteric pathogens, *Salmonella* and *Shigella,* in water. One PCR technique, called rep-PCR DNA fingerprinting, has been used by us to identify coliform bacteria, much as DNA fingerprinting techniques have been used in paternity and forensic cases (Dombek *et al.*, 2000). Organisms yielding indistinguishable DNA banding patterns can be regarded as being identical or near-identical, and as such, define the source of the fecal contamination. The rep-PCR technique has been shown to provide the necessary sensitivity and resolving power to differentiate between strains of fecal coliform bacteria originating from different human and animal sources. Of the various genetic fingerprinting strategies, rep-PCR is a

relatively simple and cost effective technique, which can be adapted for high throughput applications.

Various genomic DNA fingerprinting methods have been used to investigate epidemiologic, taxonomic, and phylogenetic relationships among microorganisms. While initial studies used classical restriction enzyme-generated DNA fingerprints of bacteria for the epidemiological analyses of nosocomial infections (Kaper *et al.*, 1982; Kuijper *et al.*, 1987; Langenberg *et al.*, 1986; Skjoid *et al.*, 1987; Tompkins *et al.*, 1987), DNA hybridizations and the rep-PCR DNA fingerprinting techniques have also found great application in the epidemiological and taxonomic analysis of yeast (Panchal *et al.*, 1987; Scherer and Stevens, 1987), mycoplasmas (Chandler *et al.*, 1982; Ruland *et al.*, 1990), fungi (Koch *et al.*, 1991), viruses (Buchman *et al.*, 1978; Christensen *et al.*, 1987), several diverse bacterial species (Langenberg *et al.*, 1986; Ramos and Harlander, 1990), and humans (Gill *et al.*, 1987). In addition to medically important organisms, DNA fingerprinting techniques have also been used to study the taxonomic relatedness of agriculturally important microorganisms. These organisms include bacterial and fungal pathogens (Lazo *et al.*, 1987) as well as plant symbionts (Brown *et al.*, 1989; Glynn *et al.*, 1985; Judd *et al.*, 1993; Kaijalainen and Lindstrom, 1989; Mielenz *et al.*, 1979; Sadowsky *et al.*, 1987; Sadowsky *et al.*, 1996; and Schmidt *et al.*, 1986).

The rep-PCR DNA fingerprinting (Versalovic *et al.*, 1991; de Bruijn, 1992; Versalovic *et al.*, 1994) uses the polymerase chain reaction and primers based on endogenous repetitive DNA to amplify specific portions of the microbial genome which are subsequently visualized following electrophoresis and ethidium bromide staining. The resulting banding patterns produced are generally unique to particular microbes and as such, can serve as a "fingerprint" for strain identification or analysis of populations. Organisms having indistinguishable banding patterns can be regarded as being identical or near-identical and those having similar banding patterns are regarded as genetically related. DNA fingerprints have been shown to be stable over many generations of microbial growth (Ramos and Harlander, 1990; Scherer and Stevens, 1987; Sadowsky *et al.*, 1996; Schneider and de Bruijn, 1996).

Endogenous repetitive DNA sequences have found wide application for the fingerprinting of prokaryotic genomes. Bacterial genomes contain a variety of repetitive DNA sequences. These repetitive elements typically are comprised of duplicated genes, such as rRNA, tRNA, and members of the *rhs* gene family (Lin *et al.*, 1984; Sadosky *et al.*, 1989), insertion sequences and transposons (Kleckner, 1981), interspersed repetitive extragenic palindromes (REP) (Lupski and Weinstock, 1992) and other palindromic unit (PU) sequences (Gilson *et al.*, 1984, Gilson *et al.*, 1987), intergenic repeat units (IRU) (Sharples *et al.*, 1990) or enterobacterial repetitive intergenic consensus (ERIC) sequences (Hulton *et al.*, 1991), bacterial interspersed mosaic elements (BIME) (Gilson *et al.*, 1991), short tandemly repeated repetitive (STRR) sequences (Mazel *et al.*, 1990), and BOX elements (Martin *et al.*, 1992). The exact function of many of these repetitive sequences is unknown, although some have been postulated to be important for genome structure and function, Most well defined

interspersed repetitive sequences are 26 to 400 bp in size (see Lupski and Weinstock, 1992 for a review). Of these repetitive elements, BOX, ERIC, and REP have been used the most often for analyzing a variety of Gram negative bacteria. Both ERIC and REP PCR primers have been useful in identifying *E. coli* strains causing mastitis in cows (Lipman *et al.*, 1995).

Repeated DNA sequences have also been found in the genomes of *Escherichia coli, Enterobacter aerogenes,* and *Salmonella typhimurium* and have been very useful for the examination of genome structure and for the fingerprinting of bacterial DNA. The repeated palindrome (REP) or palindromic unit (PU) sequences and bacterial interspersed mosaic elements (BIMEs) have been detected in a large number of bacterial strains (Lupski and Weinstock, 1992; Gilson *et al.*, 1984, Gilson *et al.*, 1991, Sharples *et al.*, 1990; Stern *et al.*, 1984). Rep-PCR DNA fingerprinting has also been used to investigate two nosocomial outbreaks in Houston, Texas, and the technique was useful in identifying the predominant clone causing disease (Gerghiouh *et al.*, 1995). Moreover, the rep-PCR DNA fingerprinting technique has proven to be effective in clustering methicillin resistant *Staphylococcus aureus* and *Legionella pneumophilia.* For a detailed description of the medical and epidemiological uses of the rep-PCR technique see Versalovic *et al.*, 1995.

Lastly, while a majority of repeat elements have been isolated from Gram negative bacteria, there have also been reports of endogenous repeat elements in the gram positive bacteria. For example, *Streptococcus pneumoniae* contains a highly conserved DNA sequence, collectively termed BOX, which is located within intergenic regions of the chromosome. The *S. pneumoniae* genome contains about 25 BOX elements which range in size from about 40 to 60 bp (Martin *et al.*, 1992). It has been postulated that BOX sequences serve a regulatory functions. In addition, the BOX element has also been used to study the epidemiology and classification of potato tuber diseases caused by *Streptomyces* sp strains (Sadowsky *et al.*, 1996).

Recently, we showed that the BOXA1R primer was useful in differentiating among *E. coli* strains from human and animal sources (Dombek *et al.*, 2000). The BOXA1R primer will be used in our current studies.

In summary, DNA fingerprinting methods using endogenous repeat elements have been used to investigate epidemiologic, taxonomic, and phylogenetic relationships among microorganisms. Endogenous repeat sequences have been found in almost all organisms thus far examined and provide a valuable tool to identify and track medically and environmentally important microorganisms (Versalovic *et al.*, 1998). In addition, although the function(s) of most of the endogenous repeat sequences remain unknown, they nevertheless have proven invaluable in phylogenetic studies and the analysis of genome structure and function. The rep-PCR technique takes advantage of these repeat elements and in doing so, provides the necessary sensitivity to differentiate between strains of fecal coliform bacteria originating from different human and animal sources. Of all the various genetic fingerprinting strategies, rep-PCR is a relatively

simple and cost effective technique which can be adapted for high throughput applications, Coupled to the use of computer-assisted pattern analysis and database software, for example BioNumerics, a large number of rep-PCR DNA fingerprints can be cataloged, analyzed, and characterized in an expeditious and statistically relevant manner (Dombek *et al.*, 2000; and Rademaker *et al.*, 1998).

The project will examine three watershed areas determined in consultation with Soil and Water Conservation District personnel, the Minnesota River Basin Joint Powers Board, and Metropolitan Council Environmental Services staff. The selected locations represent a range of water quality challenges, are well-characterized in terms of potential fecal pollution sources, and will provide an opportunity to evaluate the effectiveness of existing and planned pollution abatement efforts. The project will provide a scientifically sound basis for prioritizing pollution control efforts so that resources can be efficiently and effectively allocated to lower fecal coliform counts and achieve water quality goals in these watersheds. Beyond the specific applications described here, the project has national importance as it adds to the database of DNA fingerprints that can be applied to future pollution tracking efforts throughout the nation.

Results of our preliminary work have been published in a peer-reviewed journal:

Dombek, P.E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the polymerase chain reaction to differentiate *Escherichia coli* from human and animal sources. Appl. Environ. Microbiol.66:2572-2577.

Hypothesis: Our studies are based on the hypothesis that the rep-PCR technique using the BOXA1R primer has the necessary resolving power and robustness to differentiate between environmental *E. coli* ecotypes in natural watershed areas, and that the database of E. coli fingerprints that we previously generated can be used to identify sources of unknown *E. coli* strains isolated from watershed areas.

III. Methodology

All of the methodologies described below have been developed and refined during our 1999-2001 LCMR project and were found to provide the necessary specificity and resolving power to assign environmental *E. coli* isolates to their correct source groups.

A. Sampling sites

Three watershed areas will be intensively monitored for sources of fecal pollutants during the two-year period. Fecal coliform bacteria will be isolated from water samples collected in each of the 3 watershed areas in the spring, summer, and fall months, during baseline and critical run-off periods.

The watersheds to be sampled are: [1] Minneopa Creek, near Mankato (Blue Earth County), [2] High Island Creek, near Henderson (Sibley County), and [3] Vermillion River, near Farmington (Dakota County).

Water samples will be collected in sterile Whirl-Pac bags and kept on ice until processed. Ten sites will be sampled in each watershed. Sites were chosen with the input of project cooperators.

Sampling site descriptions are outlined below, and sampling site locations are indicated on the enclosed maps. Approximately 80 water samples will be collected from each watershed (10 sites sampled per watershed on 8 separate sampling occasions for a total of 240 samples for the three watersheds). To the extent possible, samples will be collected to capture spring flush after snow melt, as well as during baseline flow and after storm events.

Watershed 1 - Minneopa Creek

The Minneopa Creek Watershed covers an area of 85.2 square miles primarily in Blue Earth County. Minneopa Creek begins in western Blue Earth County and meanders eastward before entering into the Minnesota River approximately two miles west of Mankato. The creek has been channelized from its headwaters to Lily Lake, and ditches and tiles are prevalent in this area. The land surrounding Minneopa Creek consists of generally flat agricultural land along its western portion, rolling hills near its middle section, and a wooded area (Minneopa State Park) along its confluence with the Minnesota River.

Water quality in Minneopa Creek was intensively monitored in 1995-1996 (with partial funding from the LCMR). Fecal coliform levels in the Creek ranged from 10-23,000 fecal coliform/100 ml, with the highest levels associated with storm events. Sites frequently exceeded the bacteriological water quality standard of 200 or 1000 fecal coliform per 100 ml. The Minneopa Creek Watershed Plan, published in 1997, identified reduction of fecal coliform levels as a high priority goal for the Blue Earth SWCD. The exact source(s) of the fecal pollution could not be identified during this study, but were generally attributed to manure runoff and/or sewage from unsewered homes in the unincorporated areas.

Minneopa Creek sites were selected with input from project cooperators. Several of the sites correspond to previously monitored sites showing excessive levels of fecal coliform bacteria. A one-time sampling of the Lake Crystal Wastewater Treatment Facility effluent prior to disinfection also will be included.

Site 1. Minneopa Creek at Minneopa State Park entrance - Located above the Minneopa Falls. Signs in the park warn visitors of high fecal coliform levels in the creek.

Site 2. Minneopa Creek at County Road 114 - Located in subwatershed 28044.

Site 3. Minneopa Creek at County Road 112 – Below the Lake Crystal Wastewater Treatment Facility.

Site 4. Crystal Lake outlet - Crystal-Loon-Mills Lakes Outlet below the dam. Drains two minor watersheds: 28045 and 28046.

Site 5. Lily Lake outlet at bridge on County State Aid Highway 20 - Receives direct inflow from subwatersheds 28047 and 20848, and urban runoff from the City of Lake Crystal.

Site 6. County Ditch 56 at County Road 9 - County Ditch 56 receives runoff from the City of Lake Crystal and drains minor watershed 28045. This ditch empties into Crystal Lake.

Site 7. County Ditch 27 at County Road 6 - Located just south of the junction of County Road 111 and County-State Road 6 in subwatershed 28047. This ditch system empties into Lily Lake.

Site 8. Minneopa Creek at County Road 111 - Located on County Road 111 just north of County-State Highway 6 at the end of subwatershed 28048. Subwatershed 28048 receives water from subwatersheds 28049 and 28050. This ditch system empties into Lily Lake.

Site 9. Minneopa Creek at County Road 22 - Drains subwatershed 28050

Site 10. Judicial Ditch 48 at County Road 6 - Drains subwatershed 28049.

Watershed 2 - High Island Creek

The High Island Creek Watershed is a rural watershed that drains 153,219 acres in Sibley, McLeod and Renville counties. The High Island watershed is a minor watershed of the Lower Minnesota River basin, and is one of the few remaining rural watersheds of the Lower Minnesota. The watershed begins in eastern Renville, where approximately 10% of the watershed is located. McLeod County and Sibley County have respectively 25% and 65% of the watershed located within their county boundaries.

The Minnesota Pollution Control Agency monitored the High Island Creek from May 12, 1999 to August 23, 1999 for fecal coliform bacteria and transparency (using transparency tubes) at five sites in the watershed. Results of the monitoring revealed low transparency tube readings and high fecal coliform counts at a majority of the monitored sites. Water at the mouth of High Island Creek has high levels of fecal

coliform, phosphorus, nitrates and total suspended solids based on the Minnesota River Assessment Report published in 1994.

The Sibley County Soil and Water Conservation District has received a Phase I - Clean Water Partnership Grant from the MPCA to assess pollution sources within the watershed. The High Island Creek sites selected for this LCMR project correspond to the ten monitoring sites developed for the High Island Watershed Assessment Project. These sites will be rated (rating curves) by the USGS. Of the ten sites, five have been designated as primary and five as secondary. The primary sampling sites have equipment to record stage levels that will help determine loading rates. Sites are listed below with selection justification:

I. Primary Sites (continuous stage recorders)

Site 2. High Island Creek at County Road 7
This site is downstream of the confluence of Judicial Ditch 11 and Judicial Ditch 15 and represents 33% of the study area. A rock riffle at the downstream side of the bridge forms a good control for establishing a stage-discharge relation.

Site 3. High Island Creek below Baker's Lake
This site is located two miles to the east of Bakers Lake on county road 13. This site allows assessment of the extent of sediment trapping and other water quality modifications that may be occurring as High Island Creek flows through Bakers Lake. Comparison of water quality at this site with water quality at Site 2 will provide an assessment of the amount of load arising from the low-gradient area in the mid-reaches of the study area. The drainage area at Bakers Lake outlet is 110 mi squared, representing 46% of the study area.

Site 5. High Island Creek at County Road 9 Near Arlington
This site provides a means of assessing water quality of the main stem of the High Island Creek before addition of water form the urban setting at Arlington. Drainage area at this site is 161 mi squared, representing 67% of the study area.

Site 9. Buffalo Creek near Henderson
This site is located on a township road near the mouth of Buffalo Creek. This site will be representative of conditions in the downstream part of the study area where land is steeply sloped and there are wooded ravines. The channel at this site is rocky and has swiftly flowing water. Buffalo Creek drains 28.2 mi squared or 12% of the study area.

Site 10. High Island Creek at County Road 6 Near Henderson
This USGS gauging station site is located near the mouth of the High Island Creek. The site will measure the total discharge of water and chemical constituents from the study area. This site has a continuous record streamflow gauging station that has been operated since 1973. The stage-discharge relation for this site has been established by the USGS. Drainage area is 238 mi. squared at the gauge.

2. Secondary Sites (staff gauges)

Site 1. Judicial Ditch 11 at County Road 8/10
This site represents most (about 14 square miles) of minor watershed 3301500 in Renville County. Water quality at this site is expected to be typical of conditions in the upper part of the study area.

Site 4. High Island Lake
The only lake site in the study is located near High Island Sportmens Park in New Auburn. During high flows (on average 2 to 3 time per year) High Island Creek overflows a dam located on the south end of the lake. During these occasions water flows directly into the lake.

Site 6. Ditch 2 at County Road 17.
This site represents a portion of the watershed that drains 16.4 square miles or 12% of the study area. The site was previously sampled in 1999 for fecal coliform bacteria, at which time had high levels.

Site 7A. High Island Creek at County Road 66 above waste water treatment plant near Arlington.
This site is located about ½ mile downstream of New Auburn, just before the city's wastewater treatment plant. Water quality data obtained from this site will show the urban inputs of New Auburn.

Site 7B. High Island Creek at County Road 66 near Sportmen's Club.
This site is located about 1 mile downstream of site 7B, below Arlington's wastewater treatment plant. This site will only be monitored 3 times during baseflow in 2000 to determine if any inputs from the wastewater treatment plant are entering the creek.

Site 8. Buffalo Creek at County Road 17.
This site is below the confluence of County Ditch 59 and High Island Ditch number 5, which drains the less steeply sloped and less wooded part of the Buffalo Creek watershed.

Watershed 3 – Vermillion River

The Vermillion River Watershed encompasses 372 square miles, mostly located through central Dakota County south of the Twin Cities metropolitan area. The main stem originates in Scott County to the west and flows generally northeast to the City of Hastings. Current landuse in the watershed is still dominated by agriculture with suburban areas and smaller urban growth centers interspersed throughout the watershed.

In 1998, the Vermillion River main stem, from Empire Township to the dam in Hastings, was listed on the Federal Clean Water Act's 303(d) list of impaired waters for fecal coliform bacteria. The river was not meeting its swimming use standard due to high bacteria levels. Also in 1998, the Vermillion River was placed on the Minnesota Pollution Control Agency's (MPCA) list of waters in need of a total maximum daily load (TMDL) study for fecal coliform. In 1999 the MPCA, with the help of local agencies and citizens, collected fecal coliform samples throughout the Vermillion River watershed to begin determining the extent of the bacterial problem. These data indicate that the river and its tributaries have bacteria levels in excess of the MPCA's state standard of 200 organisms/100 ml of sample.

The Dakota County Soil and Water Conservation District has undertaken a TMDL project to conduct additional monitoring in the Vermillion River watershed to further define the extent of the problem, determine possible sources of bacteria, calculate an acceptable load of bacteria for the various reaches, and draft a series of reduction goals for future implementation. This project will include additional fecal coliform monitoring throughout the watershed to better identify highly impaired reaches and possible sources of the bacteria. Landuse and landcover GIS data will also help identify possible sources of non-point source pollution. Using GIS data and lab results, a TMDL model will be produced and various loading scenarios will be drafted. Reduction goals and an implementation strategy will be the final outcome of the project.

The monitoring sites for our project were selected with input from Dakota County SWCD staff. Site locations and information are summarized below:

1. Middle Creek Sub-watershed

Site 1. VMCwest
- Just upstream of confluence with VMCeast; north of County Hwy. 64 in Farmington
- Agricultural and suburban influence
- Will help explain very high fecal levels found in Middle Creek in 2000
- Site is located just in a cropped field with very little buffer adjacent to the waterway

Site 2. VMCeast
- Just upstream of confluence with VMCwest; north of County Hwy. 64 in Farmington
- Agricultural and suburban influence
- Will help explain very high fecal levels found in Middle Creek in 2000
- Site is located just in a cropped field with very little buffer adjacent to the waterway

Site 3. VMC
- At the mouth of Middle Creek just upstream of its confluence with North Creek and the Main Stem Vermillion River
- Take sample from foot bridge on Hwy 3 just north of Farmington in Empire Township
- During very high water, North Creek may back up into Middle Creek so take sample further upstream

- Will help explain very high fecal levels found in Middle Creek in 2000
- Middle Creek sub-watershed has both suburban and agricultural landuses and is rapidly developing

2. North Creek Sub-watershed

Site 4. VNC175
- Just downstream of wide spot in the creek near trailer court on west side of Hwy 31 in Lakeville
- Mostly suburban influence with some agriculture much further upstream
- Will help explain very high fecal levels at this site in 2000
- Past complaints of foul "sewer" odor at this site

Site 5. VNC
- At the mouth of North Creek just upstream of its confluence with Middle Creek and the Main Stem Vermillion River
- Take sample from foot bridge on Hwy 3 just north of Farmington in Empire Township
- Will help explain sources of fecal pollution in the entire sub-watershed
- North Creek sub-watershed has both suburban and agricultural landuses

3. South Branch Sub-watershed

Site 6. VSBtrib
- On the east side of Hwy 79 just south of 232nd St. in Castle Rock Township
- In a slightly wooded area upstream of confluence with South Branch
- Drains agricultural and rural residential areas
- Will help explain very high levels of fecal coliform measured in 2000 at this site

Site 7. VSB
- At the mouth of South Branch at Hwy 66 in Vermillion Township
- Slightly wooded site with large trout found here in 2000
- Will help explain sources of fecal pollution from the entire sub-watershed
- South Branch sub-watershed drains agricultural and rural residential areas
- City of Hampton sewage lagoons discharge to South Branch during periods of high flow
- Many horse pastures, feedlots and sod farms are in this sub-watershed

4. Main Stem Vermillion River

Site 8. V31
- At Hwy 31 bridge in Farmington just south of Hwy 50
- Site is downstream of upper areas of watershed prior to Middle Creek, North Creek and South Branch inlets

- Upper portions of watershed have agricultural, rural residential, quickly developing suburban areas, an industrial park, and wastewater discharges for the Elko/New Market sewage lagoons

Site 9. Vverm
- At Hwy 85 just outside the City of Vermillion in Vermillion Township
- Slightly wooded site
- Downstream of all major tributary inlets
- Downstream of Empire Wastewater treatment plant and City of Vermillion wastewater treatment plant

Site 10. V47
- At Hwy 47 just inside Hastings city limits
- Site is open and channel has been straightened
- Just upstream of Vermillion Falls below which fecal levels drop below state standard

B. Isolation and confirmation of *E. coli* from watershed samples

Water samples will be analyzed for fecal coliform bacteria by an EPA-certified laboratory using the membrane filtration method (*Standard Methods for the Examination of Water and Wastewater,* 1995). Agar plates (mFC) generated from the fecal coliform analysis will be used by project staff for isolation, identification and confirmation of *E. coli* strains. Typical fecal coliform colonies will be picked from the mFC agar plates used to enumerate fecal coliforms. Blue colonies will be restreaked for purity onto the same medium. The purified presumptive coliform bacteria will be subjected to confirmatory tests using EC broth with MUG (with Durham tubes). The EC-MUG Broth will be incubated at 44.5 ± 0.2 °C. The EC medium differentiates between coliform bacteria of fecal or other origins. Presumptive *E. coli* colonies will be streaked for isolation to MacConkey agar and plated onto ChromAgar ECC to differentiate between *E. coli* and *Klebsiella*. Confirmed coliform bacteria will be indicated by the production of gas in Durham tubes, strong fluorescence when EC-MUG tubes are examined by using ultraviolet light, and production of blue colonies on ChromAgar. The identity of coliform bacteria will be verified by using standard biochemical tests as outlined in *Standard Methods*. Only isolates confirmed as being *E. coli* will be used in our studies. The confirmed coliform isolates will be cataloged, stored, and preserved at -80°C until needed.

About 1600 *E. coli* strains will be isolated per watershed area for a project total of 4800 isolates:

No. of Watershed	Sites per Watershed	Sampling Occasions	E. coli per sample	Total E. coli Isolated
3	10	8	20	4800

C. rep-PCR DNA fingerprinting of environmental *E. coli*

All confirmed isolates from known sources will be individually streaked onto Plate Count Agar and incubated overnight at 37°C. An individual colony of each isolate will be picked with a 1 μl loop and suspended in 100 μl dH_2O. Cell suspensions will be stored at -80°C until used. The cell suspensions will be subjected to rep-PCR using BOXA1R primers. A single BOXA1R primer is needed. The sequence of the BOXAIR oligonucleotide PCR primer is 5'-CTACGGCAAGGCGACGCTGACG-3'. The PCR reactions are performed in 25 μl of PCR buffer as previously described (Dombek et al., 2000; Judd et al., 1993 and Sadowsky et al., 1996) using 2 μl of cell suspensions as template. The PCR reaction conditions for the BOXA1R primer are: denaturation for 7 min at 95°C; 35 cycles of 90°C for 30 sec, 53°C for 1 min, and 56°C for 8 min; final extension at 65°C for 16 min; and a 4°C soak. PCR products will be separated by horizontal electrophoresis on 1.5% agarose gels run at 70V for 17.5 hr at 4°C with buffer recirculation, stained with ethidium bromide, and photographed. Molecular weight and an internal standard (PCR products from one of the strains) will be included on each gel to aid in normalization and comparison across gels. The resulting DNA fingerprints will be scanned into digital images, converted to TIFF files and used for subsequent analyses.

D. Data analysis

DNA fingerprints will be normalized to molecular weight and internal standards, and compared and analyzed using BioNumerics software as described by Dombek *et al.* (2000) and Rademaker *et al.* (1998). The relatedness of isolates to each other and to those in the known-source DNA fingerprint library will be determined by cluster (using the Pearson's correlation coefficients), principal component, and discriminant statistical analyses. The number and diversity of isolates falling into each source group category, obtained during different sampling times, will be analyzed for statistical differences using analysis of variance and other applicable statistical methods. Isolates will be assigned to source groups if they have > 80% average similarity value to those isolates in the fingerprint library. Sources of unknown environmental isolates will be provided to cooperating agencies and used to evaluate existing and proposed abatement programs. Results of our will be stored on a rep-PCR web site (www.ecolirep.umn.edu) at the Department of Soil, Water, and Climate, University of Minnesota. The known source DNA fingerprints generated during our previous LCMR project will be made available for downloads from this server. The BioNumerics program contains database sharing functions which allow for the exchange of fingerprint information between laboratories. We view this as an essential resource for other laboratories to use both at the state and national levels.

IV. Results and Products

Result 1: Acquire fecal coliform bacteria from watershed areas. Fecal coliform bacteria will be isolated from water samples collected in each of the 3 watershed areas in the spring, summer, and fall months (during baseline and critical run-off periods). Approximately 80 water samples will be collected from each watershed (10 samples collected per watershed on 8 separate sampling occasions). Samples will be analyzed for fecal coliforms by an EPA-certified contract laboratory. *E. coli* bacteria from fecal coliform plates will be isolated (about 1600 isolates per watershed for a total of 4800 isolates) and the identity of bacteria will be confirmed by using selective and differential microbiological media and biochemical tests. The confirmed *E. coli* isolates will be cataloged, stored, and preserved in glycerol at -80°C until DNA fingerprinting is done.

Result 2: Generate DNA fingerprints from *E. coli* isolates obtained from watersheds. DNA from confirmed *E. coli* isolates will be subjected to rep-PCR DNA fingerprinting using BOXA1R primers. The resulting **4800** DNA fingerprints will be scanned into digital images, normalized to molecular weight and internal standards, and compared and analyzed using BioNumerics pattern recognition and statistical analysis software. The animal(s) contributing to *E. coli* in watersheds will be determined by cluster and discriminant statistical analyses by comparison to our known source DNA fingerprint library.

Result 3: Dissemination and Implementation of Results. Results from this project will be disseminated in reports made to the LCMR, in periodic update reports made to cooperators, and in scientific publications in peer-reviewed journals. In addition, results from our studies will be posted on the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) housed at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our 1999 LCMR project and includes searchable and downloadable DNA fingerprints that can be used by state and local agencies to track sources of fecal coliform bacteria. Data obtained from our studies will be utilized by the cooperating agencies to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas.

V. Timetable

Water sampling will take place during the spring, summer and fall seasons. Given that the project period begins and ends in mid-summer, sampling will take place in both years of the project period. Sampling will begin in the Summer of 2001 and extend into the Fall of 2001. Sampling will resume in the Spring of 2002 and will continue into Summer 2002. Each watershed will be sampled on a monthly basis, and will be sampled eight times in 2001-2002.

Concurrent with water sampling activities, *E. coli* bacteria will be isolated and confirmed.

DNA fingerprinting and data analysis activities will occur during the late Fall through early Spring months, when field work and isolations are not being done. Reports to the LCMR will be submitted December 31, 2001, June 30, 2002, December 31, 2002, and June 30, 2003.

	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
	2001						2002												2003					
Sampling																						?		
Isolation																						?		
Fingerprinting																						?	?	
Data Analysis																								
Dissemination						•						•						•						•

• = Progress reports to LCMR

? = Only if needed due to Spring run-off events.

VI. Budget Requirements

LCMR Recommended Funding: $275,000

	Objective/Result			
Budget Item	Result 1 Acquire Fecal Coliform Bacteria	Result 2 Generate DNA Fingerprints	Result 3 Dissemination Activities	**Row Total**
Wages, salaries & benefits				
Senior Scientist	45,600	45,600	22,800	114,000
Junior Scientist	30,400	45,600		76,000
Student worker (2@15 hr/wk)	21,000	9,000		30,000
Total: Salary & Benefits	**97,000**	**100,200**	**22,800**	**220,000**
Laboratory Supplies	20,700	13,400		34,100
Office Supplies	1,000	1,000		2,000
Equipment				
Computer		2,500		2,500
Software		4,800	1,000	5,800
Publication costs			1,000	1,000
Contracted Services (fecal coliform analysis)	4,800			4,800
Local Travel (mileage, food, lodging)	4,100		700	4,800
Column Total	**127,600**	**121,900**	**25,500**	**275,000**

VII. Principal Investigators and Cooperators

Principal Investigators:

Michael J. Sadowsky – Professor, University of Minnesota (CV attached)

LeeAnn K. Johnson – Sr. Scientist, University of Minnesota (resume attached)

Cooperators

High Island Creek
Lauren Klement – Water Plan Coordinator, Sibley County Soil and Water Conservation District
Scott Matteson – Project Coordinator High Island Watershed Assessment Project, Sibley County Soil and Water Conservation District

Minneopa Creek
Dr. Beth Proctor – Professor and Associate Director, Minnesota State University-Mankato Water Resources Center
Julie Conrad – Water Plan Coordinator, Blue Earth County Environmental Services Department

Vermillion River
Laura Jester – Dakota County Soil and Water Conservation District
Kent Johnson – Water Quality Manager, Metropolitan Council Environmental Services

CURRICULUM VITAE

Name: Michael Jay Sadowsky

Present Title: Professor

Department of Soil, Water, and Climate, Biological Process Technology Institute, Department of Microbiology, and Director of Graduate Studies Microbial Ecology and Microbial Engineering Programs

Address: Department of Soil, Water, and Climate
University of Minnesota
439 Borlaug Hall
1991 Upper Buford Circle
St. Paul, Minnesota 55108

Education:

Ph.D., 1983 University of Hawaii, Honolulu, Hawaii. Major: Microbiology
M.S., 1979 University of Wisconsin, Oshkosh, Wisconsin. Major: Biology/Microbiology
B.S, 1977 University of Wisconsin, Madison, Wisconsin. Major: Bacteriology

Professional Experience:

Professor: Department of Soil, Water, and Climate and Department of Microbiology University of Minnesota, St. Paul, Minnesota 07/96 - present.
Associate Professor: Departments of Soil Science and Microbiology University of Minnesota, St. Paul, Minnesota 07/93 - 6/96.
Assistant Professor: Departments of Soil Science and Microbiology University of Minnesota, St. Paul, Minnesota 06/89 - 6/93.
Microbiologist: U.S. Department of Agriculture-ARS; Beltsville, Maryland, 01/86 - 05/89.
Research Molecular Biologist: Allied Corporation; Syracuse, New York, 04/85 - 12/85.
Post-Doctoral Research Fellow: Department of Biology, Plant Molecular Biology Laboratory, McGill University; Montreal, Canada, 09/83 - 04/85.
Graduate Research Assistant: Department of Microbiology, University of Hawaii; Honolulu, Hawaii, 08/80 - 08/83.
Graduate Teaching Assistant: Department of Microbiology, University of Hawaii; Honolulu, Hawaii, 08/79 - 08/80 (General and Advanced Microbiology, Medical Microbiology Laboratory).

Awards and Honors:

1990	Young Investigator Award, American Society for Microbiology.
1988	Patent Award, U.S. Department of Agriculture.
1991-1992	Bush Foundation Excellence in Teaching Program.
1999	Fellow American Academy of Microbiology

Professional Activities:

Editor, Applied and Environmental Microbiology 1999-2004
Associate Editor, Molecular-Plant Microbe Interactions 1995-1998
Associate Editor, Applied Environmental Microbiology 1989-1999
Associate Editor, Symbiosis
Co-Organizer of: Agricultural Microbes Genomes I Conference. Jan. 13-14, 2000. San Diego, CA (with Stephen Heller et al.).
Co-Organizer of: 14th North American Symbiotic Nitrogen Fixation Conference. July 25-29 1993, Minneapolis, Minnesota (with Peter Graham and Carroll Vance).
Organizer of Training Workshop: Use of Molecular Biology Techniques in Biological Nitrogen Fixation Research, June 18-22 1995, Cairo, Egypt.
Panel Member, Ad Hoc, NIH, Molecular Biology 1994
Panel member, USDA/CSRS, Competitive Research Grants Program, Nitrogen Fixation Panel, 1991.
Panel member, USDA/National Research Initiative, Competitive Research Grants Program, Nitrogen Fixation Panel, 1992.
Member, U.S. EPA, Biological Safety Advisory Committee, 1988.
Panel member, U.S. AID/PSTC, Competitive Grants Program, Molecular Biology Panel, 1987.
Panel member, U.S. AID/PSTC, Competitive Grants Program, Molecular Biology Panel,1988

Patents:

Held: A Mu-dI(Kan,lac) vector system for analysis of symbiotic genes in Rhizobium; Patent awarded in Canada and U.S. through Allelix Inc.

Patent Pending: (I) Cloning, characterization, and sequence of a gene region from Pseudomonas sp. strain ADP involved in the dechlorination of atrazine; (II) Improved enzymes for biodegradation of atrazine and other xenobiotic compounds obtained by DNA shuffling. Submitted by University of Minnesota and Ciba Geigy, Inc; (III) Atrazine-degrading plants; Submitted by University of Minnesota and Novartis Crop Protection. Pending: SAFTI-Pak, Solid Appliance Fermentation Technology for Inoculants. US 193. 394

Publications: 1995 – 2000

Dombek, P.E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the polymerase chain reaction to differentiate *Escherichia coli* from human and animal sources. Appl. Environ. Microbiol. 66:2572-2577.

Strong, L. C., McTavish, H., Sadowsky, M. J., and Wackett, L. P. 2000. Field-scale remediation of atrazine-contaminated soil using recombinant *Escherichia coli* expressing atrazine chlorohydrolase. Environ. Microbiol. 2:91-98.

Montealegre,C. M., van Kessel, C., Blumenthal, Jurg, Hur, Hor-Gil, Hartwig, U. A., and M.J. Sadowsky. 2000. Elevated atmospheric CO_2 alters microbial population structure in a pasture ecosystem. Global Change Biology 6:475-482.

Seffernick, J.L., G. Johnson, M. J. Sadowsky, and L. P. Wackett. 2000. Substrate specificity of atrazine chlorohydrolase and atrazine-catabolizing bacteria. Appl. Environ. Microbiol. IN PRESS.

Loh, J., M. G. Stacey, M. J. Sadowsky, and G. Stacey. 1999. The *Bradyrhizobium japonicum* nolA gene encodes three functionally distinct proteins. J. Bacteriol. 181: 1544-1554.

Sadowsky, M. J., Z. Tong, M. de Souza, and L. P. Wackett. 1998. AtzC is a new member of the amidohydrolase protein superfamily and is homologous to other atrazine-metabolizing enzymes. J. Bacteriol. 180:152-158.

de Souza, M. L., D. Newcombe, S. Alvey, D. E. Crowley, A. Hay, M. J. Sadowsky, and L. P. Wackett. 1998. Molecular basis of a bacterial consortium: interspecies catabolism of atrazine. Appl. Environ. Microbiol. 64:178-184

de Souza, M. L., L. P. Wackett and M. J. Sadowsky. 1998. The atzABC genes encoding atrazine catabolism genes are located on a self-transmissible plasmid in *Pseudomonas* strain ADP. Appl. Environ. Microbiol. 64:2323-2326.

Lohrke, S. L., B. Day, V. K. Kolli, R. Hancock, J. P.-Y. Yuen, M. L. de Souza, G. Stacey, R. Carlson, Z. Tong, H.-G. Hur, J. H. Orf, and M. J. Sadowsky. 1998. The *Bradyrhizobium japonicum* noeD gene: a negatively-acting, genotype-specific nodulation gene for soybean. Molec. Plant-Microbe Interact. 11:476-488

de Souza, M. L., J. Seffernick, M. J. Sadowsky, and L. P. Wackett. 1998. The atrazine catabolism genes atzABC are widespread and highly conserved. J. Bacteriol. 180:1951-1954.

Sadowsky, M. J., and H.-G. Hur. 1998. Use of endogenous repeated sequences to fingerprint bacterial genomic DNA. In: J.R. Lupski, G. Weinstock, and F. J. de Bruijn (eds.). Bacterial Genomes: Structure and Analysis. Chapman and Hall.

Sadowsky, M. J., L. P. Wackett, M. L. de Souza, K. L. Boundy-Mills, and R. T. Mandelbaum. 1998. Genetics of Atrazine Degradation in *Pseudomonas* sp. Strain ADP, pp. 88-94. In: Triazine Herbicides: Risk Assessment, L. Ballantine, J. McFarland, and D. Hackett (eds.), American Chemical Society, Washington, D.C.

Wackett, L. P., M. J. Sadowsky, M. L. de Souza, and R. T. Mandelbaum. 1998.

Atrazine hydrolysis by a bacterial enzyme, pp. 82-87. In: Triazine Herbicides: Risk Assessment, L. Ballantine, J. McFarland, and D. Hackett (eds.), American Chemical Society, Washington, D.C.

Boundy-Mills, K. L., M. L. de Souza, L. P. Wackett, R. Mandelbaum, and M. J. Sadowsky. 1997. The atzB gene of *Pseudomonas* sp. strain ADP encodes hydroxyatrazine ethylaminohydrolase, the second step of a novel atrazine degradation pathway. Applied Environ. Microbiol. 63:916-923.

Daane, L. L., J. A. E. Molina, and M. J. Sadowsky. 1997. Plasmid transfer between spatially-separated donor and recipient bacteria in earthworm-containing soil microcosms. Appl. Environ. Microbiol. 63:679-686.

Daane, L. L., J. A. E. Molina, and M. J. Sadowsky. 1997. Scanning electron microscopy of the microflora in egg capsules of the earthworm *Eisenia fetida*. Pedobiologia 42:79-87.

Schortemeyer, M., U. A. Hartwig, G. R. Hendrey, and M. J. Sadowsky. 1997 Influence of free air carbon dioxide enrichment (FACE) on microbial communities in the rhizosphere of white clover and perennial ryegrass. Soil Biology and Biochemistry. 28:1717-1724.

Sadowsky, M. J., and M. Schortemeyer. 1997. Soil microbial responses to increased concentrations of atmospheric CO_2. Global Change Biology. 3:217-244.

Schortemeyer, M., H. Santruckova, and M. J. Sadowsky. 1977. Relationship between root parameters and soil microorganisms in the rhizosphere of clover and ryegrass. Comm. Soil Sci. Plant Anal. 28:1675-1682.

Boldt, Y. R., A. K. Whitting, M. L. Wagner, M. J. Sadowsky, L. Que, and L. P. Wackett. 1997. Manganese(II) active site mutants of 3,4-dihydroxyphenylacetate 2,3-dioxygenase from *Arthrobacter globiformis* strain CmM-2. Biochemistry. 36:2147-2153.

Hur, H.-G., L. M. Newman, L. P. Wackett, and M. J. Sadowsky. 1997. Toluene 2-monooxygenase-dependent growth of *Burkholderia cepacia* G4/PR1 on diethly ether. Appl. Environ. Microbiol. 63:1606-1609.

Sadowsky, M. J., and P. H. Graham. 1997. Soil Biology of the Rhizobiaceae. In: The *Rhizobiaceae*, pp. 155-172.. H. P. Spaink, A. Kondorosi, and P. J. J. Hooykaas (eds.), Kluwer, The Netherlands.

de Souza, M. L., M. J. Sadowsky, and L. P. Wackett. 1996. Atrazine chlorohydrolase from *Pseudomonas* sp. ADP: gene sequence, enzyme purification and protein characterization. J. Bacteriol. 178:4894-4900.

Daane, L. L., J. A. E. Molina, E. C. Berry, and M. J. Sadowsky. 1996. Influence of earthworm activity on gene transfer from *Pseudomonas fluorescens* to indigenous soil bacteria. Appl. Environ. Microbiol. 62:515-521.

Sadowsky, M. J., L. L. Kinkel, J. H. Bowers, and J. L. Schottel. 1996. Use of repetitive intergenic DNA sequences and fatty Acid profiles to classify pathogenic and disease-suppressive strains of Streptomyces. Appl. Environ. Microbiol. 62:3489-3493.

Lohrke, S. M., J. H. Orf, and M. J. Sadowsky. 1996. Inheritance of host-controlled restriction of nodulation by *Bradyrhizobium japonicum* strain USDA 110. Crop Sci. 36:1271-1276.

de Souza, M. L., L. P. Wackett, K. L. Boundy-Mills, R. T. Mandelbaum, and M. J. Sadowsky. 1995. Cloning, characterization, and expression of a gene region from *Pseudomonas* sp. strain ADP involved in the dechlorination of atrazine. Appl. Environ. Microbiol. 61: 3373-3378.

Sadowsky, M. J., R. M. Kosslak, B. Golinska, C. J. Madrzak, and P. B. Cregan. 1995. Restriction of nodulation by *B. japonicum* is mediated by factors present in the roots of *Glycine max*. Appl. Environ. Microbiol. 61:832-836.

Mdrzak, C. J., B. Goliska, J. Króliczak, K. Pudeko, D. Aewska, B. Lampka, and M. J. Sadowsky. 1995. Diversity among field populations of *Bradyrhizobium japonicum* in Poland. Appl. Environ. Microbiol 61:1194-1200.

Lohrke, S. M., J. H. Orf, E. Martínez-Romero, and M. J. Sadowsky. 1995. Host-controlled restriction of nodulation by *Bradyrhizobium japonicum* strains in serogroup 110. Appl. Environ. Microbiol. 61:2378-2383.

Graham, P. H., M. J. Sadowsky, S. W. Tighe, J. A. Thompson, R. A. Date, J. G. Howieson, and R. Thomas. 1995. Differences among strains of *Bradyrhizobium* in fatty acid-methyl ester (FAME) analysis. Can. J. Microbiol. 41:1038-1042.

Noyd, R. K., F. L. Pfleger, R. E. Norland, and M. J. Sadowsky. 1995. Native prairie grasses and microbial community responses to reclamation of taconite iron ore tailing. Can. J. Botany. 73:1645-1654.

LeeAnn K. Johnson

University of Minnesota
Department of Soil, Water and Climate
1991 Upper Buford Circle, 439 Borlaug Hall
St. Paul, MN 55108
Phone: (612) 625-4241
E-mail: ljohnson@soils.umn.edu

EXPERIENCE

1999-present **Senior Scientist**
Department of Soil, Water & Climate
University of Minnesota, St. Paul, MN

Manage a field- and laboratory-based research project using DNA fingerprinting methods to address fecal pollution in Minnesota watersheds:

- Develop sampling and laboratory protocols and procedures
- Manage field sampling and laboratory activities
- Maintain database and analyze data using standard and specialized software
- Hire and supervise technical support staff
- Coordinate stream sampling with local government staff
- Present project results through publications and presentations
- Write grant proposals
- Manage project budgets

1987-1999 **Senior Research Scientist**
Research and Development Section
Metropolitan Council Environmental Services, St. Paul, MN

Managed research and monitoring projects related to wastewater treatment and water quality:

- Responsibilities: Review scientific literature, design experiments, conduct field and laboratory studies, collect and analyze samples, analyze data, communicate results through written reports and oral presentations, direct technical support staff.
- Projects: activated sludge bulking control, toxicity and treatability of industrial discharges, ultraviolet disinfection, biological odor control, wastewater operations troubleshooting, automated monitoring of treatment plant processes and effluent quality, laboratory method development and troubleshooting.
- Other: Team leader for Mercury Reduction Strategy, radiation safety officer and hazardous waste coordinator, MN Class C Wastewater Treatment Facility Operator Certificate.

1984-1987 **Pollution Control Specialist**
Solid and Hazardous Waste Division
Minnesota Pollution Control Agency, St. Paul, MN

Enforced state and federal hazardous waste regulations:
- Inspected companies for compliance with hazardous waste regulations
- Took enforcement action as needed
- Provided technical assistance to the regulated community
- Assisted in program development

1984 **Pollution Control Technician**
Water Quality Division
Minnesota Pollution Control Agency, St. Paul, MN

Conducted water quality studies to determine wastewater treatment plant effluent limits and advanced treatment justifications:
- Operated and maintained field sampling and monitoring equipment
- Collected and processed water samples
- Summarized data

1979-1984 **Research Assistant**
Microbiology Department
University of Minnesota, Gray Freshwater Biological Institute, Navarre, MN

Researched biodegradability of toxic organic compounds in natural aquatic systems. Reviewed literature, designed and conducted field and laboratory studies, analyzed data, published results.

1978-1979 **Laboratory Attendant**
Department of Food Science and Nutrition
University of Minnesota, St. Paul, MN

1977-1979 **Laboratory Attendant**
Biochemistry Department
University of Minnesota, St. Paul, MN

EDUCATION

B.S. Microbiology, University of Minnesota, 1979
Graduate studies in Microbiology, University of Minnesota, 1979-1984
MPH Environmental Microbiology, University of Minnesota (anticipated 2001)

References for Research Addendum

Brown, G., Khan, Z., and Lifshitz, R. (1989). Plant promoting rhizobacteria: strain identification by restriction fragment length polymorphisms. Can J. Microbiol. 36, 242-248.

Buchman, T.G., Roizman, B., Adams, G., and Stover, B.H. (1978). Restriction endonuclease fingerprinting of *Herpes simplex* virus DNA: a novel epidemiological tool applied to a nosocomial outbreak. J. Infect. Dis. 138, 488-498.

Chandler, D.K.S., Razin, S., Stephens, E.B., Harasawa, R., and Barile, M.F. (1982). Genomic and phenotypic analysis of *Mycoplasma pneumoniae* strains. Infect. Immun. 38, 604-609.

Christensen, L.S., Soerensen, K.J., and Lei, J.C. (1987). Restriction fragment pattern (RFP) analysis of genomes from Danish isolates of suid herpes virus 1 (Aujezsky's disease virus). Arch. Virol. 97, 215-224.

de Bruijn, F.J. (1992). Use of repetitive (repetitive extragenic element and enterobacterial repetitive intergenic consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. Appl. Environ. Microbiol. 58, 2180-2187.

Dombek, P.E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the polymerase chain reaction to differentiate *Escherichia coli* from human and animal sources. Appl. Environ. Microbiol. 66:2572-2577.

Georghiou, P., Hamill, R.J., Wright, C.E., Versalovic, J., Koeuth, T., Watson, D. A., and Lupski, J.R. (1995) Molecular epidemiology of infection due to *Enterobacter aerogenes*: identification of hospital outbreak strains by molecular techniques. Clin. Infect. Dis. 20:84-94.

Gill, P., Lygo, J.E., Fowler, S.J., and Werrett, D.J. (1987). An evaluation of DNA fingerprinting for forensic purposes. Electrophoresis 8, 38-44.

Gilson, E., Clement, J.-M., Brutlag, D., and Hofnung, M. (1984). A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. EMBO J. 3, 1417-1421.

Gilson, E., Clement, J.M., Perrin, D., and Hofnung, M. (1987). Palindromic units, a case of highly repetitive DNA sequences in bacteria. Trends Genet. 3, 226-230.

Gilson, E., Saurin, W., Perrin, D., Bachellier, S., and Hofnung, M. (1991). Palindromic units are part of a new bacterial interspersed mosaic element (BIME). Nucleic Acids Res. 19, 1375-1383.

Glynn, P., Higgins, P., Squartini, A., and O'Gara, F.. (1985). Strain identification in *Rhizobium trifolii* using DNA restriction analysis, plasmid DNA profiles, and intrinsic antibiotic resistances. FEMS Microbiol. Lett. 30, 177-182.

Hulton, C.S.J., Higgins C.F., and Sharp, P.M. (1991). ERIC sequences, a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. Mol. Microbiol. 5, 825-834.

Judd, A.K., and Sadowsky, M.J. (1993). The *Bradyrhizobium japonicum* serocluster 123 hyperreiterated DNA region, HRS1, has DNA and amino acid sequence homology to IS*1380*, an insertion sequence from *Acetobacter pasteurianus*. Appl. Environ. Microbiol. 59, 1656-1661.

Judd, A. K., M. Schneider, M. J. Sadowsky and F. J. de Bruijn. (1993). The use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. Appl. Environ. Microbiol. **59:**1702-1708.

Kaijalainen, S., and Lindstrom, K. (1989). Restriction fragment length polymorphism analysis of *Rhizobium galegae* strains. J. Bacteriol. 171, 5561-5566.

Kaper, J.B., Bradford, H.B., Roberts, and Falkow, S. (1982). Molecular epidemiology of *Vibrio cholerae* in the U.S. Gulf coast. J. Clin. Microbiol. 16, 129-134.

Kleckner, N. (1981). Transposable elements in prokaryotes. Annu. Rev. Genet. 15, 341-404.

Koch, E., Song, K., Osborn, T. C., and Williams, P.H. (1991). Relationship between pathogenicity and phylogeny based on restriction fragment length polymorphism in *Leptosphaeria maculans*. Mol. Plant Microbe Interact. 4, 341-349.

Kuijper, E.J., Oudbier, J.H., Stuifbergen, W.N., Jansz, A., and Zanen, H.C. (1987). Application of whole-cell DNA restriction endonuclease profiles to the epidemiology of *Clostridium difficile*-induced diarrhea. J. Clin. Microbiol. 25, 751-753.

Langenberg, W., Rauws, E.A.J., Widjojokusumo, A., Tytgat, G.N.J., and Zanen, H.C. (1986). Identification of *Campylobacter pyloridis* isolates by restriction endonuclease DNA analysis. J. Clin. Microbiol. 24, 414-417.

Lazo, G.R., Roffey, R., and Gabriel, D.W. (1987). Pathovars of *Xanthomonas campestris* are distinguishable by restriction fragment length polymorphisms. Int. J. Syst. Bacteriol. 37, 214-221.

Lipman, L.J.A., de Nijs, A., Lam, TJG, and Gaastra, W. (1995) Identification of Escherichia coli strains from cows with clinical mastitis by serotyping and DNA polymorphism patters with REP and ERIC primers. Vet. Microbiol. 43:13-19.

Lin, R.-J., Capage, M., and Hill, C.W. (1984). A repetitive DNA sequence, *rhs*, responsible for duplications within the *Escherichia coli* K-12 chromosome. J. Mol. Biol. 177, 1-18.

Lupski, J.M., and Weinstock, G.M. (1992). Short, interspersed repetitive DNA sequences in prokaryotic genomes. J. Bacteriol. 174, 4525-4529.

Martin, B., Humbert, O., Camara, M., Guenzi, E., Walker, J., Mitchell, T., Andrew, P., Prudhome, M., Alloing, G., Hakenbeck, R., Morrison, D.A., Boulnois, G.J., and Claverys, J.-P. (1992). A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res. 20, 3479-3483.

Mazel, D., Houmard, J., Castets, A.M., and de Marsac, N.T. (1990). Highly repetitive DNA seqquences in cyanobacterial genomes. J. Bacteriol. 172, 2755-2761.

Mielenz, J.R., Jackson, L.E., O'Gara, F., and Shanmugam, K.T. (1979). Fingerprinting bacterial chromosomal DNA with restriction endonuclease *Eco*R1: comparison of *Rhizobium* spp. and identification of mutants. Can. J. Microbiol. 25, 803-807.

Panchal, C.J., Bast, L., Dowhanick, T., and Stewart, G.G. (1987). A rapid, simple, and reliable method of differentiating brewing yeast strains based on DNA restriction patterns. J. Inst. Brew. 93, 325-327.

Rademaker, J. L. W., Louws, F. J., and de Bruijn, F. J. (1998). Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting. In: Akkermans, A.D.L., van Elsas, J.D., and de Bruijn, F.J. Molecular Microbial Ecology Manual, Kluwer Academic Publishers, Dordrecht, pp. 1-26..

Rademaker, J. L. W., Louws, F. J., Rossbach, U., Vinuesa, and de Bruijn, F. J. (1998). Computer assisted pattern analysis of molecular fingerprints and database construction. In: Akkermans, A.D.L., van Elsas, J.D., and de Bruijn, F.J. Molecular Microbial Ecology Manual, Kluwer Academic Publishers, Dordrecht.

Ramos, M.S., and Harlander, S.K. (1990). DNA fingerprinting of lactococci and streptococci used in dairy fermentations. Appl. Microbiol. Biotechnol. 34, 368-374.

Ruland, K., Wenzel, R., and Herrmann, R. (1990). Analysis of three different repeated DNA elements present in the P1 operon of *Mycoplasma pneumoniae*: size, number, and distribution on the genome. Nucleic Acids Res. 18, 6311-6317.

Sadosky, A.B., Davidson, A., Lin, R.-J., and Hill, C.W. (1989). *rhs* gene family of *Escherichia coli* K-12. J. Bacteriol. 171, 636-642.

Sadowsky, M. J., and Hur, H.-G. (1998) Use of endogenous repeated sequences to fingerprint bacterial genomic DNA. *In:* J.R. Lupski, G. Weinstock, and F. J. de Bruijn (eds.). Bacterial Genomes: Structure and Analysis. Chapman and Hall, pp. 399-413..

Sadowsky, M. J., L. L. Kinkel, J. H. Bowers, and Schottel, J.L.. (1996). Use of repetitive intergenic DNA sequences and fatty acid profiles to classify pathogenic and disease-suppressive strains of *Streptomyces*. Appl. Environ. Microbiol. **62:**3489-3493.

Sadowsky, M.J., Tully, R.E., Cregan, P.B., and Keyser, H.H. (1987). Genetic diversity in *Bradyrhizobium japonicum* serogroup 123 and its relation to genotype-specific nodulation of soybeans. Appl. Environ. Microbiol. 53, 2624-2630.

Scherer, S., and Stevens, D.A. (1987). Application of DNA typing methods to epidemiology and taxonomy of *Candida* species. J. Clin. Microbiol. 25, 675-679.

Schmidt, E.L., Zidwick, M.J. and Abebe, H.H. (1986). *Bradyrhizobium japonicum* serocluster 123 and diversity among member isolates. Appl. Environ. Microbiol. 51, 1212-1215.

Schneider, M., and de Bruijn, F.J. (1996). Rep-PCR mediated genomic fingerprinting of rhizobia and computer-assisted phylogenetic analysis. World J. Microbiol. Biotechnol. 12, 163-174.

Skjold, S., Quie, P.G., Fries, L.A., Barnham, M., and Cleary, P.P. (1987). DNA fingerprinting of *Streptococcus zooepidemicus* (Lancefield group C). as an aid to epidemiological study. J. Infect. Dis. 155, 1145-1150.

Sharples, G.J., and Llod, R.G. (1990). A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. Nucleic Acids Res. 18, 6503-6508.

Standard Methods for the Examination of Water and Wastewater (1995). American Public Health Association, Washington, D.C.

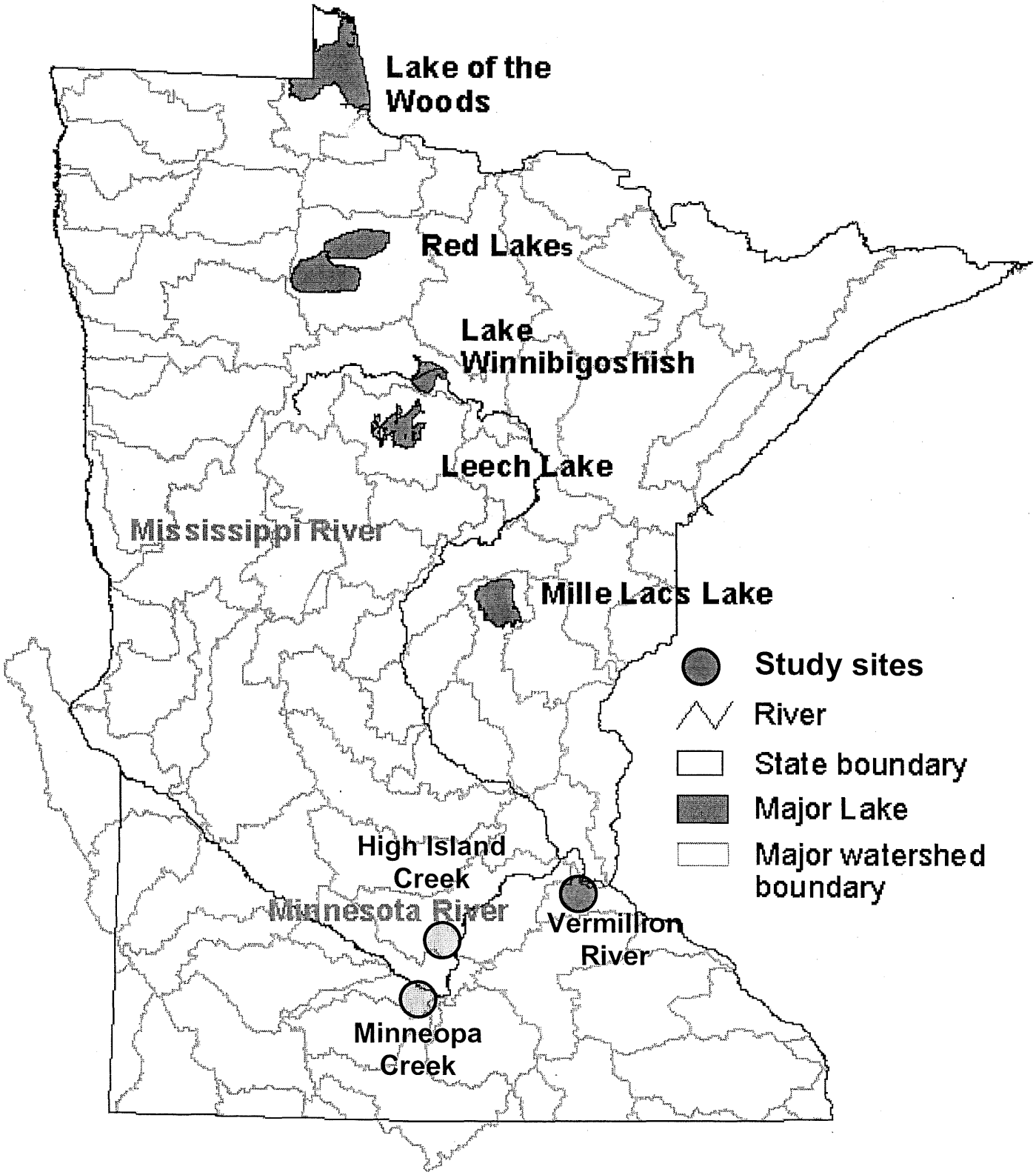
Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C., and Higgins, C.F. (1984). Repetitive extragenic palindromic sequences, a major component of the bacterial genome. Cell 37, 1015-1026.

Tompkins, L.S., Troup, N.J., Woods, T., Bibb, W., and McKinney, R.M. (1987). Molecular epidemiology of *Legionella* species by restriction endonuclease and alloenzyme analysis. J. Clin. Microbiol. 25, 1875-1880.

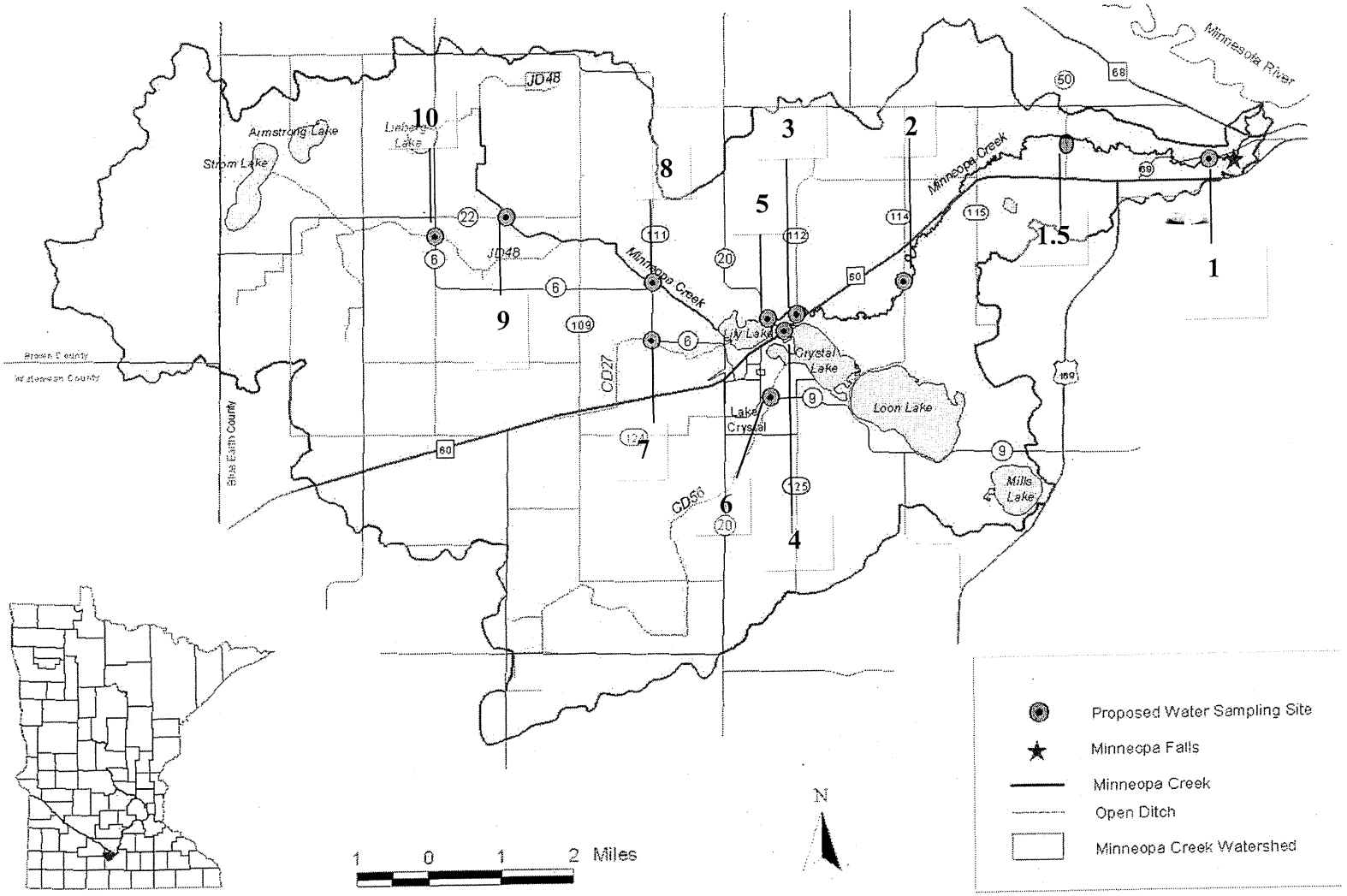
Versalovic, J., de Bruijn, F.J., and Lupski, R. (1998). Repetitive sequence-based PCR (rep-PCR) DNA fingerprinting of bacterial genomes. *In:* J.R. Lupski, G. Weinstock, and F. J. de Bruijn (eds.). Bacterial Genomes: Structure and Analysis. Chapman and Hall, pp. 437-454.

Versalovic, J., Koeuth, T., and Lupski, J.R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res. 24, 6823-6831.

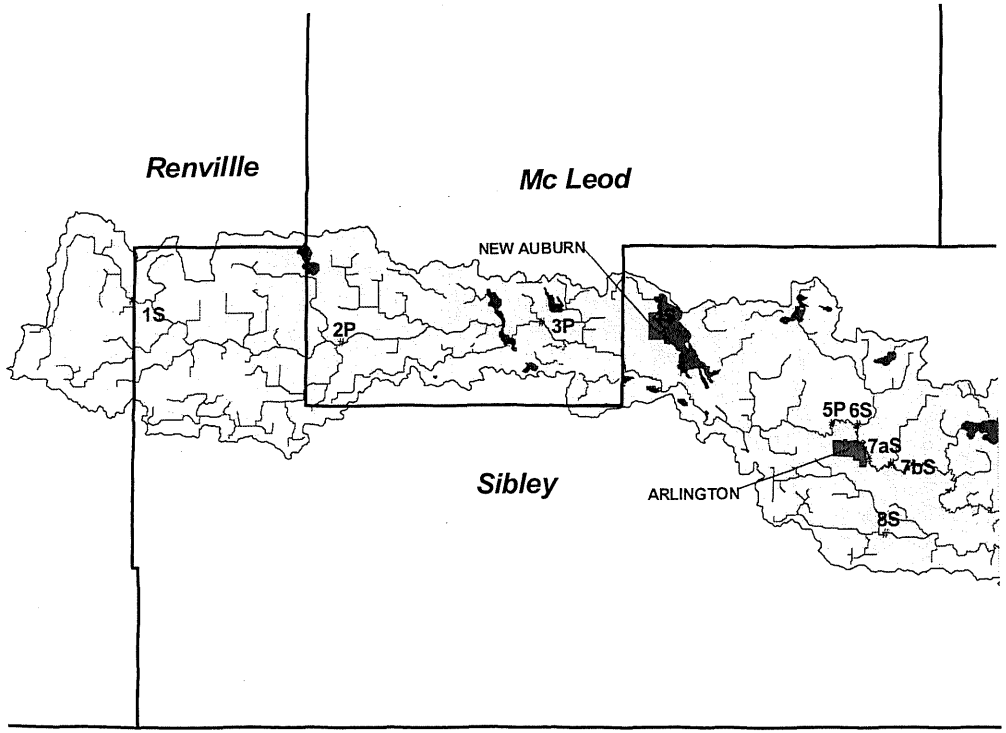
Versalovic, J., Schnieder, M., de Bruijn, F.J., and Lupski, J.R. (1994). Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Meth. Molec. Cell. Biol. 5. 25-40.


Lake of the Woods
Red Lakes
Lake Winnibigoshish
Leech Lake
Mississippi River
Mille Lacs Lake
Study sites
River
State boundary
Major Lake
Major watershed boundary
High Island Creek
Minnesota River
Vermillion River
Minneopa Creek

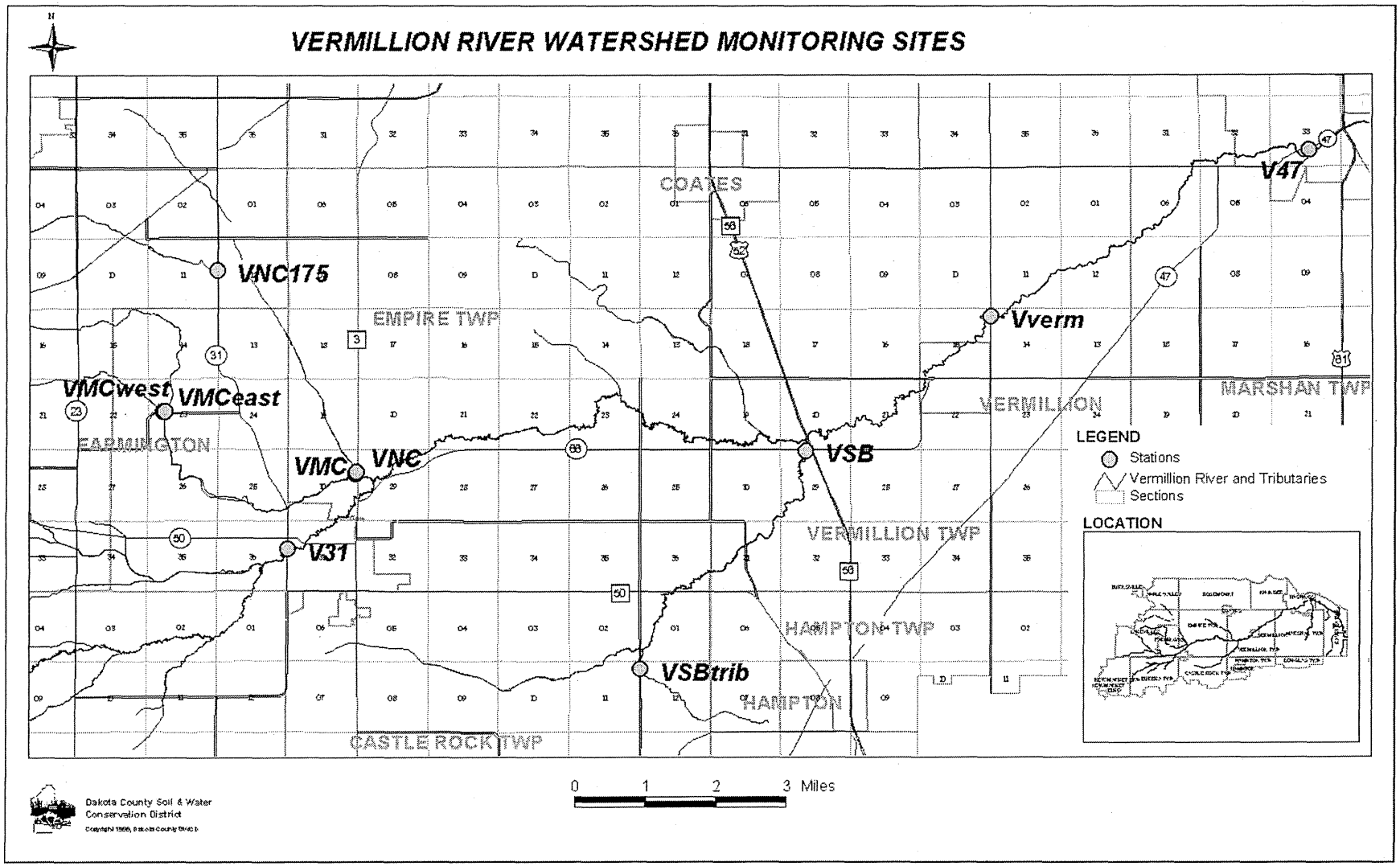
Minneopa Creek Watershed - Sampling Sites



High Island Watershed - Sample Sit



Prepared
High Islan
ArcView 3
October 6


VERMILLION RIVER WATERSHED MONITORING SITES
N
VNC175
Vverm
V47
VMCwest
VMCeast
VMC
VNC
VSB
V31
VSBtrib
COATES
EMPIRE TWP
FARMINGTON
VERMILLION
MARSHAN TWP
VERMILLION TWP
HAMPTON TWP
HAMPTON
CASTLE ROCK TWP
LEGEND
Stations
Vermillion River and Tributaries Sections
LOCATION
0 1 2 3 Miles
Dakota County Soil & Water Conservation District

REVISED AEM 2391-03 – IN PRESS

Sample Size, Library Composition, and Genotypic Diversity Influence Accuracy of Determining Sources of Fecal Pollution Among Natural Populations of *Escherichia coli* from Different Animals

LeeAnn K. Johnson[†], Mary B. Brown, Ethan A. Carruthers, John A. Ferguson, Priscilla E. Dombek and Michael J. Sadowsky[*]

Department of Soil, Water, and Climate, University of Minnesota, St. Paul, MN 55108

Running Title: Diversity Among Natural *E. coli*

[*]Author to whom correspondence should be addressed:

Michael J. Sadowsky
University of Minnesota
Dept. of Soil, Water, and Climate
1991 Upper Buford Circle, 439 Borlaug Hall
St. Paul, MN 55108
Phone: 612-624-2706
FAX: 612-625-2208
Email: Sadowsky@soils.umn.edu

[†]Current Address: Minnesota Department of Agriculture, 90 West Plato Boulevard, Saint Paul, Minnesota 55107

ABSTRACT

A horizontal, fluorophore-enhanced, rep-PCR DNA fingerprinting technique (HFERP) was developed and evaluated as a means to differentiate human from animal sources of *Escherichia coli*. Box A1R primers and PCR was used to generate 2,466 rep-PCR and 1,531 HFERP DNA fingerprints from *E. coli* strains isolated from fecal material from known human and 12 animal sources: dogs, cats, horses, deer, geese, ducks, chickens, turkeys, cows, pigs, goats, and sheep. HFERP DNA fingerprinting reduced within gel grouping of DNA fingerprints and improved alignment of DNA fingerprints between gels, relative to that achieved using rep-PCR DNA fingerprinting. Jackknife analysis of the complete rep-PCR DNA fingerprint library, done using Pearson's product-moment correlation coefficient, indicated that animal and human isolates were assigned to the correct source groups with a 82.2% average rate of correct classification. However, when only unique isolates were examined, isolates from a single animal having a unique DNA fingerprint, Jackknife analysis showed that isolates were assigned to the correct source groups with a 60.5% average rate of correct classification. The percentage of correctly classified isolates were about 15 and 17% greater for rep-PCR and HFERP, respectively, when analyses were done using the curve-based Pearson's product-moment correlation coefficient, rather than the band-based Jaccard algorithm. Rarefaction analysis indicated that despite the relatively large size of the known source database, genetic diversity in *E. coli* was very great, and is most likely accounting for our inability to correctly classify many environmental *E. coli* isolates. Our data indicate that removal of duplicate genotypes within DNA fingerprint libraries, increased database size, proper method of statistical analysis, and correct alignment of band data within and between gels improves the accuracy of microbial source tracking methods.

INTRODUCTION

Protection of humans from pathogen contamination is dependent on the purity of waters designated for recreation, drinking, and shellfish harvesting. Bacterial pathogens have been listed as major pollutants in rivers, streams, and estuaries (37). Restoration of polluted water is currently being accomplished through the development of Total Maximum Daily Loads (TMDLs). Source assessment is an important component of TMDL development in which pollutants are identified and characterized by type, magnitude, and location (38). The implementation of TMDLs has provided one of the driving forces for the development of methods to distinguish between human and animal sources of fecal pollution. Sources of fecal coliform bacteria may include runoff from feedlots and manure-amended agricultural land, wildlife, inadequate septic systems, urban runoff, and sewage discharges.

Both phenotypic and genotypic methods have been explored as means to study the ecology of fecal bacteria related to host specificity, and determining potential sources of fecal bacteria found in surface water (6,32,34). The most widely investigated bacteria for these studies have been *Escherichia coli* and *Enterococcus* sp. strains. The use of these methods is based on the hypothesis that specific strains, or a strain's phenotypic or genetic attributes, are related to specific host animals. This hypothesis, however, has only been tested in a limited manner.

The majority of phenotypic and genotypic methodologies require the construction of known-source libraries (a host origin database) to differentiate among isolates, which is subsequently used to determine the host origin of unknown environmental isolates (34). However, in most cases, the size of the host origin databases are rather limited, consisting of 35 to about 500 isolates (2-4,6,9,12,13,23,24-26,31,33,42,43), making broader comparisons to

larger populations of *E. coli* and *Enterococcus* in the environment difficult. In addition, temporal and geographic variation in bacterial genotypes within and between animal species (7,12,16,31), multiple strains within a single animal (23), and diet variation within a host animal (13) have been shown to influence the representativeness of known source libraries. Moreover, while microbial source tracking studies done using phenotypic approaches and antibiotic resistance patterns have frequently used large known-source libraries, consisting of about 1000 – 6,000 isolates (2,8,10,15,44-46), many of the strains examined were isolated from the same source material or sample, and thus libraries may be biased due to the presence of multiple replications (clones) of the same bacterial genotype.

The rep-PCR DNA fingerprinting technique uses the polymerase chain reaction and primers based on highly conserved and repetitive nucleotide sequences to amplify specific portions of the microbial genome (22,29,40,41). When the PCR products are separated by agarose gel electrophoresis and visualized following staining with ethidium bromide, the resulting banding patterns produce a "fingerprint" unique to each strain. The rep-PCR technique has proven to be a valuable tool to identify and track medically and environmentally important microorganisms (5,17,30,40), and it has also been recently evaluated for its use as a source-tracking tool (1,4,6,20,23). The rep-PCR DNA fingerprinting technique is relatively quick, easy, and inexpensive to perform, and lends itself to high throughput applications, making it an ideal method for microbial source-tracking studies.

Initial studies done in our laboratory indicated that rep-PCR done with Box A1R primers and *E. coli* yielded a more consistent and complex DNA fingerprints than did studies done using REP primers (6). However, rep-PCR reactions done with Box, ERIC, and REP primers have all been evaluated in microbial source-tracking studies (1,4,6,23). Dombek *et al.* (6) used a minimal

data set consisting of about 200 non-unique *E. coli* isolates and reported that 100% of chicken and cow isolates, and between 78-90% of human, goose, duck, pig and sheep isolates were correctly assigned to host source groups using rep-PCR DNA fingerprinting and Box AIR primers. Similarly, Carson *et al.* (4) reported that rep-PCR DNA fingerprinting done using Box A1R produced a 96.6% average rate of correct classification (ARCC) for human and non-human *E. coli* isolates, and McLellan *et al.* (23) reported a 79.3% ARCC for *E. coli* analyzed using rep-PCR done and REP primers.

While all these initial analyses indicated that the rep-PCR technique may be useful for determining animal sources of *E. coli*, these studies were done with relatively small datasets. Moreover, since rep-PCR, and most other source tracking methods, require the assembly of libraries of known-source fingerprints, which is labor-intensive and time-consuming, it is very important that the fingerprint database is unbiased, has high fidelity (36), and is representative of the diversity of *E. coli* potentially present in animal hosts and in environmental samples.

rep-PCR DNA fingerprints are usually analyzed using statistical tools. Binary similarity coefficients are used to analyze presence/absence data (19), and simple banding data obtained from DNA fingerprints can be analyzed using binary coefficients such as Dice or Jaccard band matching algorithms. However, more quantitative algorithms, such as Pearson's product-moment correlation coefficient, can also be applied to complex DNA banding patterns, such as those found using rep-PCR. In this case, fingerprints are analyzed as densitometric curves, taking into account both peak position and height (intensity) (11).

In this study we created a large, known-source, rep-PCR and horizontal fluorophore-enhanced rep-PCR (HFERP) DNA fingerprint databases from 2,466 *E. coli* isolates obtained from humans and 12 animal sources: cows, pigs, sheep, goats, turkeys, chickens, ducks, geese,

deer, horses, dogs, and cats, and evaluated the usefulness of these method to differentiate human from animal sources of fecal *E. col.*

MATERIALS AND METHODS

Isolation of *E. coli* from known animal sources. Fecal samples, representing humans and 12 animal source groups, were collected from wild and domesticated animals throughout Minnesota and western Wisconsin. Fresh fecal material was collected from individual animals as previously described (6) by swabbing the rectal or cloacal region using a Culturette7 swab transport system (BD Diagnostic System, Sparks, MD), or by collecting freshly voided feces with a sterile tongue depressor. Fecal samples were placed into sterile Whirl-Pak® bags (Nasco, Fort Atkinson, WI) and kept at 4°C until processed, usually within 6 hr. Fecal material was streaked onto mFC agar plates (Difco, BD Diagnostic Systems, Sparks, MD), and incubated at 44.5°C for 24 hours. Characteristic blue colonies (usually six) from mFC plates were picked and evaluated using selective and differential media as previously described (6). Isolates were used for subsequent studies if growth and color responses on all media were typical for *E. coli*. Isolates giving atypical responses for colony color on all media or MUG reaction were further screened using API 20E test kits (bioMerieux, Inc., St. Louis, MO). Isolates yielding a "good" to "excellent" *E. coli* identification by the API 20E kit were used for DNA fingerprinting. Three *E. coli* colonies from each individual fecal sample were used for DNA fingerprinting and were stored at –80°C in 50% glycerol.

***E. coli* preparation and rep-PCR conditions**. *E. coli* isolates were streaked onto Plate Count Agar (Difco, BD Diagnostic Systems, Sparks, MD) and grown overnight at 37°C. Single colonies were picked with a 1 µl sterile inoculating loop (Fisher Scientific, Pittsburgh, PA), suspended in 100 µl of distilled H_2O in 96-well microtiter plates, and 2 µl of the resulting suspension was used

as template for PCR. The rep-PCR fingerprints were obtained using the Box A1R primer: 5'-CTACGGCAAGG CGACGCTGACG-3', and PCR reactions were done as described previously (6,27,28). PCR was performed using an MJ Research PTC 100 (MJ Research, Waltham, MA) using the protocol specific for this thermocyclers and the Box A1R primer. PCR was initiated with an incubation at 95°C for 2 minutes, followed by 30 cycles, consisting of 94°C for 3 seconds, 92°C for 30 seconds, 50°C for 1 minute, and 65°C for 8 minutes (27). PCR reactions were terminated after an extension at 65°C for 8 min, and stored at 4°C. Reactions that were not used immediately for gel electrophoresis analysis were stored at -20°C.

Electrophoresis was done at 4°C for 17-18 hours at 70V with constant buffer recirculation (6,27). Gels were stained for 20 min in 0.5 μg/ml ethidium bromide prepared in 0.5× TAE buffer. Gel images were captured as TIF files using a FOTO/ Analyst Archiver electronic documentation system (Fotodyne Inc., Hartland, WI).

HFERP studies. Horizontal, fluorophore-enhanced, rep-PCR (HFERP) analyses were performed using a modification of the procedures of Versalovic *et al.* (39) as follows: Single *E. coli* colonies were picked with a 1 μl sterile inoculating loop (Fisher Scientific, Pittsburgh, PA), suspended in 100 μl of 0.05 M NaOH in 96-well, low profile, PCR plates (MJ Research, Waltham, MA), heated to 95 °C for 15 min, and centrifuged at 640 RPM for 10 min in a Hermle/Labnet Z383K centrifuge (Edison, NJ). A 2 μl aliquot of the supernatant in each well was used as template for PCR using the protocol described above for rep-PCR. The primer consisted of a mixture of 0.09 μg of unlabeled Box A1R primer per μl and 0.03 μg of 6-FAM fluorescently labeled Box A1R primer per μl (Integrated DNA Technologies, Coralville, IA). The primer mixture was used at a final concentration of 0.12 μg per 25 μl PCR reaction. A 6.6 μl aliquot of a mixture of 50 μl Genescan-2500 ROX internal lane standard (Applied Biosystems,

Foster City, CA) and 200 µl non-migrating loading dye (150 mg Ficoll 400 per ml, and 25 mg blue dextran per ml) was added to each 25 µl PCR reaction prior to loading the PCR reaction into agarose gels, 12 µl of the resulting mixture was loaded per gel lane. DNA fragments were separated as described for rep-PCR, and HFERP images were captured using a Typhoon 8600 Variable Mode Imager (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) operating in the fluorescence acquisition mode using the following settings: green (532 nm) excitation laser; 610 BP 30 and 526 SP emission filters in the autolink mode with 580 nm beam splitter; normal sensitivity; 200 micron/pixel scan resolution; + 3 mm focal plane; and 800 V power.

Computer-assisted rep-PCR fingerprint analysis. Separated gel images (ROX-stained standards and HFERP banding patterns) were processed using ImageQuant image analysis software (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) and converted to 256 gray scale TIF images. Gel images were normalized and analyzed using BioNumerics v.2.5 software (Applied-Maths, Sint-Martens-Latem, Belgium). Rep-PCR gel lanes were normalized using the 1 kb ladder from 298 bp to 5090 bp, as external reference standards, while HFERP gel lanes were normalized using the Genescan-2500 ROX internal lane standard from 287 bp to 14,057 bp. Band matching for rep-PCR DNA fingerprints was accomplished by using the following BioNumerics settings: minimum profiling 5%, gray zone 5%, minimum area 0%, and shoulder sensitivity of 5; while band matching for HFERP DNA fingerprints was done by using 3% minimum profiling, 0% gray zone, 0% minimum area, and 0 shoulder sensitivity. DNA fingerprint similarities were calculated by using either the curve-based cosine or Pearson's product-moment correlation coefficient, with 1% optimization, or the band-based Jaccard coefficient. Dendrograms were generated using the unweighted pair-group method using

arithmetic means (UPGMA). The percentages of known-source isolates assigned to their correct source group were calculated by using Jackknife analysis, with maximum similarities (9).

RESULTS AND DISCUSSION

Evaluation of isolates. Of the 2,672 *E. coli* strains obtained from known human and animal sources using an array of selective and differential plating media, 219 isolates gave at least one atypical result when examined by routine biochemical screening tests, the wrong color on indicator media or an incorrect MUG reaction. The biochemical characteristics of these isolates were examined further by using the API 20E system. Results of this analysis indicated the majority of these isolates, 167, were bona fide *E. coli,* while the remainder, 52, could not be confirmed as this bacterium. The latter group was not used in rep-PCR analysis or included in the DNA fingerprint database.

Influence of duplicate *E. coli* strains on classification of known source library. Since results from several studies suggest that *E. coli* is genetically diverse and clonal in origin, and that this may influence the usefulness of this bacterium for source tracking studies (7), we evaluated this technology using a large library of *E. coli* obtained from humans and 12 animal sources collected throughout Minnesota and Western Wisconsin (Table 1).

2,466 high-quality rep-PCR DNA fingerprints were generated using the Box A1R primer and template DNA from *E. coli* strains obtained from the 13 human and animal sources (Table 1). About 25-40 PCR product bands were obtained from the *E. coli* isolates using rep-PCR. Jackknife analysis performed on the 2,466 DNA fingerprints from the entire known-source rep-PCR DNA fingerprint database, using Pearson's product-moment correlation coefficient, indicated that 69-97% of animal and human *E. coli* isolates were assigned into correct source

groups (Table 2). This corresponds to an 82.2% average rate of correct classification for the 2,466 rep-PCR DNA fingerprints.

Increasing the size of the known source library to 2,466 isolates, however, did not necessarily lead to an increase in the ability to correctly assign strains to the correct source group. In fact, the average rate of correct classification decreased 4.2% using the larger library reported here, relative to what was seen using a smaller library in our previous studies (6). This may in part be due to the uncovering of increased genetic diversity among isolates, increased accumulation of errors due to gel-to-gel variation, or the presence of duplicate genotypes (DNA fingerprints) from the same individual within our original library.

Since identical DNA fingerprints from *E. coli* strains obtained from the same individual most likely represent isolates of clonal origin and can artificially bias subsequent analyses, we eliminated duplicate DNA fingerprints originating from *E. coli* strains obtained from the same individual human or source animal. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%.

Results in Table 1 show that of the 2,466 DNA fingerprints analyzed, 1,535 (62%) remained in the "unique" DNA fingerprint library. The influence of duplicate DNA fingerprints on the correct classification of library strains is shown in Table 2. When the 1,535 DNA fingerprints from the unique *E. coli* isolates were examined, Jackknife analyses indicated that only 44-74% of the isolates were assigned to the correct source group; with an average rate of correct classification of 60.5% (Table 2). Thus, there was a 21.7% reduction in the average rate of correct classification by using the unique DNA fingerprint library, relative to that seen with the complete library, and less than we and

others have previously reported using smaller libraries of *E. coli* strains containing duplicate DNA fingerprints from the same individual animal (4,6,23). Our results indicate that the clonal nature of *E. coli* (11,20,33) originating from the same source animal artificially biases the average rate of correct classification, alters the fidelity of the database, and overestimates the ability of the database to assign isolates to their correct source group.

Influence of library size on usefulness of DNA fingerprint libraries. We also determined whether *E. coli* isolates obtained in this study were sufficient to capture the genetic diversity present within the *E. coli* populations sampled. *E. coli* isolates between animal source groups with rep-PCR DNA fingerprint similarities of 90% or greater (based on cosine coefficient, 1% optimization and UPGMA) were assigned to the same genotype. By this definition, 657 genotypes were distinguished from the 1,535 unique *E. coli* isolates in the known-source database. The isolates were randomized, and a rarefaction curve was constructed by summing the number of genotypes that accumulated with the successive addition of isolates. Despite a library size of 1,535 DNA fingerprints, genetic diversity has not been saturated. This is evidenced by the apparent first order relationship between isolate numbers (sampling effort) and accumulation of new genotypes (data not shown). Moreover, 58.75% of the genotypes from isolated strains, across all animal groups, occurred only once in the database and a limited number occurred multiple times (Figure 1).

Since our rarefaction curve did not become asymptotic, our data cannot be used to predict the ultimate size that our fingerprint library needs to be. However, our data indicate that with our current library size, each new isolate added to the library has only about a 50% chance of being new. It has been suggested that a library size of 20,000 to 40,000 isolates

may be needed to capture all the genetic diversity present in *E. coli* (Mansour Samadpour, personal communication). Taken together, our data show that the use of relatively small-sized libraries, that do not take into account the tremendous genetic diversity present in *E. coli* (7,14,23,35) and enterococci, will make broader comparisons to larger populations of these organisms in the environment difficult.

One suggested strategy to avoid this under-representation problem in large regional or national libraries, is to develop moderate sized libraries for a highly confined geographical region, wherein isolates are only obtained from the animals in the study area. In this way only animals pertinent to the study site, and those likely to have an impact on the targeted watershed, need to be examined in detail. However, it is also important to note that in some cases animals thought to be important or prevalent to the study site may vary over time, depending on agricultural practices and migration. Thus, a careful inventory of potential animals in the study site needs to be made prior to, and during, sampling and analysis.

HFERP DNA fingerprinting. In our studies we noted that cluster analysis of rep-PCR DNA fingerprint data often produced groupings that were more related to the gels from which they originated, than the host animal from which they were isolated. We hypothesized that within-gel clustering of DNA fingerprints was in part due to intrinsic gel-to-gel variation, differential DNA migration in repeated runs of the same and different PCR samples, and the inability to correct for heat and buffer-induced gel distortion across and between single and multiple gels. Since DNA fingerprint libraries are assembled from many different gels, this could have a major impact on the fidelity of

DNA fingerprint libraries and their subsequent use for tracking sources of unknown fecal bacteria.

To overcome these major limitations, we developed and evaluated the use of a horizontal, fluorophore-enhanced, rep-PCR (HFERP) technique as a means to differentiate human from animal sources of fecal bacteria. In this method, alignment, correction, and normalization of fluorescently-labeled, rep-PCR DNA fingerprint bands within and between gels is facilitated by the use of internal ROX-labeled molecular weight markers that are present in each lane. The technique is similar to that previously described for use with a DNA sequencer (27,39), but instead uses a standard horizontal agarose gel and a dual-wavelength scanner. An example of an unseparated HFERP gel displaying the ROX-labeled internal lane standard and 6-FAM-labeled Box A1R DNA fingerprints is shown in Figure 2A, and the separated gel images are shown in Figures 2B and 2C. Typically, and with our *E. coli* strains, 12 to 20 DNA bands per strain were revealed using the HFERP technique.

To test whether HFERP reduced within-gel groupings of DNA fingerprints, we analyzed DNA fingerprints from 40 *E. coli* strains obtained from dogs on 2 different gels using Pearson's product-moment coefficient. Results of these studies indicated that rep-PCR DNA fingerprints from strains run on the same gel were, on average, 50% (range 29 – 57%) more likely to be grouped together as the same strains analyzed by using the HFERP technique (data not shown). This indicates that HFERP method considerably reduces within gel grouping of DNA fingerprints. In addition, the HFERP method reduced alignment difficulties due to within- and between-gel variation in band migration found with rep-PCR gels (Figure 3).

The repeatability of the rep-PCR and HFERP DNA fingerprinting methods was also examined by fingerprinting a single, reference, control *E. coli* strain (pig isolate number 294) that was included on each gel. DNA fingerprints from 29 and 41 repetitions of *E. coli* control pig strain 294, each from a separate gel, were generated by using the rep-PCR and HFERP methods, respectively. When analyzed using the curve-based Pearson's correlation coefficient, the rep-PCR DNA fingerprints had an average similarity of 88%, whereas the HFERP-derived DNA fingerprints had an average similarity of 92%. Previously, Versalovic, *et al.* (39) and Rademaker, *et al.* (27) reported on the use of fluorophore-enhanced rep-PCR (FERP), whereby polyacrylamide gel electrophoresis and automated DNA sequencers were used to separate and detect bands generated by the FERP protocol. While the more automated method presented by these authors has some advantages, the increased cost of analyses and the limited dynamic range of fragment size separation on sequencing gels did not make this technique useful in our applications. In contrast, the HFERP method described here is relatively inexpensive to perform, can be done on standard electrophoresis apparatus, has high throughput, and allows for the separation of a large range of DNA band sizes. It should be noted, however, that the intensity of HFERP bands are more variable than those generated by rep-PCR, and that some of the gains achieved by more precise alignment of bands may be offset by more variation in band intensity. We found that this variation in intensity can be overcome by the careful mixing of all reagents in the PCR master mix and greater pipetting precision when loading gels (data not presented). Further improvements to increasing the intensity of HFERP-generated DNA fingerprints may also be obtained by varying the ratio of labeled to unlabeled primer and the final concentration of the primer mixture in PCR

reactions. Nevertheless, our results clearly show that HFERP-derived DNA fingerprint bands are more precisely aligned than the rep-PCR bands and reduce within gel groupings of fingerprints, which can have profound ramifications for the assembly of libraries and the analysis of unknown environmental isolates. This technology will have application to other DNA fingerprinting methods that rely on the use of PCR primers.

Assignment of *E. coli* isolates to source groups using HFERP DNA fingerprints. Of the 1,535 previously selected unique *E. coli* isolates from animals and humans (Table 1), 1,531 were subjected to HFERP DNA fingerprinting using a combination of fluorescently labeled and unlabeled Box A1R PCR primers. Jackknife analyses of HFERP gels done using the curve-based Pearson's correlation coefficient indicated that 38-73% of the isolates were assigned to the correct source group using this technique (Table 2). For the curve-based analysis, the HFERP technique had the lowest percent of correctly classified strain in cases where the numbers of analyzed fingerprints were relatively small (for sheep, horses, and goats). The average rate of correct classification for the unique HFERP-generated DNA fingerprints was 59.9%.

In contrast, Jackknife analyses of HFERP-generated DNA fingerprints done using the band-based Jaccard analysis showed that only 8-56% of the *E. coli* isolates were assigned to the correct source group, with a 43.0% average rate of correct classification. This indicates that for this type of data, the Pearson's product-moment correlation coefficient was superior to Jaccard's band matching algorithm for assigning known isolates to the correct source groups. Interestingly, results in Table 2 also show that despite problems associated with within- and between-gel variation, within-gel grouping of isolates, and repeatability issues, Jackknife analysis of rep-PCR DNA fingerprints, analyzed using

Pearson's correlation coefficient, indicated that 48-74% of the isolates were assigned to the correct source group, a 60.9% average rate of correct classification.

While band-matching data obtained from DNA fingerprints can be analyzed using binary similarity coefficients, which are mostly used to analyze presence/absence data (19), quantitative similarity coefficients, which require a measure of relative abundance (18), can also be applied to DNA fingerprints if they are analyzed as densitometric curves that take into account both peak position and intensity (peak height). Results of our analysis of rep-PCR DNA fingerprint data indicated that the Jaccard band-based method was not as useful in separating *E. coli* isolates into their correct source group as was the curve-based quantitative method. This is similar to results reported by Häne, et al. (11) who demonstrated that for complex DNA fingerprints, such as those produced with the techniques we used here, a curve-based method such as Pearson's product-moment correlation coefficient more reliably identified similar or identical DNA fingerprints than band matching formulas, such as simple matching, Dice, or Jaccard. Similarly, Louws and co-workers (21) reported that curve-based statistical methods worked best for analysis of complex banding profiles generated by rep-PCR, since comparison of curve data is less dependent on DNA concentration in loaded samples and is relatively insensitive to background differences in gels. More recently, Albert et al. (1) performed a statistical evaluation of rep-PCR DNA fingerprint data and reported that k-nearest neighbor's classification was similar to Pearson's product-moment coefficient in its ability to correctly classify fingerprints of 584 E. coli isolates.

Groupings of fingerprint data. In some instances, it may be sufficient to identify unknown watershed *E. coli* isolates to larger groupings, rather than to individual animal

types. To determine if the HFERP-generated DNA fingerprint data from our library of unique *E. coli* isolates grouped well into larger categories, we assembled DNA fingerprints from pets (dogs and cats), domesticated animals (chickens, cows, goats, horses, pigs, sheep, and turkeys), wild-life (deer, ducks, and geese), and humans, and used Jackknife analysis to assess the percent of correctly classified strains. Results in Table 3 show that the HFERP DNA fingerprints, analyzed using Pearson's product-moment correlation coefficient, correctly classified about 83, 54, 71, and 59% of the isolates into the domesticated, human, wildlife, and pet categories, respectively. The average rate of correct classification for these groups was 74.3%. In contrast, when DNA fingerprints were analyzed using Jaccard's coefficient, the average rate of correct classification was 66.2%. As before, the least precision was found in categories having the smallest number of fingerprints, pets and humans, suggesting that there is an apparent relationship between the number of fingerprints analyzed and the percentage of correctly classified isolates.

In microbial source tracking studies it may often be useful to determine if unknown isolates belong to either animal or human source groups, rather than to more specific categories. Results in Table 4 show that about 94% and 54% of *E. coli* from animals and humans, respectively, were assigned to the correct source groups using HFERP-generated DNA fingerprints and Pearson's correlation coefficient. The average rate of correct classification was

88.2 and 86.1% for analyses done using Pearson's and Jaccard's algorithms, respectively. The lower percentage of correctly classified human isolates may, in part, be due to the smaller size of fingerprints analyzed for this category. Taken together, these results

indicated that 1, broader classifications of source groups should be used when appropriate, or 2, a targeted subset of the DNA fingerprint database should be used to more precisely determine sources of fecal pollutants in watersheds where specific source groups are known to be present. The pooling of source groups into a more limited number of categories has previously been shown to increase the average rate of correct classification following discriminant analysis of antibiotic resistance (10,15,45), ribotype (3,4), and rep-PCR DNA fingerprint analyses (4).

In summary, our results suggest that HFERP-generated Box A1R DNA fingerprints of *E. coli* are useful to differentiate between different *E. coli* subtypes of human and animal origin and that this method reduces within gel groupings of DNA fingerprints, and ensures more proper alignment and normalization of fingerprint data. However, our results further indicate that other important issues must also be resolved to more fully understand the potential applications and limitations of this and other library-based microbial source tracking methodologies. Among these are questions concerning the inclusion of identical DNA fingerprints from the same animal in the library, the number of fingerprints that must be included in an *E. coli* known source library to adequately capture the diversity of *E. coli* genotypes that exist among potential host animals, and ultimately, whether *E. coli* exhibits a sufficient level of host specificity to allow unambiguous assignment of unknown environmental *E. coli* to specific host animals.

ACKNOWLEDGEMENTS

This work was supported, in part, by funding from the Legislative Commission on Minnesota Resources through the Environment and Natural Resources Trust Fund and the MN Future Resource Fund, the Metropolitan Council Environmental Services, and The University of Minnesota Agricultural Experiment Station (to M.J.S.).

We would like to thank Priscilla Dombek for DNA fingerprinting assistance, Linda Kinkel and Anita Davelos for help with rarefaction analysis, and Vivek Kapur, Todd Markowski and Bruce Witthun for help with laser image analysis.

REFERENCES

1. **Albert, J. M., J. Munkata-Marr, L. Tenorio, and R. L. Siegrist.** 2003. Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli.* Environ. Sci. technol. **37:**4554-4560.

2. **Burnes, B. S.** 2003. Antibiotic resistance analysis of fecal coliforms to determine fecal pollution sources in a mixed-use watershed. Environ. Monit. Assess. **85:**87-98.

3. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and A. Asfaw.** 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. Appl. Environ. Microbiol. **67:**1503-1507.

4. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and J. D. Schnell.** 2003. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. Appl Environ Microbiol **69:**1836-1839.

5. **de Bruijn, F. J.** 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. Appl. Environ. Microbiol. **58:**2180-2187.

6. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66:**2572-2577.

7. **Gordon, D. M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiol. **147:**1079-1085.

8. **Graves, A. K., A. T. Hagedorn, M. Mahal, A. M. Booth, and R. B. Reneau Jr.** 2002. Antibiotic resistance profiles to determine sources of fecal contamination in a rural Virginia watershed. J. Environ. Qual. **31:**1300-1308.

9. **Guan, S., R. Xu, S. Chen, J. Odumeru, and C. Gyles.** 2002. Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources. Appl. Environ. Microbiol. **68:**2690-2698.

10. **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.

11. **Häne, B.G., K. Jäger, and H. Drexler.** 1993. The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprinting profiles than band matching algorithms. Electrophor. **14:**967-972.

12. **Hartel, P. G., J. D. Summer, J. L. Hill, J. Collins, J. A. Entry, W. and I. Segars.** 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. Environ. Qual. **31:**1273-1278.

13. **Hartel, P. G., J. D. Summer, and W. I. Segars.** 2003. Deer diet affects ribotypo diversity of *Escherichia coli* for bacterial source tracking. Water Res. **37:**3263-3268.

14. **Hartl, D. L., and D. E. Dykhuizen.** 1984. The population genetics of *Escherichia coli.* Annu. Rev. Genet. **18:**31-68.

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

16. **Jenkins, M. B., P. G. Hartel, T. J. Olexa, and J. A. Stuedemann.** 2003. Putative temporal variability of *Escherichia coli* ribotypes from yearling steers. J Environ Qual. **32:**305-309.

17. **Judd, A. K., M. Schneider, M. J. Sadowsky, and F. J. de Bruijn.** 1993. Use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. Appl. Environ. Microbiol. **59:**1702-1708.

18. **Krebs, Charles J**. 1999. Ecological methodology. Benjamin/Cummings, Menlo Park, CA.

19. **Legendre, P., and L. Legendre.** 1998. Numerical Ecology, 2nd English edition. Elsevier Science, Amsterdam, The Netherlands.

20. **Lipman, J. A., A. de Nijs, T. J. G. M. Lam, and W. Gaastra.** 1995. Identification of *Escherichia coli* strain from cows with clinical mastitis by serotyping and DNA polymorphism patterns with REP and ERIC primers. Vet. Microbiol. **43:**13-19.

21. **Louws, F. J, J. L. W. Rademaker, and F. J. de Bruijn.** 1999. The three Ds of PCR-based genomic analysis of phytobacteria: diversity, detection, and disease diagnosis. Ann. Rev. Phytopath. **37:**81-125.

22. **Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D. A. Morrison, G. J. Boulnois, and J.-P. Claverys.** 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucl. Acids Res. **20:**3479-3483.

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

24. **Nebra, Y., X. Bonjoch, A. R. Blanch.** 2003. Use of *Bifidobacterium dentium* as an indicator of the origin of fecal water pollution. Appl. Environ. Microbiol. **69:**2651-2003.

25. **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.* Water Res. **35:**379-386.

26. **Parveen, S., R. L. Murphree, L. Edmiston, C. W. Kasper, K. M. Portier, and M. L. Tamplin.** 1997. Association of multiple-antibiotic-resistance profiles with point and

nonpoint sources of *Escherichia coli* in Apalachicola bay. Appl. Environ. Microbiol. **63:**2607-2612.

27. **Rademaker, J. L. W., F. J. Louws, and F. J. deBruijn.** 1998. Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting, suppl. 3, chapter 3.4.3, p.1-26. *In* A. D. L. Akkermans, J. D. van Elsas and F. J. de Bruijn (ed.), Molecular microbial ecology manual. Luwer Academic Publishers, Dordrecht, The Netherlands.

28. **Rademaker, J. L. W., F. J. Louws, U. Rossbach, P. Vinuesa, and F. J. deBruijn.** 1999. Computer-assisted pattern analysis of molecular fingerprints and database construction, suppl. 4, chapter 7.1.3, p.1-33. *In* A. D. L. Akkermans, J. D. van Elsas and F. J. de Bruijn (ed.), Molecular microbial ecology manual. Kluwer Academic Publishers, Dordrecht, The Netherlands.

29. **Sadowsky, M. J., and H.-G. Hur**. 1998. Use of endogenous repeated sequences to fingerprint bacterial genomic DNA, 399-413. *In* J.R. Lupski, G. Weinstock, and F. J. de Bruijn (ed.), Bacterial genomes: structure and analysis. Chapman & Hall, New York, NY.

30. **Sadowsky, M. J., L. L. Kinkel, J. H. Bowers, and J. L. Schottel.** 1996. Use of repetitive intergenic DNA sequences to classify pathogenic and disease-suppressive *Streptomyces* strains. Appl. Environ. Microbiol. **62:**3489-3493.

31. **Scott, T. M., S. Parveen, K. M. Portier, J. B. Rose, M. L. Tamplin, S. R. Farrah, A. Koo, and J. Lukasik.** 2003. Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. **69:**1089-1092.

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

33. **Seurinck, S., W. Verstraete, and S. D. Siciliano.** 2003. Use of 16S-23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify nonpoint fecal sources. Appl. Environ. Microbiol. **69:**4942-4950.

34. **Simpson, J. M., J. W. Santo Domingo and D. J. Reasoner.** 2003. Microbial source tracking: state of the science. Environ. Sci. Tech. **36:**5280-5288.

35. **Souza, V., M. Rocha, A. Valera, and L. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. **65:**3373-3385.

36. **Tyler, K. D., G. Wang, S. D. Tyler, and W. M. Johnson.** 1997. Factors affecting reliability and reproducibility of amplification-based DNA fingerprinting of representative bacterial pathogens. J. Clin. Microbiol. **35:**339-346.

37. **U.S. Environmental Protection Agency.** 2000. National Water Quality Inventory: 1998 Report to Congress. EPA-841-R-00-001. Office of Water, Washington, D.C.

38. **U.S. Environmental Protection Agency.** 2001. Protocol for Developing Pathogen TMDLs. EPA 841-R-00-002. Office of Water, Washington, DC.

39. **Versalovic, J., V. Kapur, T. Koeuth, G. H. Mazurek, T. S. Whittam, J. M. Musser, and J. R. Lupski.** 1995. DNA fingerprinting of pathogenic bacteria by fluorophore-enhanced repetitive sequence-based polymerase chain reaction. Arch. Pathol. Lab. Med. **119:**23-29.

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

41. **Versalovic, J., M. Schneider, F. J. de Bruijn, and J. R. Lupski.** 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol. Cell. Biol. **5:**25-40.

42. **Wheeler, A. L., P. G. Hartel, D. G. Godfrey, J. L. Hill, and W. I. Segars.** 2002. Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking. J. Environ. Qual. **31:**1286-1293.

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

44. **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.

45. **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. L. Torrijos, and M. E. Zimmerman.** 1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Appl. Environ. Microbiol. **65:**3483-3486.

46. **Wiggins, B.A., P. W. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M.**

Smith, E. N. Stevens, and A. K. Varner. 2003. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. Appl. Environ. Microbiol. **69:**3399-3405.

Table 1. Animal source groups and rep-PCR DNA fingerprints generated from *E. coli* isolates.

Animal Source Group	Individuals Sampled	Total Fingerprints	Unique Fingerprints[a]
Cat	37	108	48
Chicken	86	231	144
Cow	115	299	191
Deer	64	179	96
Dog	71	196	106
Duck	42	122	81
Goat	36	104	42
Goose	73	200	135
Horse	44	114	79
Human	197	307	211
Pig	111	303	215
Sheep	37	101	61
Turkey	69	202	126
Total	**982**	**2,466**	**1,535**

[a]Identical *E. coli* genotypes from each individual animal were removed.

Table 2. Total and unique *E. coli* isolates correctly classified into source groups by rep-PCR and HFERP DNA fingerprinting methods[a].

Source group	All Fingerprints (n=2,466)	Unique Fingerprints (n=1,535)			
	Percent Correctly Classified[a]	Percent Correctly Classified			
	rep-PCR	rep-PCR		HFERP	
	Pearson	Pearson	Jaccard	Pearson	Jaccard
Pet[b]	91.8 (279)[d]	61.7 (95)[d]	45.5 (70)	59.1 (91)	44.8 (69)
Chicken	81.4 (188)	59.7 (86)	38.9 (56)	63.2 (91)	31.9 (46)
Cow	79.6 (238)	55.0 (104)	47.6 (90)	62.0 (117)	48.2 (91)
Deer	85.5 (145)	55.2 (53)	36.5 (35)	62.2 (60)	42.6 (41)
Waterfowl[c]	81.4 (262)	66.2 (150)	52.8 (114)	70.4 (152)	56.5 (122)
Goat	97.1 (101)	66.7 (27)	59.5 (25)	47.6 (20)	42.9 (18)
Horse	69.3 (79)	44.3 (35)	34.2(27)	52.6 (41)	32.1 (25)
Human	78.3 (240)	59.2 (124)	47.4(100)	53.8 (113)	45.2 (95)
Pig	77.9 (236)	63.7 (137)	43.7 (94)	54.4 (117)	36.3 (78)
Sheep	79.0 (80)	7.5 (29)	39.3 (24)	37.7 (23)	8.2 (5)
Turkey	88.6 (179)	73.8 (93)	52.4 (66)	73.0 (92)	54.8 (69)
Overall	82.2 (2,027)	60.9 (933)	45.8 (701)	59.9 (917)	43.0 (659)

[a]Based on Jackknife analysis with 1% optimization and maximum similarities using curve-based (Pearson's product moment correlation coefficient) or band-based (Jaccard's coefficient) similarity calculations.
[b]Pet group consists of cats and dogs.
[c]Waterfowl group consists of ducks and geese.
[d]Values in parentheses are number of isolates correctly classified.

Table 3. Percentage of *E. coli* isolates correctly classified into domestic, human and wildlife source groups by using the HFERP DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]	
		Pearson	Jaccard
Domesticated[b]	855	83.2 (711)[e]	77.5 (663)
Human	210	53.8 (113)	45.2 (95)
Wildlife[c]	312	71.4 (223)	59.6 (186)
Pets[d]	154	59.1 (91)	44.8 (69)
Overall	1,531	74.3 (1,138)	66.2 (1,013)

[a]Done using Jackknife analysis with 1% optimization and maximum similarities using curve-based Pearson's product-moment correlation coefficient and band-based Jaccard similarity calculations.
[b]Domesticated group includes, chickens, cows, goats, horses, pigs, sheep and turkeys.
[c]Wildlife group includes deer, ducks and geese.
[d]Pet group includes dos and cats.
[e]Values in parentheses are number of isolates correctly classified.

Table 4. Percentage of *E. coli* isolates correctly classified into human and animal source groups by using the HFERP DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]	
		Pearson	Jaccard
Animal	1321	93.7 (1,237)[b]	92.6 (1,223)
Human	210	53.8 (113)	45.2 (95)
Overall	1,531	88.2 (1,350)	86.1 (1,318)

[a]Done using Jackknife analysis with 1% optimization and maximum similarities using curve-based Pearson's product moment correlation coefficient and band-based Jaccard's similarity calculations.
[b]Values in parentheses are number of isolates correctly classified.

Figure Legends

Figure 1. Frequency of occurrence of genotypes among rep-PCR DNA fingerprints from unique *E. coli* isolates. Analysis was limited to the 657 genotypes identified among the 1,535 unique *E. coli* isolates with rep-PCR DNA fingerprint similarities of 90% or greater.

Figure 2. Representative examples of HFERP DNA fingerprint images. Genomic DNAs from 24 *E. coli* strains were subjected to HFERP DNA fingerprint analysis using a mixture of unlabeled Box A1R and 6-FAM fluorescently labeled Box A1R primers. Each lane contained Genescan-2500 ROX internal lane standards and HFERP DNA fingerprints. The combined, dual colored, HFERP image (A) was captured using a Typhoon Imager and two emission filters. Values in margin are in base pairs. Individual images of the HFERP DNA fingerprints (B) and Genescan-2500 ROX internal lane standard (C) were acquired using one filter at a time.

Figure 3. Comparison of DNA fingerprint patterns of a reference *E. coli* strain generated using rep-PCR and HFERP. (A) rep-PCR DNA fingerprint patterns were assembled from 29 individual PCR reactions, each of which was run on a separate agarose gel. Fingerprints were generated using *E. coli* isolate P294 as template DNA and the Box A1R primer. (B) HFERP DNA fingerprint patterns were assembled from 29 individual PCR reactions each, of which was run on a separate agarose gel. Fingerprints were generated using *E. coli* isolate P294 as template DNA and a mixture of unlabeled Box A1R and 6-FAM fluorescently labeled Box A1R primers. Bands were aligned using Genescan-2500 ROX internal standards, which were present in each lane. Similarities were determined using the cosine algorithm of Bionumerics and dendrograms were generated using the unweighted pair-group method using arithmetic means (UPGMA).

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Apr. 2004, p. 2503–2507 Vol. 70, No. 4
0099-2240/04/$08.00+0 DOI: 10.1128/AEM.70.4.2503–2507.2004

Frequency and Distribution of Tetracycline Resistance Genes in Genetically Diverse, Nonselected, and Nonclinical *Escherichia coli* Strains Isolated from Diverse Human and Animal Sources

Andrew Bryan,[1]† Nir Shapir,[1,2,3,4] and Michael J. Sadowsky[1,2,3]*

Department of Soil, Water, and Climate,[1] BioTechnology Institute,[2] Center for Microbial and Plant Genomics,[3] and Department of Biochemistry, Molecular Biology, and Biophysics,[4] University of Minnesota, St. Paul, Minnesota 55108

Received 22 September 2003/Accepted 6 January 2004

Nonselected and natural populations of *Escherichia coli* from 12 animal sources and humans were examined for the presence and types of 14 tetracycline resistance determinants. Of 1,263 unique *E. coli* isolates from humans, pigs, chickens, turkeys, sheep, cows, goats, cats, dogs, horses, geese, ducks, and deer, 31% were highly resistant to tetracycline. More than 78, 47, and 41% of the *E. coli* isolates from pigs, chickens, and turkeys were resistant or highly resistant to tetracycline, respectively. Tetracycline MICs for 61, 29, and 29% of *E. coli* isolates from pig, chickens, and turkeys, respectively, were ≥233 μg/ml. Muliplex PCR analyses indicated that 97% of these strains contained at least 1 of 14 tetracycline resistance genes [*tetA*, *tetB*, *tetC*, *tetD*, *tetE*, *tetG*, *tetK*, *tetL*, *tetM*, *tetO*, *tetS*, *tetA(P)*, *tetQ*, and *tetX*] examined. While the most common genes found in these isolates were *tetB* (63%) and *tetA* (35%), *tetC*, *tetD*, and *tetM* were also found. *E. coli* isolates from pigs and chickens were the only strains to have *tetM*. To our knowledge, this represents the first report of *tetM* in *E. coli*.

Problems associated with the presence of antibiotic-resistant bacteria have reached epidemic proportions in recent years, with cost estimates exceeding $4 billion in the United States alone (6, 12). The spread of antibiotic-resistant bacteria in the environment is dependent on the presence and transfer of resistance genes among microorganisms, mutations, and selection pressure to keep these genes in a population. Selection pressure has been neatly provided by the approximately 50 million pounds of antibiotics that are produced and used each year in the United States (14). Only half of these antibiotics are used for humans, while the remainder are administered to animals or other organisms (8). The causes and effects of antibiotic overuse are varied. One of the most controversial applications of antibiotics, however, is for growth promotion in livestock, and this application has raised concerns about its contribution to the presence of resistant bacteria in humans (1, 25).

Tetracyclines have become the drugs of choice to treat *Mycoplasma*- and *Chlamydia*-induced pneumonia (13) and have been used to treat other atypical pneumonias, rickettsial infections, Lyme disease, ehrlichiosis, and other diseases and cancers (23). The clinically useful chlortetracycline was introduced in 1948 (24). Only a year later, it was shown that young chickens fed tetracyclines had enhanced growth characteristics (10). However, by 1953, it was reported that *Shigella dysenteriae* had developed resistance to tetracycline antibiotics, and by 1955, a *Shigella* sp. strain had developed multidrug resistance (20). Because of that history and the broad clinical use of tetracycline, this antibiotic was chosen, along with commensal strains of *Escherichia coli*, to provide a prototypical view of the use of antibiotics and their effects on bacterial populations (21). Tetracycline is a broad-spectrum antibiotic that inhibits bacterial protein synthesis by preventing aminoacyl-tRNA from binding to the bacterial ribosome (20). Resistance to the antibiotic is conferred by 1 or more of the 36 currently described *tet* genes, which encode one of three mechanisms of resistance: an efflux pump, a method of ribosomal protection, or direct enzymatic inactivation of the drug (7). Efflux mechanisms appear to be more abundant among gram-negative microorganisms, while ribosomal protection mechanisms are more common among gram-positive organisms (7). Generally speaking, the rapid spread of tetracycline resistance among bacteria is due to the localization of *tet* genes on plasmids, transposons, and integrons (7, 15, 21).

While several studies have examined tetracycline resistance among bacteria, most have employed clinically isolated bacteria (4, 11, 17) or populations specifically isolated for their ability to grow in the presence of tetracyclines (5, 22). These studies, while useful, do not give an unbiased appraisal of the presence and types of *tet* genes that are present in natural (nonclinical), nonselected populations of bacteria in the environment.

Only a limited number of studies have examined tetracycline resistance determinants in bacteria isolated from a large variety of animal species with different histories of exposure to tetracyclines or in environmental samples (11). While Sengeløv and coworkers (22) examined 100 *E. coli* isolates for the presence of five *tet* resistance determinants and Blake et al. (5) used PCR to examine 200 tetracycline-resistant *E. coli* strains

* Corresponding author. Mailing address: Department of Soil, Water, and Climate, University of Minnesota, 1991 Upper Buford Circle, 439 Borlaug Hall, St. Paul, MN 55108. Phone: (612) 624-2706. Fax: (612) 625-2208. E-mail: sadowsky@umn.edu.

† Present address: Department of Medical Microbiology and Immunology, University of Wisconsin, Madison, WI 53706.

TABLE 1. *E. coli* isolates used in this study and their animal sources

Animal source of *E. coli*	No. of isolates used for MIC analysis	No. of isolates used for multiplex PCR
Cat	46	9
Cow	158	24
Deer	74	1
Turkey	82	30
Duck	70	1
Human	176	30
Sheep	48	15
Goose	122	3
Dog	47	9
Pig	182	131
Horse	66	3
Chicken	151	66
Goat	41	3
Total	1,263	325

for seven *tet* genes, few have examined a large number of *tet* determinants in nonclinical *E. coli* isolates from a variety of animal species. To better understand the distribution of resistance genes in the environment and to provide insight into selection pressures involved with the use of antibiotics in animal feed, we investigated tetracycline resistance among natural and unselected populations of *E. coli* from 12 animal sources and humans and determined which resistance genes were present in this population.

Isolates and determination of MIC. In order to characterize tetracycline resistance in natural, nonclinical *E. coli* strains from both human and animal sources, 1,263 unique isolates were obtained from humans, cats, cows, deer, turkeys, ducks, sheep, geese, dogs, pigs, horses, chickens, and goats (Table 1). Fecal materials were collected by swabbing the rectal or cloacal region of individual wild and domesticated animals located throughout Minnesota and western Wisconsin as previously described (9). Fecal samples were kept at 4°C and analyzed within 6 h of swabbing. Fecal material was streaked onto mFC agar plates (Difco, BD Diagnostic Systems, Sparks, Md.) and incubated at 44.5°C for 24 h, and six blue colonies from the mFC agar plates were picked and evaluated by using selective and differential growth media as previously described (9). Only isolates giving growth and color responses on all media that were typical for *E. coli* were used in these studies. Three *E. coli* colonies from each individual fecal sample were used for DNA fingerprinting. All isolates were subjected to DNA fingerprint analysis using rep-PCR and BOXA1R primers (9), and identical clones from the same animal were eliminated from analyses. Unique isolates were grown overnight in 150 μl of Luria-Bertani liquid medium in microtiter plates and were spot inoculated, with a multiple inoculator, onto tryptic soy agar (Difco Laboratories, Detroit, Mich.) supplemented with 0, 5, 10, 20, 40, 70, 93, 117, 175, and 233 μg of tetracycline per ml (Sigma Chemicals, St. Louis, Mo.). The plates were incubated overnight at 37°C and visually examined for growth. MICs were determined from growth patterns, and average values are shown in Fig. 1. If the tetracycline MIC for an isolate was <5 μg /ml, the isolate was considered sensitive to the antibiotic; if it was 10 to 70 or >90 μg/ml, the isolate was considered resistant or highly resistant, respectively. For statistical analysis, a MIC of >233 μg/ml was considered to be 233 μg/ml.

Of the 1,263 *E. coli* isolates examined, 31% were resistant to tetracycline (MICs, >10 μg/ml). Forty-two, 21, 17, and 4% of the isolates from livestock, humans, companion animals (cats, dogs, and horses), and wild animals, respectively, were resistant to tetracycline. More than 78, 47, and 41% of the *E. coli* isolates from pigs, chickens, and turkeys were resistant or highly resistant to tetracycline, respectively. Together these resistant isolates represent about 20% of the 1,263 isolates examined. In contrast, about 22, 30, 3, 3, 21, 33, 7, 23, 6, and 12.2% of the *E. coli* isolates from cats, cows, deer, duck, humans, sheep, geese, dogs, horses, and goats were resistant or highly resistant to tetracycline, respectively. Moreover, the tetracycline MICs for 61, 29, and 29% of *E. coli* isolates from pigs, chickens, and turkeys, respectively, were ≥233 μg/ml. In contrast, the lowest numbers of *E. coli* strains showing resistance or a high level of resistance to tetracycline were those from goats, horses, ducks, geese, and deer. Our results may be explained by the potential exposure of livestock, humans, and companion and wild animals to tetracyclines. Tetracycline is often continuously fed to livestock at subtherapeutic levels for the purpose of growth promotion. For example, up to 70% of U.S. cattle and pig operations use feeds supplemented with antibiotics for growth promotion, and the majority are tetracyclines (2). In contrast, humans and companion animals are most often treated therapeutically, for a limited time, for bac-

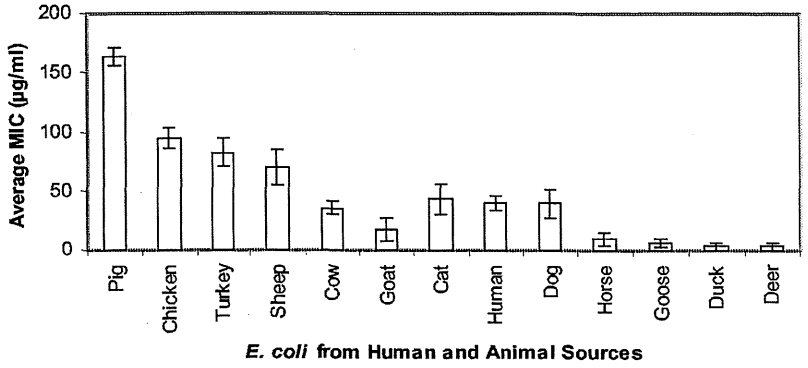


FIG. 1. Average MICs of tetracycline for *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, cats, humans, dogs, horses, geese, ducks, and deer, as determined by the plate dilution method.

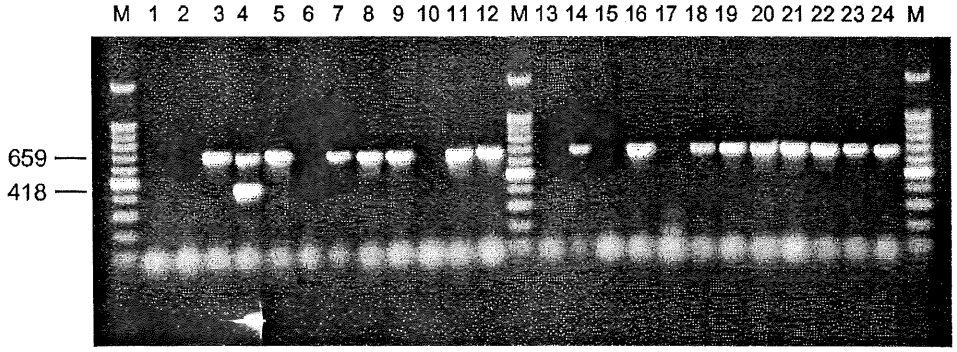


FIG. 2. Representative agarose gel of PCR products from nonclinical *E. coli* isolates, using primer group I, containing primers for *tetB*, *tetC*, and *tetD*. Lanes: 1, no template control; 2, *E. coli* H25; 3, *E. coli* H45; 4, *E. coli* H77; 5, *E. coli* P282; 6, *E. coli* P284; 7, *E. coli* P285; 8, *E. coli* P286; 9, *E. coli* P289; 10, *E. coli* P290; 11, *E. coli* P291; 12, *E. coli* P293; 13, *E. coli* P294; 14, *E. coli* P295; 15, *E. coli* P296; 16, *E. coli* P297; 17, *E. coli* P298; 18, *E. coli* P300; 19, *E. coli* P304; 20, *E. coli* P307; 21, *E. coli* P308; 22, *E. coli* P309; 23, *E. coli* P310; and 24, *E. coli* P312. *E. coli* isolate numbers beginning with P and H were isolated from pigs and horses, respectively. Lane M, molecular weight markers (100 bp ladder). The sizes of the amplicons in base pairs are indicated on the left.

terial infections, perhaps reflecting the intermediate level of resistance to tetracycline (average MICs, 10 to 70 μg/ml) of the isolates from these organisms. This resistance level may be changing, however, as other uses of antibiotics become more common, such as the treatment of parasitic and noninfectious diseases (21). The low level of occurrence of tetracycline resistance among isolates from wild animals is presumably due to their low exposure to these antibiotics. Most of these isolates either had a high level of resistance or none at all, suggesting that the acquisition of a mobile genetic element accounts for resistance.

Epidemiology of *tet* genes. All isolates for which the tetracycline MIC was ≥93 μg/ml (which we considered to indicate a high level of resistance) (n = 325) were examined further by use of a multiplex PCR for the presence of the *tetA*, *tetB*, *tetC*, *tetD*, *tetE*, *tetG*, *tetK*, *tetL*, *tetM*, *tetO*, *tetS*, *tetA(P)*, *tetQ*, and *tetX* genes (18). Single-colony isolates were streaked onto plate count agar (Difco), picked using disposable 10-μl sterile loops, and suspended in 50 μl of sterile H_2O. One microliter of the standardized cell suspension served as a template DNA for colony-based multiplex PCR. The primers used for PCR amplification of the 14 tetracycline resistance genes were as described by Ng et al. (18). The primers were aliquoted into four groups: group I contained primers for *tetB*, *tetC*, and *tetD*; group II contained primers for *tetA*, *tetE*, and *tetG*; group III contained primers for *tetK*, *tetL*, *tetM*, *tetO*, and *tetS*; and group IV contained primers for *tetA(P)*, *tetQ*, and *tetX*. PCR was performed with 96-well plates and an MJ Research (Waltham, Mass.) model PTC100 thermocycler, by using the following conditions as described previously (18): 5 min of initial denaturation at 94°C, followed by 35 cycles of 94°C for 1 min, 55°C for 1 min, and 72°C for 1.5 min. The PCR products were separated by gel electrophoresis in 1% (wt/vol) agarose gels in 1× Tris-acetate-EDTA buffer, stained with ethidium bromide, and visualized under UV illumination. The validity of multiplex PCRs and product sizes was ascertained by using the following positive control plasmids: pSL18, pRT11, pBR322, pSL106, pSL1504, pJA8122, pAT102, pVB.A15, pJ13, pUOA1, pAT451, pJIR39, pNFD13-2, and pBS5, for the genes *tetA*, *tetB*, *tetC*, *tetD*, *tetE*, *tetG*, *tetK*, *tetL*, *tetM*, *tetO*, *tetS*, *tetA(P)*, *tetQ*, and *tetX*, respectively (18). The sizes of the PCR products were determined by comparison to the migration of a 100-bp ladder (Gibco BRL). The identity of all *tet* genes in a representative sample of nonclinical isolates was ascertained by DNA sequencing of the PCR products, following extraction from agarose gels. A representative agrose gel of PCR products obtained using primer group I, amplifying *tetB*, *tetC*, and *tetD*, is shown in Fig. 2.

Of the 325 strains analyzed by PCR, 97% contained at least 1 of 14 [*tetA*, *tetB*, *tetC*, *tetD*, *tetE*, *tetG*, *tetK*, *tetL*, *tetM*, *tetO*, *tetS*, *tetA(P)*, *tetQ*, and *tetX*] tetracycline resistance determinants. The most common determinants were Tet B (63% of isolates) and Tet A (35% of isolates) (Fig. 3). However, Tet C, Tet D, and Tet M were also found with various frequencies. The frequencies of *tetA*, *tetB*, *tetC*, and *tetD* in the tested isolates (Fig. 3) were consistent with those previously reported for lactose-fermenting coliforms based on colony hybridization (11). In contrast, Sengeløv and coworkers (22) reported that 71 and 25% of 100 isolates from the diseased and healthy pigs, cattle, and chickens that they tested for five tetracycline resistance determinants contained *tetA* and *tetB*, respectively. None

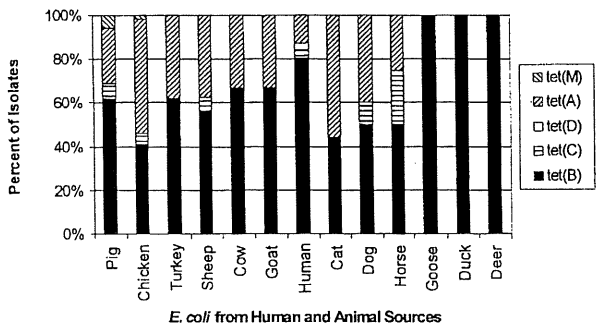


FIG. 3. Frequency of *tetM*, *tetA*, *tetD*, *tetC*, and *tetB* in *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, humans, cats,dogs, horses, geese, ducks, and deer, as determined by colony multiplex PCR. The tetracycline genes *tetE*, *tetG*, *tetK*, *tetL*, *tetO*, *tetS*, *tetA(P)*, *tetQ*, and *tetX* were not found among any of the 325 *E. coli* isolates tested.

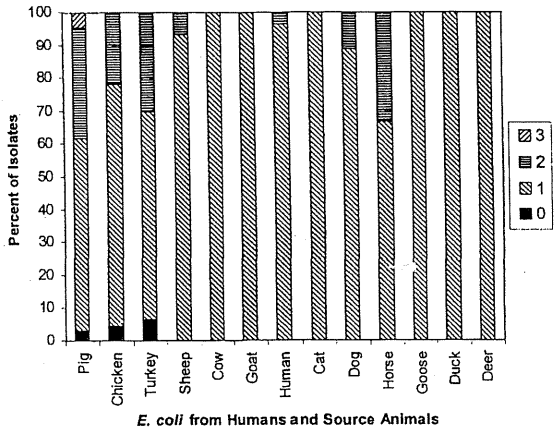


FIG. 4. Percentages of *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, humans, cats, dogs, horses, geese, ducks, and deer, containing multiple tetracycline resistance genes as determined by multiplex PCR using primers for *tetA*, *tetB*, *tetC*, *tetD*, *tetE*, *tetG*, *tetK*, *tetL*, *tetM*, *tetO*, *tetS*, *tetA(P)*, *tetQ*, and *tetX*.

of the tested strains contained *tetE*, *tetG*, *tetK*, *tetL*, *tetO*, *tetS*, *tetA(P)*, *tetQ*, or *tetX*. Since our studies analyzed only highly resistant isolates by PCR, it is possible that additional resistance genes were present in the *E. coli* populations but were nonfunctional or provided only intermediate or low-level resistance.

Isolates from pigs and chickens were the only strains to contain *tetM* and commonly had more than one tetracycline resistance determinant per strain (Fig. 4). The greatest number of strains for which the MICs were high were *E. coli* isolates from these animals. Over 30% of *E. coli* isolates from turkeys, pigs, and horses contained two tetracycline resistance determinants, and 4.5% of the pig isolates contained three *tet* genes. However, the presence of more than one resistance determinant did not lead to noticeably higher MICs. It is possible that strong selection pressures provided by environments containing elevated levels of tetracycline lead to the acquisition of more than one tetracycline gene in a given strain due to their prevalence in the environment, rather than to a selective advantage. The results of our studies also showed that 22.2 and 1.9% of the isolates contained two and three *tet* genes, respectively. This is in contrast to results from previous studies, in which only 3.5% (16) and 5.4% (22) of isolates had two genes, perhaps due to our use of a larger number and variety of isolates and to the greater number of genes examined.

To our knowledge, this is the first report documenting the presence of the *tetM* gene in *E. coli* (7). Due to the uniqueness of these results, the presence of *tetM* in one of our *E. coli* isolates from pigs was verified by sequencing the PCR product produced using *tetM*-specific primers. BLAST analysis (3) indicated that of the 386 bp of high-quality and continuous sequence examined, there was 98% nucleotide sequence identity to the *tetM* gene from *Enterococcus faecalis* (GenBank accession number M85225). The *tetM* gene, which imparts resistance to tetracyclines by encoding a ribosomal protection mechanism, commonly occurs in transposons Tn*916* and Tn*1545*. The *tetM* gene is widely dispersed among various gram-positive organisms, but it has only rarely been documented in gram-negative bacteria (19, 21). The presence of *tetM* in *E. coli* is most likely due to genetic transfer from *Enterococcus*, a common carrier of *tetM* (8). Evidence for this possibility is provided by the studies of Poyart et al. (19), who demonstrated the in vitro transfer of Tn*916* from *E. faecalis* to *E. coli* (16).

In summary, by examining the frequency and distribution of tetracycline resistance genes among diverse natural *E. coli* populations present in different animal species, a picture of the selection pressures in the various host animals can be inferred. Not only did those animal hosts that presumably had continuous exposure to tetracycline have a higher percentage of tetracycline-resistant *E. coli* isolates, but also those isolates carried a greater diversity of resistance genes. Moreover, these isolates often had more than one tetracycline resistance determinant and contained a *tet* gene previously thought not to be present in *E. coli*. This suggests that human activity provides environments that select for resistant strains and encourages the transfer of genetic information from unrelated bacterial species. Although this study examined only nonclinical *E. coli* isolates, the prevalence of tetracycline resistance genes among these unrelated bacteria, and circumstantial and direct evidence of horizontal gene transfer, suggests that these same resistance determinants may also be present in animal and human pathogens.

We thank LeeAnn Johnson, John Ferguson, Mary Brown, and Priscilla Dombek for help with the *E. coli* strains. We also thank Lai-King Ng for the generous gift of positive control strains and cloned *tet* resistance genes.

This study was supported in part by grants from the University of Minnesota Agricultural Experiment Station and from the Legislative Commission on Minnesota Resources, through the Environment and Natural Resources Trust Fund and the Minnesota Future Resource Fund (to M.J.S.), and by funding from the Life Sciences Summer Undergraduate Research Program (to A.B.).

REFERENCES

1. **Aarestrup, F. M., A. M. Seyfarth, H. D. Emborg, K. Pedersen, R. S. Hendriksen, and F. Bager.** 2001. Effect of abolishment of the use of antimicrobial agents for growth promotion on occurrence of antimicrobial resistance in fecal enterococci from food animals in Denmark. Antimicrob. Agents Chemother. **45**:2054–2059.
2. **Akkina, J. E., and R. Johnson.** 1999. Antibiotic use in U.S. livestock production, p. 17–29. *In* Antimicrobial resistance issues in animal agriculture. USDA/APHIS/VS/CEAHI/CEI, Washington, D.C.
3. **Altschul, S. F., T. L. Madden, A Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25**:3389–3402.
4. **Arzese, A. R., L. Tomasetig, and G. A. Botta.** 2000. Detection of *tetQ* and *ermF* antibiotic resistance genes in *Prevotella* and *Porphyromonas* isolates from clinical specimens and resident microbiota of humans. J. Antimicrob. Chemother. **45**:577–582.
5. **Blake, D. P. R. W. Humphry, K. P. Scott, K. Hillman, D. R. Fenlon, and J. C. Low.** 2003. Influence of tetracycline exposure on tetracycline resistance and the carriage of tetracycline resistance genes within commensal *Escherichia coli* populations. J. Appl. Microbiol. **94**:1087–1097.
6. **Boyce, J. M.** 2001. Consequences of inaction: importance of infection control practices. Clin. Infect. Dis. **33**(Suppl.):S133–S137.
7. **Chopra, I., and M. Roberts.** 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol. Mol. Biol. Rev. **65**:232–260.
8. **DeFlaun, M. F., and S. B. Levy.** 1989. Genes and their varied hosts, p. 1–32. *In* S. B. Levy and R. V. Miller (ed.), Gene transfer in the environment. McGraw-Hill, New York, N.Y.
9. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66**:2572–2577.
10. **DuPont, H. L., and J. H. Steele.** 1987. The human health implication of the use of antimicrobial agents in animal feeds. Vet. Q. **9**:309–320.
11. **Guillaume, G., D. Verbrugge, M.-L. Chasseur-Libotte, W. Moens, and J.-M.**

Collard. 2003. PCR typing of tetracycline resistance determinants (Tet A-E) in *Salmonella enterica* serotype Hadar and in the microbial community of activated sludges from hospital and urban wastewater treatment facilities in Belgium. FEMS Microbiol. Ecol. **32:**77–85.

12. **Jones, M. E., J. A. Karlowsky, D. C. Draghi, C. Thornsberry, D. F. Sahm, and D. Nathwani.** 2003. Epidemiology and antibiotic susceptibility of bacteria causing skin and soft tissue infections in the USA and Europe: a guide to appropriate antimicrobial therapy. Int. J. Antimicrob. Agents **22:**406–419.
13. **Lenart, J., A. A. Andersen, and D. D. Rockey.** 2002. Growth and development of tetracycline-resistant *Chlamydia suis*. Antimicrob. Agents Chemother. **45:** 2198–2203.
14. **Levy, S. G.** 2001. Antibiotic resistance: consequences of inaction. Clin. Infect. Dis. **33**(Suppl.)**:**S124–S129.
15. **Levy, S. B., G. B. Fitzgerald, and A. B. Macone.** 1976. Changes in intestinal flora of farm personnel after introduction of tetracycline-supplemented feed on a farm. N. Engl. J. Med. **295:**583–588.
16. **Marshall, B., C. Tachibana, and S. B. Levy.** 1983. Frequency of tetracycline resistance determinant classes among lactose-fermenting coliforms. Antimicrob. Agents Chemother. **24:**835–840.
17. **Mathew, A. G., D. B. Arnett, P. Cullen, and P. D. Ebner.** 2003. Characterization of resistance patterns and detection apramycin resistance genes in *Escherichia coli* isolated from swine exposed to various environmental conditions. Int. J. Food Microbiol. **89:**11–20.
18. **Ng, L. K., I. Martin, M. Alfa, and M. Mulvey.** 2001. Multiplex PCR for the detection of tetracycline resistant genes. Mol. Cell. Probes **15:**209–215.
19. **Poyart, C., J. Celli, and P. Trieu-Cuot.** 1995. Conjugative transposition of Tn*916*-related elements from *Enterococcus faecalis* to *Escherichia coli* and *Pseudomonas fluorescens*. Antimicrob. Agents Chemother. **39:**500–506.
20. **Roberts, M. C.** 1996. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. FEMS Microbiol. Rev. **19:**1–24.
21. **Roberts, M. C.** 2003. Tetracycline therapy: update. Clin. Infect. Dis. **36:**462–467.
22. **Sengeløv, G., B. Halling-Sørensen, and F. M. Aarestrup.** 2003 Susceptibility of *Escherichia coli* and *Enterococcus faecium* isolated from pigs and broiler chickens to tetracycline degradation products and distribution of tetracycline resistance determinants in *E. coli* from food animals. Vet. Microbiol. **95:**91–101.
23. **Smilack, J. D.** 1999. The tetracyclines. Mayo Clin. Proc. **74:**727–729.
24. **U.S. Congress Office of Technology Assessment.** 1995. Impacts of antibiotic-resistant bacteria. OTA-H-629, stock no. GPO 052–003–01446–7. Government Printing Office, Washington, D.C.
25. **Wegener, H. G., F. M. Aarestrup, P. Gerner-Smidt, and F. Bager.** 1999. Transfer of antibiotic resistant bacteria from animals to man. Acta Vet. Scand. Suppl. **92:**51–57.

ATTACHMENT C. SUPPORTING DATA

LCMR WATERSHED FECAL COUNTS:

Table 1

VERMILLION RIVER FECAL COUNTS								
SITE	**07/11/01**	**08/08/01**	**09/05/01**	**10/03/01**	**03/27/02**	**05/01/02**	**06/05/02**	**07/02/02**
VMCwest	770	4300	1000	830	73	8	1000	5600
VMCeast	1800	560	500	200	150	35	930	1900
VMC	360	610	270	170	3	39	1400	420
VNC175	150	200	150	120	97	44	1100	520
VNC	590	390	240	180	25	8	560	210
VSBtrib	220	240	73	140	20	20	610	97
VSB	410	767	450	210	60	22	2000	1100
V31	360	733	93	63	10	10	1400	340
Vverm	190	1633	290	83	970	10	2400	250
V47	560	290	800	150	190	23	2800	330

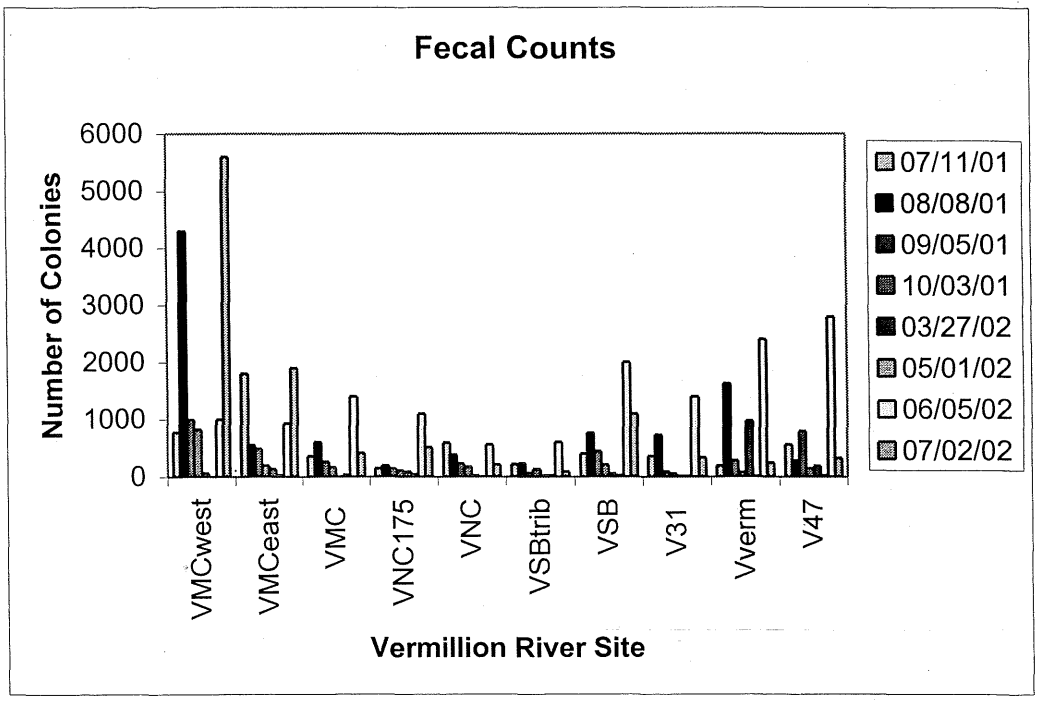


Figure 1

Table 2

HIGH ISLAND CREEK FECAL COUNTS								
SITE	04/09/01	07/18/01	04/10/02	04/24/02	05/08/02	05/29/02	06/19/02	07/10/02
IS	unknown	40	60	1650	360	280	460	5500
2P		620	520	10	400	190	250	1100
3P		40	10	32	270	160	200	1200
5P		280	10	10	81	600	670	460
6S		80	73	34	240	5200	1100	1500
7S		610	140	55	1360	X	7300	1600
8P		490	3	2	1500	X	2400	5900
9P		90	10	10	748	X	1800	2200
10P		600	9	20	240	4300	11200	1300
T		670	640	2	2000	2100	900	1000

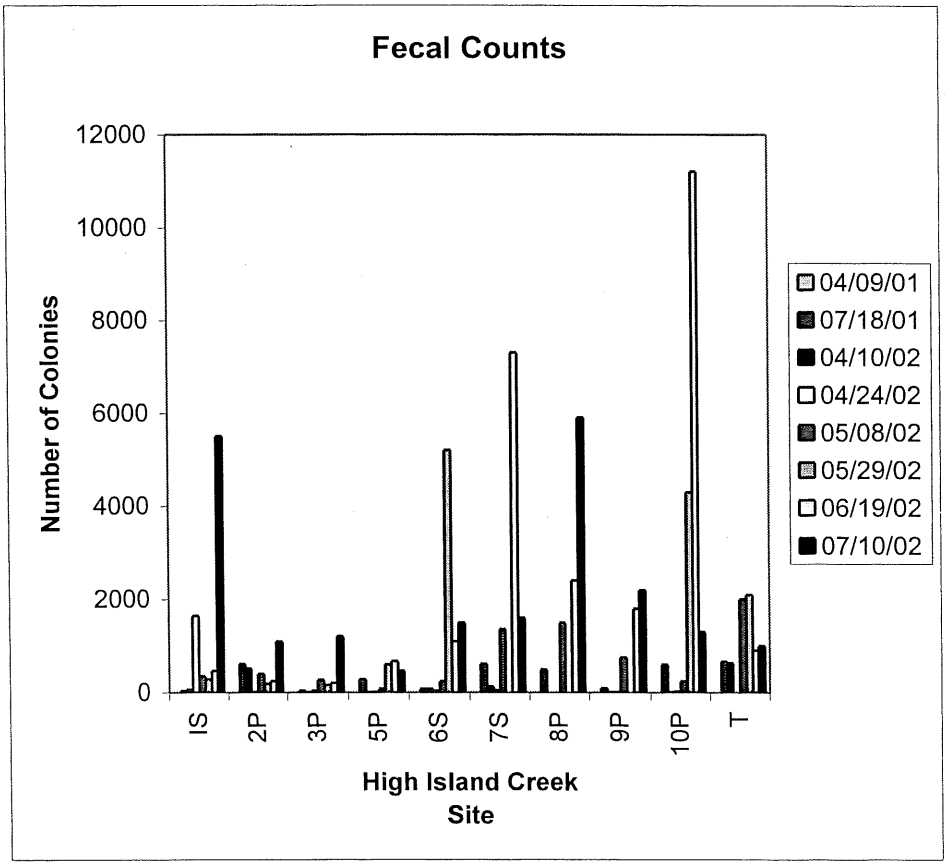


Figure 2

Table 3

MINNEOPA CREEK FECAL COUNTS									
SITE	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
1	unknown	330	680	140	700	110	75	240	660
2		150	980	600	930	830	180	100	580
3		60	370	170	1300	1320	200	90	420
4		20	74	10	80	20	210	130	90
5		67	370	280	900	5	82	46	70
6		240	250	1100	800	110	400	590	490
7		350	67	91	100	6	7	X	130
8		160	110	9	230	62	64	X	730
9		67	140	56	9	10	53	X	84
10		70	67	6	39	4	45	X	180

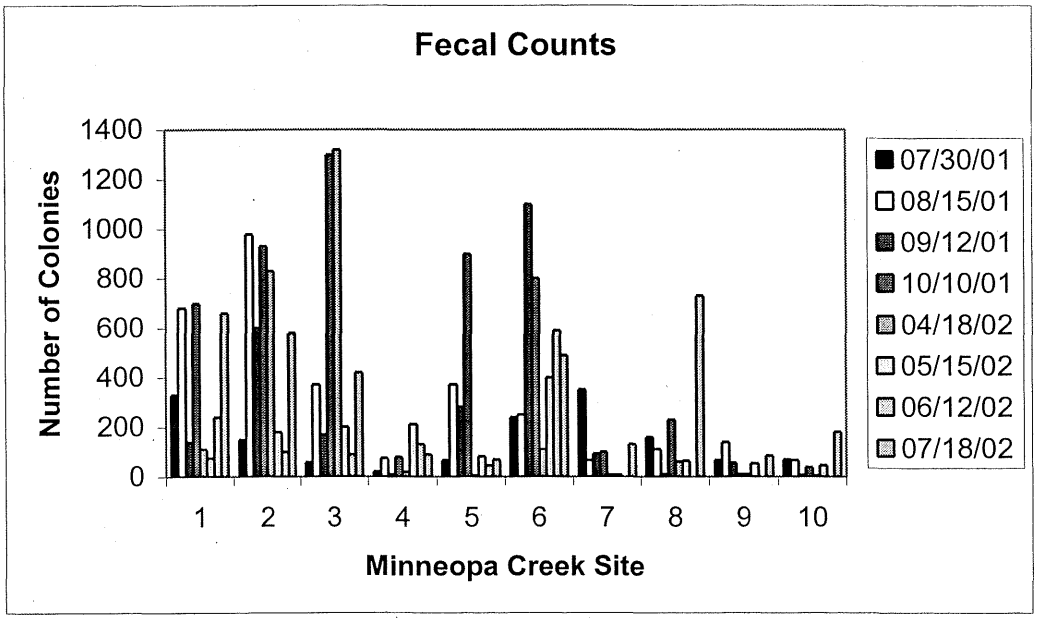


Figure 3

FINAL REPORT

2001 Project Abstract
For the Period Ending June 30, 2004 (extended from June 30, 2003)

TITLE: Determination of Fecal Pollution Sources in Minnesota Watersheds
PROJECT MANAGER: Dr. Michael J. Sadowsky
ORGANIZATION: University of Minnesota
ADDRESS: Department of Soil, Water & Climate, 1991 Upper Buford Circle, 439 Borlaug Hall St. Paul, MN 55108
WEB SITE ADDRESS: www.ecolirep.umn.edu
FUND: Future Resources Fund
LEGAL CITATION: ML 2001, 1st Special Session, Ch. 2, Sec.14, Subd. 6 (d)

APPROPRIATION AMOUNT: $275,000

Overall Project Outcome and Results

We used a library of DNA fingerprints, created using the rep-PCR and HFERP techniques, in an attempt to define sources of fecal bacterial pollution, *E. coli*, in three Minnesota watersheds, Minneopa Creek (Blue Earth County), High Island Creek (Sibley County), and Vermillion River (Dakota County). Sampling from 10 sites per watershed took place in 2001 and 2002. Approximately 25 *E. coli* isolates were obtained from each site per sampling date. About 1,776, 1,651, and 1,762 E. coli were DNA fingerprinted from the Vermillion River, High Island Creek, Minneopa Creek Watersheds, respectively. The most reliable results from data came from bootstrap analyses of fecal bacteria segregated into Human vs. Non-human categories, or into groupings consisting of Humans, Pets (dogs and cats), Waterfowl (geese, ducks), Wildlife (deer), and Domesticated animals (chickens, cows, goats, horses, pigs, sheep, turkeys). Analysis of the Vermillion River showed that 93 and 6.1 % of the isolates identified were of Non-Human and Human origin, respectively. The greatest potential contributors to fecal pollution in this watershed were domesticated animals (23 %), pets (45%), and deer (19%). Similar results were found with the Minneopa Creek isolates, where 90 and 10% of the isolates were from non-human and human origin, respectively. Of these 23% were from Domesticated animals, 36% from Pets, and 21% from deer. In contrast, while 84 and 16% of High Island Creek isolates were Non-Human and Human sources, respectively, the majority came from domesticated animals (42%, mostly from cows), with the remainder contributed by geese, 14%, and humans 16%. It should be noted however, that our research showed that much larger database of DNA fingerprints is needed for more accurate assignments to the animal level. A reliable bacterial source tracking method would aid watershed managers tremendously, giving them another tool to efficiently direct efforts clean watersheds of bacterial pollutants.

Project Results Use and Dissemination

Results from this project have been disseminated in reports made to the LCMR, in periodic update reports made to cooperators, in seminars given throughout the state, nationally and internationally, and in scientific publications in peer-reviewed journals. In addition, results from our studies will be posted and updated on the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) which is housed on computers at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our previous LCMR projects. Data obtained from our studies will be utilized by cooperating agencies to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas.

Date of Report: January 21, 2004

JAN 2 7 2004

Date of Workprogram Approval: June 29, 2001

Project Completion Date: December 31, 2003

LCMR Final Work Program Report

I. PROJECT TITLE: Determination of Fecal Pollution Sources in Minnesota Watersheds

Project Manager: Dr. Michael J. Sadowsky

Affiliation: University of Minnesota
Mailing Address: Department of Soil, Water & Climate
1991 Upper Buford Circle
439 Borlaug Hall
St. Paul, MN 55108

Telephone Number: (612) 624-2706
E-Mail: sadowsky@soils.umn.edu
Fax: (612) 625-6725
Web Page address: http://www.ecolirep.umn.edu

Total Biennial Project Budget:

$ LCMR Appropriation	- $ Amount Spent	= $ Balance:
$275,000	$271,581	$3,419

Legal Citation: ML 2001, 1st Special Session, Ch. 2, Sec.14, Subd. 6 (d).

Appropriation Language:

Determination of Fecal Pollution Sources in Minnesota Watersheds $275,000 is from the future resources fund to the University of Minnesota for the second biennium to determine sources of fecal pollution in three impacted watersheds utilizing DNA fingerprinting techniques, and evaluate the efficacy of implemented and proposed abatement procedures to remediate fecal contamination.

The availability of the appropriation for the following project is extended to June 30, 2004, unless an earlier date is specified in the work program: ML 2003, Art. 1,Ch.128, Sec. 9, Subd. 20(a): 6 (d) Determination of fecal pollution sources in Minnesota.

II. and III. FINAL PROJECT SUMMARY

We used a library of DNA fingerprints, created using the rep-PCR and HFERP techniques, in an attempt to define sources of fecal bacterial pollution, *E. coli*, in three Minnesota watersheds, Minneopa Creek (Blue Earth County), High Island Creek (Sibley County), and Vermillion River (Dakota County). Sampling from 10 sites per watershed took place in 2001 and 2002. Approximately 25 *E. coli* isolates were obtained from each site per sampling date. About 1,776, 1,651, and 1,762 *E. coli* were DNA fingerprinted from the Vermillion River, High Island Creek, Minneopa Creek Watersheds, respectively. The most reliable results from data came from bootstrap analyses of fecal bacteria segregated into Human vs. Non-human categories, or into groupings consisting of Humans, Pets (dogs and cats), Waterfowl (geese, ducks), Wildlife (deer), and Domesticated animals (chickens, cows, goats, horses, pigs, sheep, turkeys). Analysis of the Vermillion River showed that 93 and 6.1 % of the isolates identified were of Non-Human and Human origin, respectively. The greatest potential contributors to fecal pollution in this watershed were domesticated animals (23 %), pets (45%), and deer (19%). Similar results were found with the Minneopa Creek isolates, where 90 and 10% of the isolates were from non-human and human origin, respectively. Of these 23% were from Domesticated animals, 36% from Pets, and 21% from deer. In contrast, while 84 and 16% of High Island Creek isolates were Non-Human and Human sources, respectively, the majority came from domesticated animals (42%, mostly from cows), with the remainder contributed by geese, 14%, and humans 16%. It should be note however, that our research showed that much larger database of DNA fingerprints is needed for more accurate assignments. A reliable bacterial source tracking method would aid watershed managers tremendously, giving them another tool to efficiently direct efforts clean watersheds of bacterial pollutants.

IV. OUTLINE OF PROJECT RESULTS:

Result 1: Acquisition of fecal coliform bacteria from watershed areas.

Fecal coliform bacteria were isolated from water samples collected in each of the 3 watershed areas in the Spring, Summer, and Fall months (during baseline and critical run-off periods). Approximately 80 water samples were collected from each watershed (10 samples collected per watershed on at least eight separate sampling occasions). Samples were analyzed for fecal coliform bacteria by an EPA-certified contract laboratory (see below). *E.coli* bacteria from fecal coliform plates were isolated (about 1600 isolates per watershed for a total of approximately 4800 isolates) and the identity of bacteria was confirmed by using selective and differential microbiological media and biochemical tests. The confirmed *E. coli* isolates were cataloged, stored, and preserved in glycerol at -80°C until DNA fingerprinting was completed.

LCMR Budget:	**$101,200**
Balance:	**$846**

Personnel:	$80,000	(Sr. Scientist /Assistant Scientist [40%], Jr. Scientist [40%], and Student lab Techs [70%])
Supplies:	$14,900	(Consumables: $2.38/isolate x 5232 isolates = $12,400,Pipetors: $1000, Miscellaneous lab supplies: $1000, Sampling Supplies $500)
Local Travel:	$1,500	Approx. 5 trips @ $180/trip mileage, lodging and meals; 10 trips @ $50/trip mileage)
Contracted Services:	$4,800	(Fecal coliform counts, 80 samples/ watershed x 3 watersheds @ $20/sample)

Total $101,200

Water Sample Collection - Collection of water samples from the three watersheds was completed in July 2002. Each watershed was sampled on at least eight sampling dates according to our original plan. Approximately 18 ml of water was collected from the center of each sample site in sterile Whirl-Pak® bags (Nasco, Fort Atkinson, WI) attached to a homemade pole with an end clamp. Samples were labeled and stored on ice at 4°C until delivered to the laboratory for analysis.

a. The isolation of bacteria from the three watersheds generally proceeded as originally planned. However, it should be noted that due to drought conditions during the summer of 2001, there was insufficient water flow at the High Island Creek watershed to obtain samples once per month. Nevertheless, we did collect snowmelt water samples from High Island on 4/9/01 (before this current project started) and a set of samples on 7/18/01. Wet weather conditions and substantial water flows in the late Spring and early Summer in 2002 allowed us to obtain additional water samples from the High Island Creek watershed. Likewise, we collected a set of samples from Minneopa Creek during Spring flooding on 4/3/01. In the other watersheds, Vermillion and Minneopa, water samples were collected once per month as originally proposed (see Table 1 below).

All water samples obtained were processed according to our original plan. Water samples were analyzed for fecal coliform bacteria by an EPA-certified laboratory at Metropolitan Council Environmental Services laboratory in St. Paul MN, using the membrane filtration method and mFC Agar plates (*Standard Methods for the Examination of Water and Wastewater,* 1995). The mFC agar plates used to enumerate fecal coliform bacteria were used by project staff for the isolation, identification and confirmation of *E. coli* strains. Typical fecal coliform colonies (blue in color) were picked and restreaked for purity onto the same medium. The purified presumptive coliform bacteria were subjected to confirmatory tests using EC broth with MUG (with Durham tubes). The EC-MUG Broth was incubated at 44.5 ± 0.2 °C. The EC medium differentiates between coliform bacteria of fecal or other origin. Presumptive *E. coli* colonies were streaked for isolation on MacConkey agar and plated onto ChromAgar ECC to differentiate between *E. coli* and *Klebsiella*. Confirmed

coliform bacteria are indicated by the production of gas in Durham tubes, strong fluorescence when EC-MUG tubes are examined by using ultraviolet light, and the production of blue colonies on ChromAgar. Isolates giving atypical responses with any test were further screened using API 20E test kits (bioMerieux, Inc., St. Louis, MO). Isolates yielding a "good" to "excellent" *E. coli* identification by the API 20E kit were used for DNA fingerprinting. Three *E. coli* colonies from each individual fecal sample were used for DNA fingerprinting and were stored at –80°C in 50% glycerol.

b. The isolation of bacteria from the three watersheds was completed in August 2002. Fecal *E. coli* bacteria were isolated from each water sample (see Table 1) using selective and differential microbiological plating media (as described above). The final number of *E. coli* isolates from each watershed exceeded the original goal of 1600 (see Table 1). In total, 5,232 *E. coli* isolates were obtained from all three watersheds. Of these, 5,189 were confirmed as *E. coli* and subjected to DNA fingerprinting.

Table 1. Water samples obtained and bacteria isolated.

Watershed	Date Sampled	*E. coli* Isolates Obtained	Total Isolates (% of Goal)
Minneopa Creek	4/3/01	100	
	7/30/01	204	
	8/15/01	216	1783
	9/12/01	185	(111)
	10/10/01	217	
	4/18/02	177	
	5/15/02	224	
	6/12/02	239	
	7/17/02	221	
High Island Creek	4/9/01	208	
	7/18/01	214	
	4/10/02	178	1651
	4/24/02	196	(103)
	5/8/02	225	
	5/29/02	180	
	6/19/02	233	
	7/10/02	217	
Vermillion River	7/11/01	237	
	8/8/01	229	
	9/5/01	238	1798
	10/3/01	238	(112)
	3/27/02	215	
	5/1/02	200	
	6/5/02	226	
	7/2/02	215	

Result 2: Generate DNA fingerprints from *E. coli* isolates obtained from watersheds.

DNAs from confirmed *E. coli* isolates were subjected to rep-PCR and HFERP DNA fingerprinting using BOXA1R primers. The resulting 5,189 DNA fingerprints were captured as digital images, band migration on gels was normalized to internal molecular weight standards, and compared and analyzed using BioNumerics pattern recognition and statistical analysis software. The animal(s) and animal groups contributing to *E. coli* in watersheds were determined by cluster and discriminant statistical analyses, by comparison to our known source DNA fingerprint library.

	LCMR Budget:	**$143,553**
	Balance:	**$1,228**
Personnel:	~~$121,053~~	(Sr. Scientist /Assistant Scientist [40%],
	$121,622	Jr. Scientist [60%], and Student Lab Techs [30%])
Supplies:	~~$15,200~~	(Consumables; $2.92 /isolate x 5211 isolates)
	$14,631	
Equipment:	$7,300	(Computer, $2500; Gel Analysis Software; $4,800)

Total: $143,553

Completion Date: December 31, 2003

Results Status:

We previously used the rep-PCR DNA fingerprinting technique, with BOXA1R primers, to generate a DNA fingerprint library from *E. coli* bacteria obtained from 12 known animal sources and humans. This known source library was used in this current LCMR project to determine the sources of fecal bacteria in the three watersheds. Before we analyzed for the potential sources of *E. coli* isolates from the three watersheds, we performed cluster analysis of 29 DNA fingerprints generated from the same control *E. coli* strain. This was used to determine the reproducibility of rep-PCR DNA technique over a large number of gels. We found an average similarity of about 88 % between the 29 fingerprints (Fig. 1). While this level of reproducibility is sufficient to examine genetic diversity among the known source bacteria and place these bacteria into their respective source groups, it may not be adequate to assign unknown *E. coli* to the correct animal source group with a high degree of certainty. That is because the unknown isolates may be very closely related to several animal source groups at levels greater than 90%. As such, only small differences separate some isolates from different source groups, and this may reduce the statistical certainty of source group assignment. Consequently, in order assign source groups to the unknown *E. coli* bacteria from the watersheds with a greater degree of statistical certainty, and to reduced within gel grouping of DNA fingerprints and improved alignment of DNA fingerprints between gels, we modified the rep-PCR fingerprinting technique to improve

precision. The modifications used fluorescently labeled BOX primers in the rep-PCR reaction, to generate labeled PCR products, and included the use of molecular weight standards in each fingerprint lane that are labeled with a second fluorophore. This allowed accurate normalization of DNA bands in each fingerprint lane, and allows for more precise assignment of DNA bands within and across several fingerprint gels. The revised HFERP DNA fingerprinting protocol is listed below:

***E. coli* preparation and PCR conditions for HFERP Fingerprinting**. *E. coli* isolates were streaked onto Plate Count Agar (Difco, BD Diagnostic Systems, Sparks, MD) and grown overnight at 37°C. Colonies were picked with a 1 µl sterile inoculating loop (Fisher Scientific, Pittsburgh, PA), suspended in 100 µl of 0.05 M NaOH in 96-well, low profile, PCR plates (MJ Research, Waltham, MA), heated to 95 °C for 15 min, and centrifuged at 640 RPM for 10 min in a Hermle/Labnet Z383K centrifuge. A 2µl aliquot of the supernatant in each well was used as template for PCR. The PCR master mix (described here in µl per reaction), consisted of 12.65µl ddi H_2O, 5µl 5x Gitscher buffer, 2.5µl DMSO, 1.0µl 6FAM-BOX primers, 1.25µl dNTP's, 0.2µl, and 0.4µl TAQ polymerase, The primer consisted of a mixture of 0.09 µg of unlabeled Box A1R primer per µl and 0.03 µg of 6-FAM fluorescently labeled Box A1R primer per µl (Integrated DNA Technologies, Coralville, IA). The PCR was performed using an MJ Research PTC 100 (MJ Research, Waltham, MA) using the protocol specific for this thermocyclers and the Box A1R primer. PCR was initiated with an incubation at 95°C for 2 minutes, followed by 30 cycles, consisting of 94°C for 3 seconds, 92°C for 30 seconds, 50°C for 1 minute, and 65°C for 8 minutes (40). PCR reactions were terminated after an extension at 65°C for 8 min, and stored at 4°C. Reactions that were not used immediately for gel electrophoresis analysis were stored at -20°C. A 6.6 µl aliquot of a mixture of 50 µl Genescan-2500 ROX internal lane standard (Applied Biosystems, Foster City, CA) and 200 µl non-migrating loading dye (150 mg Ficoll 400 per ml, and 25 mg blue dextran per ml) was added to each 25 µl PCR reaction prior to loading the PCR reaction into agarose gels, 12 µl of the resulting mixture was loaded per gel lane. DNA fragments were separated by electrophoresis, which was done at 4°C for 17-18 hours at 70V with constant buffer recirculation. Gel Images were captured as TIF files, using a Typhoon 8600 Variable Mode Imager (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) operating in the fluorescence acquisition mode using the following settings: green (532 nm) excitation laser; 610 BP 30 and 526 SP emission filters in the autolink mode with 580 nm beam splitter; normal sensitivity; 200 micron/pixel scan resolution; + 3 mm focal plane; and 800 V power.

Using this modification we now show an average similarity of about 92% between fingerprints from the same control *E. coli* strain (Fig. 2). This improvement will also reduce between gel variability and increase the overall precision of our results. While this took some time to do, we think it was well worth the effort. Efforts to further refine this modification continued into January 2002, at which time we began to DNA fingerprint the watershed isolates obtained in

Result 1.

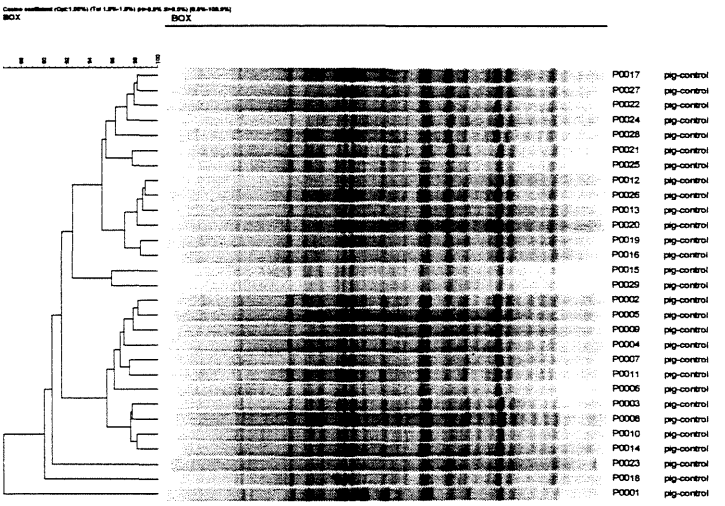


Figure 1. Similarity of DNA fingerprints from 29 control strains using old rep-PCR method.

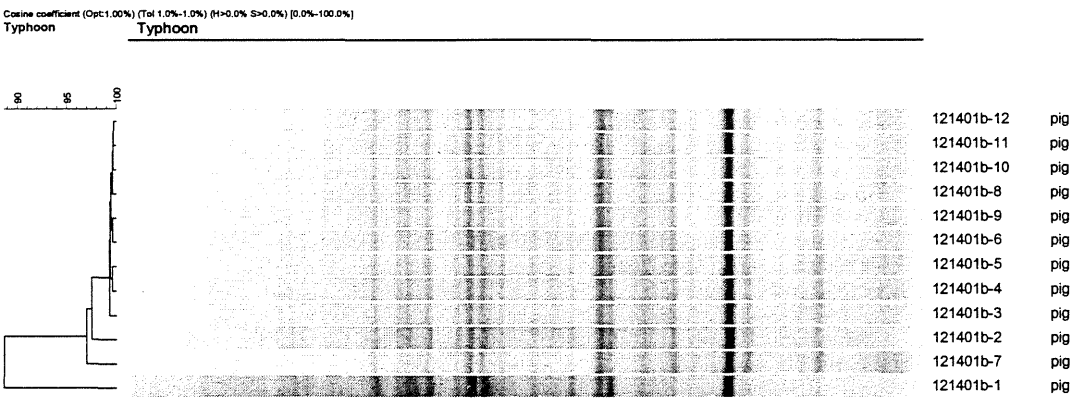


Figure 2. Similarity of DNA fingerprints from 12 control strains using New Modified rep-PCR method.

To test whether HFERP reduced within-gel groupings of DNA fingerprints, we analyzed DNA fingerprints from 40 *E. coli* strains obtained from dogs on 2 different gels using Pearson's product-moment coefficient. Results of these studies indicated that rep-PCR DNA fingerprints from strains run on the same gel were, on average, 50% (range 29 – 57%) more likely to be grouped together as the same strains analyzed by using the HFERP technique. This indicates that HFERP method considerably reduces within gel grouping of DNA fingerprints. In addition, the

HFERP method reduced alignment difficulties due to within- and between-gel variation in band migration found with rep-PCR gels.

Due to increased method precision, our data obtained using the fluorophore-enhanced technique reduced inter- and intra-gel variability contributing to error in the correct classification of known (and presumably unknown) isolates. The precision of fluorophore-enhanced DNA fingerprinting method was determined by repeated fingerprinting of a reference *E. coli* strain (pig isolate number 294). Fifty-eight (58) positive controls were generated using this method. Cluster analysis of the control strains revealed the level at which the DNA fingerprints could be repeated between experimental runs. When analyzed using the curve-based Pearson correlation coefficient and 1% optimization, the DNA fingerprints obtained using the fluorophore-enhanced technique had a 91.2% average similarity.

The result of our new normalization process is that fingerprint patterns from different gels can be accurately compared. It should be noted, however, that the intensity of HFERP bands are more variable than those generated by rep-PCR, and that some of the gains achieved by more precise alignment of bands may be offset by more variation in band intensity. We found that this variation in intensity can be overcome by the careful mixing of all reagents in the PCR master mix and greater pipetting precision when loading gels (data not presented). Further improvements to increasing the intensity of HFERP-generated DNA fingerprints may also be obtained by varying the ratio of labeled to unlabeled primer and the final concentration of the primer mixture in PCR reactions. Nevertheless, our results clearly show that HFERP-derived DNA fingerprint bands are more precisely aligned than the rep-PCR bands. In addition, we show that HFERP DNA fingerprints generated by our method reduce within gel groupings of fingerprints, which can have major ramifications for the assembly of libraries and the analysis of unknown environmental isolates.

While we previously described the use of rep-PCR DNA fingerprinting to determine sources of fecal bacteria (Dombek et al. 2000), our initial studies, and many others by most researchers, used libraries consisting of a relatively small number of samples, some of which were obtained from the same individual animal. To test the influence of library size and duplication of samples in libraries, 2,466 high-quality rep-PCR DNA fingerprints were generated using the Box A1R primer and template DNA from *E. coli* strains obtained from the 13 human and animal sources (Table 1). Of the 2,466 DNA fingerprints analyzed, 1,535 (62%) remained in the "unique" DNA fingerprint library (Table 2). The influence of duplicate DNA fingerprints on the correct classification of library strains is shown in Table 3.

Table 2. Animal source groups and rep-PCR DNA fingerprints generated from *E. coli* isolates.

Animal Source Group	Individuals Sampled	Total Fingerprints	Unique Fingerprints[a]
Cat	37	108	48
Chicken	86	231	144
Cow	115	299	191
Deer	64	179	96
Dog	71	196	106
Duck	42	122	81
Goat	36	104	42
Goose	73	200	135
Horse	44	114	79
Human	197	307	211
Pig	111	303	215
Sheep	37	101	61
Turkey	69	202	126
Total	**982**	**2,466**	**1,535**

[a]Identical *E. coli* genotypes from each individual animal were removed.

Jackknife analysis performed on the 2,466 DNA fingerprints from the entire known-source rep-PCR DNA fingerprint database, using Pearson's product-moment correlation coefficient, indicated that 69-97% of animal and human *E. coli* isolates were assigned into correct source groups (Table Y). This corresponds to an 82.2% average rate of correct classification for the 2,466 rep-PCR DNA fingerprints. However, since identical DNA fingerprints from *E. coli* strains obtained from the same individual most likely represent isolates of clonal origin, and can artificially bias subsequent analyses of strain groupings (e.g. increase the average rate of correct classification) and the fidelity of the database, we eliminated duplicate DNA fingerprints originating from *E. coli* strains obtained from the same individual animal or human. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%.

Of the 2,466 DNA fingerprints analyzed, 1,535 (62%) remained in the "unique" DNA fingerprint library (Table 2). The influence of duplicate DNA fingerprints on the correct classification of library strains is shown in Table 2. When the 1,535 DNA fingerprints from the unique *E. coli* isolates were examined, Jackknife analyses indicated that only 44-74% of the isolates were assigned to the correct source group (Table 3). The average rate of correct classification for these 1,535 unique rep-PCR DNA fingerprints was 60.5%. Taken together, these results indicate that

inclusion of duplicate DNA fingerprints in the library can artificially influence strain groupings and increase percentages of strains correctly assigned to source groups.

Table 3. Percentage of known-source rep-PCR DNA fingerprints assigned to the correct source group by Jackknife analysis[a].

Animal Source	**All Fingerprints (n=2,466)**	**Unique Fingerprints (n=1,535)**
	Percent Correctly Classified Isolates	
Pets[b]	91.8 (279)[d]	61.7 (95)
Chicken	81.4 (188)	59.7 (86)
Cow	79.6 (238)	55.0 (105)
Deer	85.5 (145)	55.2 (53)
Waterfowl[c]	81.4 (262)	66.2 (143)
Goat	97.1 (101)	66.7 (28)
Horse	69.3 (79)	44.3 (35)
Human	78.3 (240)	59.2 (125)
Pig	77.9 (236)	63.7 (137)
Sheep	79.0 (80)	47.5 (29)
Turkey	88.6 (179)	73.8 (93)
Overall	82.2 (2,027)	60.5 (929)

In addition, our studies reported here we show that increasing the size of the known source library to 2,466 isolates does not necessarily lead to an increase in the ability to correctly assign strains to the correct source group. In fact, the average rate of correct classification decreased 4.2% using the larger library reported here, relative to what was seen using a smaller library in our previous studies. This may in part be due to the uncovering of increased genetic diversity among isolates, increased accumulation of errors due to gel-to-gel variation, or the presence of duplicate genotypes (DNA fingerprints) from the same individual within our original library. Reduction in the percentage of known-source *E. coli* isolates that were correctly classified was especially apparent when our unique library of 1535 *E. coli* isolates was examined. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%. Since DNA fingerprints from *E. coli* strains obtained from the same individual represent isolates of clonal origin, these duplicate strains (or fingerprints) can artificially bias the average rate of correct classification and the fidelity of the database. Results in Table 3 show that there was a 21.7% reduction in the average rate of correct classification by using the unique DNA fingerprint library, relative to that seen with the complete library. More importantly, our results show that failure to remove identical fingerprints from analyses resulted in an overestimation of the ability of the database to assign isolates to their correct source group, perhaps in part due to the clonal composition of *E. coli*

populations. Taken together, our results indicate that inclusion of duplicate DNA fingerprints in the library can artificially influence strain groupings and incorrectly increases percentages of strains correctly assigned to source groups.

The fluorophore-enhanced DNA fingerprinting method was then applied to 1,531 *E. coli* isolates included in the "unique" isolate subset. Table 4 summarizes the ability of this method to correctly classify *E. coli* isolates into each of the thirteen human and animal source groups based on Pearson correlation calculations. The ability to correctly classify isolates into their source groups was 57.6% overall correct classification for the fluorophore-enhanced technique.

Table 4. *E. coli* isolates correctly classified into thirteen source groups using the fluorophore-enhanced DNA fingerprinting (HFERP) method.

Source group	Number of DNA Fingerprints	Percent correctly classified[1]
Cat	48	37.5 (18)
Chicken	144	63.2 (91)
Cow	189	62.4 (118)
Deer	96	49.0 (47)
Dog	106	58.5 (62)
Duck	81	60.5 (49)
Goat	42	50.0 (21)
Goose	135	65.2 (88)
Horse	78	52.6 (41)
Human	210	55.7 (117)
Pig	215	54.0 (116)
Sheep	61	39.3 (24)
Turkey	126	71.4 (90)
Overall	1531	57.6 (882)

[1]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) calculation.
[2]Values in parentheses are number of isolates (n) correctly classified.

Further refinements to the jackknife analysis, including pooling of animal source groups (Tables 5 and 6) and limiting the analysis to relevant source groups, were found to improve the ability of the DNA fingerprint library to correctly classify isolate sources by both DNA fingerprinting methods. When all animal sources were pooled into one group, the overall correct classification rate for humans and animals by the fluorophore-enhanced technique was 88.1% (Table 5). Accordingly, these results indicated that (1) broader classifications of source groups should be used when appropriate, or (2) a targeted subset of the DNA fingerprint database should be used to more precisely determine sources of fecal pollutants in watersheds where specific source groups are known to be present or absent. When isolates were separated into domesticated animals, wildlife, and humans, the average rate of correct classification was 78.5%, with domesticated animals being most successfully classified (Table 6).

Table 5. Percentages of *E. coli* isolates correctly classified into human and animal source groups by the fluorophore-enhanced DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent correctly classified[1]
Animal	1321	93.3 (1232)[2]
Human	210	55.7 (117)
Overall	1531	88.1 (1349)

[1]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) similarity calculation.
[2]Values in parentheses are number of isolates (n) correctly classified.

Table 6. Percentages of *E. coli* isolates correctly classified into domesticated, human and wildlife source groups by the fluorophore-enhanced DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent correctly classified[1]
Domesticated[3]	1009	86.2 (870)
Human	210	55.7 (117)
Wildlife[4]	312	68.9 (215)
Overall	1531	78.5 (1202)

[1]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) similarity calculation.
[2]Values in parentheses are number of isolates (n) correctly classified.
[3]The domesticated group includes cat, chicken, cow, dog, goat, horse, pig, sheep and turkey.
[4]The wildlife group includes deer, duck and goose.

To estimate genetic diversity of the *E. coli* comprising the known source database, an accumulation curve was constructed using fluorophore-enhanced fingerprints from the known-source DNA fingerprint library. To do this, each *E. coli* isolate in the library was assigned to a genotype. A genotype was defined as a cluster of DNA fingerprints with similarity of 90% or greater (based on Pearson correlation, 1% optimization and UPMGA). Using this definition, 657 genotypes were identified from the 1,531 unique *E. coli* isolates in the known-source database. The isolates were randomized, and an accumulation curve was constructed by summing the number of genotypes represented by the isolates. The resulting accumulation curve is shown in Figure 3, and is essentially linear over its entire range. The linear, non-asymptotic nature of the curve is indicative of high genetic diversity. Based on these results it is not possible to predict how many new isolates would be required before *E. coli* no new genotypes are acquired. It is clear, however, that the 1531 unique DNA fingerprints contained in the known-source database do not adequately capture the genotypic diversity that exists among naturally-occurring *E.coli*, and that significantly more samples would be required to obtain a library that is adequate enough

to identify all naturally-occurring unknown isolates in watersheds. This problem is not unique to our technique, as all library-based source tracking methods using *E. coli* have the same inherent problem.

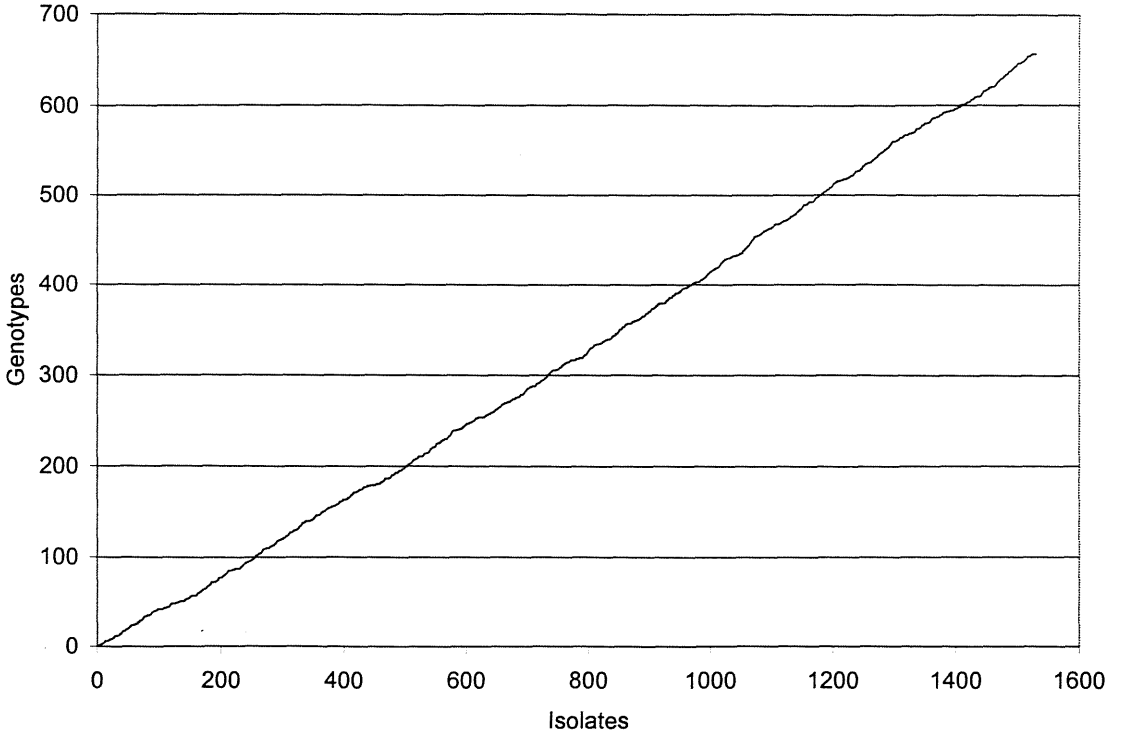


Figure 3. Diversity among *E. coli* genotypes in the unique isolate subset of the known-source library (n=1531). A genotype was defined as a cluster of DNA fingerprints with similarity of 90% or greater (based on Pearson correlation coefficient, 1% optimization and UPMGA).

Results presented here also show that despite having a known source library or over 1,500 unique isolates, the number of genotypes uncovered by DNA fingerprinting continued to increase at a constant rate. Moreover, across all animal hosts, the majority of these fingerprints occurred only once. For a library to be truly representative it needs to be large enough to capture all the unknowns present in an environmental sample, otherwise strain assignment will most likely be incorrect, or a large number of isolates will be characterized as being unknowns or cosmopolitan. Since the rarefaction curve in Figure 1 has not become asymptotic, our data cannot be used to predict the ultimate size that this library needs to be. However, data presented in Figure 3 indicates that with our current library size, each new isolate added to the library only has a greater than 50% chance of being new. It has been suggested that a library size of 20,000 to 40,000 isolates may be needed to capture all the genetic diversity present in *E. coli* (Mansour Samadpour, personal communication). One suggested strategy to avoid this under-representation problem in large regional or national libraries, is to develop moderate sized libraries for a highly confined geographical region, wherein isolates are only obtained from the animals in the study

area. In this way only animals pertinent to the study site, and those likely to have an impact on the targeted watershed, need to be examined in detail

In October 2002, work on DNA fingerprinting the remaining *E. coli* isolates from the three watersheds was temporarily suspended to allow us to participate in a "round-robin" study of source tracking methods (see below).

The Southern California Coastal Water Research Project (SCCWRP) and the U.S. EPA asked us to participate in a "round-robin" study to evaluate methods for determining sources of fecal pollution in waterways. SCCWRP and U.S. EPA provided funds for personnel and supplies for us to evaluate the use of our LCMR-funded rep-PCR DNA fingerprinting method to determine sources of fecal pollutants in water samples. About 22 laboratories are participating in this study, and are examining the usefulness of 8 different methods to determine sources of fecal bacteria in replicated and identical samples. All participants in the study received samples on October 8, 2002 and have about 4 months to complete their analyses. The study offers a unique opportunity to verify and validate our DNA fingerprinting methods that were developed from our LCMR-funded project. We will complete work on the SCCWRP-EPA project by the end of February 2003, and at that time we will resume work on our LCMR-sponsored project. We requested and were granted a no-cost extension of our LCMR appropriation, so that we could participate in the SCCWRP-EPA round-robin study.

Results from this round robin study were published in the November issue of the Journal of Water and Health. We contributed to two manuscripts in this publication: one dealing with genotypic methods for source tracking (Comparison of genotypic-based microbial source tracking methods requiring a host origin database by Samuel P. Myoda, C. Andrew Carson, Jeffry J. Fuhrmann, Byoung-Kwon Hahm, Peter G. Hartel, Helen Yampara-Iquise, LeeAnn Johnson, Robin L. Kuntz, Cindy H. Nakatsu, Michael J. Sadowsky and Mansour Samadpour, pp. 167-180), and the other a paper (Assessment of statistical methods used in library-based approaches to microbial source tracking by Kerry J. Ritter, Ethan Carruthers, C. Andrew Carson, R. D. Ellender, Valerie J. Harwood, Kyle Kingsley, Cindy Nakatsu, Michael Sadowsky, Brian Shear, Brian West, John E. Whitlock, Bruce A. Wiggins and Jayson D. Wilbur, pp. 209-223) on the proper statistical analyses to use to analyze DNA fingerprint data generated by rep-PCR. When our data was re-analyzed with stringent assignments to source groups, our rep-PCR based fingerprinting method was found to be far superior to almost all other methods used by participants of the study.

Our data analysis was improved, due to our collaboration with the SCCWRP statistician, Ms. Kerry Ritter, the development of a new software module, ID bootstrap, produced by Bionumerics, and the use of quality factors in our data analysis. Briefly, the ID bootstrap software addition applies reiterative analysis of the integrity of known source groups and applies the resulting correlation statistic to assign identities to unknown isolates. This determines the fidelity of our assignments of unknown isolates, and reduces the number of false positive results from 48% to 6% and false negative results from 4% to 2%.

In February 2003, following completion of the SCCWRP project, our staff resumed DNA fingerprinting of environmental *E. coli* isolates from the three watersheds using the newly developed fluorophore-enhanced modification of the rep-PCR DNA fingerprinting technique (FERP). Fingerprinting and analysis proceeded until December 2003.

Reanalysis of known source isolates - Due to improved statistical analysis techniques that we learned working on the SCCWRP project and with our consultation with Ms. Kerry Ritter, a statistician at SCCWRP, the inclusion of an amended grouping system, and the accumulation of a larger number of positive control strains, in the Fall of 2003 we reanalyzed the isolates and groups previously identified in Tables 4, 5 and 6 (see new analysis results in Tables 7, 8 and 9).

To put this analysis in perspective, it is important to note that a variety of similarity measures exist. Binary similarity coefficients are mostly used to analyze presence/absence data and band-matching data obtained from DNA fingerprints can be analyzed using binary coefficients. However, quantitative similarity coefficients require a measure of relative abundance. Quantitative coefficients can be applied to DNA fingerprints when the fingerprints are analyzed as densitometric curves that take into account both peak position and intensity (peak height). For complex DNA fingerprints, such as those produced with the techniques we used here, a curve-based method such as Pearson's product-moment correlation coefficient more reliably identified similar or identical DNA fingerprints than band matching formulas, such as simple matching, Dice, or Jaccard. Results presented here confirm that the curve-based Pearson's product-moment correlation coefficient was superior to the band-based Jaccard algorithm is correctly assigning isolates to the correct source group. The influence of analysis method on the classification of source group isolates is shown in Table 7.

Table 7. Unique *E. coli* isolates correctly classified into source groups by rep-PCR and HFERP DNA fingerprinting methods.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]			
		rep-PCR		HFERP	
		Pearson	Jaccard	Pearson	Jaccard
Pets[b]	154	61.7 (95)[d]	45.5 (70)	59.1 (91)	44.8 (69)
Chicken	144	59.7 (86)	38.9 (56)	63.2 (91)	31.9 (46)
Cow	189	55.0 (104)	47.6 (90)	62.0 (117)	48.2 (91)
Deer	96	55.2 (53)	36.5 (35)	62.2 (60)	42.6 (41)
Waterfowl[c]	216	66.2 (150)	52.8 (114)	70.4 (152)	56.5 (122)
Goat	42	66.7 (27)	59.5 (25)	47.6 (20)	42.9 (18)
Horse	78	44.3 (35)	34.2(27)	52.6 (41)	32.1 (25)
Human	210	59.2 (124)	47.4(100)	53.8 (113)	45.2 (95)
Pig	215	63.7 (137)	43.7 (94)	54.4 (117)	36.3 (78)
Sheep	61	7.5 (29)	39.3 (24)	37.7 (23)	8.2 (5)
Turkey	126	73.8 (93)	52.4 (66)	73.0 (92)	54.8 (69)
Overall	1,531	60.9 (933)	45.8 (701)	59.9 (917)	43.0 (659)

[a]Based on Jackknife analysis with 1% optimization and maximum similarities using curve-based (Pearson's product moment correlation coefficient) or band-based (Jaccard's coefficient) similarity calculations.
[b]Pet group consists of cats and dogs.
[c]Waterfowl group consists of ducks and geese.
[d]Values in parentheses are number of isolates correctly classified.

The 1,535 previously selected unique *E. coli* isolates from animals and humans were subjected to HFERP DNA fingerprinting using a combination of fluorescently labeled and unlabeled Box A1R PCR primers. Jackknife analyses of HFERP gels done using the curve-based Pearson's correlation coefficient indicated that 38-73% of the isolates were assigned to the correct source group using this technique (Table 7). For the curve-based analysis, the HFERP technique had the lowest percent of correctly classified strain in cases where the numbers of analyzed fingerprints were relatively small (for sheep, horses, and goats). The average rate of correct classification for the unique HFERP-generated DNA fingerprints was 59.9%. In contrast, Jacknife analyses of HFERP-generated DNA fingerprints done using the band-based Jaccard analysis showed that only 8-56% of the *E. coli* isolates were assigned to the correct source group, with a 43.0% average rate of correct classification. This indicates that for this type of data, the Pearson's product-moment correlation coefficient was superior to Jaccard's band matching algorithm for assigning known isolates to the correct source groups. Interestingly, results in Table 3 also show that despite problems associated with within- and between-gel variation, within-gel grouping of isolates, and repeatability issues, Jacknife analysis of rep-PCR DNA fingerprints, analyzed using Pearson's correlation coefficient, indicated that 48-74% of the isolates were assigned to the correct source group, a 60.9% average rate of correct classification. Analysis of rep-PCR DNA fingerprint data using the Jaccard band-based method was not as useful in separating *E. coli* isolates into their correct source group as was the curve-based method.

In some instances, it may be sufficient to identify unknown watershed *E. coli* isolates to larger groupings, rather than to individual animal types. To determine if the HFERP-generated DNA fingerprint data from our library of unique *E. coli* isolates grouped well into larger categories, we assembled DNA fingerprints from pets (dogs and cats), domesticated animals (chickens, cows, goats, horses, pigs, sheep, and turkeys), wild-life (deer, ducks, and geese), and humans, and used Jacknife analysis to assess the percent of correctly classified strains. Results in Table 8 show that the HFERP DNA fingerprints, analyzed using Pearson's product-moment correlation coefficient, correctly classified 83.2, 53.8, 71.4, and 59.1% of the isolates into the domesticated, human, wildlife, and pet categories, respectively. The average rate of correct classification for these groups was 74.3%. However, when DNA fingerprints were analyzed using Jaccard's coefficient, the average rate of correct classification was 66.2%. As before, the least precision was found in categories having the smallest number of fingerprints, pets and humans, suggesting that there is an apparent relationship between the number of fingerprints analyzed and the percentage of correctly classified isolates.

In microbial source tracking studies it may often be useful to determine if unknown isolates belong to either animal or human source groups, rather than to more specific categories. Results in Table 9 show that about 94% and 54% of *E. coli* from animals and humans, respectively, were

assigned to the correct source groups using HFERP-generated DNA fingerprints and Pearson's correlation coefficient. The average rate of correct classification was 88.2 and 86.1% for analyses done using Pearson's and Jaccard's algorithms, respectively. The lower percentage of correctly classified human isolates may, in part, be due to the smaller size of fingerprints analyzed for this category.

Table 8. Percentages of *E. coli* isolates correctly reclassified into domesticated, human, pets and wildlife source groups by the fluorophore-enhanced DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]
Domesticated[b]	855	83.2 (711)[d]
Human	210	53.8 (113)
Wildlife[c]	312	71.4 (223)
Pets	154	59.1 (91)
Overall	1531	78.5 (1202)

[a]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient)
[b]The domesticated group includes cat, chicken, cow, dog, goat, horse, pig, sheep and turkey.
[c]The wildlife group includes deer, duck and goose.
[d]Values in parentheses are number of isolates (n) correctly classified.

Table 9. Percentages of *E. coli* isolates correctly reclassified into human and animal source groups by the fluorophore-enhanced DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]
Animal	1321	93.7 (1237)[b]
Human	210	53.8 (113)
Overall	1531	88.2 (1350)

[a]Done using jackknife analysis with 1% optimization and maximum similarities using a curve-based (Pearson correlation coefficient) similarity calculation.
[b]Values in parentheses are number of isolates (n) correctly classified.

Results of our studies indicated that further refinements to the Jackknife analysis, including the pooling of source groups into domesticated , human, and wild-life categories, were found to improve the ability to correctly classify isolate to their respective source groups. Over 83, 53, and 71% of domesticated animals, humans, and wild-life animals, respectively, were correctly classified using this approach with the unique DNA fingerprint library analyzed by HFERP.

When all animal sources were pooled into one group, the overall correct classification rate for humans and animals by HFERP was improved to about 94 and 54%, respectively, when analyzed using the curve-based Pearson's correlation coefficient. Accordingly, these results indicated that (1) broader classifications of source groups should be used when appropriate, or (2) a targeted subset of the DNA fingerprint database should be used to more precisely determine sources of fecal pollutants in watersheds where specific source groups are known to be present. The pooling of source groups into a more limited number of categories has previously been shown to increase the average rate of correct classification following discriminant analysis of antibiotic resistance, ribotype, and rep-PCR DNA fingerprint analyses.

In summary, our results suggest that HFERP-generated Box A1R DNA fingerprints of *E. coli* are useful to differentiate between different *E. coli* subtypes of human and animal origin and that this method reduces within gel groupings of DNA fingerprints, and ensures more proper alignment and normalization of fingerprint data. However, our results further indicate that other important issues must also be resolved to more fully understand the potential applications and limitations of this and other library-based microbial source tracking methodologies. Among these are questions concerning the inclusion of identical DNA fingerprints from the same animal in the library, the number of fingerprints that must be included in an *E. coli* known source library to adequately capture the diversity of *E. coli* genotypes that exist among potential host animals, and ultimately, whether *E. coli* exhibits a sufficient level of host specificity to allow unambiguous assignment of unknown environmental *E. coli* to specific host animals.

Results of these studies were recently submitted to Applied and Environmental Microbiology for consideration of publication (see Appendix B). The manuscript is entitled: "Duplication of Genotypes in DNA Fingerprint Libraries and a High Degree of Genetic Diversity Among Natural Populations of Escherichia coli from Different Animals Influences Accuracy of Determining Sources of Fecal Pollution" by LeeAnn K. Johnson, Mary B. Brown, Ethan A. Carruthers, John A. Ferguson, Priscilla E. Dombek and Michael J. Sadowsky.

WATERSHED ANALYSES

Computer-assisted HFERP DNA fingerprint analysis.

DNA fingerprinting of environmental *E. coli* isolates, using the newly developed HFERP fluorophore-enhanced modification of the rep-PCR DNA fingerprinting technique, was initiated in September 2002. The watershed *E. coli* isolates, or "unknowns", were compared with a previously described existing known source DNA fingerprint library, consisting of 12 animal species and humans, combined into either 11, 4 or 2 groups. Results in Table 10 report the total number of isolates obtained from each watershed, and the number of bonefide *E. coli* isolated fingerprinted. The final number of *E. coli* isolates from each watershed exceeded the original goal of 1600 by 3 – 12%. In total, 5,232 *E. coli* isolates were obtained from all three watersheds. Of these, 5,189 (99%) were confirmed as *E. coli* and subjected to DNA fingerprinting.

Below are results our analysis of the identity of *E. coli* isolates obtained from each watershed. At the end of each watershed is a summary of the analysis.

Table 10. *E.coli* isolates from watersheds that were DNA fingerprinted.

Watershed	Number *E. coli* Isolated	Number *E. coli* Fingerprinted
Vermillion River	1,798	1,776
High Island Creek	1,651	1,651
Minneopa Creek	1,783	1,762

SUMMARY OF HIGH ISLAND CREEK ANALYSES

We analyzed 1,651 *E.coli* isolates obtained from the High Island Creek Watershed. Sampling took place on 4/9/2001, 7/18/2001, 4/10/2002, 4/24/2002, 5/8/2002, 5/29/2002, 6/19/2002, and 7/10/2002. Approximately 25 *E.coli* isolates were obtained from each sampling site on each date. Sites sampled were: 1s, 2p, 3p, T (tile-line), 5p, 6s, 7s, 8s, 9p, 10p, and 11. (Site 11 was sampled only on 4/9/2001 (see Appendix B for sample Map).

As reported in the July 2003 status report, we did a preliminary study in which we applied statistical analysis to unknown environmental *E. coli* isolates from High Island Creek. The Pearson cosine coefficient analysis and ID Bootstrap analysis was used on DNA fingerprint fragments sizes of 287 to 14,051 basepairs. Please note that the following figures and tables are a result of the use of improved settings and parameters, as described in the method section above. While we initially reported in the July 2003 progress report that it was our plan to exclude cat isolates from analyses due to difficulties they present, upon further examination we decided to include these isolates in the combined pet category.

Results presented in Table 11 show our analysis of the probable identity of *E. coli* isolates in High Island Creek, using Pearson's Correlation Coefficient with 1% optimization. Similar tables and figures are shown on a watershed-by-watershed basis, and discussed at the end.

Table 11. Percent Identity of *E. coli* isolates from Separate Animals, by Date, obtained from High Island Creek.

	Sample Date							
Animal	**04/09/01**	**07/18/01**	**04/10/02**	**04/24/02**	**05/08/02**	**05/29/02**	**06/19/02**	**07/10/02**
	Percent Isolates							
Cat	2.9	1.8	16.2	4.8	1.7	3.7	2.7	1.4
Chicken	4.6	1.8	0.7	3.4	6.3	8.1	12.4	10.3
Cow	13.9	18.8	4.2	15.0	16.5	12.5	9.2	12.3
Deer	28.3	29.4	23.9	25.2	6.3	16.9	18.9	13.7
Dog	2.9	5.9	14.8	6.1	6.8	8.1	2.7	3.4
Duck	3.5	2.4	0.7	3.4	5.7	1.5	3.8	2.1
Goat	0.6	1.2	0.0	1.4	2.3	5.2	1.6	5.5
Goose	8.1	11.8	2.8	9.5	10.8	14.0	13.5	17.1
Horse	0.0	0.0	0.0	1.4	6.8	2.9	0.5	0.7
Human	6.4	5.3	19.0	6.8	8.5	6.6	10.3	6.2
Pig	18.5	15.3	10.6	12.9	19.9	11.8	10.3	8.2
Sheep	7.5	4.1	3.5	2.7	2.3	1.5	2.2	4.1
Turkey	2.9	2.4	3.5	7.5	6.3	7.4	11.9	15.1

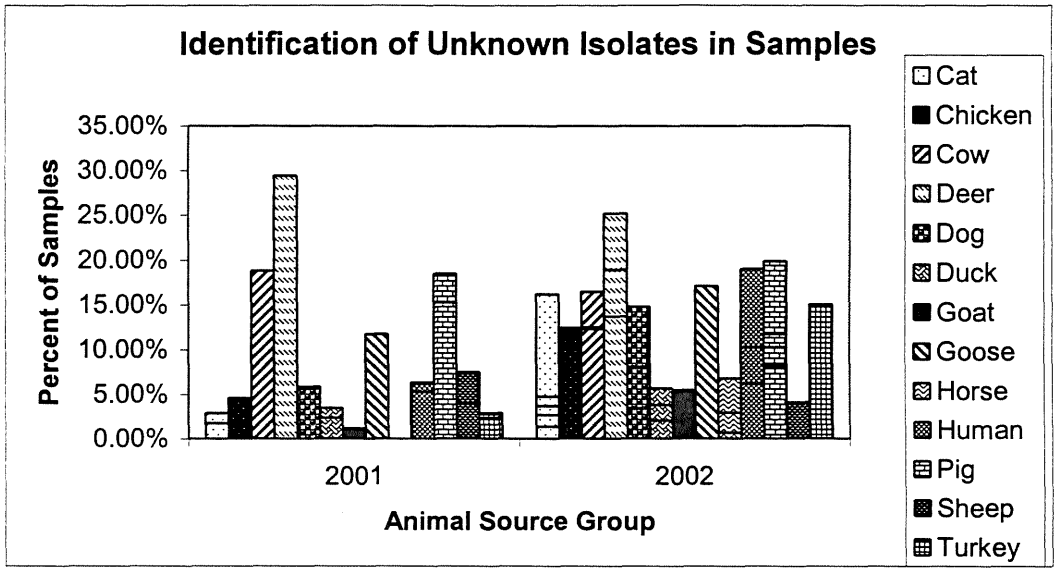


Figure 4. Percent of samples from each animal source group recovered in 2001 and 2002.

Table 12. Probable Identity of *E. coli* isolates from Separate Animals, by Site, obtained from High Island Creek.

Animal Type	Site										
	1s	2p	3p	T	5p	6s	7s	8s	9p	10p	11
	Percent of Samples										
Cat	1.4	2.6	4.4	6.3	3.7	2.2	5.0	10.4	5.9	2.7	0.0
Chicken	3.6	7.1	6.6	4.5	5.9	6.6	6.7	8.5	6.9	5.3	0.0
Cow	7.9	16.2	21.2	13.4	10.3	19.0	15.1	9.4	5.9	8.9	5.0
Deer	25.2	18.2	15.3	16.1	16.2	24.8	15.1	11.3	24.5	35.4	30.0
Dog	7.9	7.1	4.4	7.1	8.8	2.2	10.1	1.9	6.9	4.4	5.0
Duck	4.3	1.3	2.9	1.8	2.9	0.7	1.7	3.8	8.8	3.5	0.0
Goat	3.6	2.6	0.7	5.4	5.9	1.5	0.0	0.0	0.0	0.9	0.0
Goose	13.7	9.1	8.0	10.7	12.5	8.8	10.1	15.1	9.8	12.4	15.0
Horse	2.2	0.7	1.5	1.8	0.7	1.5	2.5	2.8	1.0	1.8	0.0
Human	10.8	17.5	2.9	8.0	4.4	5.1	10.1	10.4	11.8	4.4	5.0
Pig	15.1	9.1	17.5	16.1	15.4	13.9	11.8	17.9	7.8	10.6	20.0
Sheep	1.4	2.6	6.6	3.6	4.4	3.7	5.9	2.8	2.0	0.9	10.0
Turkey	2.9	5.8	8.0	5.4	8.8	10.2	5.9	5.7	8.8	8.9	10.0

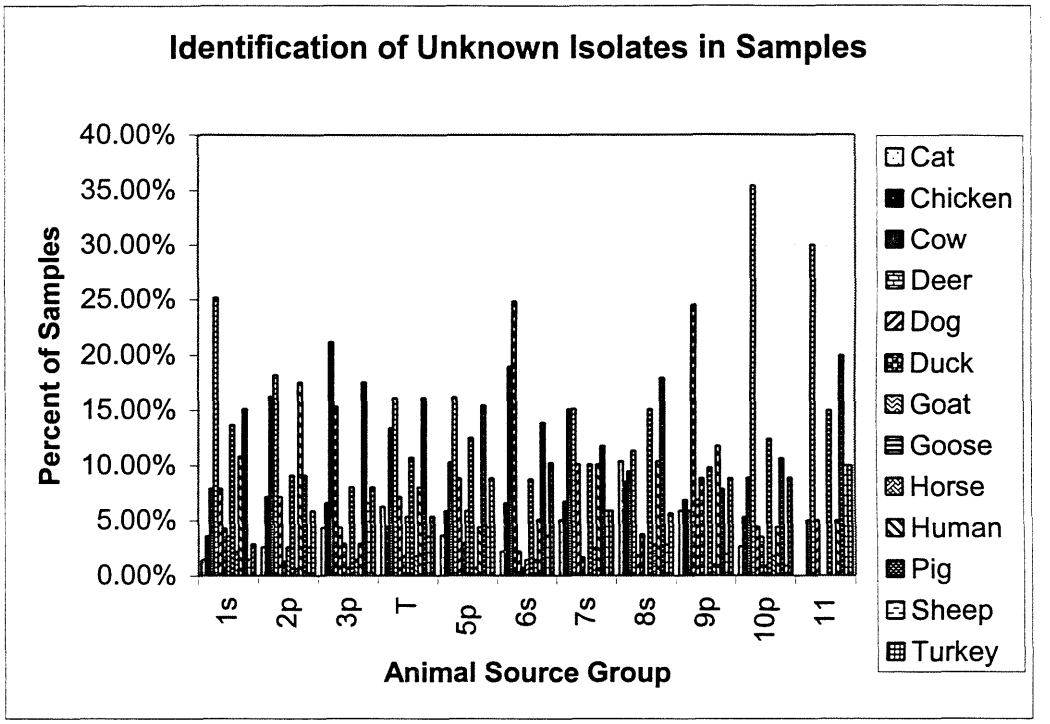


Figure 5. Percent of samples having isolates in each animal source group recovered by Site, obtained from High Island Creek.

Table 13. Probable Identity of *E. coli* isolates from Domestic, Human, Pets, Waterfowl, Wild source groups, by Sample Date, obtained from High Island Creek.

Animal Type	Date							
	04/09/01	07/18/01	04/10/02	04/24/02	05/08/02	05/29/02	06/19/02	07/10/02
	Percent of Isolates							
Domesticated	48.0	43.5	22.5	47.8	60.2	49.3	48.1	56.2
Human	6.4	5.3	19.0	6.4	8.5	6.6	10.3	6.2
Pets	5.8	7.7	31.0	10.2	8.5	11.8	5.4	4.8
Waterfowl	11.6	14.1	3.5	12.1	16.5	15.4	17.3	19.2
Wild Animal (Deer)	28.3	29.4	23.9	23.6	6.3	16.9	18.9	13.7

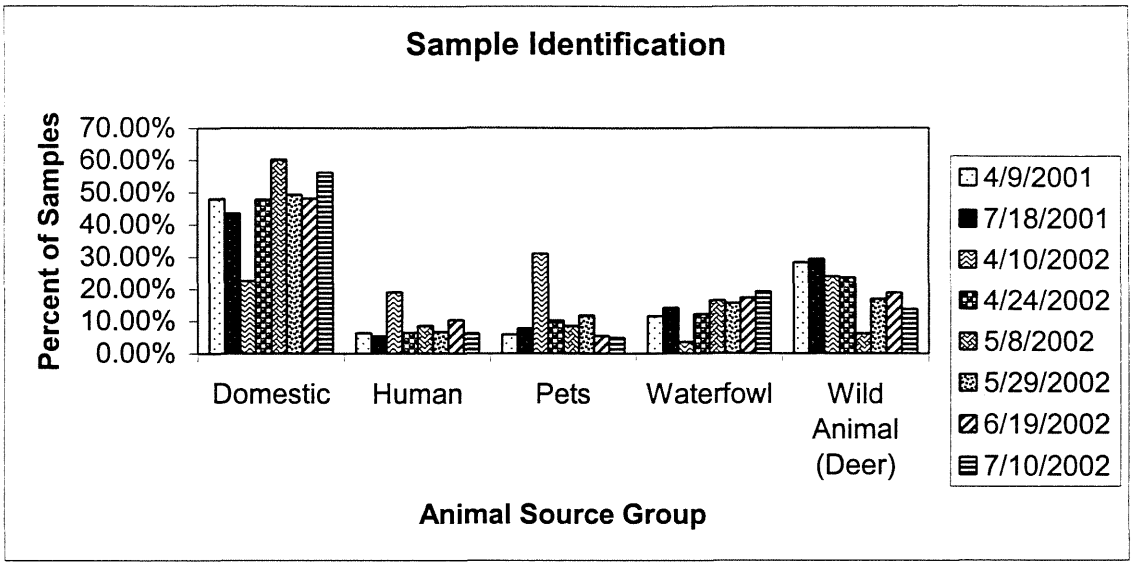


Figure 6. Percent of samples having isolates in animal source groups by sample date.

Table 14. Probable Identity of *E. coli* isolates from Domestic, Human, Pets, Waterfowl, Wild source groups, by Site, obtained from High Island Creek.

Animal Type	1s	2p	3p	T	5p	6s	7s	8s	9p	10p	11
	Percent of Samples										
Domesticated	36.7	44.2	62.0	50.0	51.5	56.2	47.9	47.2	32.4	37.2	45.0
Human	10.8	17.5	2.9	8.0	4.4	5.1	10.1	10.4	11.8	4.4	5.0
Pets	9.4	9.7	8.8	13.4	12.5	4.4	15.1	12.3	12.8	7.1	5.0
Waterfowl	18.0	10.4	11.0	12.5	15.4	9.5	11.8	18.9	18.6	15.9	15.0
Wild Animal (Deer)	25.2	18.2	15.3	16.1	16.2	24.8	15.1	11.3	24.5	35.4	30.0

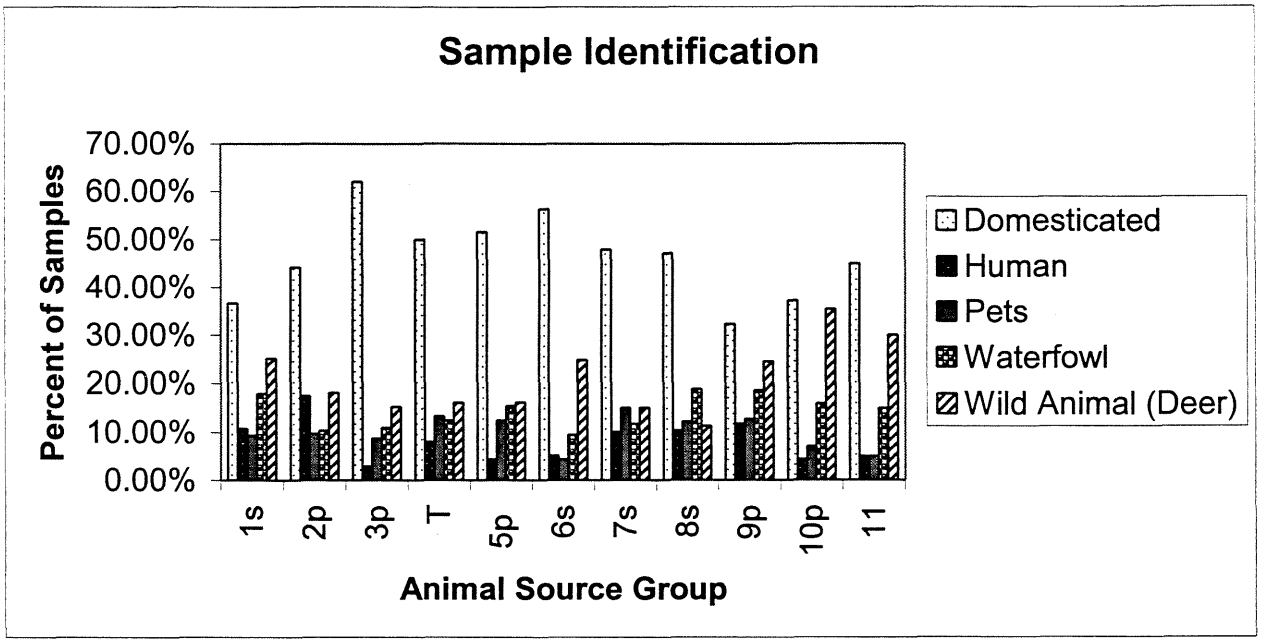


Figure 7. Percent of samples having isolates in Domestic, Human, Pets, Waterfowl, and Wild animal source groups by site.

Table 15. Probable Identity of *E. coli* isolates from Household (human) and Non-Households, by Sample Date, obtained from High Island Creek.

Animal Type	Date							
	04/09/01	07/18/01	04/10/02	04/24/02	05/08/02	05/29/02	06/19/02	07/10/02
	Percent of Samples							
Household	12.1	12.9	50.0	17.7	17.1	18.4	15.7	11.0
Non-Household	87.9	87.1	50.0	82.3	83.0	81.6	84.3	89.0

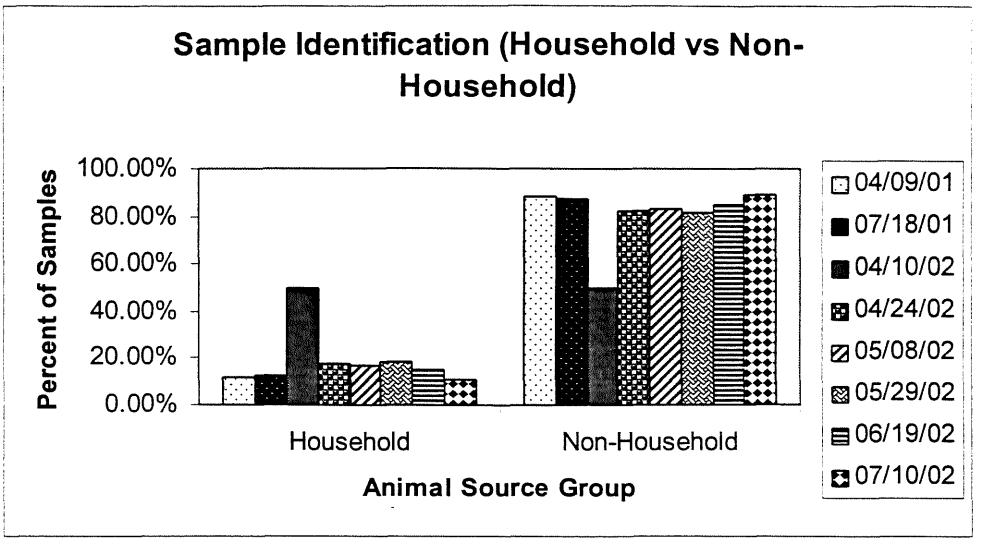


Figure 8. Percent of samples having isolates from Household (human) and Non-Households, by Sample Date, obtained from High Island Creek.

Table 16. Probable Identity of *E. coli* isolates from Household and Non-Households, by Site, obtained from High Island Creek.

Animal Type	Site										
	1s	2p	3p	T	5p	6s	7s	8s	9p	10p	11
	Percent of Samples										
Household	20.1	27.3	11.7	21.4	16.9	9.5	25.2	22.6	24.5	11.5	10.0
Non-Household	79.9	72.7	88.3	78.6	83.1	90.5	74.8	77.4	75.5	88.5	90.0

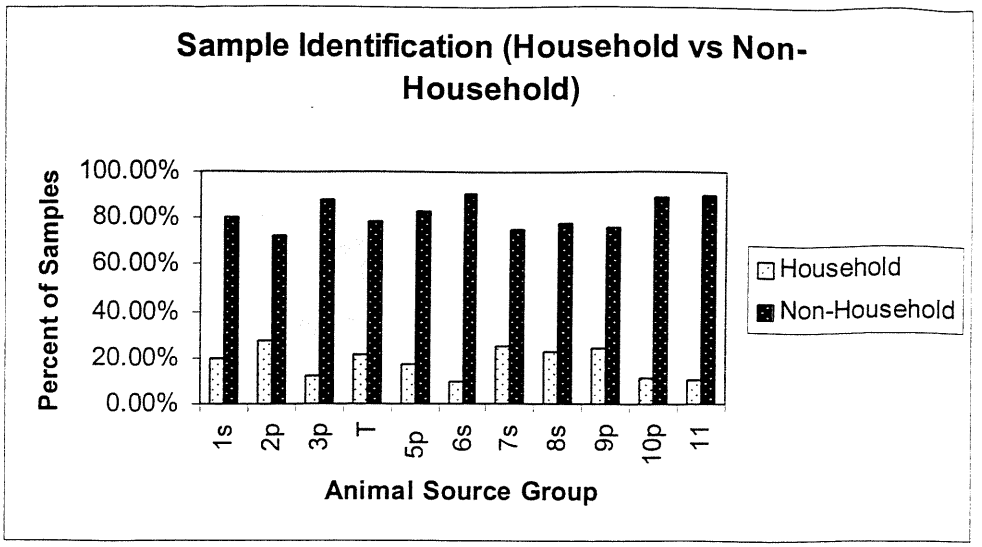


Figure 9. Percent of samples having isolates from Household and Non-Households, by Sample Site, obtained from High Island Creek.

Table 17. Probable Identity of *E. coli* isolates from Humans and Non-Humans, by Sample Date, obtained from High Island Creek.

Animal Type	Date							
	04/09/01	07/18/01	04/10/02	04/24/02	05/08/02	05/29/02	06/19/02	07/10/02
	Percent of Samples							
Human	6.4	5.3	19.0	6.8	8.5	6.6	10.3	6.2
Non-Human	93.6	94.7	81.0	93.2	91.5	93.4	89.7	93.8

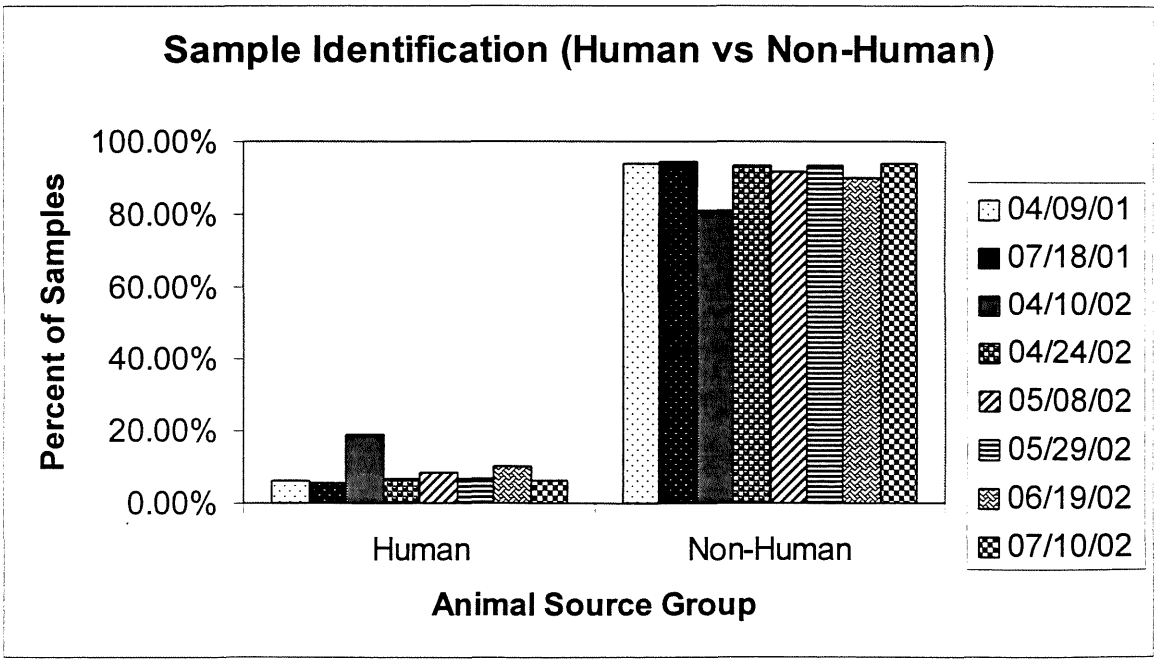


Figure 10. Percent of samples having isolates from Humans and Non-Humans, by Sample Date, obtained from High Island Creek.

Table 18. Probable Identity of *E. coli* isolates from Humans and Non-Humans, by Site, obtained from High Island Creek.

Animal Type	Site										
	1s	2p	3p	T	5p	6s	7s	8s	9p	10p	11
	Percent of Samples										
Human	10.8	17.5	3.0	8.0	4.4	5.1	10.1	10.4	11.8	4.4	5.0
Non-Human	89.2	82.5	97.1	92.0	95.6	94.9	89.9	89.6	88.2	95.6	95.0

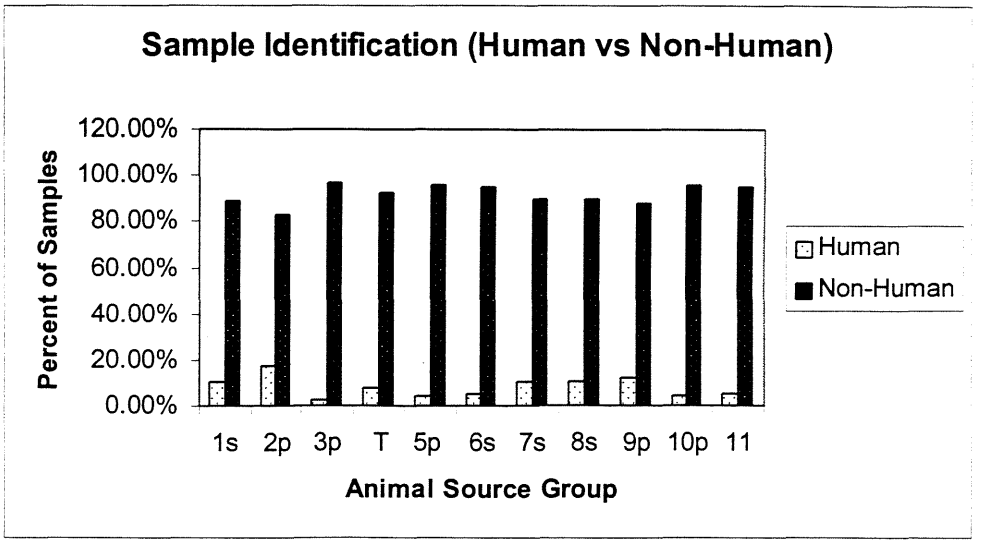


Figure 11. Percent of samples having isolates from Humans and Non-Humans, by Site, obtained from High Island Creek.

ID Bootstrap Analysis of High Island Creek Isolates

While the above analyses show the probable identities of various *E. coli* isolates from the High Island Creek watershed, our initial studies and subsequent analyses done during the SCCWRP project indicated that despite the use of quality factors in our analyses, our library size limitations and the presence of isolates with similar ID values makes definitive classification of isolates somewhat tenuous. To overcome this limitation, we used ID bootstrap software provided by Bionumerics to make more definitive assignment of the unknown isolates. ID boostrapping applies reiterative analysis of the integrity of known source groups and applies the resulting correlation statistic to assign identities to unknown isolates. This analysis reduces the number of false positive results by 42% and false negative results by 2%. Moreover, while the analysis results in us discarding isolates from our final assignments, the isolates that remain are assured (at a $\geq$90% confidence level) more correct classification. Below is the ID bootstrap analysis for High Island Creek Isolates. Given that isolates had to be discarded from consideration due to non-statistical assignment reasons, only analyses by animal group (rather than by date or site) is given. Nevertheless, this gives a better picture of the probable source of isolates present in the watershed.

Table 19. Probable Identity of *E. coli* isolates obtained from High Island Creek, by Animals Group, using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Cat	9	3.9
Chicken	8	3.5
Cow	49	21.3
Deer	24	10.4
Dog	21	9.1
Duck	9	3.9
Goat	4	1.7
Goose	33	14.4
Horse	3	1.3
Human	37	16.1
Pig	3	1.3
Sheep	15	6.5
Turkey	15	6.5
TOTAL	**230**	

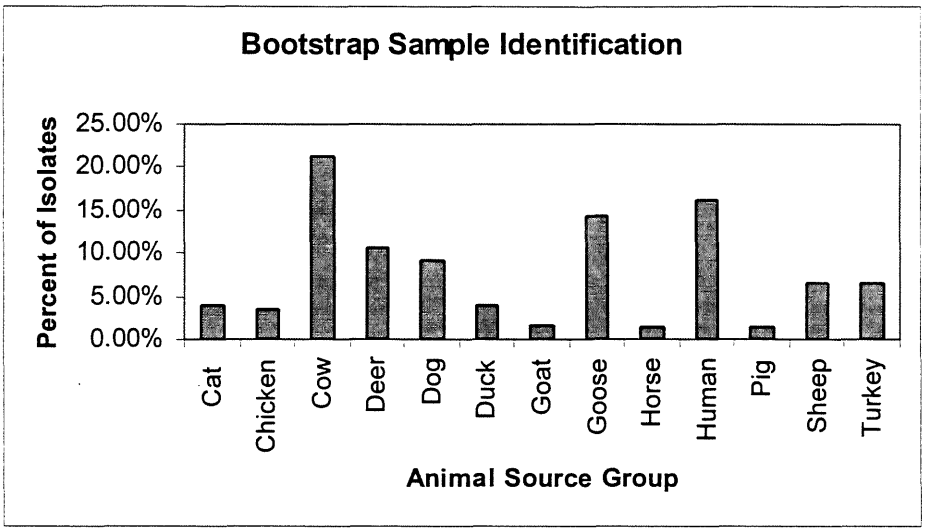


Figure 12. Percent of Isolates from obtained from High Island Creek in Animals Group by using ID Bootstrap Analysis.

Table 20. Probable Identity of *E. coli* Isolates Obtained from High Island Creek, in Domestic, Human, Pet, Waterfowl, and Wild Animal Groups, Using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Domesticated	97	42.2
Human	37	16.1
Pets	30	13.0
Waterfowl	42	18.3
Wild Animal (Deer)	24	10.4
TOTAL	230	

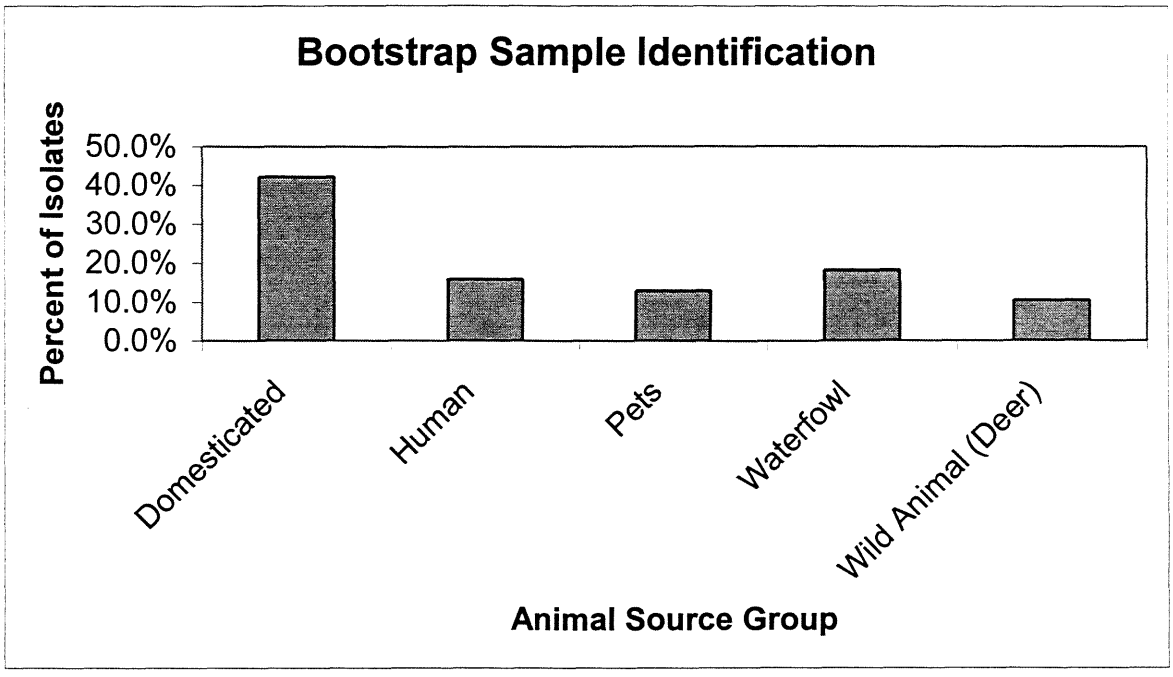


Figure 12. Percent of Isolates from obtained from High Island Creek in Domestic, Human, Pet, Waterfowl, and Wild Animal Groups, using ID Bootstrap Analysis.

Table 21. Probable Identity of *E. coli* isolates from Household and Non-Households obtained from High Island Creek, Using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Household	67	29.1
Non-Household	163	70.9
TOTAL	230	

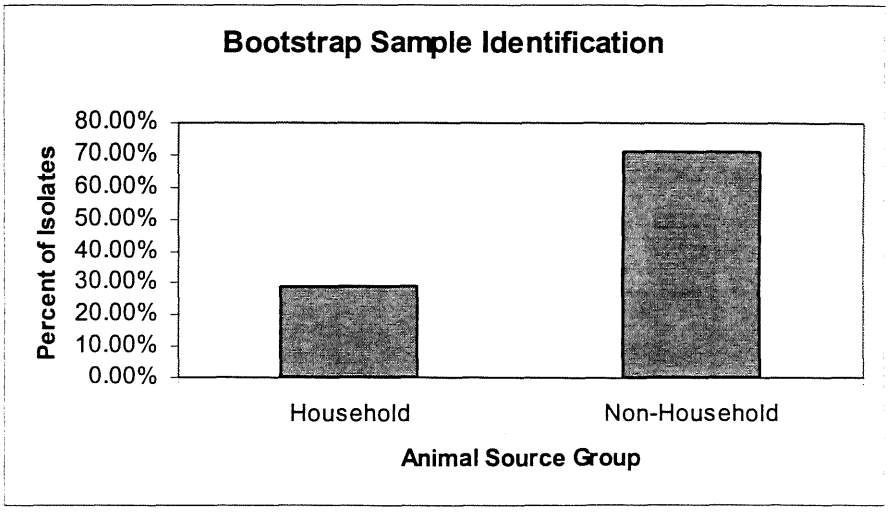


Figure 13. Percent of Isolates Obtained from High Island Creek in Household and Non-Household groups Using ID Bootstrap Analysis.

Table 22. Probable Identity of *E. coli* isolates obtained from High Island Creek in Humans and Non-Humans groups Using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Human	37	16.1
Non-Human	193	83.9
TOTAL	230	

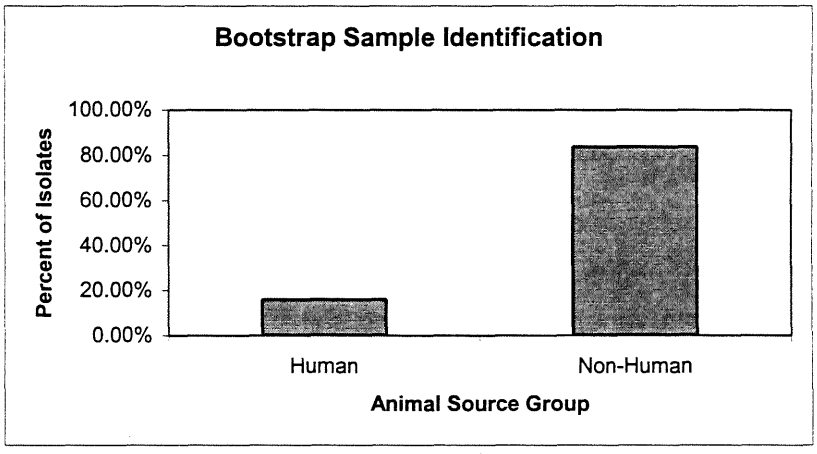


Figure 14. Percent of Isolates Obtained from High Island Creek in Human and Non-Human groups Using ID Bootstrap Analysis.

DISCUSSION OF HIGH ISALND CREEK ANALYSES

Results presented in Tables 11-18 and Figures 4-11 report on our analyses of the identities of watershed isolates obtained the High Island Creek Watershed. In total 1,651 *E. coli* isolates were analyzed. As we indicated above, the most stringent analysis for the potential identification of these isolates comes from ID bootstrap studies (Tables 19-22 and Figures 12-14). Thus while the trends for Pearson's analyses Tables 11-18 are similar, we believe that table data from ID bootstrap analyses more accurately reflect the identity of isolates to a 90% certainty. Consequently, we will discuss these results in more detail. Results in Table 19 show that the majority of isolates were removed from the study using ID bootstrap analysis. However, of the remaining isolates, the majority have cow as their source. Our data however, points to the fact that there is some input into this watershed from Geese (14% of isolates) and humans (16%). Results in Table 20 and Figure 13 show that when we break down the isolates into larger groups, that domesticated animals (chickens, cows, goats, horses, pigs, sheep, and turkeys) contribute 42% of the *E. coli* isolates to this watershed, the remainder mostly being contributed by humans and waterfowl (ducks and geese). Thus, the majority of isolates in the watershed come from non-household sources (Table 21 and Figure 14). On an even larger scale, this is further reflected in Table 22 and Figure 14 which show that over 80% of the *E. coli* isolates in the High Island Creek watershed come from non-human sources.

SEE ATTACHEMNT C FOR ADDITIONAL RAW DATA BY SITE AND DATE

SUMMARY OF VERMILLION RIVER ANALYSIS – Analyses done using Pearson's cosine coefficient analysis with 1% optimization.

We analyzed 1,776 *E. coli* isolates obtained from the Vermillion River Watershed. Sampling took place on 7/11/01, 08/08/01, 09/05/01, 10/03/01, 03/27/02, 05/01/02, 06/05/02, and 07/02/02. Approximately 25 *E.coli* isolates were obtained from each site at each date. The sites sampled were: VMCwest, VMCeast, VMC, VNC175, VNC, VSBtrib, VSB, V31, Vverm, and V47.

Table 23. **Percent Identity of *E. coli* isolates from Separate Animals, by Date, obtained from the Vermillion River watershed.**

Animal Type	Date							
	07/11/01	08/08/01	09/05/01	10/03/01	03/27/02	05/01/02	06/05/02	07/02/02
	Percent of Isolates							
Cat	15.2	11.6	12.4	16.8	3.3	10.8	13.0	11.2
Chicken	0.5	0.0	2.3	3.2	3.3	8.5	1.6	0.6
Cow	7.3	2.9	4.5	7.9	16.7	18.2	10.9	11.8
Deer	14.7	11.6	16.9	8.4	3.9	5.7	5.2	4.1
Dog	5.2	12.1	9.0	6.3	17.2	12.5	4.7	8.3
Duck	5.8	1.7	5.6	2.1	2.2	2.8	1.6	6.5
Goat	0.5	0.0	0.0	1.1	0.6	0.6	1.0	0.0
Goose	14.1	17.3	9.0	17.4	8.3	11.9	22.4	13.6
Horse	1.6	0.0	0.0	0.0	0.0	0.0	2.6	0.6
Human	3.1	13.9	6.2	4.2	21.7	8.0	6.3	4.7
Pig	16.2	13.9	13.5	12.1	11.1	8.0	15.1	7.7
Sheep	5.8	6.4	6.7	14.2	8.3	6.3	9.9	13.6
Turkey	10.0	8.7	14.0	6.3	3.3	6.8	5.7	17.2

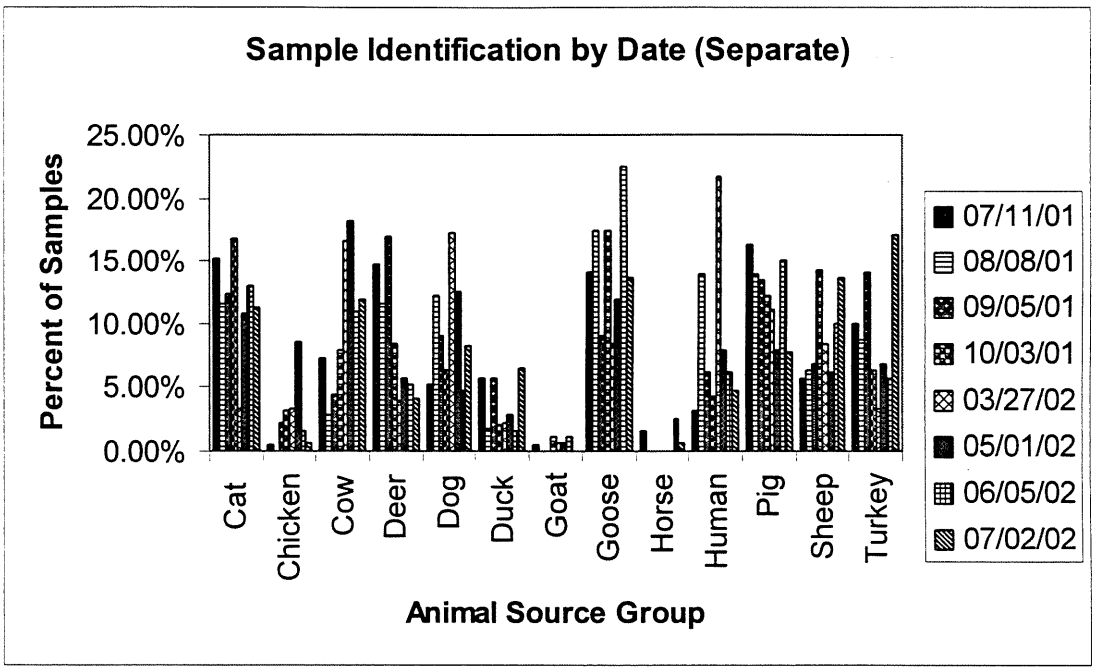


Figure 15. Percent of samples in each animal source group recovered, by Sample date, in the Vermillion River Watershed.

Table 24. Percent Identity of *E. coli* isolates from Separate Animals, by Site, obtained from the Vermillion River watershed.

Animal Type	Site									
	VMC west	VMCeast	VMC	VNC175	VNC	VSBtrib	VSB	V31	Vverm	VMC west
	Percent of Samples									
Cat	15.3	12.1	13.0	10.7	8.2	16.8	8.4	9.0	12.3	13.3
Chicken	0.8	0.0	0.7	0.7	2.2	7.7	5.4	3.8	2.1	1.3
Cow	3.1	4.7	18.8	8.1	4.4	6.3	16.2	7.5	15.8	13.3
Deer	6.1	9.4	16.0	10.7	9.6	9.8	6.6	5.3	6.9	8.2
Dog	9.9	4.0	4.4	13.4	11.9	9.1	9.0	15.8	6.9	9.5
Duck	4.6	2.7	5.1	11.4	0.7	0.7	1.8	2.3	3.4	2.5
Goat	0.0	0.7	0.0	0.7	0.7	0.0	0.6	2.3	0.0	0.0
Goose	9.2	16.1	8.7	10.7	17.0	11.9	16.8	18.8	16.4	17.1
Horse	0.0	0.0	0.7	0.0	0.7	0.7	0.6	3.0	0.7	0.0
Human	11.5	6.0	6.5	9.4	6.7	13.3	6.0	9.0	6.2	10.1
Pig	16.0	16.8	13.8	10.1	17.8	4.9	14.4	7.5	9.6	12.0
Sheep	6.1	7.4	7.3	4.7	11.9	14.0	7.8	9.8	11.0	9.5
Turkey	17.6	20.1	5.1	9.4	8.2	4.9	6.6	6.0	8.9	3.2

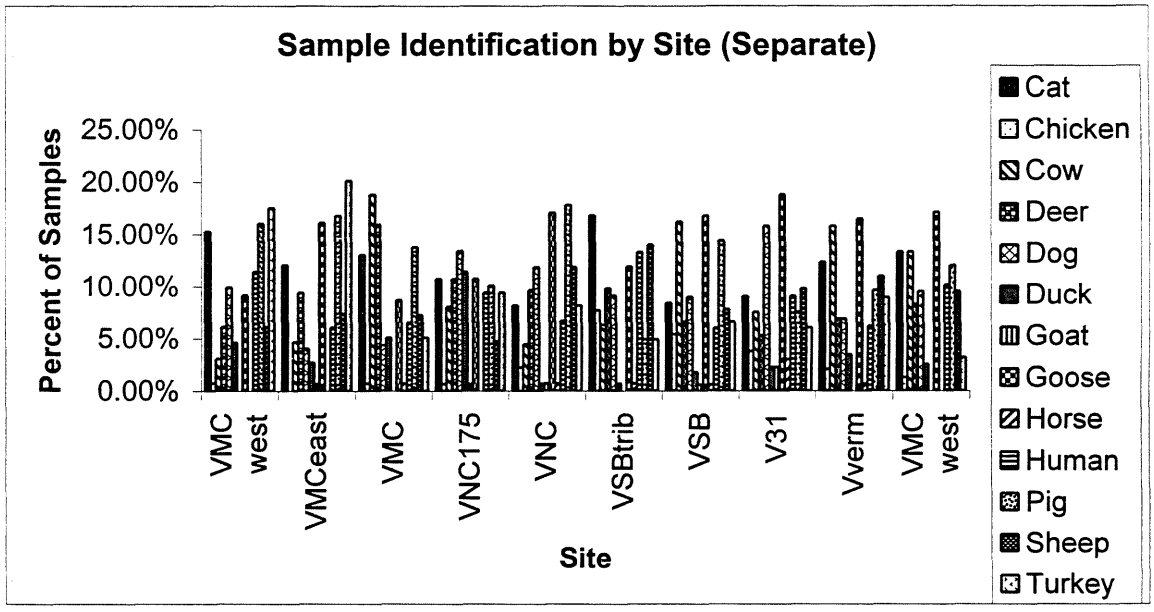


Figure 16. Percent of samples in each animal source group recovered by Site, in the Vermillion River Watershed.

Table 25. Percent of *E. coli* samples obtained from the Vermillion River watershed in Domesticated, Human, Pets, Waterfowl, and Wild animal Source Groups, by Date.

Animal Type	Date							
	07/11/01	08/08/01	09/05/01	10/03/01	03/27/02	05/01/02	06/05/02	07/02/02
	Percent of Samples							
Domesticated	41.9	31.8	41.0	44.7	43.3	48.3	46.9	51.5
Human	3.1	13.9	6.2	4.2	21.7	8.0	6.3	4.7
Pets	20.4	23.7	21.4	23.2	20.6	23.3	17.7	19.5
Waterfowl	19.9	19.1	14.6	19.5	10.6	14.8	24.0	20.1
Wild Animal (Deer)	14.7	11.6	16.9	8.4	3.9	5.7	5.2	4.1

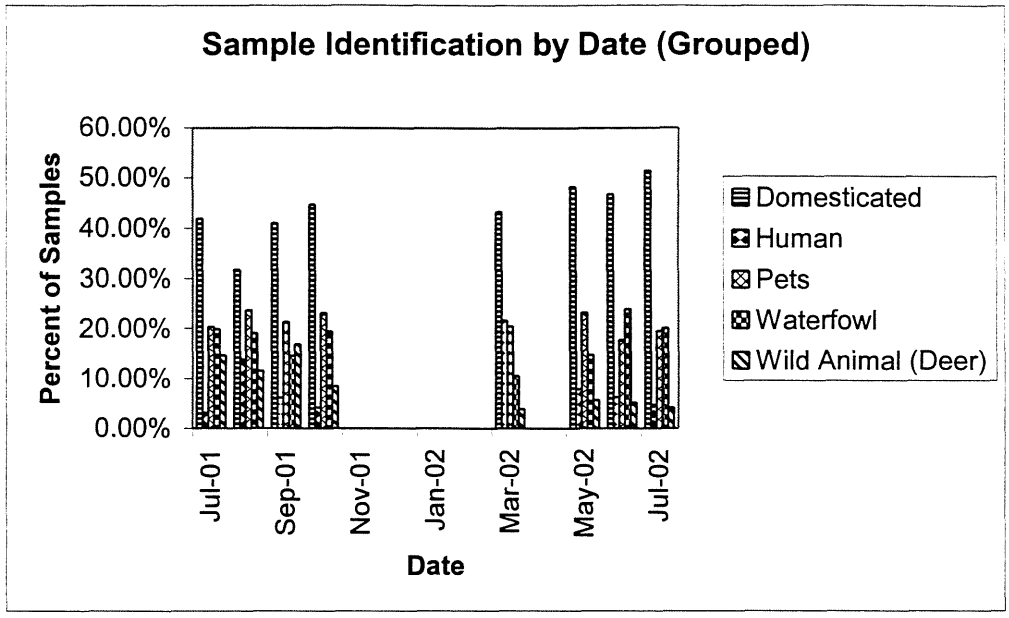


Figure 17. Percent of samples in Domesticated, Human, Pets, Waterfowl, and Wild animal Source Groups recovered by Date, in the Vermillion River Watershed.

Table 26. Percent of samples obtained from the Vermillion River watershed in Domesticated, Human, Pets, Waterfowl, and Wild animal Source Groups, by Site.

	Site									
Animal Type	**VMC west**	**VMCeast**	**VMC**	**VNC175**	**VNC**	**VSBtrib**	**VSB**	**V31**	**Vverm**	**VMC west0**
	Percent of Samples									
Domesticated	43.5	49.7	46.4	33.6	45.9	44.4	51.5	39.9	48.0	39.2
Human	11.5	6.0	6.5	9.4	6.7	15.3	6.0	9.0	6.2	10.1
Pets	25.2	16.1	17.4	24.2	20.0	14.5	17.4	24.8	19.2	22.8
Waterfowl	13.7	18.8	13.8	22.2	17.8	14.5	18.6	21.1	19.9	19.6
Wild Animal (Deer)	6.1	9.4	15.9	10.7	9.6	11.3	6.6	5.3	6.9	8.2

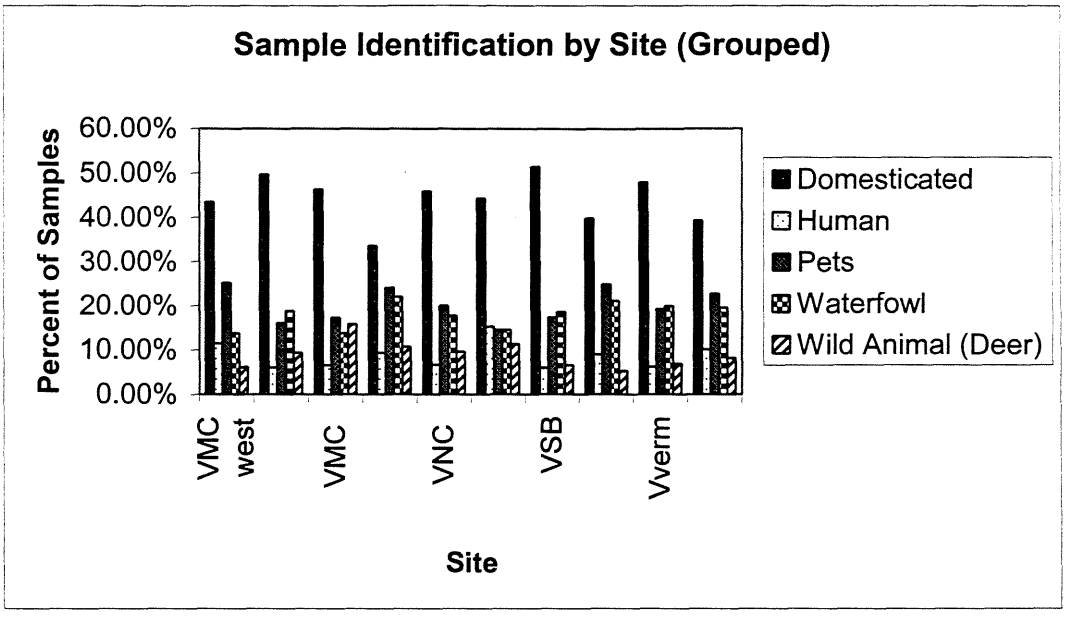


Figure 18. Percent of samples in Domesticated, Human, Pets, Waterfowl, and Wild animal Source Groups recovered by Site, in the Vermillion River Watershed.

Table 23. Percent of samples obtained from the Vermillion River watershed in Household and non-household Source Groups, by Date.

	Date							
Animal Type	**07/11/01**	**08/08/01**	**09/05/01**	**10/03/01**	**03/27/02**	**05/01/02**	**06/05/02**	**07/02/02**
	Percent of Samples							
Household	23.6	37.6	27.5	27.4	42.2	31.3	24.0	24.3
Non-Household	76.4	62.4	72.5	72.6	57.8	68.8	76.0	75.7

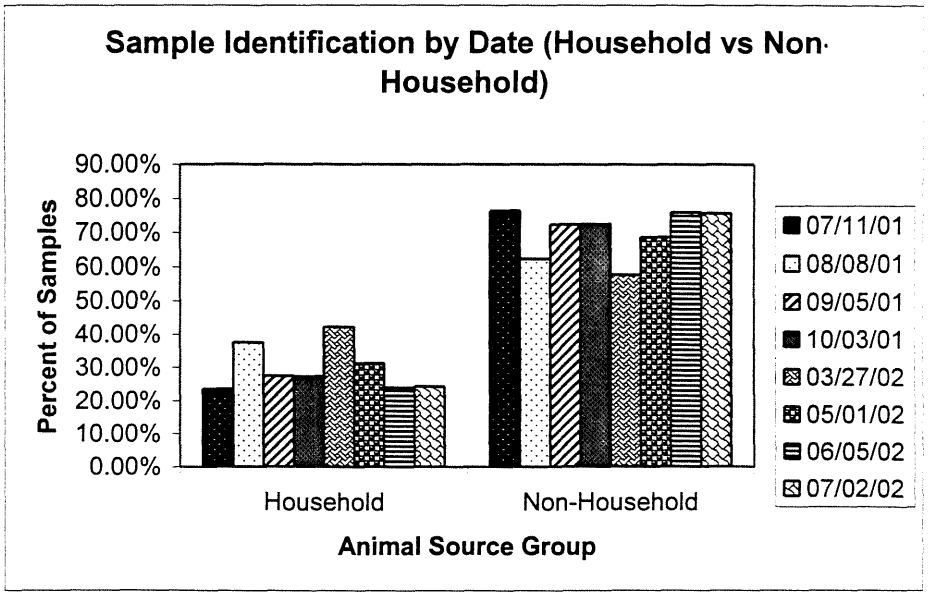


Figure 19. Percent of samples in Household and non-Household Source Groups recovered by Date, in the Vermillion River Watershed.

Table 24. Percent of samples obtained from the Vermillion River watershed in Household and non-household Source Groups, by Site.

	Site									
Animal Type	**VMC west**	**VMCeast**	**VMC**	**VNC175**	**VNC**	**VSBtrib**	**VSB**	**V31**	**Vverm**	**VMC west**
	Percent of Samples									
Household	36.6	22.2	23.9	33.6	26.7	39.2	23.4	33.8	25.3	32.9
Non-Household	63.4	77.9	76.1	66.4	73.3	60.8	76.7	66.2	74.7	67.1

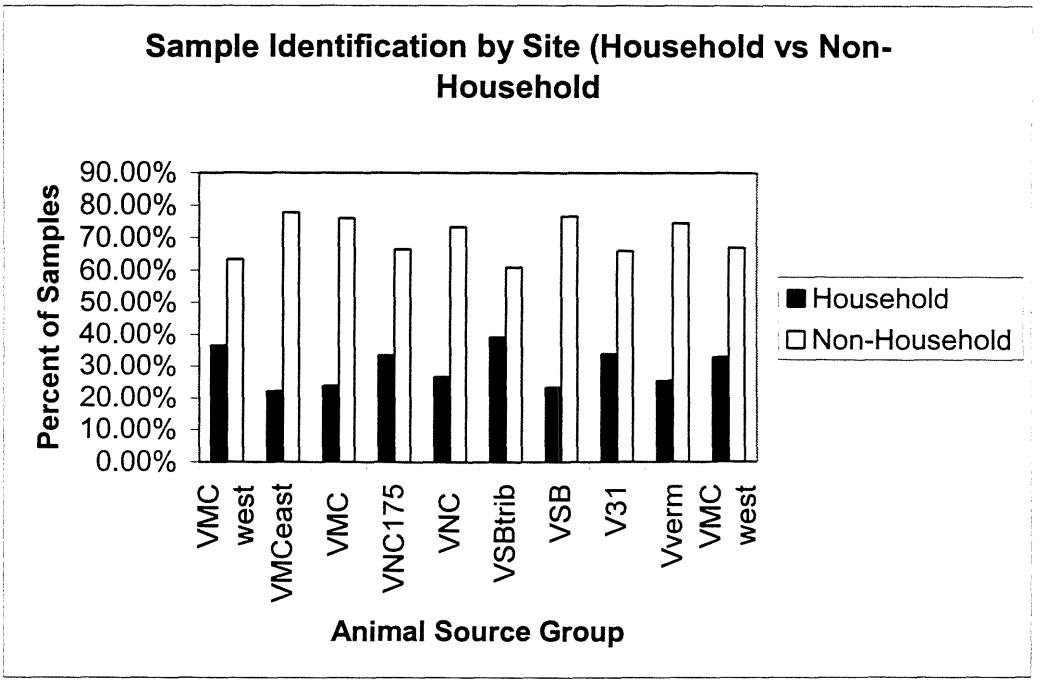


Figure 20. Percent of samples in Household and non-Household Source Groups recovered by Site, in the Vermillion River Watershed

Table 25. Percent of samples obtained from the Vermillion River watershed in Human and Non-Human Source Groups, by Date.

Animal Type	Date							
	07/11/01	08/08/01	09/05/01	10/03/01	03/27/02	05/01/02	06/05/02	07/02/02
	Percent of Samples							
Human	3.1	16.8	6.7	8.1	22.5	11.5	9.6	14.4
Non-Human	96.9	83.2	93.3	91.9	77.5	88.5	90.5	85.6

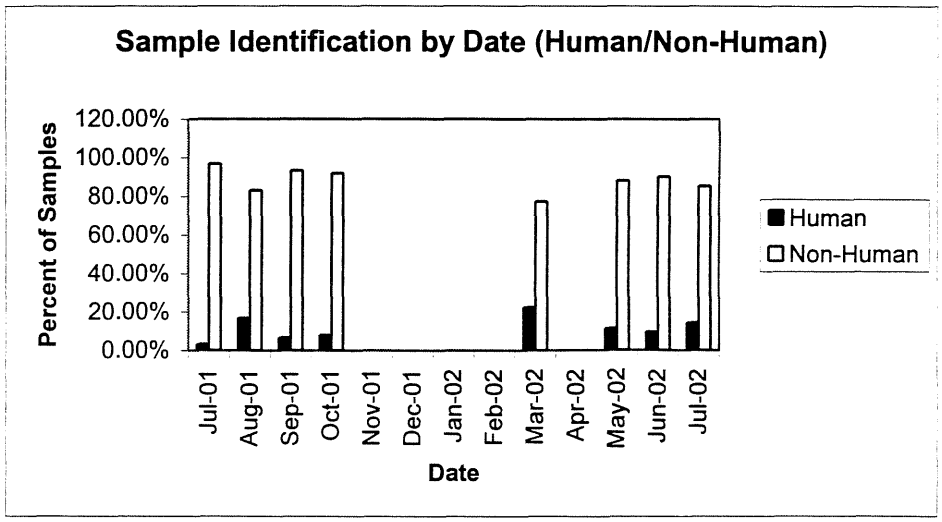


Figure 21. Percent of samples in Human and non-Human Source Groups recovered by Date, in the Vermillion River Watershed

Table 26. Percent of samples obtained from the Vermillion River watershed in Human and Non-Human Source Groups, by Site.

Animal Type	Site									
	VMCwest	VMCeast	VMC	VNC175	VNC	VSBtrib	VSB	V31	Vverm	V47
	Percent of Samples									
Human	20.6	9.7	9.2	14.6	8.7	16.2	7.1	14.8	7.4	11.8
Non-Human	79.5	90.3	90.9	85.4	91.3	83.8	92.9	85.2	92.6	88.2

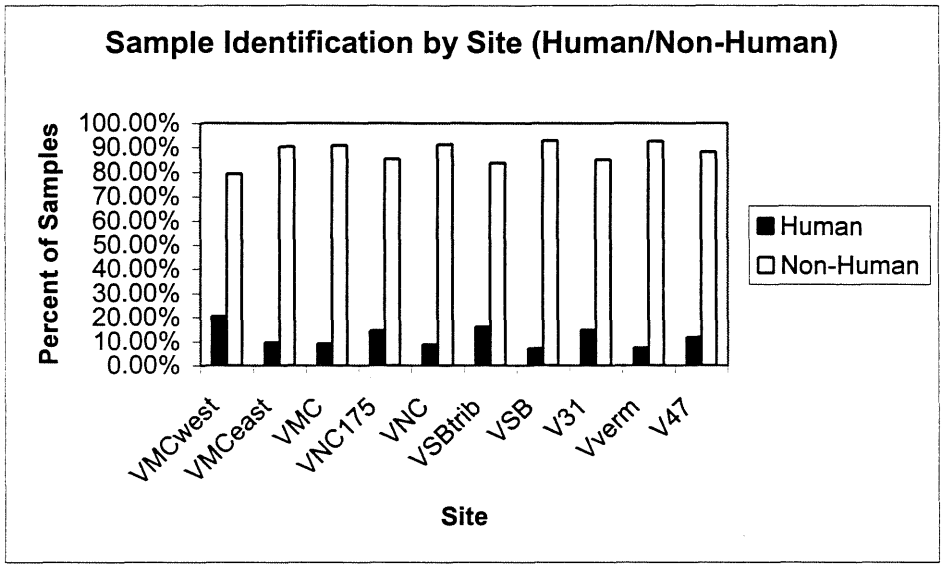


Figure 22. Percent of samples in Human and non-Human Source Groups recovered by Site in the Vermillion River Watershed

BOOTSTRAP ANALYSIS OF VERMILLION RIVER WATERSHED

Table 31. Bootstrap analysis of isolates from the Vermillion River Watershed, percentage and frequency by animals.

Animal Type	Frequency	Percentage
Cat	49	23.0
Chicken	6	2.8
Cow	28	13.2
Deer	41	19.3
Dog	47	22.2
Duck	3	1.4
Goat	4	1.9
Goose	12	5.6
Horse	0	0.0
Human	13	6.1
Pig	0	0.0
Sheep	6	2.8
Turkey	4	1.9
TOTAL	213	

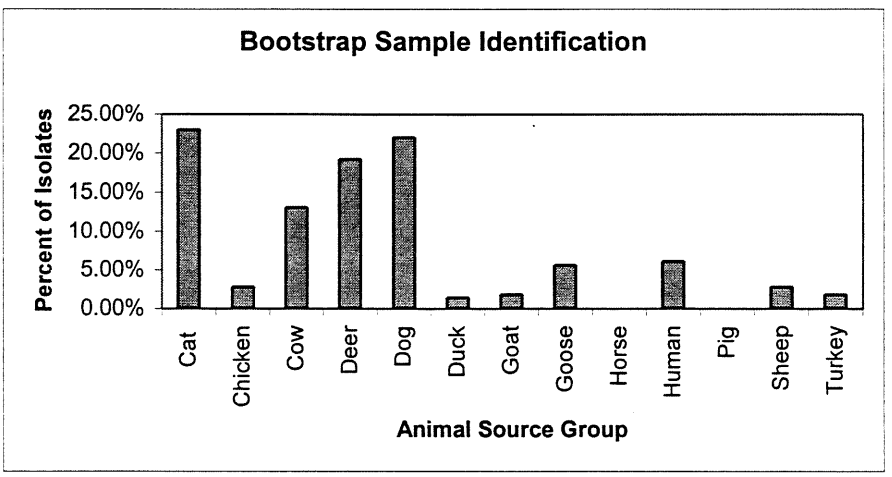


Figure 23. Percent of isolates in Animal Source Groups recovered in the Vermillion River Watershed, ID Bootstrap Analysis.

Table 32. Bootstrap analysis of isolates from the Vermillion River Watershed, percentage and frequency of isolates in domesticated, human, pet, waterfowl, and wild animal source groups.

Animal Type	Frequency	Percentage
Domesticated	48	22.5
Human	13	6.1
Pets	96	45.1
Waterfowl	15	7.0
Wild Animal (Deer)	41	19.3
TOTAL	213	

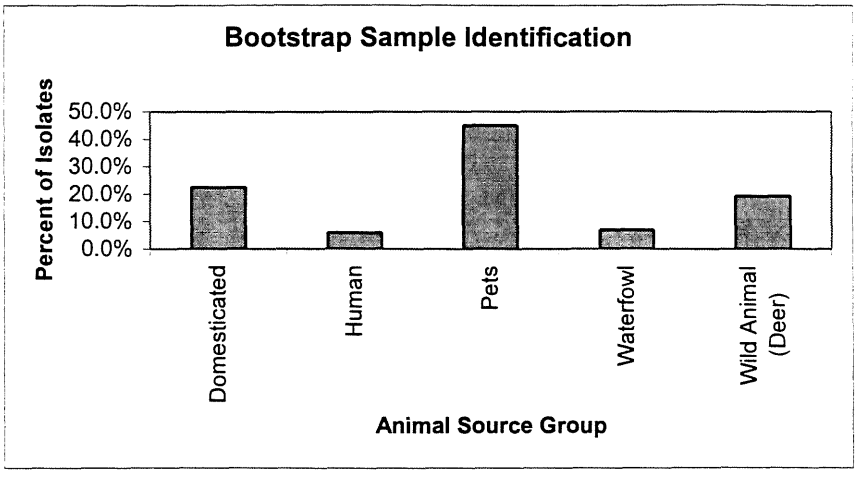


Figure 24. Percent of isolates in domesticated, human, pet, waterfowl, and wild animal source groups in the Vermillion River Watershed, ID Bootstrap Analysis

Table 33. Bootstrap analysis of isolates from the Vermillion River Watershed, percentage and frequency of isolates in Household and Non-household source groups.

Animal Type	Frequency	Percentage
Household	109	51.2
Non-Household	104	48.8
TOTAL	213	

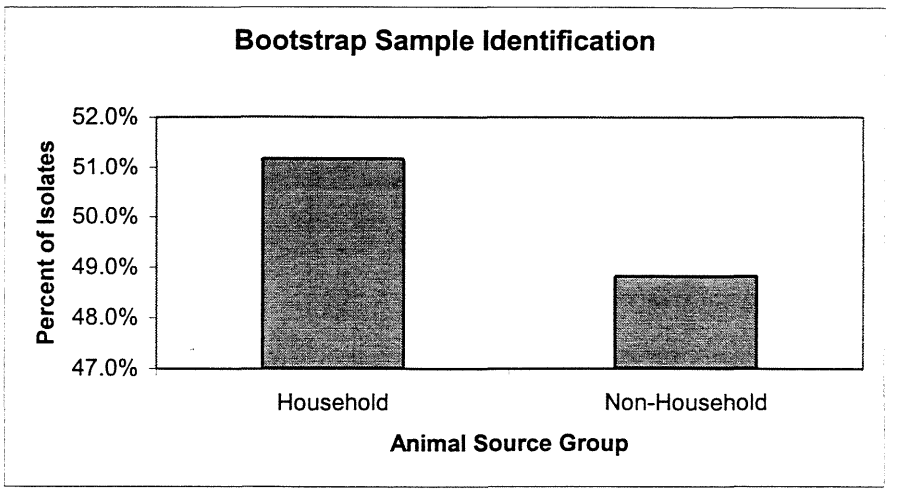


Figure 25. Percent of isolates in Household and Non-Household source groups in the Vermillion River Watershed, ID Bootstrap Analysis

Table 34. Bootstrap analysis of isolates from the Vermillion River Watershed, percentage and frequency of isolates in Human and Non-Human source groups.

Animal Type	Frequency	Percentage
Human	13	6.1
Non-Human	200	93.9
TOTAL	213	

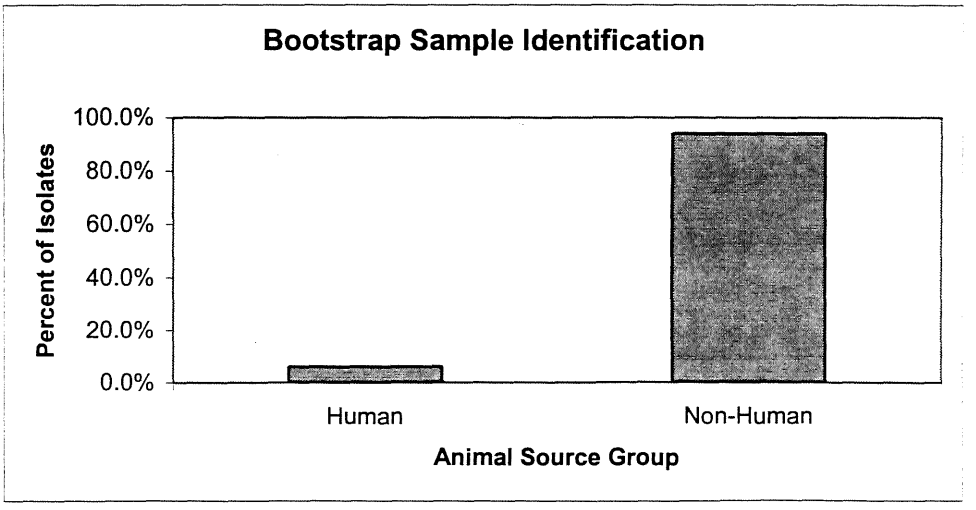


Figure 26. Percent of isolates in Human and Non-Human source groups in the Vermillion River Watershed, ID Bootstrap Analysis.

SEE ATTACHMENT C FOR ADDITIONAL RAW DATA BY SITE AND DATE

DISCUSSION OF VERMILLION RIVER ANALYSES

Results presented in Tables 23-30 and Figures 15-22 report on our analyses of the identities of watershed isolates obtained from the Vermillion River Watershed. In total 1,798 *E. coli* isolates were analyzed. As we indicated above, the most stringent analysis for the potential identification of these isolates comes from ID bootstrap studies (Tables 31-34 and Figures 23-26). Thus while the trends for Pearson's analyses Tables 23-30 are similar, we believe that table data from ID bootstrap analyses more accurately reflect the identity of isolates to a 90% certainty. There was a distinct clustering of data by sample date and site, suggesting that climatic and land use factors affected the origin of isolates present (Figures 17 and 18). However, generally speaking, the sites were dominated by bacteria from domesticated, non-household, non-human animals, regardless of the date or site. Since ID bootstrap gives more reliable interpretation, we will discuss these results in more detail. Results in Table 31 show that the majority of isolates were removed from the study using ID bootstrap analysis. However, of the remaining isolates, the majority have cow and pets as their source. Our data however, points to the fact that there is some input into this watershed from Deer (19% of isolates), but very few humans. As we previously discussed, we are unsure why pets contribute so many isolates in this watershed, but we and others have noted similarities between human and pet bacteria in the past. It may also however be due to the fact that we have a limited number of these isolates to analyze and thus their contribution to bacterial load in the watershed may be overestimated. Nevertheless, results in Table 32 show that 45, 22, and 19% of the isolates come from pets, domesticated animals, and wild animals (deer), respectively. This result is also reflected in Table 33 showing a roughly equal contribution of household and non-household bacteria to this watershed. However, as was presented for the High-Island Creek watershed, the Vermillion River is impacted mostly by non-human sources. Thus, on a larger scale, results in Tables 34 and Figure 26 show that over 93% of the *E. coli*

isolates in the Vermillion River watershed come from non-human sources. This will have surely guide remediation efforts.

SUMMARY OF MINNEOPA CREEK ANALYSIS - Analyses done using Pearson's cosine coefficient analysis with 1% optimization.

We analyzed 1,762 *E. coli* isolates obtained from the Minneopa Creek Watershed. Sampling was done on 4/03/2001, 7/30/2001, 8/15/2001, 9/12/2001, 10/10/2001, 4/18/2002, 5/15/2002, 6/12/2002, and 7/17/2002. Approximately 25 *E.coli* isolates were obtained from each site on each date. The sites sampled were labeled 1-10. Site 1.5 was sampled once only (this was not an official project site).

Table 35. Percent Identity of *E. coli* isolates from Separate Animals, by Date, obtained from the Minneaopa Creek Watershed.

Animal Type	Date								
	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
	Percent Isolates								
Cat	8.5	7.2	10.0	12.7	10.2	5.1	4.6	5.8	2.4
Chicken	4.2	0.6	2.4	2.7	10.2	5.7	6.3	4.2	0.6
Cow	2.8	10.2	5.9	10.7	3.7	5.1	13.1	7.4	28.3
Deer	53.5	30.7	34.1	30.7	22.5	26.8	21.7	31.1	18.7
Dog	2.8	6.6	3.5	4.0	6.4	8.9	13.7	3.7	4.8
Duck	0.0	2.4	1.8	2.7	10.7	1.3	1.7	4.7	2.4
Goat	0.0	2.4	0.6	4.7	0.0	1.9	0.0	0.0	4.8
Goose	1.4	14.5	15.9	7.3	10.2	20.4	12.0	10.0	9.6
Horse	0.0	0.0	0.6	2.0	0.0	0.0	0.0	2.6	0.6
Human	7.0	3.0	8.2	4.7	8.0	7.0	11.4	4.7	3.0
Pig	4.2	4.2	9.4	13.3	67.0	15.3	6.9	15.8	12.7
Sheep	5.6	12.1	2.4	0.7	0.5	0.6	8.0	3.7	7.8
Turkey	9.9	6.0	5.3	4.0	10.7	1.9	0.6	6.3	4.2

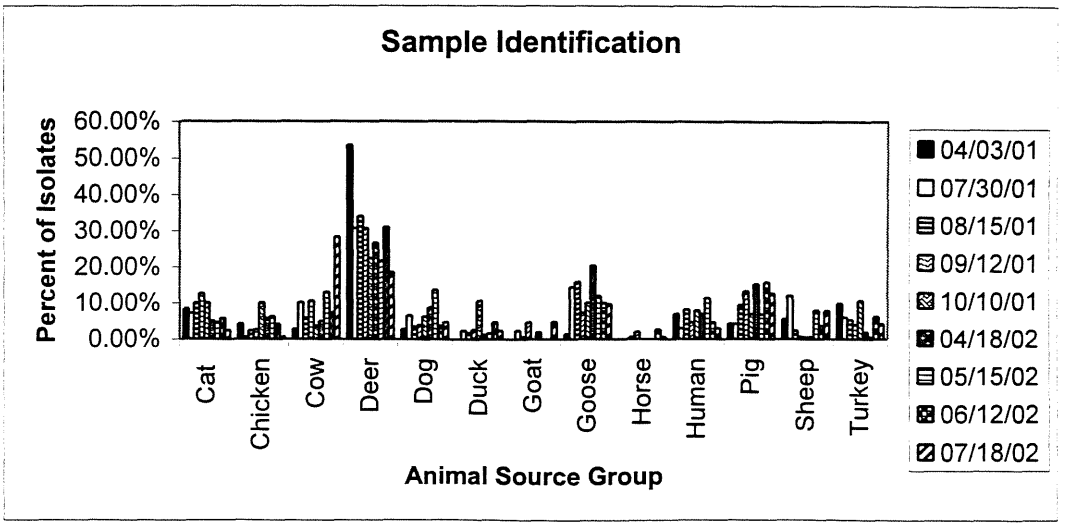


Figure 27. Percent of isolates in All Animal source groups in the Minneopa Creek Watershed, by Date.

Table 36. Percent Identity of *E. coli* isolates from Separate Animals, by Site, obtained from the Minneaopa Creek Watershed.

	Site										
Animal Type	**Site 1**	**Site 2**	**Site 3**	**Site 4**	**Site 5**	**Site 6**	**Site 7**	**Site 8**	**Site 9**	**Site 10**	**Site 1.5**
	Percent Isolates										
Cat	8.9	5.6	10.5	7.6	8.4	1.9	4.7	4.6	9.0	13.5	0.0
Chicken	4.8	4.9	3.0	8.5	2.4	0.0	5.4	0.0	3.0	10.3	16.7
Cow	11.3	7.6	5.3	10.4	10.8	11.5	8.1	12.9	10.5	13.5	0.0
Deer	33.3	32.6	24.1	23.6	27.1	26.1	36.9	28.0	21.1	23.8	50.0
Dog	4.2	5.6	9.0	9.4	12.6	7.6	3.4	3.8	3.0	4.8	0.0
Duck	1.8	2.1	5.3	0.9	1.2	12.1	4.7	1.5	0.8	3.2	0.0
Goat	0.6	2.1	3.0	0.0	7.2	0.0	0.7	0.8	0.8	0.0	0.0
Goose	13.7	9.7	13.5	7.6	7.2	10.2	13.4	18.9	16.5	9.5	0.0
Horse	0.6	0.7	0.8	0.0	1.2	0.0	2.0	0.8	0.8	0.0	0.0
Human	2.4	11.8	6.8	12.3	4.2	8.9	4.7	3.8	9.0	0.8	11.1
Pig	6.0	6.9	7.5	7.6	10.2	14.7	6.7	14.4	18.1	11.1	5.6
Sheep	4.8	5.6	4.5	4.7	4.2	3.2	4.0	6.8	1.5	4.8	16.7
Turkey	7.7	4.9	6.8	7.6	3.0	3.8	5.4	3.8	6.0	4.8	0.0

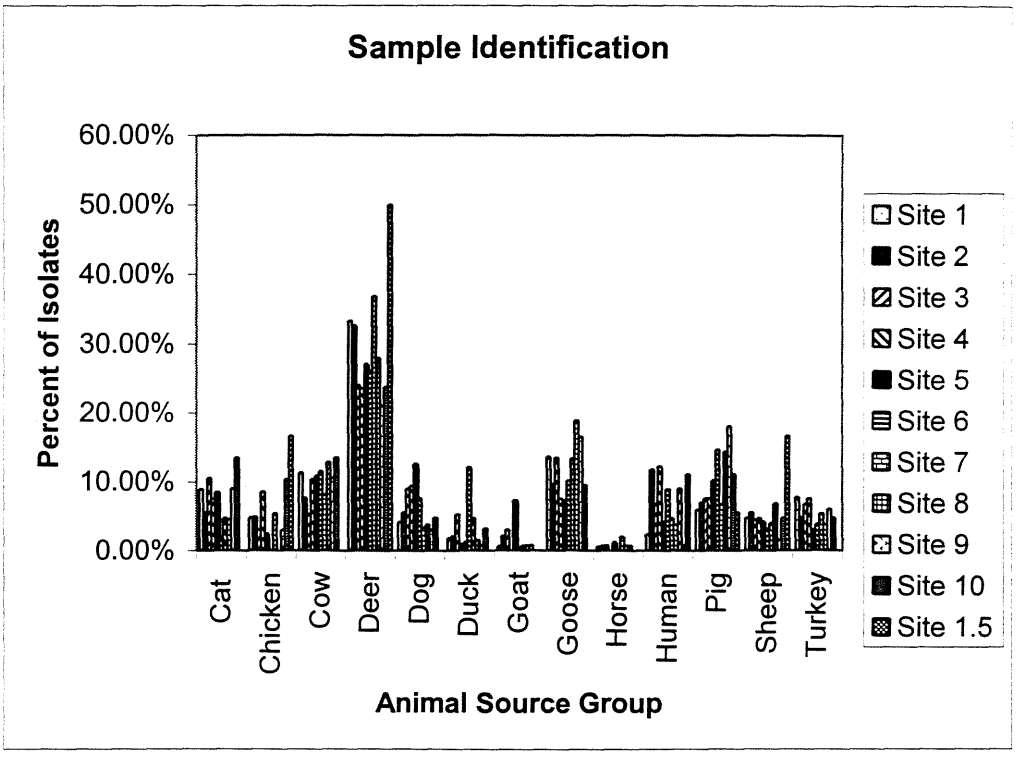


Figure 28. Percent of isolates in All Animal source groups in the Minneopa Creek Watershed, by Site.

Table 37. Percent Identity of *E. coli* isolates from Domesticated, Human, Pets, Waterfowl, and Wild-Animal Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Animal Type	Date								
	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
	Percent of Isolates								
Domesticated	26.8	35.5	26.5	41.1	32.1	30.6	34.9	40.0	59.0
Human	7.0	3.0	8.2	4.4	8.0	7.0	11.4	4.7	3.0
Pets	11.3	13.9	13.5	15.8	16.6	14.0	18.3	9.5	7.2
Waterfowl	1.4	16.9	17.7	9.5	20.9	21.7	13.7	14.7	12.1
Wild Animal (Deer)	53.5	30.7	34.1	29.1	22.5	26.8	21.7	31.1	18.7

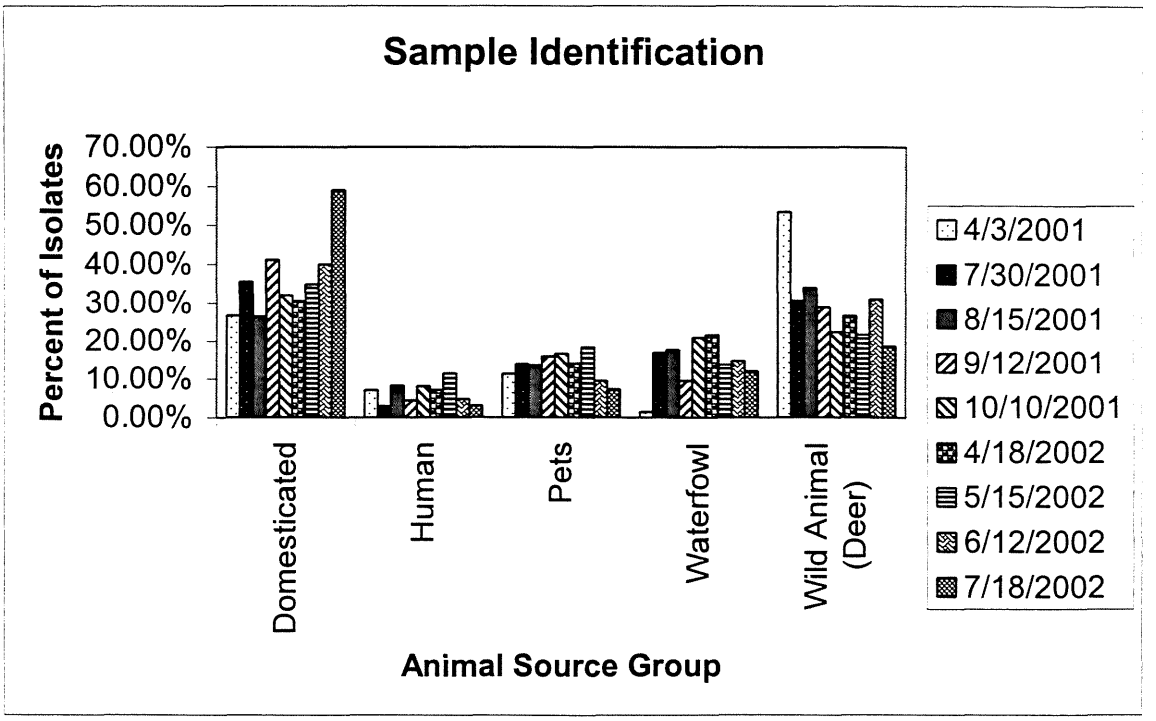


Figure 29. Percent of *E. coli* isolates in Domesticated, Human, Pets, Waterfowl, and Wild-Animal Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Table 38. Percent Identity of *E. coli* isolates in Domesticated, Human, Pets, Waterfowl, and Wild-Animal Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

Animal Type	Site										
	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 1.5
	Percent Isolates										
Domesticated	35.7	32.6	30.8	38.7	39.2	33.1	32.2	39.4	40.6	44.4	38.9
Human	2.4	11.8	6.8	12.3	4.2	8.9	4.7	3.8	9.0	0.8	11.1
Pets	13.1	11.1	19.6	17.0	21.1	9.6	8.1	8.3	12.0	18.3	0.0
Waterfowl	15.5	11.8	18.8	8.5	8.4	22.3	18.1	20.5	17.3	12.7	0.0
Wild Animal (Deer)	33.3	32.6	24.1	23.6	27.1	26.1	36.9	28.0	21.1	23.8	50.0

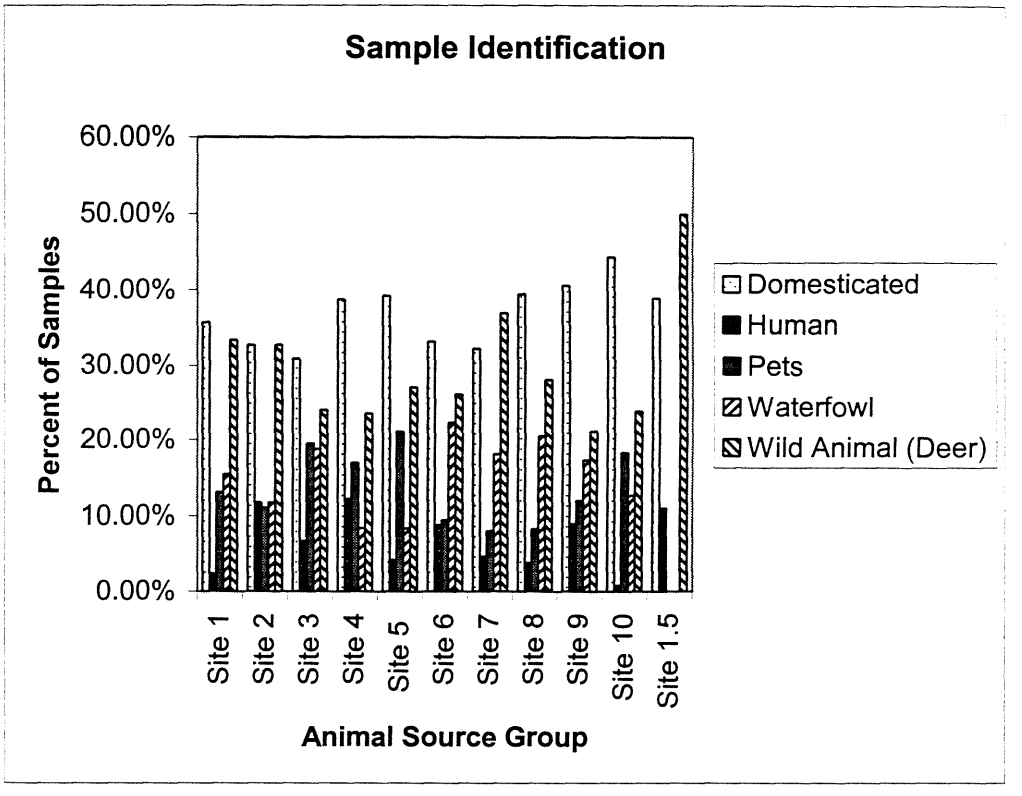


Figure 30. Percent of samples in Domesticated, Human, Pets, Waterfowl, and Wild-Animal Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

Table 39. Percent Identity of *E. coli* isolates in Household and Non-Household Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Animal Type	Date								
	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
	Percent Isolates								
Household	18.3	16.9	21.8	21.3	24.6	21.0	29.7	14.2	10.2
Non-Household	81.7	83.1	78.2	78.7	75.4	79.0	70.3	86.0	90.0

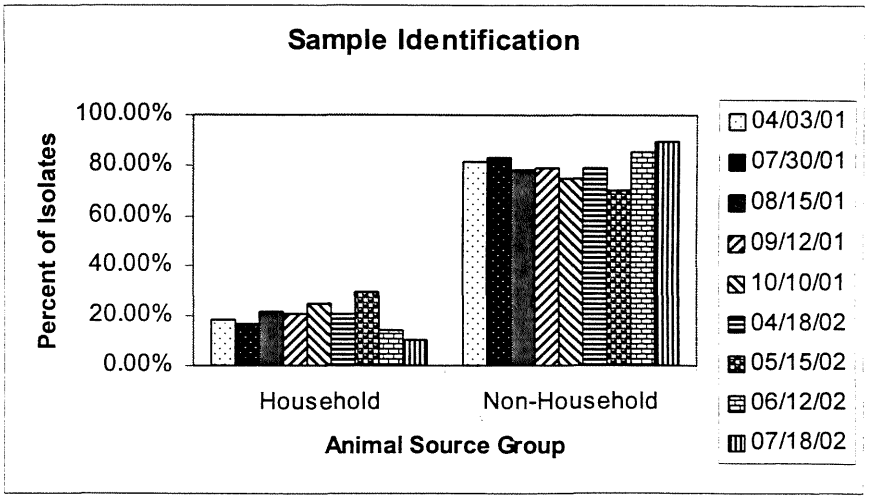


Figure 31. Percent of *E. coli* isolates in Household and Non-Household Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Table 40. Percent Identity of *E. coli* isolates in Household and Non-Household Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

Animal Type	Site										
	Site 1	**Site 2**	**Site 3**	**Site 4**	**Site 5**	**Site 6**	**Site 7**	**Site 8**	**Site 9**	**Site 10**	**Site 1.5**
	Percent Isolates										
Household	15.5	22.9	26.3	29.3	25.3	18.5	12.8	12.1	21.1	19.1	11.1
Non-Household	84.5	77.1	73.7	70.8	74.7	81.5	87.3	87.9	79.0	81.0	89.0

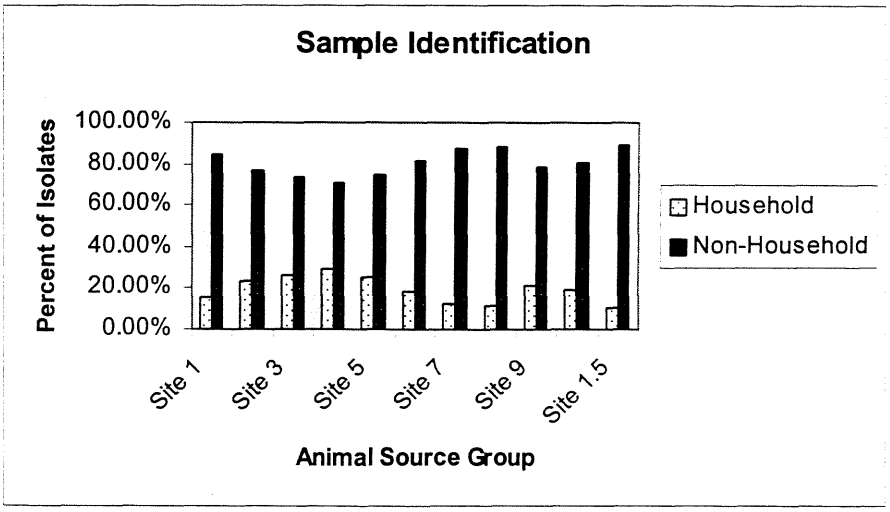


Figure 32. Percent of *E. coli* isolates in Household and Non-Household Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

Table 41. Percent Identity of *E. coli* isolates in Human and Non-Human Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Animal Type	Date								
	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
	Percent Isolates								
Human	7.0	3.0	8.2	4.7	8.0	7.0	11.4	4.7	3.0
Non-Human	93.0	97.0	92.0	95.3	92.0	93.0	88.6	95.3	97.0

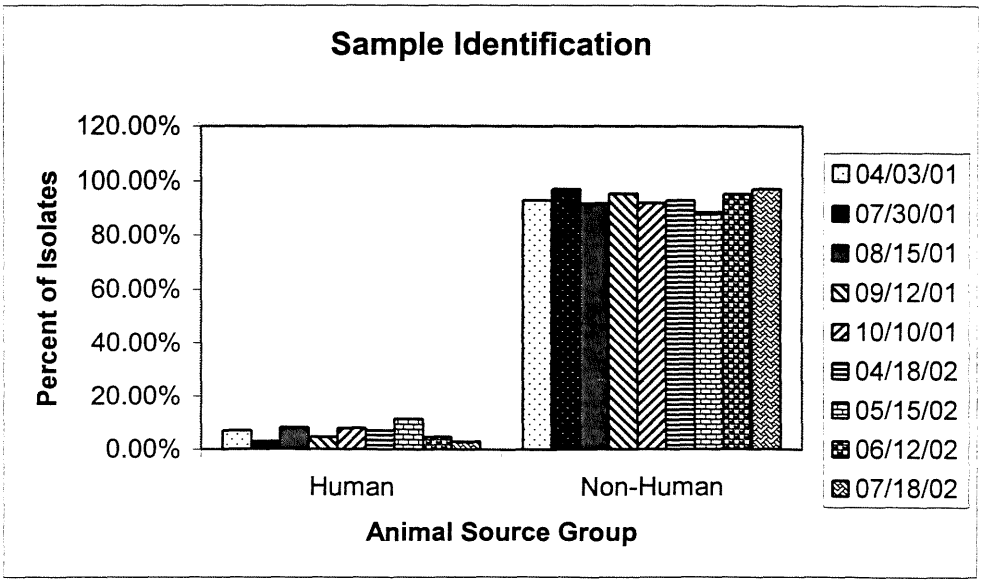


Figure 33. Percent of *E. coli* isolates in Human and Non-Human Source Groups, by Date, obtained from the Minneaopa Creek Watershed.

Table 42. Percent Identity of *E. coli* isolates in Human and Non-Human Source Groups, by Site, obtained from the Minneaopa Creek Watershed

	Site										
Animal Type	**Site 1**	**Site 2**	**Site 3**	**Site 4**	**Site 5**	**Site 6**	**Site 7**	**Site 8**	**Site 9**	**Site 10**	**Site 1.5**
	Percent Isolates										
Human	2.4	11.8	6.8	12.3	4.2	8.9	4.7	3.8	9.0	0.8	11.1
Non-Human	97.6	88.2	93.2	87.7	95.8	91.1	95.3	96.2	91.0	99.2	88.9

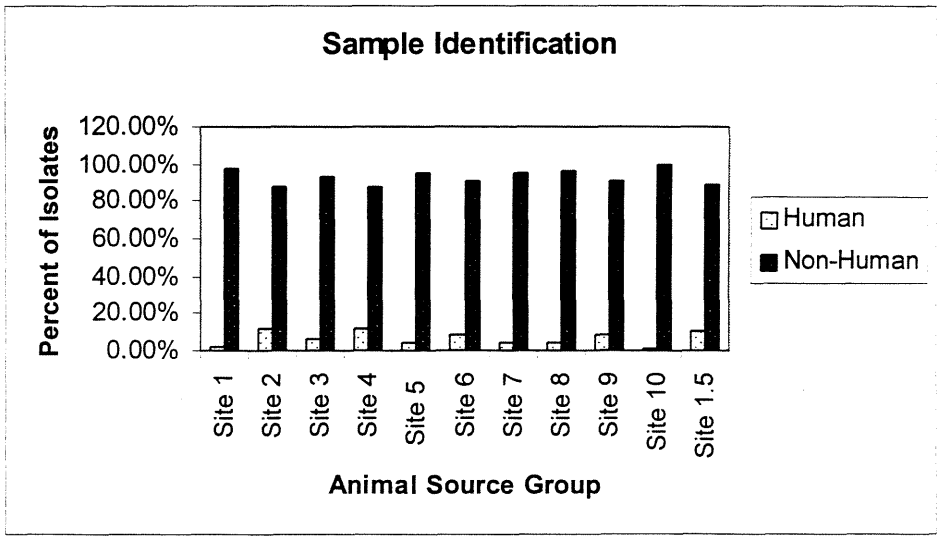


Figure 34. Percent of *E. coli* isolates in Human and Non-Human Source Groups, by Site, obtained from the Minneaopa Creek Watershed.

SEE ATTACHMENT C FOR ADDITIONAL RAW DATA BY SITE AND DATE

BOOTSTRAP ANALYSIS OF MINNEOPA CREEK WATERSHED

Table 43. Bootstrap analysis of isolates from the Minneopa Creek Watershed, percentage and frequency by animals.

Animal Type	Frequency	Percentage
Cat	36	19.4
Chicken	4	2.2
Cow	7	3.8
Deer	40	21.5
Dog	31	16.7
Duck	1	0.5
Goat	14	7.5
Goose	17	9.1
Horse	1	0.5
Human	18	9.7
Pig	2	1.1
Sheep	11	5.9
Turkey	4	2.2
TOTAL	186	

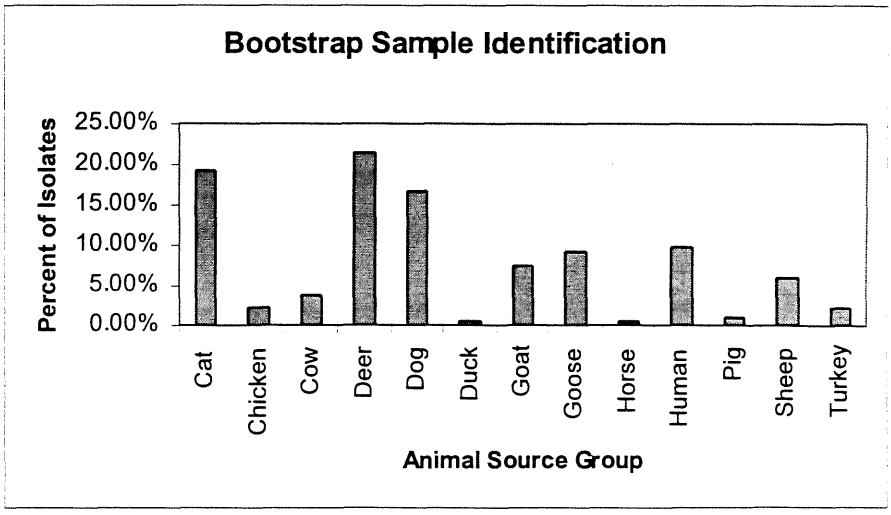


Figure 35. Percent of isolates in all Animal Source Groups recovered in the Minneopa Creek Watershed by ID Bootstrap Analysis.

Table 44. Probable Identity of *E. coli* isolates from Domestic, Human, Pets, Waterfowl, and Wild Animal Groups in the Minneopa Creek Watershed using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Domestic	43	23.1
Human	18	9.7
Pets	67	36.0
Waterfowl	18	9.7
Wild Animal (Deer)	40	21.5
TOTAL	186	

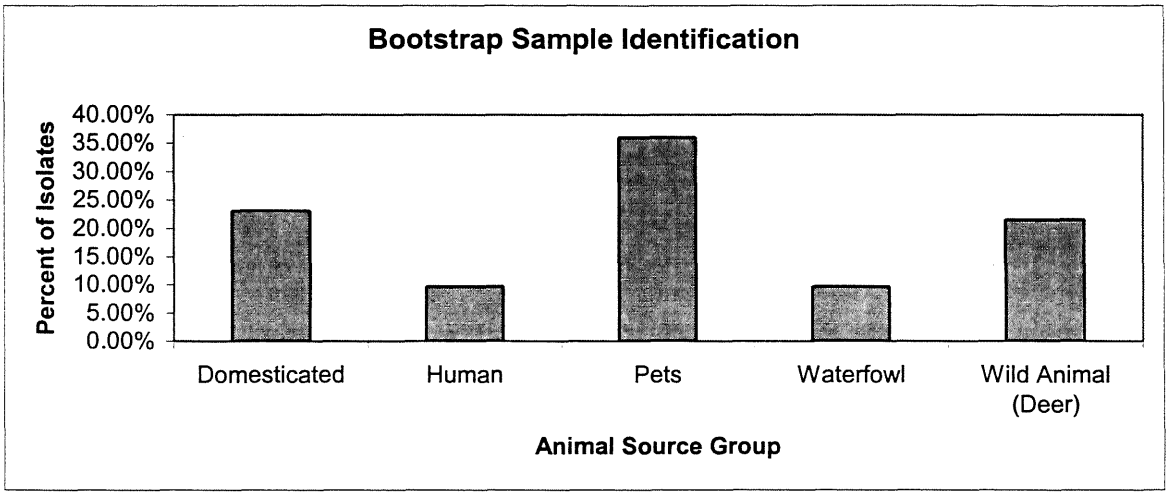


Figure 36. Percent of isolates in domesticated, human, pet, waterfowl, and wild animal source groups in the Minneopa Creek Watershed using ID Bootstrap Analysis

Table 45. Frequency and percentage of *E. coli* isolates from Household and Non-Household Groups in the Minneopa Creek Watershed using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Household	85	45.7
Non-Household	101	54.3
TOTAL	186	

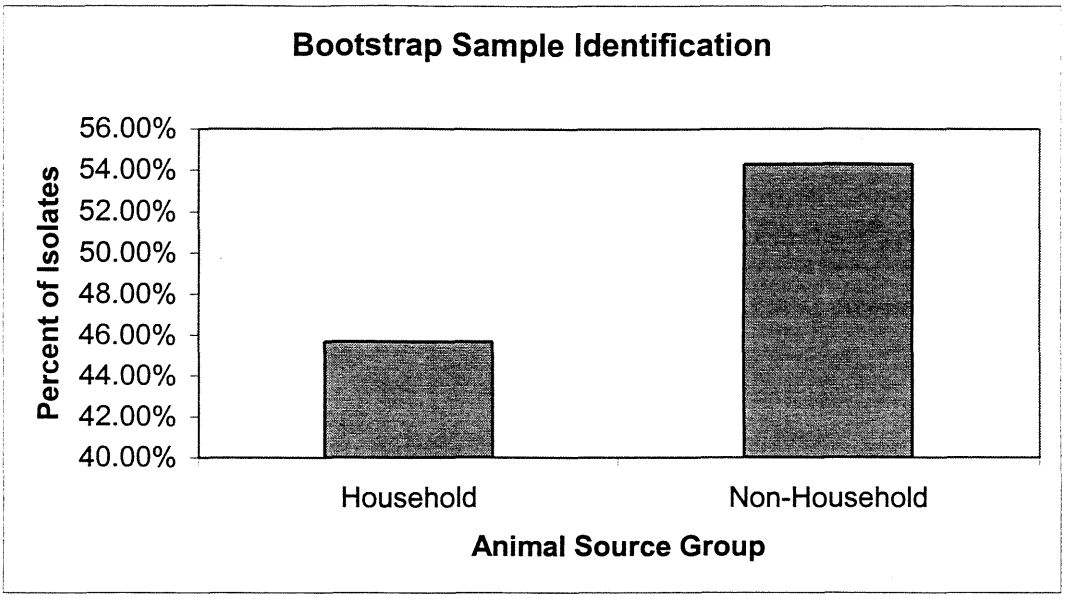


Figure 37. Percent of isolates in Household and Non-Household source groups in the Minneopa Creek Watershed using ID Bootstrap Analysis.

Table 46. Frequency and percentage of *E. coli* isolates from Human and Non-Human source groups in the Minneopa Creek Watershed using ID Bootstrap Analysis.

Animal Type	Frequency	Percentage
Human	18	9.7
Non-Human	168	90.3
TOTAL	186	

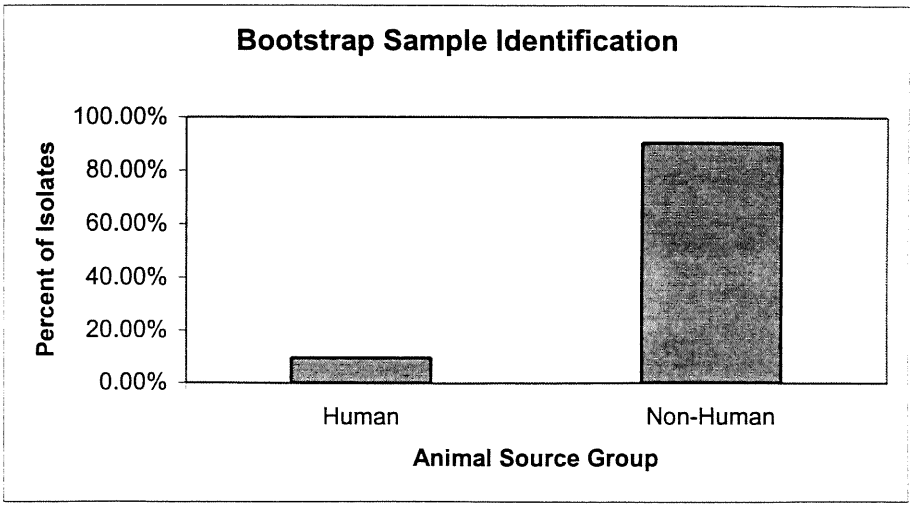


Figure 38. Percent of isolates in Human and Non-Human source groups in the Minneopa Creek Watershed using ID Bootstrap Analysis

SEE ATTACHMENT C FOR ADDITIONAL RAW DATA BY SITE AND DATE

DISCUSSION OF MINNEOPA CREEK WATERSHED ANALYSES

Results presented in Tables 35-42 and Figures 27-34 report on our analyses of the identities of watershed isolates obtained from the Minneopa Creek Watershed. In total 1,762 *E. coli* isolates were analyzed. As we indicated above, the most stringent analysis for the potential identification of these isolates comes from ID bootstrap studies (Tables 44-46 and Figures 35-38). Thus while the trends for Pearson's analyses Tables 35-42 are similar, we believe that table data from ID bootstrap analyses more accurately reflect the identity of isolates to a 90% certainty. As before, there was a distinct clustering of data by sample date and site, suggesting that climatic and land use factors affected the origin of isolates present. However, generally there was a dominance of deer and perhaps geese *E. coli* in the watershed across all sites and dates (Tables 35 and 36, Figures 27 and 28). However, generally speaking, the sites were dominated by bacteria from domesticated, non-household, non-human animals, regardless of the date or site. Since ID bootstrap gives more reliable interpretation, we will discuss these results in more detail. Results in Table 43 show that the majority of isolates were removed from the study using ID bootstrap analysis. However, of the remaining isolates, the majority have E. coli from deer and pets as their source. Our data however, points to the fact that there is some lower input into this watershed from human and geese (about 9% of isolates), but very few from other animals. As we previously discussed, we are unsure why pets contribute so many isolates in this watershed, but we, and others, have noted similarities between human and pet bacteria in the past. It may also however be due to the fact that we have a limited number of these isolates to analyze and thus their contribution to bacterial load in the watershed may be overestimated. Nevertheless, results in Table 44 and Figure 36 show that 36, 23, and 21% of the isolates come from pets, domesticated animals, and wild animals (deer), respectively. This is similar to what was seen in the Vermillion Rover watershed. Our overall results are also reflected in Table 45 showing a roughly equal contribution of household and non-household bacteria to this watershed. However,

as was presented for the High-Island Creek and Vermillion River Watersheds, the Minneopa Creek watershed is impacted mostly by non-human and non-household sources. Results in Tables 45 and 46 and Figure 37 and 38 show that over 90% of the *E. coli* isolates in the Minneopa Creek Watershed come from non-human sources.

Result 3: Dissemination and Implementation of Results.

Results from this project have been disseminated in reports made to the LCMR, in periodic update reports made to cooperators, in seminars given throughout the state, nationally and internationally, and in scientific publications in peer-reviewed journals. In addition, results from our studies will be posted and updated on the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) which is housed on computers at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our previous LCMR projects. Data obtained from our studies will be utilized by cooperating agencies to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas.

Our 1999 LCMR project generated a great deal of interest among water resources management and pollution control professionals, farm organizations, scientific researchers, and citizen groups throughout Minnesota and elsewhere. As a result we were invited to present information and findings of our research project at several State and local government-sponsored conferences and at national meetings. In all cases, we were able to accommodate the requests for presentations. The 2001 LCMR project generated a similar level of interest, and we were invited to give several presentations of our research findings.

	LCMR Budget:	**$30,247**
	Balance:	**$1,851**
Personnel:	~~$27,547~~ $28,288	(Sr. Scientist /Assistant Scientist [20%])
Software:	~~$1,000~~ $495	
Publication Costs:	$1,000	
Local Travel:	~~$700~~ $464	(2 trips @ $200 mileage, food & lodging; 3 trips @ $100 mileage)

Total: $30,247

Completion Date: December 31, 2003

Dissemination activities include:

- 8/16/01 - Newspaper article in Farmington Independent, entitled "Source of Pollution in Vermillion River Sought by University".

- 9/5/01 - "Environmental Journal" television series segment recorded, with subsequent airing of "Bacteria Busters" on Cable TV.

- 9/13/01 - Presentation to Metropolitan Council Environmental Services Environmental Planning and Evaluation managers, St. Paul, MN.

- 9/27/01 - Presentation to Sibley County Commissioners, Soil and Water Conservation District staff, and Environmental Services staff, Gaylord, MN.

- 12/18/01 - Presentation to Rice County Extension Service staff, DNR staff, Rice County SWCD staff, and Rice County Commissioners and township officials, Faribault, MN.

- Maintain the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) which is housed on computers at the University of Minnesota, Department of Soil, Water, and Climate. The Website specific for this project was developed as part of our previous LCMR project and will be updated through this project period.

- 2/5-7/02 – Invited speaker and participant at US EPA and Southern California Coastal Water Research Project (SCCWRP)-sponsored workshop: "Microbiological Source Tracking Workshop," Irvine, CA

- 4/17/02 – Presentation at Minnesota Water 2002 Conference, St. Cloud, MN

- 5/3/02 – Presentation at Minnesota Environmental Health Association Annual Spring Conference, Nisswa, MN

- 5/8/02 – Presentation to University of Minnesota Extension Natural Resources Planning Group, St. Paul, MN

- 5/21/02 – Invited speaker at American Society for Microbiology General Meeting symposium entitled "Development and Application of Methods to Identify Sources of Fecal Pollution in Water," Salt Lake City, UT

- 5/22/02 - Convened Colloquium at American Society for Microbiology General Meeting entitled "Tracking Sources and Sinks of Microorganisms", Salt Lake City, UT

- 6/20/02 – Presentation at American Farm Bureau Federation Watershed Heroes Conference, St. Peter, MN

- 7/267/02 – MPH thesis seminar and defense, University of Minnesota School of Public Health, Minneapolis, MN

- 10/14/02 – Presentation at Minnesota Department of Agriculture Water Quality Seminar, St. Paul, MN

- 11/16/02 – Invited speaker at the Water Environment Federation 2002 National TMDL Science and Policy Conference, Phoenix, AZ

- 2/19-2/21/03 – Invited speaker and participant at US EPA and Southern California Coastal Water Research Project (SCCWRP)-sponsored workshop: "Microbiological Source Tracking Workshop," Irvine, CA

- 6/19-6/20/03 - Invited participant in Health Canada workshop on Microbial Source Tracking, Ottawa, Ontario, Canada

- 8/7/03 – Presentation and planning meeting with Vermillion River collaborator, St. Paul, MN.

- 8/8/03 – Presentation and planning meeting with High Island collaborators and MNPCA, St. Paul, MN.

- 9/19/03- Presentation and meeting with Vermillion River collaborator, St. Paul, MN.

Formal Review of Project

The methods, approach, and overall impact of this project have been formally peer reviewed as part of documentation provided to Sea Grant for two research proposals and to U.S. EPA for a research proposal/contract. These projects are:

Title: Identifying Sources of Fecal Coliform Bacteria in Coastal Ecosystems and Their Relationship to Land Use", Sea Grant, 01/01/01 - 01/31/04.

Title: Sources and Impacts of "Naturalized" Escherichia coli in Coastal Environments", Sea Grant, 02/01/03 - 01/31/05.

Title: "Comparative Evaluation of Microbiological Source Tracking Techniques: rep-per DNA Fingerprinting", U.S. EPA, 03/03/03 - 03/02/04.

V. TOTAL PROJECT BUDGET:

All Results: Personnel:	~~**$228,600**~~ **$229,910**
All Results: Supplies:	~~**$30,100**~~ **$29,531**
All Results: Computer & Software:	~~**$8,300**~~ **$7,795**
All Results: Publication Costs	**$1,000**
All Results: Contracted Services:	**$4,800**
All Results: Local Travel:	~~**$2,200**~~ **$1,964**
TOTAL BUDGET:	**$275,000**

Budget Details: See Attachment A

1. A requested funds shifts between results was approved on July 31, 2003. Those changes included the following:

- Transfer of $17,000 from personnel in result 1 to personnel in result 2.
- Transfer of $500 office supplies, $600 mileage, and $2,000 travel expenses, from result 1, to personnel in result 2.
- Transfer of $800 from lab supplies in result 1 to lab supplies in result 2.
- Transfer of $4,747 from personnel in result 2 to personnel in result 3.

2. **On January 12, 2004 we requested permission to shift funds within results to pay for severance and vacation buy-out of personnel. An explanation of this request for shifting funds in indicated below:**

As the majority of work on the current project was terminated on December 31, 2003, I ended the employment of two employees, Mary Brown and Ethan Carruthers. As mandated by University policy, I was required to pay severance benefits to Mary Brown ($2,307, $1,384 from Result #2 and $923 from Result 3) and vacation buy-out to Ethan Carruthers ($550 from Result 2). Accordingly, since the personnel categories did not have sufficient funds to cover these cost, we requested that funds were transferred, within a result, from the following categories to cover these required salary payments:

$568 from Laboratory Supplies from Result 2 to personnel Result 2
$236 from Local mileage from Result 3 to personnel Result 3
$505 from Software Result 3 software to personnel Result 3
$270 from Salary Student Worker Result 2 to Salary Junior Scientist Result 2

Additional End of Project Costs

Before the end of the project period, June 30, 2004, I anticipate the following additional charges to the budget: $500 for professional/technical costs associated with the final web site update, and approximately $1,000 in publication costs.

Anticipated Final Project Balance

The current project balance is estimated as of 12/31/03 to be $3,419. After taking Into consideration the additional end of project costs listed above, the final balance will be approximately $1,919. This will be returned to the LCMR.

VI. PAST, PRESENT AND FUTURE SPENDING:

A. Past Spending:

The LCMR funded project W13 in the 1999-2001 biennium (for $300,000) to develop DNA fingerprinting tools for tracking human and animal sources of fecal pollution. The proposed project will leverage the results and resources of that project by utilizing a library of DNA fingerprints generated from known human and animal sources for the identification of unknown environmental isolates. Laboratory equipment purchased under the current LCMR-sponsored project also will be fully utilized for the proposed project.

In March 2000 the Metropolitan Council Environmental Services provided us with about $38,000 in funds and $22,000 in equipment and supplies, in lieu of initially promised in-kind services, to support our LCMR–funded project. These funds were used to provide personnel to aid in DNA fingerprinting efforts and to create antibiotic and metabolic profiles of the isolated *E. coli* bacteria.

B. Current and Future Spending:

We entered into an agreement with Bacterial Bar Codes, Inc. (Houston, Texas) to sell them up to 1500 *E. coli* bacteria that were isolated during 1999 W13 LCMR-sponsored project. The sale will generate up to $15,000 in program income. These funds were deposited in an auditable account managed by the Sponsored Projects Administration at the University of Minnesota. As per our discussions with LCMR staff, these funds were only used to offset projected increases in fringe benefit rates for project personnel and to conduct additional biochemical testing of atypical *E. coli* bacteria. A total of 1010 isolates were sold to Bacterial Barcodes, Inc., which generated $10,100 in programmatic income. As approved by the LCMR, these funds were used to cover fringe benefit shortfalls (due to increases and the UM) and salary for personnel.

C. Project Partners: Salary costs for all project partners is at no cost to the project

High Island Creek

Lauren Klement – Water Plan Coordinator, Sibley County Soil and Water Conservation District, Gaylord, MN

Scott Matteson – Project Coordinator High Island Watershed Assessment
Project, Sibley County Soil and Water Conservation District, Gaylord, MN

Minneopa Creek

Dr. Beth Proctor – Professor and Associate Director, Minnesota State University-Mankato Water Resources Center, Mankato, MN

Julie Conrad – Water Plan Coordinator, Blue Earth County Environmental Services Department, Mankato, MN

Vermillion River

Laura Jester – Watershed Conservationist, Dakota County Soil and Water Conservation District, Farmington, MN

D. Time: July 1, 2001 – December 31, 2003

VII. DISSEMINATION: Results from this project were disseminated in reports made to the LCMR, in periodic update reports made to cooperators, in scientific publications in peer-reviewed journals, and in scientific presentations. In addition, results from our studies will be posted on the *E. coli* rep-PCR web page (see http://www.ecolirep. umn.edu/) at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our 1999 LCMR project. Data obtained from our studies will be utilized by the cooperating agencies to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas. We anticipate that we will continue to receive several invitations to present our research results at local and regional conferences and meetings.

VIII. LOCATION: The project was conducted in the following areas (see maps in Attachment B):

Minneopa Creek Watershed in Blue Earth County
High Island Creek Watershed in Sibley County
Vermillion River Watershed in Dakota County

IX. REPORTING REQUIREMENTS: Periodic workprogram progress reports will be submitted not later than December 31, 2001, June 30, 2002, December 31, 2002, and July 25, 2003. A final workprogram report and associated products will be submitted by December 31, 2003. We have requested, and received a no-cost extension for this project, and have specifed a December 31, 2003 completion date.

X. RESEARCH PROJECTS: See Attachment B

ATTACHMENT B: Research Format

I. Abstract

Many of Minnesota's rivers and streams do not achieve the Clean Water Act "swimmable" goal due to elevated numbers of fecal coliform bacteria. Sources of fecal coliform bacteria include runoff from feedlots and manure-amended agricultural land, wildlife, inadequate septic systems, urban runoff, and sewage discharges. In this project we propose to define sources of fecal pollution in waters with excessive levels of fecal coliform bacteria. To achieve our goals, we will use a DNA fingerprinting technique to differentiate between strains of fecal coliform bacteria originating from animal and human sources. In our research studies we will use the polymerase chain reaction (PCR) technique, coupled with the use of a specific nucleic acid primer, BOXA1R, to characterize *E. coli* strains in three Minnesota watersheds. The DNA fingerprints that are generated will be compared to a library of DNA fingerprints from known human and animals sources. The library was created with funding from the Legislative Commission on Minnesota Resources during the 1999-2001 biennium. Fingerprints from the watershed isolates will be compared to those in the DNA fingerprint library using a sophisticated pattern recognition and statistical analysis software, BioNumerics. Our previous studies have demonstrated that the rep-PCR technique has the necessary sensitivity, specificity, and resolving power to differentiate between strains of fecal coliform bacteria originating from different human and animal sources. Results of these studies will be used to identify the sources of fecal pollution and target appropriate water pollution abatement efforts.

II. Background and Hypothesis

Currently, many of Minnesota's rivers and streams do not meet the water quality standard for fecal coliform bacteria. Sources of fecal coliform bacteria include runoff from feedlots and manure-amended agricultural land, wildlife, inadequate septic systems, urban runoff, and sewage discharges. High levels of fecal bacteria in Minnesota's rivers, lakes, streams, and aquifers threaten the use of these water resources for swimming and drinking. The State's water quality standard for fecal coliform bacteria is 200 microorganisms per 100 milliliters of water (as a monthly average). This number is used as an indicator of the possible presence of human pathogenic microorganisms. According to the 1996 report to the U.S. Congress on the condition of Minnesota's rivers, lakes and streams (as reported by the Minnesota Pollution Control Agency), 47% of the river miles assessed could not support swimming due to high levels of fecal bacteria.

In our 1999 LCMR-supported project, we generated a database of DNA fingerprints from a large number (greater than 2400) of fecal coliform bacteria. These DNA fingerprints have proven useful in differentiating between fecal coliform bacteria of animal and human origin. **In this project we will leverage the existing database and**

resources to determine and track sources of fecal contamination in three watershed areas. We will systematically sample multiple locations in three watershed areas, during baseline and critical run-off periods, to determine likely sources and locations of fecal contamination. Many of these watersheds have been intensively characterized with respect to potential sources of fecal pollution, feedlot locations and size, water quality, and status of septic systems. In an effort to validate local, state, and federal-supported abatement efforts, we will coordinate our monitoring program with cooperators prior to and following the implementation of best management practices in the watershed areas.

A better understanding of the source of fecal contamination will be a valuable tool in efforts to minimize the deleterious environmental consequences of fecal pollution. The human health risks associated with the ingestion of water contaminated with human fecal materials is well documented. Also, there is increasing concern about possible pathogens associated with fecal material from animal sources (e.g. the *Cryptosporidium* outbreak in Milwaukee and the *E. coli* outbreak in Canada).

The ability to distinguish between human and animal sources of fecal contamination is an important assessment tool. From a public health perspective, fecal contamination originating from human sources poses a greater human health risk than that originating from animal sources. Armed with knowledge about contamination sources, agencies could respond more quickly and more directly to inform that segment of the population at the greatest risk, without unnecessarily alarming people at low or insignificant risk. From a water quality perspective the ability to narrow the source of fecal contamination among the many potential sources will facilitate more tailored and cost effective pollution abatement efforts.

Conventional microbiological methods cannot differentiate between sources of fecal pollution giving rise to elevated coliform counts. While various methods have been proposed to determine the source of water-borne fecal contamination, many problems with these procedures are yet to be satisfactorily resolved. However, recent developments in molecular biology have provided some of the answers. Modern molecular biological approaches have been used to detect and track coliform bacteria and specific microbial pathogens in water. The polymerase chain reaction (PCR) technique, coupled with the use of specific nucleic acid primers and gene probes, has been used successfully to detect *E. coli* and the enteric pathogens, *Salmonella* and *Shigella,* in water. One PCR technique, called rep-PCR DNA fingerprinting, has been used by us to identify coliform bacteria, much as DNA fingerprinting techniques have been used in paternity and forensic cases (Dombek *et al.*, 2000). Organisms yielding indistinguishable DNA banding patterns can be regarded as being identical or near-identical, and as such, define the source of the fecal contamination. The rep-PCR technique has been shown to provide the necessary sensitivity and resolving power to differentiate between strains of fecal coliform bacteria originating from different human and animal sources. Of the various genetic fingerprinting strategies, rep-PCR is a

relatively simple and cost effective technique, which can be adapted for high throughput applications.

Various genomic DNA fingerprinting methods have been used to investigate epidemiologic, taxonomic, and phylogenetic relationships among microorganisms. While initial studies used classical restriction enzyme-generated DNA fingerprints of bacteria for the epidemiological analyses of nosocomial infections (Kaper *et al.*, 1982; Kuijper *et al.*, 1987; Langenberg *et al.*, 1986; Skjoid *et al.*, 1987; Tompkins *et al.*, 1987), DNA hybridizations and the rep-PCR DNA fingerprinting techniques have also found great application in the epidemiological and taxonomic analysis of yeast (Panchal *et al.*, 1987; Scherer and Stevens, 1987), mycoplasmas (Chandler *et al.*, 1982; Ruland *et al.*, 1990), fungi (Koch *et al.*, 1991), viruses (Buchman *et al.*, 1978; Christensen *et al.*, 1987), several diverse bacterial species (Langenberg *et al.*, 1986; Ramos and Harlander, 1990), and humans (Gill *et al.*, 1987). In addition to medically important organisms, DNA fingerprinting techniques have also been used to study the taxonomic relatedness of agriculturally important microorganisms. These organisms include bacterial and fungal pathogens (Lazo *et al.*, 1987) as well as plant symbionts (Brown *et al.*, 1989; Glynn *et al.*, 1985; Judd *et al.*, 1993; Kaijalainen and Lindstrom, 1989; Mielenz *et al.*, 1979; Sadowsky *et al.*, 1987; Sadowsky *et al.*, 1996; and Schmidt *et al.*, 1986).

The rep-PCR DNA fingerprinting (Versalovic *et al.*, 1991; de Bruijn, 1992; Versalovic *et al.*, 1994) uses the polymerase chain reaction and primers based on endogenous repetitive DNA to amplify specific portions of the microbial genome which are subsequently visualized following electrophoresis and ethidium bromide staining. The resulting banding patterns produced are generally unique to particular microbes and as such, can serve as a "fingerprint" for strain identification or analysis of populations. Organisms having indistinguishable banding patterns can be regarded as being identical or near-identical and those having similar banding patterns are regarded as genetically related. DNA fingerprints have been shown to be stable over many generations of microbial growth (Ramos and Harlander, 1990; Scherer and Stevens, 1987; Sadowsky *et al.*, 1996; Schneider and de Bruijn, 1996).

Endogenous repetitive DNA sequences have found wide application for the fingerprinting of prokaryotic genomes. Bacterial genomes contain a variety of repetitive DNA sequences. These repetitive elements typically are comprised of duplicated genes, such as rRNA, tRNA, and members of the *rhs* gene family (Lin *et al.*, 1984; Sadosky *et al.*, 1989), insertion sequences and transposons (Kleckner, 1981), interspersed repetitive extragenic palindromes (REP) (Lupski and Weinstock, 1992) and other palindromic unit (PU) sequences (Gilson *et al.*, 1984, Gilson *et al.*, 1987), intergenic repeat units (IRU) (Sharples *et al.*, 1990) or enterobacterial repetitive intergenic consensus (ERIC) sequences (Hulton *et al.*, 1991), bacterial interspersed mosaic elements (BIME) (Gilson *et al.*, 1991), short tandemly repeated repetitive (STRR) sequences (Mazel *et al.*, 1990), and BOX elements (Martin *et al.*, 1992). The exact function of many of these repetitive sequences is unknown, although some have been postulated to be important for genome structure and function, Most well defined

interspersed repetitive sequences are 26 to 400 bp in size (see Lupski and Weinstock, 1992 for a review). Of these repetitive elements, BOX, ERIC, and REP have been used the most often for analyzing a variety of Gram negative bacteria. Both ERIC and REP PCR primers have been useful in identifying *E. coli* strains causing mastitis in cows (Lipman *et al.*, 1995).

Repeated DNA sequences have also been found in the genomes of *Escherichia coli, Enterobacter aerogenes,* and *Salmonella typhimurium* and have been very useful for the examination of genome structure and for the fingerprinting of bacterial DNA. The repeated palindrome (REP) or palindromic unit (PU) sequences and bacterial interspersed mosaic elements (BIMEs) have been detected in a large number of bacterial strains (Lupski and Weinstock, 1992; Gilson *et al.*, 1984, Gilson *et al.*, 1991, Sharples *et al.*, 1990; Stern *et al.*, 1984). Rep-PCR DNA fingerprinting has also been used to investigate two nosocomial outbreaks in Houston, Texas, and the technique was useful in identifying the predominant clone causing disease (Gerghiouh *et al.*, 1995). Moreover, the rep-PCR DNA fingerprinting technique has proven to be effective in clustering methicillin resistant *Staphylococcus aureus* and *Legionella pneumophilia.* For a detailed description of the medical and epidemiological uses of the rep-PCR technique see Versalovic *et al.*, 1995.

Lastly, while a majority of repeat elements have been isolated from Gram negative bacteria, there have also been reports of endogenous repeat elements in the gram positive bacteria. For example, *Streptococcus pneumoniae* contains a highly conserved DNA sequence, collectively termed BOX, which is located within intergenic regions of the chromosome. The *S. pneumoniae* genome contains about 25 BOX elements which range in size from about 40 to 60 bp (Martin *et al.*, 1992). It has been postulated that BOX sequences serve a regulatory functions. In addition, the BOX element has also been used to study the epidemiology and classification of potato tuber diseases caused by *Streptomyces* sp strains (Sadowsky *et al.*, 1996).

Recently, we showed that the BOXA1R primer was useful in differentiating among *E. coli* strains from human and animal sources (Dombek *et al.*, 2000). The BOXA1R primer will be used in our current studies.

In summary, DNA fingerprinting methods using endogenous repeat elements have been used to investigate epidemiologic, taxonomic, and phylogenetic relationships among microorganisms. Endogenous repeat sequences have been found in almost all organisms thus far examined and provide a valuable tool to identify and track medically and environmentally important microorganisms (Versalovic *et al.*, 1998). In addition, although the function(s) of most of the endogenous repeat sequences remain unknown, they nevertheless have proven invaluable in phylogenetic studies and the analysis of genome structure and function. The rep-PCR technique takes advantage of these repeat elements and in doing so, provides the necessary sensitivity to differentiate between strains of fecal coliform bacteria originating from different human and animal sources. Of all the various genetic fingerprinting strategies, rep-PCR is a relatively

simple and cost effective technique which can be adapted for high throughput applications, Coupled to the use of computer-assisted pattern analysis and database software, for example BioNumerics, a large number of rep-PCR DNA fingerprints can be cataloged, analyzed, and characterized in an expeditious and statistically relevant manner (Dombek *et al.*, 2000; and Rademaker *et al.*, 1998).

The project will examine three watershed areas determined in consultation with Soil and Water Conservation District personnel, the Minnesota River Basin Joint Powers Board, and Metropolitan Council Environmental Services staff. The selected locations represent a range of water quality challenges, are well-characterized in terms of potential fecal pollution sources, and will provide an opportunity to evaluate the effectiveness of existing and planned pollution abatement efforts. The project will provide a scientifically sound basis for prioritizing pollution control efforts so that resources can be efficiently and effectively allocated to lower fecal coliform counts and achieve water quality goals in these watersheds. Beyond the specific applications described here, the project has national importance as it adds to the database of DNA fingerprints that can be applied to future pollution tracking efforts throughout the nation.

Results of our preliminary work have been published in a peer-reviewed journal:

Dombek, P.E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the polymerase chain reaction to differentiate *Escherichia coli* from human and animal sources. Appl. Environ. Microbiol.66:2572-2577.

Hypothesis: Our studies are based on the hypothesis that the rep-PCR technique using the BOXA1R primer has the necessary resolving power and robustness to differentiate between environmental *E. coli* ecotypes in natural watershed areas, and that the database of E. coli fingerprints that we previously generated can be used to identify sources of unknown *E. coli* strains isolated from watershed areas.

III. Methodology

All of the methodologies described below have been developed and refined during our 1999-2001 LCMR project and were found to provide the necessary specificity and resolving power to assign environmental *E. coli* isolates to their correct source groups.

A. Sampling sites

Three watershed areas will be intensively monitored for sources of fecal pollutants during the two-year period. Fecal coliform bacteria will be isolated from water samples collected in each of the 3 watershed areas in the spring, summer, and fall months, during baseline and critical run-off periods.

The watersheds to be sampled are: [1] Minneopa Creek, near Mankato (Blue Earth County), [2] High Island Creek, near Henderson (Sibley County), and [3] Vermillion River, near Farmington (Dakota County).

Water samples will be collected in sterile Whirl-Pac bags and kept on ice until processed. Ten sites will be sampled in each watershed. Sites were chosen with the input of project cooperators.

Sampling site descriptions are outlined below, and sampling site locations are indicated on the enclosed maps. Approximately 80 water samples will be collected from each watershed (10 sites sampled per watershed on 8 separate sampling occasions for a total of 240 samples for the three watersheds). To the extent possible, samples will be collected to capture spring flush after snow melt, as well as during baseline flow and after storm events.

Watershed 1 - Minneopa Creek

The Minneopa Creek Watershed covers an area of 85.2 square miles primarily in Blue Earth County. Minneopa Creek begins in western Blue Earth County and meanders eastward before entering into the Minnesota River approximately two miles west of Mankato. The creek has been channelized from its headwaters to Lily Lake, and ditches and tiles are prevalent in this area. The land surrounding Minneopa Creek consists of generally flat agricultural land along its western portion, rolling hills near its middle section, and a wooded area (Minneopa State Park) along its confluence with the Minnesota River.

Water quality in Minneopa Creek was intensively monitored in 1995-1996 (with partial funding from the LCMR). Fecal coliform levels in the Creek ranged from 10-23,000 fecal coliform/100 ml, with the highest levels associated with storm events. Sites frequently exceeded the bacteriological water quality standard of 200 or 1000 fecal coliform per 100 ml. The Minneopa Creek Watershed Plan, published in 1997, identified reduction of fecal coliform levels as a high priority goal for the Blue Earth SWCD. The exact source(s) of the fecal pollution could not be identified during this study, but were generally attributed to manure runoff and/or sewage from unsewered homes in the unincorporated areas.

Minneopa Creek sites were selected with input from project cooperators. Several of the sites correspond to previously monitored sites showing excessive levels of fecal coliform bacteria. A one-time sampling of the Lake Crystal Wastewater Treatment Facility effluent prior to disinfection also will be included.

Site 1. Minneopa Creek at Minneopa State Park entrance - Located above the Minneopa Falls. Signs in the park warn visitors of high fecal coliform levels in the creek.

Site 2. Minneopa Creek at County Road 114 - Located in subwatershed 28044.

Site 3. Minneopa Creek at County Road 112 – Below the Lake Crystal Wastewater Treatment Facility.

Site 4. Crystal Lake outlet - Crystal-Loon-Mills Lakes Outlet below the dam. Drains two minor watersheds: 28045 and 28046.

Site 5. Lily Lake outlet at bridge on County State Aid Highway 20 - Receives direct inflow from subwatersheds 28047 and 20848, and urban runoff from the City of Lake Crystal.

Site 6. County Ditch 56 at County Road 9 - County Ditch 56 receives runoff from the City of Lake Crystal and drains minor watershed 28045. This ditch empties into Crystal Lake.

Site 7. County Ditch 27 at County Road 6 - Located just south of the junction of County Road 111 and County-State Road 6 in subwatershed 28047. This ditch system empties into Lily Lake.

Site 8. Minneopa Creek at County Road 111 - Located on County Road 111 just north of County-State Highway 6 at the end of subwatershed 28048. Subwatershed 28048 receives water from subwatersheds 28049 and 28050. This ditch system empties into Lily Lake.

Site 9. Minneopa Creek at County Road 22 - Drains subwatershed 28050

Site 10. Judicial Ditch 48 at County Road 6 - Drains subwatershed 28049.

Watershed 2 - High Island Creek

The High Island Creek Watershed is a rural watershed that drains 153,219 acres in Sibley, McLeod and Renville counties. The High Island watershed is a minor watershed of the Lower Minnesota River basin, and is one of the few remaining rural watersheds of the Lower Minnesota. The watershed begins in eastern Renville, where approximately 10% of the watershed is located. McLeod County and Sibley County have respectively 25% and 65% of the watershed located within their county boundaries.

The Minnesota Pollution Control Agency monitored the High Island Creek from May 12, 1999 to August 23, 1999 for fecal coliform bacteria and transparency (using transparency tubes) at five sites in the watershed. Results of the monitoring revealed low transparency tube readings and high fecal coliform counts at a majority of the monitored sites. Water at the mouth of High Island Creek has high levels of fecal

coliform, phosphorus, nitrates and total suspended solids based on the Minnesota River Assessment Report published in 1994.

The Sibley County Soil and Water Conservation District has received a Phase I - Clean Water Partnership Grant from the MPCA to assess pollution sources within the watershed. The High Island Creek sites selected for this LCMR project correspond to the ten monitoring sites developed for the High Island Watershed Assessment Project. These sites will be rated (rating curves) by the USGS. Of the ten sites, five have been designated as primary and five as secondary. The primary sampling sites have equipment to record stage levels that will help determine loading rates. Sites are listed below with selection justification:

I. Primary Sites (continuous stage recorders)

Site 2. High Island Creek at County Road 7
This site is downstream of the confluence of Judicial Ditch 11 and Judicial Ditch 15 and represents 33% of the study area. A rock riffle at the downstream side of the bridge forms a good control for establishing a stage-discharge relation.

Site 3. High Island Creek below Baker's Lake
This site is located two miles to the east of Bakers Lake on county road 13. This site allows assessment of the extent of sediment trapping and other water quality modifications that may be occurring as High Island Creek flows through Bakers Lake. Comparison of water quality at this site with water quality at Site 2 will provide an assessment of the amount of load arising from the low-gradient area in the mid-reaches of the study area. The drainage area at Bakers Lake outlet is 110 mi squared, representing 46% of the study area.

Site 5. High Island Creek at County Road 9 Near Arlington
This site provides a means of assessing water quality of the main stem of the High Island Creek before addition of water form the urban setting at Arlington. Drainage area at this site is 161 mi squared, representing 67% of the study area.

Site 9. Buffalo Creek near Henderson
This site is located on a township road near the mouth of Buffalo Creek. This site will be representative of conditions in the downstream part of the study area where land is steeply sloped and there are wooded ravines. The channel at this site is rocky and has swiftly flowing water. Buffalo Creek drains 28.2 mi squared or 12% of the study area.

Site 10. High Island Creek at County Road 6 Near Henderson
This USGS gauging station site is located near the mouth of the High Island Creek. The site will measure the total discharge of water and chemical constituents from the study area. This site has a continuous record streamflow gauging station that has been operated since 1973. The stage-discharge relation for this site has been established by the USGS. Drainage area is 238 mi. squared at the gauge.

2. Secondary Sites (staff gauges)

Site 1. Judicial Ditch 11 at County Road 8/10
This site represents most (about 14 square miles) of minor watershed 3301500 in Renville County. Water quality at this site is expected to be typical of conditions in the upper part of the study area.

Site 4. High Island Lake
The only lake site in the study is located near High Island Sportmens Park in New Auburn. During high flows (on average 2 to 3 time per year) High Island Creek overflows a dam located on the south end of the lake. During these occasions water flows directly into the lake.

Site 6. Ditch 2 at County Road 17.
This site represents a portion of the watershed that drains 16.4 square miles or 12% of the study area. The site was previously sampled in 1999 for fecal coliform bacteria, at which time had high levels.

Site 7A. High Island Creek at County Road 66 above waste water treatment plant near Arlington.
This site is located about ½ mile downstream of New Auburn, just before the city's wastewater treatment plant. Water quality data obtained from this site will show the urban inputs of New Auburn.

Site 7B. High Island Creek at County Road 66 near Sportmen's Club.
This site is located about 1 mile downstream of site 7B, below Arlington's wastewater treatment plant. This site will only be monitored 3 times during baseflow in 2000 to determine if any inputs from the wastewater treatment plant are entering the creek.

Site 8. Buffalo Creek at County Road 17.
This site is below the confluence of County Ditch 59 and High Island Ditch number 5, which drains the less steeply sloped and less wooded part of the Buffalo Creek watershed.

Watershed 3 – Vermillion River

The Vermillion River Watershed encompasses 372 square miles, mostly located through central Dakota County south of the Twin Cities metropolitan area. The main stem originates in Scott County to the west and flows generally northeast to the City of Hastings. Current landuse in the watershed is still dominated by agriculture with suburban areas and smaller urban growth centers interspersed throughout the watershed.

In 1998, the Vermillion River main stem, from Empire Township to the dam in Hastings, was listed on the Federal Clean Water Act's 303(d) list of impaired waters for fecal coliform bacteria. The river was not meeting its swimming use standard due to high bacteria levels. Also in 1998, the Vermillion River was placed on the Minnesota Pollution Control Agency's (MPCA) list of waters in need of a total maximum daily load (TMDL) study for fecal coliform. In 1999 the MPCA, with the help of local agencies and citizens, collected fecal coliform samples throughout the Vermillion River watershed to begin determining the extent of the bacterial problem. These data indicate that the river and its tributaries have bacteria levels in excess of the MPCA's state standard of 200 organisms/100 ml of sample.

The Dakota County Soil and Water Conservation District has undertaken a TMDL project to conduct additional monitoring in the Vermillion River watershed to further define the extent of the problem, determine possible sources of bacteria, calculate an acceptable load of bacteria for the various reaches, and draft a series of reduction goals for future implementation. This project will include additional fecal coliform monitoring throughout the watershed to better identify highly impaired reaches and possible sources of the bacteria. Landuse and landcover GIS data will also help identify possible sources of non-point source pollution. Using GIS data and lab results, a TMDL model will be produced and various loading scenarios will be drafted. Reduction goals and an implementation strategy will be the final outcome of the project.

The monitoring sites for our project were selected with input from Dakota County SWCD staff. Site locations and information are summarized below:

1. Middle Creek Sub-watershed

Site 1. VMCwest
- Just upstream of confluence with VMCeast; north of County Hwy. 64 in Farmington
- Agricultural and suburban influence
- Will help explain very high fecal levels found in Middle Creek in 2000
- Site is located just in a cropped field with very little buffer adjacent to the waterway

Site 2. VMCeast
- Just upstream of confluence with VMCwest; north of County Hwy. 64 in Farmington
- Agricultural and suburban influence
- Will help explain very high fecal levels found in Middle Creek in 2000
- Site is located just in a cropped field with very little buffer adjacent to the waterway

Site 3. VMC
- At the mouth of Middle Creek just upstream of its confluence with North Creek and the Main Stem Vermillion River
- Take sample from foot bridge on Hwy 3 just north of Farmington in Empire Township
- During very high water, North Creek may back up into Middle Creek so take sample further upstream

- Will help explain very high fecal levels found in Middle Creek in 2000
- Middle Creek sub-watershed has both suburban and agricultural landuses and is rapidly developing

2. North Creek Sub-watershed

Site 4. VNC175
- Just downstream of wide spot in the creek near trailer court on west side of Hwy 31 in Lakeville
- Mostly suburban influence with some agriculture much further upstream
- Will help explain very high fecal levels at this site in 2000
- Past complaints of foul "sewer" odor at this site

Site 5. VNC
- At the mouth of North Creek just upstream of its confluence with Middle Creek and the Main Stem Vermillion River
- Take sample from foot bridge on Hwy 3 just north of Farmington in Empire Township
- Will help explain sources of fecal pollution in the entire sub-watershed
- North Creek sub-watershed has both suburban and agricultural landuses

3. South Branch Sub-watershed

Site 6. VSBtrib
- On the east side of Hwy 79 just south of 232nd St. in Castle Rock Township
- In a slightly wooded area upstream of confluence with South Branch
- Drains agricultural and rural residential areas
- Will help explain very high levels of fecal coliform measured in 2000 at this site

Site 7. VSB
- At the mouth of South Branch at Hwy 66 in Vermillion Township
- Slightly wooded site with large trout found here in 2000
- Will help explain sources of fecal pollution from the entire sub-watershed
- South Branch sub-watershed drains agricultural and rural residential areas
- City of Hampton sewage lagoons discharge to South Branch during periods of high flow
- Many horse pastures, feedlots and sod farms are in this sub-watershed

4. Main Stem Vermillion River

Site 8. V31
- At Hwy 31 bridge in Farmington just south of Hwy 50
- Site is downstream of upper areas of watershed prior to Middle Creek, North Creek and South Branch inlets

- Upper portions of watershed have agricultural, rural residential, quickly developing suburban areas, an industrial park, and wastewater discharges for the Elko/New Market sewage lagoons

Site 9. Vverm
- At Hwy 85 just outside the City of Vermillion in Vermillion Township
- Slightly wooded site
- Downstream of all major tributary inlets
- Downstream of Empire Wastewater treatment plant and City of Vermillion wastewater treatment plant

Site 10. V47
- At Hwy 47 just inside Hastings city limits
- Site is open and channel has been straightened
- Just upstream of Vermillion Falls below which fecal levels drop below state standard

B. Isolation and confirmation of *E. coli* from watershed samples

Water samples will be analyzed for fecal coliform bacteria by an EPA-certified laboratory using the membrane filtration method (*Standard Methods for the Examination of Water and Wastewater,* 1995). Agar plates (mFC) generated from the fecal coliform analysis will be used by project staff for isolation, identification and confirmation of *E. coli* strains. Typical fecal coliform colonies will be picked from the mFC agar plates used to enumerate fecal coliforms. Blue colonies will be restreaked for purity onto the same medium. The purified presumptive coliform bacteria will be subjected to confirmatory tests using EC broth with MUG (with Durham tubes). The EC-MUG Broth will be incubated at 44.5 ± 0.2 °C. The EC medium differentiates between coliform bacteria of fecal or other origins. Presumptive *E. coli* colonies will be streaked for isolation to MacConkey agar and plated onto ChromAgar ECC to differentiate between *E. coli* and *Klebsiella*. Confirmed coliform bacteria will be indicated by the production of gas in Durham tubes, strong fluorescence when EC-MUG tubes are examined by using ultraviolet light, and production of blue colonies on ChromAgar. The identity of coliform bacteria will be verified by using standard biochemical tests as outlined in *Standard Methods*. Only isolates confirmed as being *E. coli* will be used in our studies. The confirmed coliform isolates will be cataloged, stored, and preserved at -80°C until needed.

About 1600 *E. coli* strains will be isolated per watershed area for a project total of 4800 isolates:

No. of Watershed	Sites per Watershed	Sampling Occasions	E. coli per sample	Total E. coli Isolated
3	10	8	20	4800

C. rep-PCR DNA fingerprinting of environmental *E. coli*

All confirmed isolates from known sources will be individually streaked onto Plate Count Agar and incubated overnight at 37°C. An individual colony of each isolate will be picked with a 1 μl loop and suspended in 100 μl dH_2O. Cell suspensions will be stored at -80°C until used. The cell suspensions will be subjected to rep-PCR using BOXA1R primers. A single BOXA1R primer is needed. The sequence of the BOXAIR oligonucleotide PCR primer is 5'-CTACGGCAAGGCGACGCTGACG-3'. The PCR reactions are performed in 25 μl of PCR buffer as previously described (Dombek et al., 2000; Judd et al., 1993 and Sadowsky et al., 1996) using 2 μl of cell suspensions as template. The PCR reaction conditions for the BOXA1R primer are: denaturation for 7 min at 95°C; 35 cycles of 90°C for 30 sec, 53°C for 1 min, and 56°C for 8 min; final extension at 65°C for 16 min; and a 4°C soak. PCR products will be separated by horizontal electrophoresis on 1.5% agarose gels run at 70V for 17.5 hr at 4°C with buffer recirculation, stained with ethidium bromide, and photographed. Molecular weight and an internal standard (PCR products from one of the strains) will be included on each gel to aid in normalization and comparison across gels. The resulting DNA fingerprints will be scanned into digital images, converted to TIFF files and used for subsequent analyses.

D. Data analysis

DNA fingerprints will be normalized to molecular weight and internal standards, and compared and analyzed using BioNumerics software as described by Dombek *et al.* (2000) and Rademaker *et al.* (1998). The relatedness of isolates to each other and to those in the known-source DNA fingerprint library will be determined by cluster (using the Pearson's correlation coefficients), principal component, and discriminant statistical analyses. The number and diversity of isolates falling into each source group category, obtained during different sampling times, will be analyzed for statistical differences using analysis of variance and other applicable statistical methods. Isolates will be assigned to source groups if they have > 80% average similarity value to those isolates in the fingerprint library. Sources of unknown environmental isolates will be provided to cooperating agencies and used to evaluate existing and proposed abatement programs. Results of our will be stored on a rep-PCR web site (www.ecolirep.umn.edu) at the Department of Soil, Water, and Climate, University of Minnesota. The known source DNA fingerprints generated during our previous LCMR project will be made available for downloads from this server. The BioNumerics program contains database sharing functions which allow for the exchange of fingerprint information between laboratories. We view this as an essential resource for other laboratories to use both at the state and national levels.

IV. Results and Products

Result 1: Acquire fecal coliform bacteria from watershed areas. Fecal coliform bacteria will be isolated from water samples collected in each of the 3 watershed areas in the spring, summer, and fall months (during baseline and critical run-off periods). Approximately 80 water samples will be collected from each watershed (10 samples collected per watershed on 8 separate sampling occasions). Samples will be analyzed for fecal coliforms by an EPA-certified contract laboratory. *E. coli* bacteria from fecal coliform plates will be isolated (about 1600 isolates per watershed for a total of 4800 isolates) and the identity of bacteria will be confirmed by using selective and differential microbiological media and biochemical tests. The confirmed *E. coli* isolates will be cataloged, stored, and preserved in glycerol at -80°C until DNA fingerprinting is done.

Result 2: Generate DNA fingerprints from *E. coli* isolates obtained from watersheds. DNA from confirmed *E. coli* isolates will be subjected to rep-PCR DNA fingerprinting using BOXA1R primers. The resulting **4800** DNA fingerprints will be scanned into digital images, normalized to molecular weight and internal standards, and compared and analyzed using BioNumerics pattern recognition and statistical analysis software. The animal(s) contributing to *E. coli* in watersheds will be determined by cluster and discriminant statistical analyses by comparison to our known source DNA fingerprint library.

Result 3: Dissemination and Implementation of Results. Results from this project will be disseminated in reports made to the LCMR, in periodic update reports made to cooperators, and in scientific publications in peer-reviewed journals. In addition, results from our studies will be posted on the *E. coli* rep-PCR web page (see http://www.ecolirep.umn.edu/) housed at the University of Minnesota, Department of Soil, Water, and Climate. A Website specific for this project was developed as part of our 1999 LCMR project and includes searchable and downloadable DNA fingerprints that can be used by state and local agencies to track sources of fecal coliform bacteria. Data obtained from our studies will be utilized by the cooperating agencies to prioritize pollution abatement efforts, implement best management practices, and validate existing pollution prevention efforts in the three watershed areas.

V. Timetable

Water sampling will take place during the spring, summer and fall seasons. Given that the project period begins and ends in mid-summer, sampling will take place in both years of the project period. Sampling will begin in the Summer of 2001 and extend into the Fall of 2001. Sampling will resume in the Spring of 2002 and will continue into Summer 2002. Each watershed will be sampled on a monthly basis, and will be sampled eight times in 2001-2002.

Concurrent with water sampling activities, *E. coli* bacteria will be isolated and confirmed.

DNA fingerprinting and data analysis activities will occur during the late Fall through early Spring months, when field work and isolations are not being done. Reports to the LCMR will be submitted December 31, 2001, June 30, 2002, December 31, 2002, and June 30, 2003.

	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
	2001						2002												2003					
Sampling																						?		
Isolation																						?		
Fingerprinting																						?	?	
Data Analysis																								
Dissemination						•						•						•						•

• = Progress reports to LCMR
? = Only if needed due to Spring run-off events.

VI. Budget Requirements

LCMR Recommended Funding: $275,000

	Objective/Result			
Budget Item	Result 1 Acquire Fecal Coliform Bacteria	Result 2 Generate DNA Fingerprints	Result 3 Dissemination Activities	**Row Total**
Wages, salaries & benefits				
Senior Scientist	45,600	45,600	22,800	114,000
Junior Scientist	30,400	45,600		76,000
Student worker (2@15 hr/wk)	21,000	9,000		30,000
Total: Salary & Benefits	**97,000**	**100,200**	**22,800**	**220,000**
Laboratory Supplies	20,700	13,400		34,100
Office Supplies	1,000	1,000		2,000
Equipment				
Computer		2,500		2,500
Software		4,800	1,000	5,800
Publication costs			1,000	1,000
Contracted Services (fecal coliform analysis)	4,800			4,800
Local Travel (mileage, food, lodging)	4,100		700	4,800
Column Total	**127,600**	**121,900**	**25,500**	**275,000**

VII. Principal Investigators and Cooperators

Principal Investigators:

Michael J. Sadowsky – Professor, University of Minnesota (CV attached)

LeeAnn K. Johnson – Sr. Scientist, University of Minnesota (resume attached)

Cooperators

High Island Creek
Lauren Klement – Water Plan Coordinator, Sibley County Soil and Water Conservation District
Scott Matteson – Project Coordinator High Island Watershed Assessment Project, Sibley County Soil and Water Conservation District

Minneopa Creek
Dr. Beth Proctor – Professor and Associate Director, Minnesota State University-Mankato Water Resources Center
Julie Conrad – Water Plan Coordinator, Blue Earth County Environmental Services Department

Vermillion River
Laura Jester – Dakota County Soil and Water Conservation District
Kent Johnson – Water Quality Manager, Metropolitan Council Environmental Services

CURRICULUM VITAE

Name: Michael Jay Sadowsky

Present Title: Professor

Department of Soil, Water, and Climate, Biological Process Technology Institute, Department of Microbiology, and Director of Graduate Studies Microbial Ecology and Microbial Engineering Programs

Address: Department of Soil, Water, and Climate
University of Minnesota
439 Borlaug Hall
1991 Upper Buford Circle
St. Paul, Minnesota 55108

Education:

Ph.D., 1983 University of Hawaii, Honolulu, Hawaii. Major: Microbiology
M.S., 1979 University of Wisconsin, Oshkosh, Wisconsin. Major: Biology/Microbiology
B.S, 1977 University of Wisconsin, Madison, Wisconsin. Major: Bacteriology

Professional Experience:

Professor: Department of Soil, Water, and Climate and Department of Microbiology University of Minnesota, St. Paul, Minnesota 07/96 - present.
Associate Professor: Departments of Soil Science and Microbiology University of Minnesota, St. Paul, Minnesota 07/93 - 6/96.
Assistant Professor: Departments of Soil Science and Microbiology University of Minnesota, St. Paul, Minnesota 06/89 - 6/93.
Microbiologist: U.S. Department of Agriculture-ARS; Beltsville, Maryland, 01/86 - 05/89.
Research Molecular Biologist: Allied Corporation; Syracuse, New York, 04/85 - 12/85.
Post-Doctoral Research Fellow: Department of Biology, Plant Molecular Biology Laboratory, McGill University; Montreal, Canada, 09/83 - 04/85.
Graduate Research Assistant: Department of Microbiology, University of Hawaii; Honolulu, Hawaii, 08/80 - 08/83.
Graduate Teaching Assistant: Department of Microbiology, University of Hawaii; Honolulu, Hawaii, 08/79 - 08/80 (General and Advanced Microbiology, Medical Microbiology Laboratory).

Awards and Honors:

1990 Young Investigator Award, American Society for Microbiology.
1988 Patent Award, U.S. Department of Agriculture.
1991-1992 Bush Foundation Excellence in Teaching Program.
1999 Fellow American Academy of Microbiology

Professional Activities:

Editor, Applied and Environmental Microbiology 1999-2004
Associate Editor, Molecular-Plant Microbe Interactions 1995-1998
Associate Editor, Applied Environmental Microbiology 1989-1999
Associate Editor, Symbiosis
Co-Organizer of: Agricultural Microbes Genomes I Conference. Jan. 13-14, 2000. San Diego, CA (with Stephen Heller et al.).
Co-Organizer of: 14th North American Symbiotic Nitrogen Fixation Conference. July 25-29 1993, Minneapolis, Minnesota (with Peter Graham and Carroll Vance).
Organizer of Training Workshop: Use of Molecular Biology Techniques in Biological Nitrogen Fixation Research, June 18-22 1995, Cairo, Egypt.
Panel Member, Ad Hoc, NIH, Molecular Biology 1994
Panel member, USDA/CSRS, Competitive Research Grants Program, Nitrogen Fixation Panel, 1991.
Panel member, USDA/National Research Initiative, Competitive Research Grants Program, Nitrogen Fixation Panel, 1992.
Member, U.S. EPA, Biological Safety Advisory Committee, 1988.
Panel member, U.S. AID/PSTC, Competitive Grants Program, Molecular Biology Panel, 1987.
Panel member, U.S. AID/PSTC, Competitive Grants Program, Molecular Biology Panel,1988

Patents:

Held: A Mu-dI(Kan,lac) vector system for analysis of symbiotic genes in Rhizobium; Patent awarded in Canada and U.S. through Allelix Inc.

Patent Pending: (I) Cloning, characterization, and sequence of a gene region from Pseudomonas sp. strain ADP involved in the dechlorination of atrazine; (II) Improved enzymes for biodegradation of atrazine and other xenobiotic compounds obtained by DNA shuffling. Submitted by University of Minnesota and Ciba Geigy, Inc; (III) Atrazine-degrading plants; Submitted by University of Minnesota and Novartis Crop Protection. Pending: SAFTI-Pak, Solid Appliance Fermentation Technology for Inoculants. US 193. 394

Publications: 1995 – 2000

Dombek, P.E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the polymerase chain reaction to differentiate *Escherichia coli* from human and animal sources. Appl. Environ. Microbiol. 66:2572-2577.

Strong, L. C., McTavish, H., Sadowsky, M. J., and Wackett, L. P. 2000. Field-scale remediation of atrazine-contaminated soil using recombinant *Escherichia coli* expressing atrazine chlorohydrolase. Environ. Microbiol. 2:91-98.

Montealegre,C. M., van Kessel, C., Blumenthal, Jurg, Hur, Hor-Gil, Hartwig, U. A., and M.J. Sadowsky. 2000. Elevated atmospheric CO_2 alters microbial population structure in a pasture ecosystem. Global Change Biology 6:475-482.

Seffernick, J.L., G. Johnson, M. J. Sadowsky, and L. P. Wackett. 2000. Substrate specificity of atrazine chlorohydrolase and atrazine-catabolizing bacteria. Appl. Environ. Microbiol. IN PRESS.

Loh, J., M. G. Stacey, M. J. Sadowsky, and G. Stacey. 1999. The *Bradyrhizobium japonicum* nolA gene encodes three functionally distinct proteins. J. Bacteriol. 181: 1544-1554.

Sadowsky, M. J., Z. Tong, M. de Souza, and L. P. Wackett. 1998. AtzC is a new member of the amidohydrolase protein superfamily and is homologous to other atrazine-metabolizing enzymes. J. Bacteriol. 180:152-158.

de Souza, M. L., D. Newcombe, S. Alvey, D. E. Crowley, A. Hay, M. J. Sadowsky, and L. P. Wackett. 1998. Molecular basis of a bacterial consortium: interspecies catabolism of atrazine. Appl. Environ. Microbiol. 64:178-184

de Souza, M. L., L. P. Wackett and M. J. Sadowsky. 1998. The atzABC genes encoding atrazine catabolism genes are located on a self-transmissible plasmid in *Pseudomonas* strain ADP. Appl. Environ. Microbiol. 64:2323-2326.

Lohrke, S. L., B. Day, V. K. Kolli, R. Hancock, J. P.-Y. Yuen, M. L. de Souza, G. Stacey, R. Carlson, Z. Tong, H.-G. Hur, J. H. Orf, and M. J. Sadowsky. 1998. The *Bradyrhizobium japonicum* noeD gene: a negatively-acting, genotype-specific nodulation gene for soybean. Molec. Plant-Microbe Interact. 11:476-488

de Souza, M. L., J. Seffernick, M. J. Sadowsky, and L. P. Wackett. 1998. The atrazine catabolism genes atzABC are widespread and highly conserved. J. Bacteriol. 180:1951-1954.

Sadowsky, M. J., and H.-G. Hur. 1998. Use of endogenous repeated sequences to fingerprint bacterial genomic DNA. In: J.R. Lupski, G. Weinstock, and F. J. de Bruijn (eds.). Bacterial Genomes: Structure and Analysis. Chapman and Hall.

Sadowsky, M. J., L. P. Wackett, M. L. de Souza, K. L. Boundy-Mills, and R. T. Mandelbaum. 1998. Genetics of Atrazine Degradation in *Pseudomonas* sp. Strain ADP, pp. 88-94. In: Triazine Herbicides: Risk Assessment, L. Ballantine, J. McFarland, and D. Hackett (eds.), American Chemical Society, Washington, D.C.

Wackett, L. P., M. J. Sadowsky, M. L. de Souza, and R. T. Mandelbaum. 1998.

Atrazine hydrolysis by a bacterial enzyme, pp. 82-87. In: Triazine Herbicides: Risk Assessment, L. Ballantine, J. McFarland, and D. Hackett (eds.), American Chemical Society, Washington, D.C.

Boundy-Mills, K. L., M. L. de Souza, L. P. Wackett, R. Mandelbaum, and M. J. Sadowsky. 1997. The atzB gene of *Pseudomonas* sp. strain ADP encodes hydroxyatrazine ethylaminohydrolase, the second step of a novel atrazine degradation pathway. Applied Environ. Microbiol. 63:916-923.

Daane, L. L., J. A. E. Molina, and M. J. Sadowsky. 1997. Plasmid transfer between spatially-separated donor and recipient bacteria in earthworm-containing soil microcosms. Appl. Environ. Microbiol. 63:679-686.

Daane, L. L., J. A. E. Molina, and M. J. Sadowsky. 1997. Scanning electron microscopy of the microflora in egg capsules of the earthworm *Eisenia fetida*. Pedobiologia 42:79-87.

Schortemeyer, M., U. A. Hartwig, G. R. Hendrey, and M. J. Sadowsky. 1997 Influence of free air carbon dioxide enrichment (FACE) on microbial communities in the rhizosphere of white clover and perennial ryegrass. Soil Biology and Biochemistry. 28:1717-1724.

Sadowsky, M. J., and M. Schortemeyer. 1997. Soil microbial responses to increased concentrations of atmospheric CO_2. Global Change Biology. 3:217-244.

Schortemeyer, M., H. Santruckova, and M. J. Sadowsky. 1977. Relationship between root parameters and soil microorganisms in the rhizosphere of clover and ryegrass. Comm. Soil Sci. Plant Anal. 28:1675-1682.

Boldt, Y. R., A. K. Whitting, M. L. Wagner, M. J. Sadowsky, L. Que, and L. P. Wackett. 1997. Manganese(II) active site mutants of 3,4-dihydroxyphenylacetate 2,3-dioxygenase from *Arthrobacter globiformis* strain CmM-2. Biochemistry. 36:2147-2153.

Hur, H.-G., L. M. Newman, L. P. Wackett, and M. J. Sadowsky. 1997. Toluene 2-monooxygenase-dependent growth of *Burkholderia cepacia* G4/PR1 on diethly ether. Appl. Environ. Microbiol. 63:1606-1609.

Sadowsky, M. J., and P. H. Graham. 1997. Soil Biology of the Rhizobiaceae. In: The *Rhizobiaceae*, pp. 155-172.. H. P. Spaink, A. Kondorosi, and P. J. J. Hooykaas (eds.), Kluwer, The Netherlands.

de Souza, M. L., M. J. Sadowsky, and L. P. Wackett. 1996. Atrazine chlorohydrolase from *Pseudomonas* sp. ADP: gene sequence, enzyme purification and protein characterization. J. Bacteriol. 178:4894-4900.

Daane, L. L., J. A. E. Molina, E. C. Berry, and M. J. Sadowsky. 1996. Influence of earthworm activity on gene transfer from *Pseudomonas fluorescens* to indigenous soil bacteria. Appl. Environ. Microbiol. 62:515-521.

Sadowsky, M. J., L. L. Kinkel, J. H. Bowers, and J. L. Schottel. 1996. Use of repetitive intergenic DNA sequences and fatty Acid profiles to classify pathogenic and disease-suppressive strains of Streptomyces. Appl. Environ. Microbiol. 62:3489-3493.

Lohrke, S. M., J. H. Orf, and M. J. Sadowsky. 1996. Inheritance of host-controlled restriction of nodulation by *Bradyrhizobium japonicum* strain USDA 110. Crop Sci. 36:1271-1276.

de Souza, M. L., L. P. Wackett, K. L. Boundy-Mills, R. T. Mandelbaum, and M. J. Sadowsky. 1995. Cloning, characterization, and expression of a gene region from *Pseudomonas* sp. strain ADP involved in the dechlorination of atrazine. Appl. Environ. Microbiol. 61: 3373-3378.

Sadowsky, M. J., R. M. Kosslak, B. Golinska, C. J. Madrzak, and P. B. Cregan. 1995. Restriction of nodulation by *B. japonicum* is mediated by factors present in the roots of *Glycine max*. Appl. Environ. Microbiol. 61:832-836.

Mdrzak, C. J., B. Goliska, J. Króliczak, K. Pudeko, D. Aewska, B. Lampka, and M. J. Sadowsky. 1995. Diversity among field populations of *Bradyrhizobium japonicum* in Poland. Appl. Environ. Microbiol 61:1194-1200.

Lohrke, S. M., J. H. Orf, E. Martínez-Romero, and M. J. Sadowsky. 1995. Host-controlled restriction of nodulation by *Bradyrhizobium japonicum* strains in serogroup 110. Appl. Environ. Microbiol. 61:2378-2383.

Graham, P. H., M. J. Sadowsky, S. W. Tighe, J. A. Thompson, R. A. Date, J. G. Howieson, and R. Thomas. 1995. Differences among strains of *Bradyrhizobium* in fatty acid-methyl ester (FAME) analysis. Can. J. Microbiol. 41:1038-1042.

Noyd, R. K., F. L. Pfleger, R. E. Norland, and M. J. Sadowsky. 1995. Native prairie grasses and microbial community responses to reclamation of taconite iron ore tailing. Can. J. Botany. 73:1645-1654.

LeeAnn K. Johnson

University of Minnesota
Department of Soil, Water and Climate
1991 Upper Buford Circle, 439 Borlaug Hall
St. Paul, MN 55108
Phone: (612) 625-4241
E-mail: ljohnson@soils.umn.edu

EXPERIENCE

1999-present **Senior Scientist**
Department of Soil, Water & Climate
University of Minnesota, St. Paul, MN

Manage a field- and laboratory-based research project using DNA fingerprinting methods to address fecal pollution in Minnesota watersheds:

- Develop sampling and laboratory protocols and procedures
- Manage field sampling and laboratory activities
- Maintain database and analyze data using standard and specialized software
- Hire and supervise technical support staff
- Coordinate stream sampling with local government staff
- Present project results through publications and presentations
- Write grant proposals
- Manage project budgets

1987-1999 **Senior Research Scientist**
Research and Development Section
Metropolitan Council Environmental Services, St. Paul, MN

Managed research and monitoring projects related to wastewater treatment and water quality:

- Responsibilities: Review scientific literature, design experiments, conduct field and laboratory studies, collect and analyze samples, analyze data, communicate results through written reports and oral presentations, direct technical support staff.
- Projects: activated sludge bulking control, toxicity and treatability of industrial discharges, ultraviolet disinfection, biological odor control, wastewater operations troubleshooting, automated monitoring of treatment plant processes and effluent quality, laboratory method development and troubleshooting.
- Other: Team leader for Mercury Reduction Strategy, radiation safety officer and hazardous waste coordinator, MN Class C Wastewater Treatment Facility Operator Certificate.

1984-1987 **Pollution Control Specialist**
Solid and Hazardous Waste Division
Minnesota Pollution Control Agency, St. Paul, MN

Enforced state and federal hazardous waste regulations:
- Inspected companies for compliance with hazardous waste regulations
- Took enforcement action as needed
- Provided technical assistance to the regulated community
- Assisted in program development

1984 **Pollution Control Technician**
Water Quality Division
Minnesota Pollution Control Agency, St. Paul, MN

Conducted water quality studies to determine wastewater treatment plant effluent limits and advanced treatment justifications:
- Operated and maintained field sampling and monitoring equipment
- Collected and processed water samples
- Summarized data

1979-1984 **Research Assistant**
Microbiology Department
University of Minnesota, Gray Freshwater Biological Institute, Navarre, MN

Researched biodegradability of toxic organic compounds in natural aquatic systems. Reviewed literature, designed and conducted field and laboratory studies, analyzed data, published results.

1978-1979 **Laboratory Attendant**
Department of Food Science and Nutrition
University of Minnesota, St. Paul, MN

1977-1979 **Laboratory Attendant**
Biochemistry Department
University of Minnesota, St. Paul, MN

EDUCATION

B.S. Microbiology, University of Minnesota, 1979
Graduate studies in Microbiology, University of Minnesota, 1979-1984
MPH Environmental Microbiology, University of Minnesota (anticipated 2001)

References for Research Addendum

Brown, G., Khan, Z., and Lifshitz, R. (1989). Plant promoting rhizobacteria: strain identification by restriction fragment length polymorphisms. Can J. Microbiol. 36, 242-248.

Buchman, T.G., Roizman, B., Adams, G., and Stover, B.H. (1978). Restriction endonuclease fingerprinting of *Herpes simplex* virus DNA: a novel epidemiological tool applied to a nosocomial outbreak. J. Infect. Dis. 138, 488-498.

Chandler, D.K.S., Razin, S., Stephens, E.B., Harasawa, R., and Barile, M.F. (1982). Genomic and phenotypic analysis of *Mycoplasma pneumoniae* strains. Infect. Immun. 38, 604-609.

Christensen, L.S., Soerensen, K.J., and Lei, J.C. (1987). Restriction fragment pattern (RFP) analysis of genomes from Danish isolates of suid herpes virus 1 (Aujezsky's disease virus). Arch. Virol. 97, 215-224.

de Bruijn, F.J. (1992). Use of repetitive (repetitive extragenic element and enterobacterial repetitive intergenic consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. Appl. Environ. Microbiol. 58, 2180-2187.

Dombek, P.E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the polymerase chain reaction to differentiate *Escherichia coli* from human and animal sources. Appl. Environ. Microbiol. 66:2572-2577.

Georghiou, P., Hamill, R.J., Wright, C.E., Versalovic, J., Koeuth, T., Watson, D. A., and Lupski, J.R. (1995) Molecular epidemiology of infection due to *Enterobacter aerogenes*: identification of hospital outbreak strains by molecular techniques. Clin. Infect. Dis. 20:84-94.

Gill, P., Lygo, J.E., Fowler, S.J., and Werrett, D.J. (1987). An evaluation of DNA fingerprinting for forensic purposes. Electrophoresis 8, 38-44.

Gilson, E., Clement, J.-M., Brutlag, D., and Hofnung, M. (1984). A family of dispersed repetitive extragenic palindromic DNA sequences in *E. coli*. EMBO J. 3, 1417-1421.

Gilson, E., Clement, J.M., Perrin, D., and Hofnung, M. (1987). Palindromic units, a case of highly repetitive DNA sequences in bacteria. Trends Genet. 3, 226-230.

Gilson, E., Saurin, W., Perrin, D., Bachellier, S., and Hofnung, M. (1991). Palindromic units are part of a new bacterial interspersed mosaic element (BIME). Nucleic Acids Res. 19, 1375-1383.

Glynn, P., Higgins, P., Squartini, A., and O'Gara, F.. (1985). Strain identification in *Rhizobium trifolii* using DNA restriction analysis, plasmid DNA profiles, and intrinsic antibiotic resistances. FEMS Microbiol. Lett. 30, 177-182.

Hulton, C.S.J., Higgins C.F., and Sharp, P.M. (1991). ERIC sequences, a novel family of repetitive elements in the genomes of *Escherichia coli*, *Salmonella typhimurium* and other enterobacteria. Mol. Microbiol. 5, 825-834.

Judd, A.K., and Sadowsky, M.J. (1993). The *Bradyrhizobium japonicum* serocluster 123 hyperreiterated DNA region, HRS1, has DNA and amino acid sequence homology to IS*1380*, an insertion sequence from *Acetobacter pasteurianus*. Appl. Environ. Microbiol. 59, 1656-1661.

Judd, A. K., M. Schneider, M. J. Sadowsky and F. J. de Bruijn. (1993). The use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. Appl. Environ. Microbiol. **59:**1702-1708.

Kaijalainen, S., and Lindstrom, K. (1989). Restriction fragment length polymorphism analysis of *Rhizobium galegae* strains. J. Bacteriol. 171, 5561-5566.

Kaper, J.B., Bradford, H.B., Roberts, and Falkow, S. (1982). Molecular epidemiology of *Vibrio cholerae* in the U.S. Gulf coast. J. Clin. Microbiol. 16, 129-134.

Kleckner, N. (1981). Transposable elements in prokaryotes. Annu. Rev. Genet. 15, 341-404.

Koch, E., Song, K., Osborn, T. C., and Williams, P.H. (1991). Relationship between pathogenicity and phylogeny based on restriction fragment length polymorphism in *Leptosphaeria maculans*. Mol. Plant Microbe Interact. 4, 341-349.

Kuijper, E.J., Oudbier, J.H., Stuifbergen, W.N., Jansz, A., and Zanen, H.C. (1987). Application of whole-cell DNA restriction endonuclease profiles to the epidemiology of *Clostridium difficile*-induced diarrhea. J. Clin. Microbiol. 25, 751-753.

Langenberg, W., Rauws, E.A.J., Widjojokusumo, A., Tytgat, G.N.J., and Zanen, H.C. (1986). Identification of *Campylobacter pyloridis* isolates by restriction endonuclease DNA analysis. J. Clin. Microbiol. 24, 414-417.

Lazo, G.R., Roffey, R., and Gabriel, D.W. (1987). Pathovars of *Xanthomonas campestris* are distinguishable by restriction fragment length polymorphisms. Int. J. Syst. Bacteriol. 37, 214-221.

Lipman, L.J.A., de Nijs, A., Lam, TJG, and Gaastra, W. (1995) Identification of Escherichia coli strains from cows with clinical mastitis by serotyping and DNA polymorphism patters with REP and ERIC primers. Vet. Microbiol. 43:13-19.

Lin, R.-J., Capage, M., and Hill, C.W. (1984). A repetitive DNA sequence, *rhs*, responsible for duplications within the *Escherichia coli* K-12 chromosome. J. Mol. Biol. 177, 1-18.

Lupski, J.M., and Weinstock, G.M. (1992). Short, interspersed repetitive DNA sequences in prokaryotic genomes. J. Bacteriol. 174, 4525-4529.

Martin, B., Humbert, O., Camara, M., Guenzi, E., Walker, J., Mitchell, T., Andrew, P., Prudhome, M., Alloing, G., Hakenbeck, R., Morrison, D.A., Boulnois, G.J., and Claverys, J.-P. (1992). A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucleic Acids Res. 20, 3479-3483.

Mazel, D., Houmard, J., Castets, A.M., and de Marsac, N.T. (1990). Highly repetitive DNA seqquences in cyanobacterial genomes. J. Bacteriol. 172, 2755-2761.

Mielenz, J.R., Jackson, L.E., O'Gara, F., and Shanmugam, K.T. (1979). Fingerprinting bacterial chromosomal DNA with restriction endonuclease *Eco*R1: comparison of *Rhizobium* spp. and identification of mutants. Can. J. Microbiol. 25, 803-807.

Panchal, C.J., Bast, L., Dowhanick, T., and Stewart, G.G. (1987). A rapid, simple, and reliable method of differentiating brewing yeast strains based on DNA restriction patterns. J. Inst. Brew. 93, 325-327.

Rademaker, J. L. W., Louws, F. J., and de Bruijn, F. J. (1998). Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting. In: Akkermans, A.D.L., van Elsas, J.D., and de Bruijn, F.J. Molecular Microbial Ecology Manual, Kluwer Academic Publishers, Dordrecht, pp. 1-26..

Rademaker, J. L. W., Louws, F. J., Rossbach, U., Vinuesa, and de Bruijn, F. J. (1998). Computer assisted pattern analysis of molecular fingerprints and database construction. In: Akkermans, A.D.L., van Elsas, J.D., and de Bruijn, F.J. Molecular Microbial Ecology Manual, Kluwer Academic Publishers, Dordrecht.

Ramos, M.S., and Harlander, S.K. (1990). DNA fingerprinting of lactococci and streptococci used in dairy fermentations. Appl. Microbiol. Biotechnol. 34, 368-374.

Ruland, K., Wenzel, R., and Herrmann, R. (1990). Analysis of three different repeated DNA elements present in the P1 operon of *Mycoplasma pneumoniae*: size, number, and distribution on the genome. Nucleic Acids Res. 18, 6311-6317.

Sadosky, A.B., Davidson, A., Lin, R.-J., and Hill, C.W. (1989). *rhs* gene family of *Escherichia coli* K-12. J. Bacteriol. 171, 636-642.

Sadowsky, M. J., and Hur, H.-G. (1998) Use of endogenous repeated sequences to fingerprint bacterial genomic DNA. *In:* J.R. Lupski, G. Weinstock, and F. J. de Bruijn (eds.). Bacterial Genomes: Structure and Analysis. Chapman and Hall, pp. 399-413..

Sadowsky, M. J., L. L. Kinkel, J. H. Bowers, and Schottel, J.L.. (1996). Use of repetitive intergenic DNA sequences and fatty acid profiles to classify pathogenic and disease-suppressive strains of *Streptomyces*. Appl. Environ. Microbiol. **62:**3489-3493.

Sadowsky, M.J., Tully, R.E., Cregan, P.B., and Keyser, H.H. (1987). Genetic diversity in *Bradyrhizobium japonicum* serogroup 123 and its relation to genotype-specific nodulation of soybeans. Appl. Environ. Microbiol. 53, 2624-2630.

Scherer, S., and Stevens, D.A. (1987). Application of DNA typing methods to epidemiology and taxonomy of *Candida* species. J. Clin. Microbiol. 25, 675-679.

Schmidt, E.L., Zidwick, M.J. and Abebe, H.H. (1986). *Bradyrhizobium japonicum* serocluster 123 and diversity among member isolates. Appl. Environ. Microbiol. 51, 1212-1215.

Schneider, M., and de Bruijn, F.J. (1996). Rep-PCR mediated genomic fingerprinting of rhizobia and computer-assisted phylogenetic analysis. World J. Microbiol. Biotechnol. 12, 163-174.

Skjold, S., Quie, P.G., Fries, L.A., Barnham, M., and Cleary, P.P. (1987). DNA fingerprinting of *Streptococcus zooepidemicus* (Lancefield group C). as an aid to epidemiological study. J. Infect. Dis. 155, 1145-1150.

Sharples, G.J., and Llod, R.G. (1990). A novel repeated DNA sequence located in the intergenic regions of bacterial chromosomes. Nucleic Acids Res. 18, 6503-6508.

Standard Methods for the Examination of Water and Wastewater (1995). American Public Health Association, Washington, D.C.

Stern, M.J., Ames, G.F.-L., Smith, N.H., Robinson, E.C., and Higgins, C.F. (1984). Repetitive extragenic palindromic sequences, a major component of the bacterial genome. Cell 37, 1015-1026.

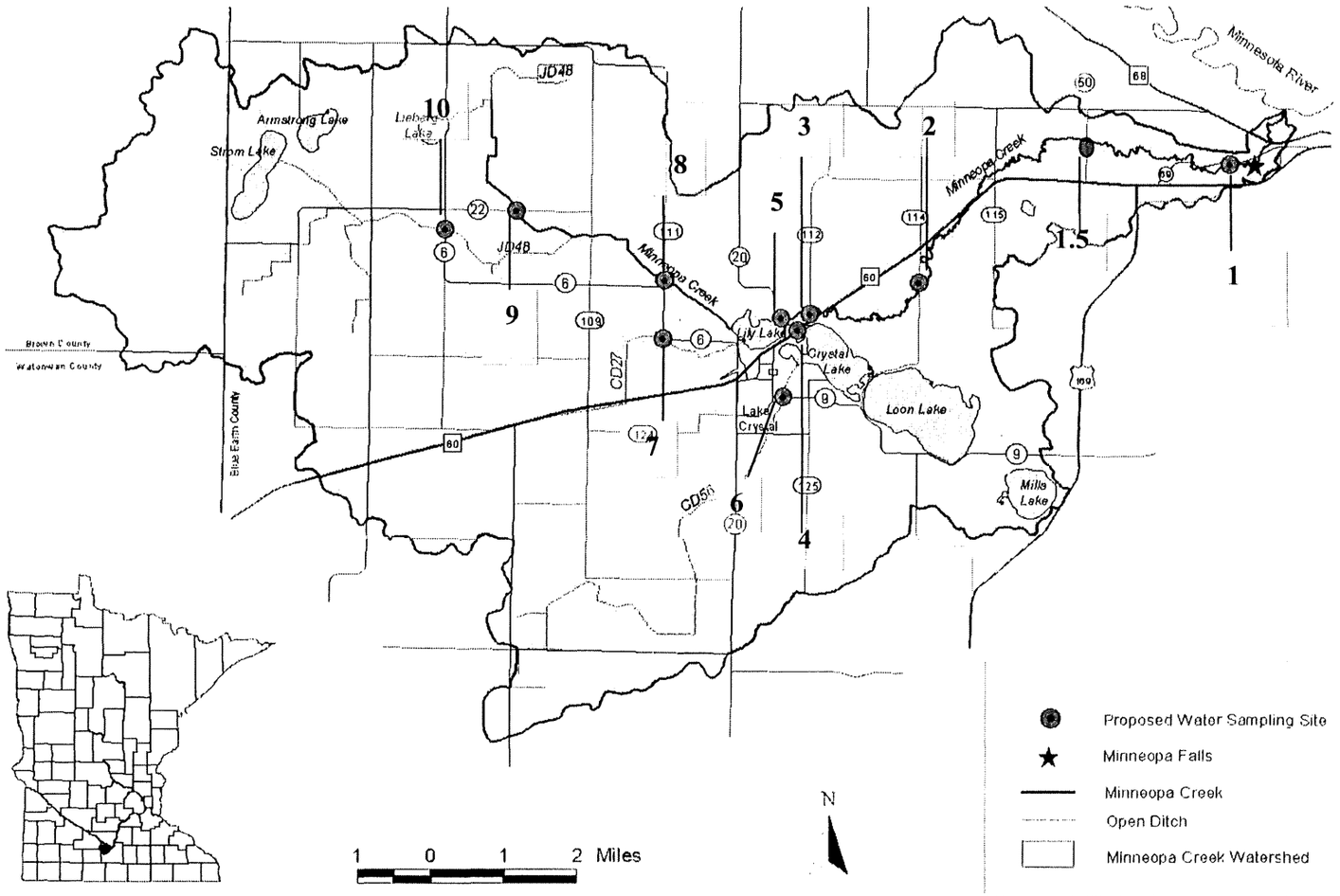
Tompkins, L.S., Troup, N.J., Woods, T., Bibb, W., and McKinney, R.M. (1987). Molecular epidemiology of *Legionella* species by restriction endonuclease and alloenzyme analysis. J. Clin. Microbiol. 25, 1875-1880.

Versalovic, J., de Bruijn, F.J., and Lupski, R. (1998). Repetitive sequence-based PCR (rep-PCR) DNA fingerprinting of bacterial genomes. *In:* J.R. Lupski, G. Weinstock, and F. J. de Bruijn (eds.). Bacterial Genomes: Structure and Analysis. Chapman and Hall, pp. 437-454.

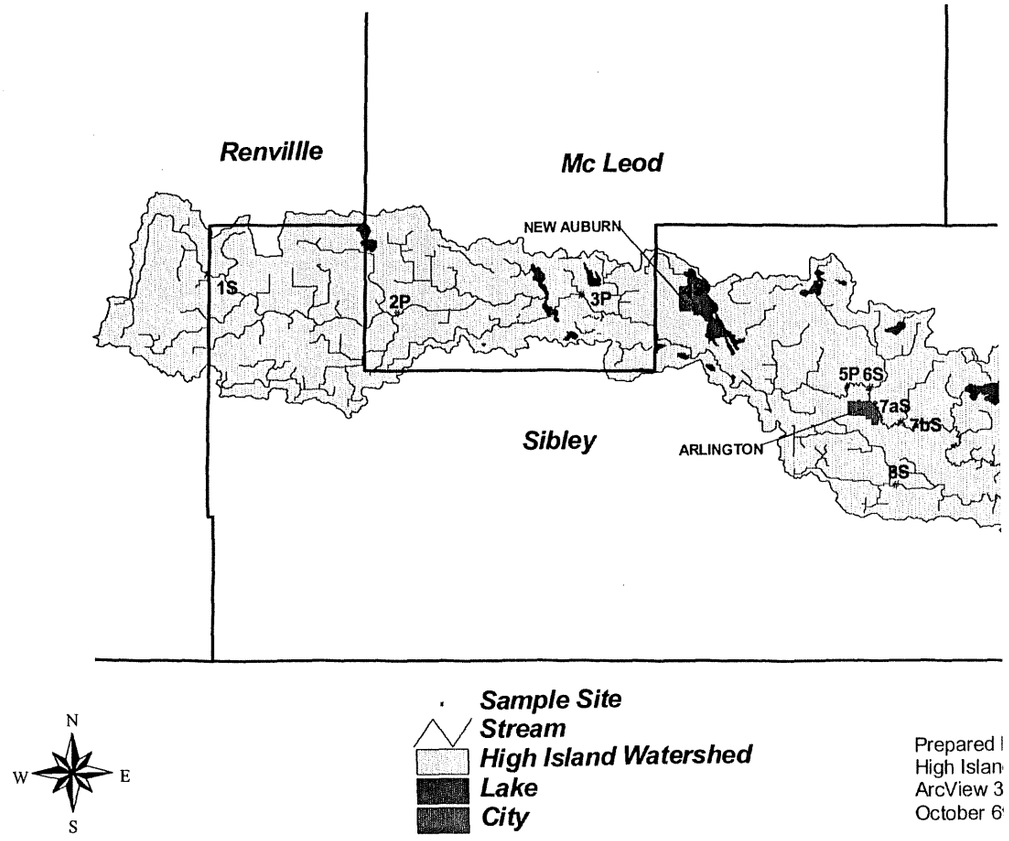
Versalovic, J., Koeuth, T., and Lupski, J.R. (1991). Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. Nucleic Acids Res. 24, 6823-6831.

Versalovic, J., Schnieder, M., de Bruijn, F.J., and Lupski, J.R. (1994). Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Meth. Molec. Cell. Biol. 5. 25-40.

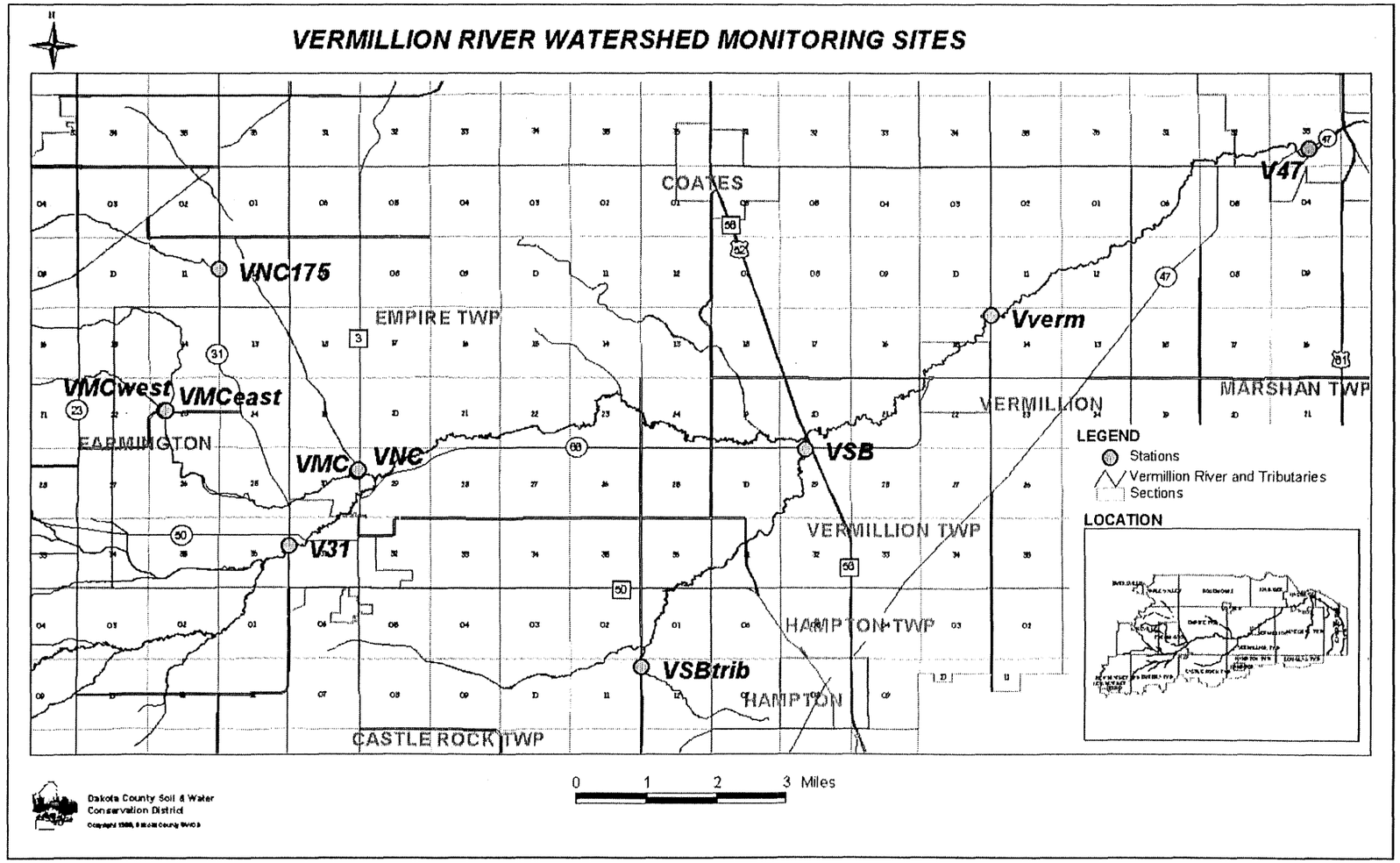
Minneopa Creek Watershed - Sampling Sites



High Island Watershed - Sample Sit



Prepared
High Islan
ArcView 3
October 6


VERMILLION RIVER WATERSHED MONITORING SITES
COATES
EMPIRE TWP
VERMILLION
MARSHAN TWP
VERMILLION TWP
HAMPTON TWP
HAMPTON
CASTLE ROCK TWP
EARMINGTON
VNC175
VMCwest
VMCeast
VMC
VNC
V31
VSB
VSBtrib
Vverm
V47
LEGEND
Stations
Vermillion River and Tributaries
Sections
LOCATION
0 1 2 3 Miles
Dakota County Soil & Water Conservation District

ATTACHMENT C. SUPPORTING DATA

LCMR WATERSHED FECAL COUNTS:

Table 1

VERMILLION RIVER FECAL COUNTS								
SITE	**07/11/01**	**08/08/01**	**09/05/01**	**10/03/01**	**03/27/02**	**05/01/02**	**06/05/02**	**07/02/02**
VMCwest	770	4300	1000	830	73	8	1000	5600
VMCeast	1800	560	500	200	150	35	930	1900
VMC	360	610	270	170	3	39	1400	420
VNC175	150	200	150	120	97	44	1100	520
VNC	590	390	240	180	25	8	560	210
VSBtrib	220	240	73	140	20	20	610	97
VSB	410	767	450	210	60	22	2000	1100
V31	360	733	93	63	10	10	1400	340
Vverm	190	1633	290	83	970	10	2400	250
V47	560	290	800	150	190	23	2800	330

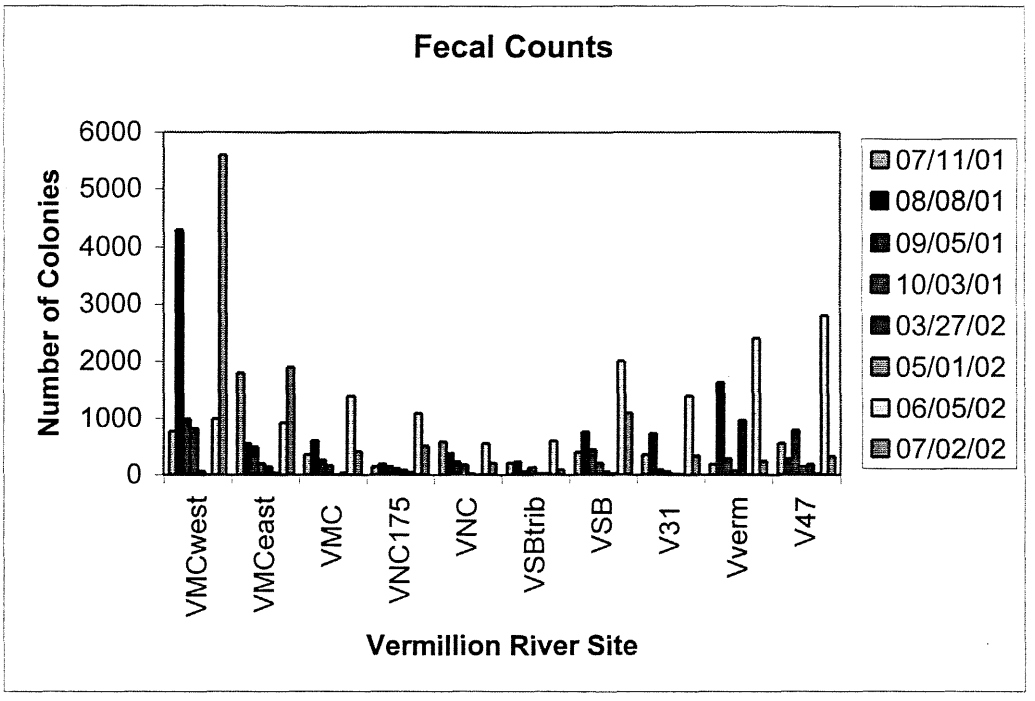


Figure 1

Table 2

HIGH ISLAND CREEK FECAL COUNTS								
SITE	04/09/01	07/18/01	04/10/02	04/24/02	05/08/02	05/29/02	06/19/02	07/10/02
IS	unknown	40	60	1650	360	280	460	5500
2P		620	520	10	400	190	250	1100
3P		40	10	32	270	160	200	1200
5P		280	10	10	81	600	670	460
6S		80	73	34	240	5200	1100	1500
7S		610	140	55	1360	X	7300	1600
8P		490	3	2	1500	X	2400	5900
9P		90	10	10	748	X	1800	2200
10P		600	9	20	240	4300	11200	1300
T		670	640	2	2000	2100	900	1000

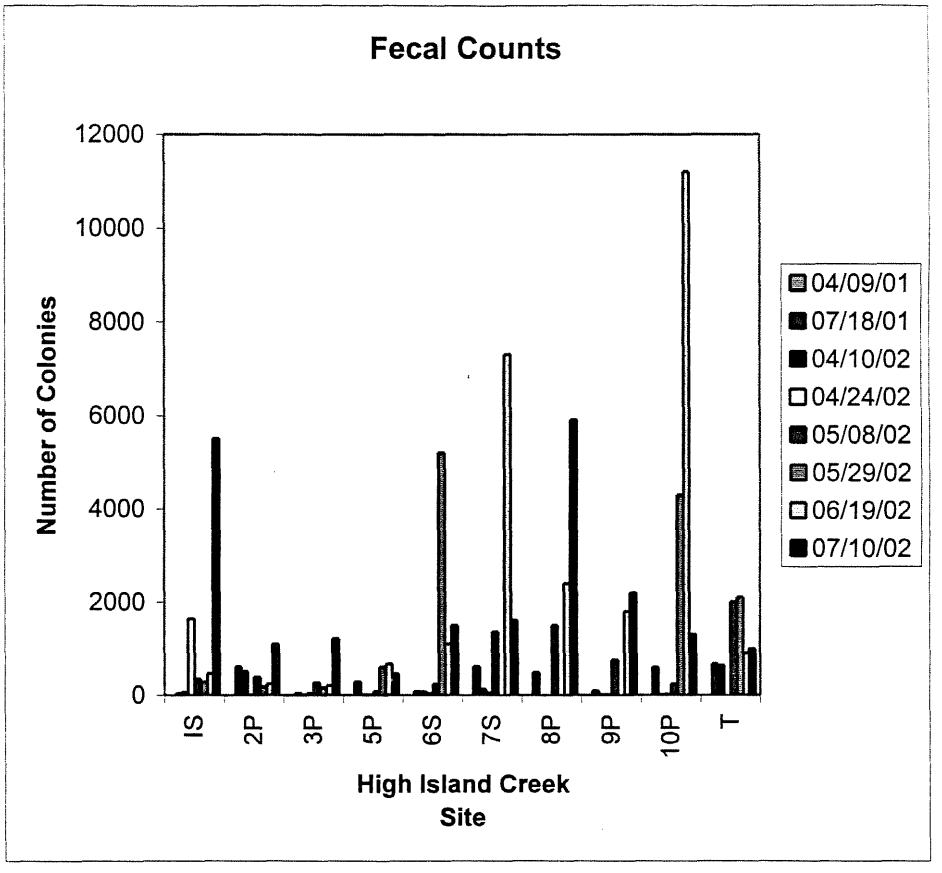


Figure 2

Table 3

MINNEOPA CREEK FECAL COUNTS									
SITE	04/03/01	07/30/01	08/15/01	09/12/01	10/10/01	04/18/02	05/15/02	06/12/02	07/18/02
1	unknown	330	680	140	700	110	75	240	660
2		150	980	600	930	830	180	100	580
3		60	370	170	1300	1320	200	90	420
4		20	74	10	80	20	210	130	90
5		67	370	280	900	5	82	46	70
6		240	250	1100	800	110	400	590	490
7		350	67	91	100	6	7	X	130
8		160	110	9	230	62	64	X	730
9		67	140	56	9	10	53	X	84
10		70	67	6	39	4	45	X	180

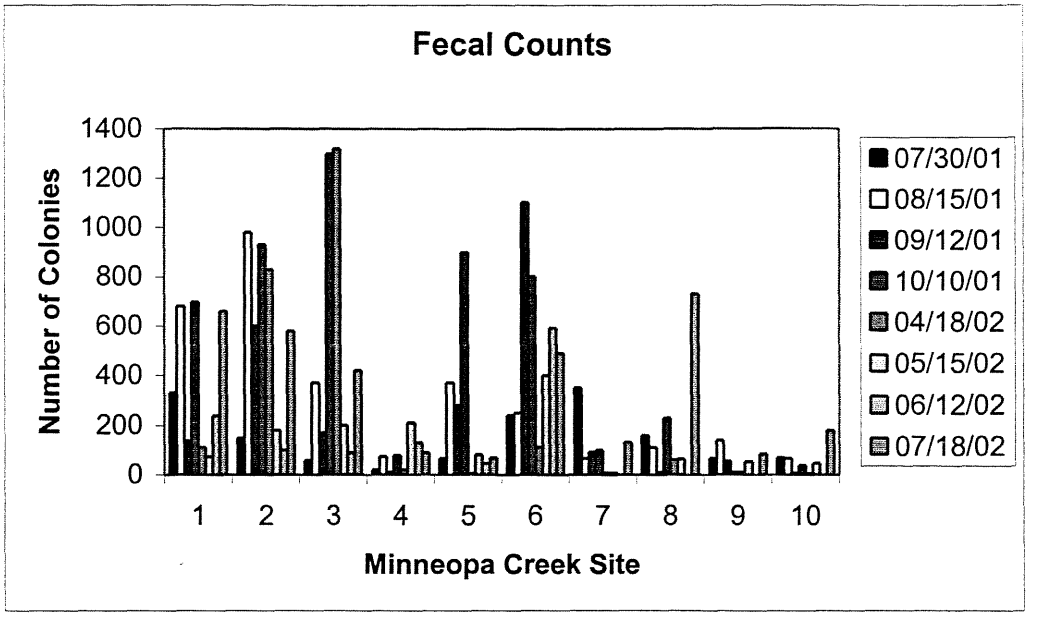


Figure 3

In Press: 2004. Applied and Environmental Microbiology

Frequency and Distribution of Tetracycline Resistance Genes in Genetically-Diverse, Non-Selected, and Non-Clinical *Escherichia coli* Isolated From Diverse Human and Animal Sources

Andrew Bryan[1†], Nir Shapir[1,2,3,4], and Michael J. Sadowsky[1,2,3,*]

[1]Department of Soil, Water, and Climate, [2]BioTechnology Institute, [3]Center for Microbial and Plant Genomics, and [4]Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, St. Paul, MN 55108

Running Title: Tet Genes in Diverse Environmental *E. coli*

*Address of corresponding author: Department of Soil, Water, and Climate, University of Minnesota, 1991 Upper Buford Circle, 439 Borlaug Hall, St. Paul, MN 55108. Phone: (612) 624-2706, Fax: (612) 625-2208, E-mail: sadowsky@umn.edu

†Current Address: Department of Medical Microbiology and Immunology, University of Wisconsin, 1300 University Avenue, 481 MSC, Madison, WI 53706.

Non-selected and natural populations of *Escherichia coli* from 12 animal sources and humans were examined for the presence and types of 14 tetracycline resistance determinants. Of 1,263 unique *E. coli* isolates from humans, pigs, chickens, turkeys, sheep, cows, goats, cats, dogs, horses, geese, ducks, and deer, 31% were highly resistant to tetracycline. Over 78, 47, and 41% of the *E. coli* isolates from pigs, chickens, and turkeys were resistant or highly resistant to tetracycline, respectively, and 61, 29, and 29% of *E. coli* isolates from pig, chickens, and turkeys, respectively, had MIC values ≥233 μg tetracycline per ml. Muliplex PCR analyses indicated that 97% of these strains contained at least one of 14 tetracycline resistance determinants (*tetA, tetB, tetC, tetD, tetE, tetG, tetK, tetL, tetM, tetO, tetS, tetAP, tetQ,* and *tetX*) examined. While the most common determinants found in these isolates were *tetB* (63%) and *tetA* (35%), *tetC*, *tetD*, and *tetM* were also found. *E. coli* isolates from pigs and chickens were the only strains to have *tetM*. To our knowledge, this represents the first report of *tetM* in *E. coli*.

Problems associated with the presence of antibiotic resistant bacteria has reached epidemic proportions in recent years, with cost estimates exceeding $4 billion in the United States alone (6). The spread of antibiotics resistant bacteria in the environment is dependent on the presence and transfer of resistance genes among microorganisms, mutations, and selection pressure to keep these genes in a population, the later neatly provided by the approximately 50 million pounds of antibiotics that are produced and used each year in the United States (14). Only half of these antibiotics are used for humans, while the remainder are administered to animals or other organisms (8). The causes and effects of antibiotic overuse are varied. One of

the most controversial applications of antibiotics, however, is their use for growth promotion in livestock, and this has raised concerns that such use contributes to the presence of resistant bacteria in humans (1, 25).

Tetracyclines have become the drugs of choice to treat *Mycoplasma*- and *Chlamydia*-induced pneumonia (13), and have been used to treat other atypical pneumonias, rickettsial infections, Lyme disease, ehrlichiosis, and other diseases and cancers (23). The clinically useful tetracycline (Tet), chlortetracycline, was introduced in 1948 (24). Only a year later, it was shown that young chickens fed tetracyclines had enhanced growth characteristics (10). However, by 1953, it was reported that *Shigella dysenteriae* had developed resistance to tetracycline antibiotics and by 1955, a *Shigella* sp. strain had developed multidrug resistance (20). Because of that history and the broad clinical use of tetracycline, this antibiotic was chosen, along with commensal strains of *E. coli,* to provide a prototypical view of the use of antibiotics and their effects on bacterial populations (21). Tetracycline is a broad-spectrum antibiotic that inhibits bacterial protein synthesis by preventing aminoacyl-tRNA from binding to the bacterial ribosome (20). Resistance to the antibiotic is conferred by one or more of the 36 currently described *tet* genes, which encode for one of three mechanisms of resistance: an efflux pump, a method of ribosomal protection, or direct enzymatic inactivation of the drug (7). Efflux mechanisms appear to be more abundant among gram-negative microorganisms, while ribosomal protection mechanisms are more common among gram-positives (7). Generally speaking, the rapid spread of tetracycline resistance among bacteria is due to the localization of *tet* genes on plasmids, transposons, and integrons (7,15, 21).

While several studies have examined Tet resistance among bacteria, most have employed clinically-isolated bacteria (4,11,17) or populations specifically isolated for their ability to grow

in the presence of tetracyclines (5,22). These studies, while useful, do not give an unbiased appraisal of the presence and types of *tet* genes that are present in natural (non-clinical), non-selected, populations of bacteria in the environment.

Only a limited number of studies have examined tetracycline resistance determinants in bacteria isolated from a large variety of animal species with different exposure histories to tetracyclines, or in environmental samples (11). While Sengeløv and coworkers (22) examined 100 *E. coli* isolates for the presence of five *tet* resistance determinants and Blake *et al.* (5) used PCR to examine 200 Tet resistant *E. coli* for seven *tet* genes, few have examined a large number of *tet* determinants in non-clinical *E. coli* isolated from a variety of animal species. To better understand the distribution of resistance genes in the environment and to provide insight into selection pressures involved with the use of antibiotics in animal feed, we investigated Tet resistance among natural and unselected populations of *Escherichia coli* from 12 animal sources and humans and determined which resistance genes were present in this population.

Isolates and determination of minimum inhibitory concentration. In order to characterize tetracycline resistance in natural, non-clinical *E. coli* strains from both human and animal sources, 1263 unique isolates were obtained from humans, cat, cow, deer, turkey, duck, sheep, goose, dog, pig, horse, chicken, and goat (Table 1). Fecal materials were collected by swabbing the rectal or cloacal region of individual wild and domesticated animals located throughout Minnesota and western Wisconsin as previously described (9). Fecal samples were kept at 4°C and analyzed within 6 hr of swabbing. Fecal material was streaked onto mFC agar plates (Difco, BD Diagnostic Systems, Sparks, MD), incubated at 44.5°C for 24 hours, and six blue colonies from mFC plates were picked and evaluated using selective and differential growth media as previously described (9). Only isolates giving growth and color responses, on

all media, that were typical for *E. coli* were used in these studies. Three *E. coli* colonies from each individual fecal sample were used for DNA fingerprinting. All isolates were DNA fingerprinted using rep-PCR and BOXA1R primers (9) and identical clones from the same animal were eliminated from analyses. Unique isolates were grown overnight in 150 µl of Luria-Bertani liquid medium in microtiter plates and were spot inoculated, using a multiple inoculator, onto Tryptic Soy Agar (Difco Laboratories, Detroit, MI) supplemented with 0, 5, 10, 20, 40, 70, 93, 117, 175, and 233 µg tetracycline per ml (Sigma Chemicals, St. Louis, MO). Plates were incubated at 37°C overnight and visually examined for growth. Minimum inhibitory concentrations (MIC) were determined from growth patterns and average values are shown in Figure 1. MIC values of <5 µg tetracycline per ml were considered sensitive to the antibiotic, while those having MIC values of 10 – 70, or >90 µg tetracycline per ml were considered resistant or highly resistant, respectively. A MIC of >233 µg/ml was considered as MIC of 233 µg/ml for statistical analysis.

Of the 1263 *E. coli* isolates examined, 31% were resistant to tetracycline (MIC values >10 µg/ml); consisting of 42% from livestock, 21% from humans, 17% from companion animals (cats, dogs, horses), and 4% from wild animals. Over 78, 47, and 41% of the *E. coli* isolates from pigs, chickens, and turkeys were resistant or highly resistant to tetracycline, respectively. Together these resistant isolates represent about 20% of the 1263 isolates examined. In contrast, about 22, 30, 3, 3, 21, 33, 7, 23, 6, and 12.2% of the *E. coli* isolates from cats, cows, deer, duck, humans, sheep, geese, dogs, horses, and goats were resistant or highly resistant to tetracycline, respectively. Moreover, 61, 29, and 29% of *E. coli* isolates from pig, chickens, and turkeys, respectively, had MIC values ≥233 µg tetracycline per ml. In contrast, goats, horses, ducks, geese, and deer had the least number of *E. coli* strains showing resistance or a high level of

resistance to tetracycline. Our results may be explained by the potential exposure of livestock, humans, and companion- and wild-animals to tetracyclines. Tetracycline is often continuously fed to livestock at sub-therapeutic levels for the purpose of growth promotion. For example, up to 70% of U.S. cattle and pig operations use feeds supplemented with antibiotics for growth promotion and the majority are tetracyclines (2). In contrast, humans and companion animals are most often treated therapeutically, for a limited time, for bacterial infections, perhaps reflecting the intermediate level (MIC 10- 70 μg/ml) of resistance to tetracycline. This may be changing, however, as other uses of antibiotics become more common, such as treatment of parasitic and non-infectious diseases (21). The low level of occurrence of tetracycline resistance among isolates from wild animals is presumably due to their low exposure to these antibiotics. Most isolates either had a high level of resistance or none at all, suggesting that the acquisition of a mobile genetic element accounts for resistance.

Epidemiology of *tet* genes. All isolates (325) with a tetracycline MIC of ≥93 μg/ml (which we considered to constitute a high level of resistance) were examined further using multiplex PCR for the presence of *tetA, tetB, tetC, tetD, tetE, tetG, tetK, tetL, tetM, tetO, tetS, tetAP, tetQ,* and *tetX* genes (18). Single-colony isolates were streaked onto Plate Count Agar (Difco, Detroit, MI), picked using disposable 10 μl sterile loops, and were suspended in 50 μl sterile H_2O. One μl of the standardized cell suspensions served as template DNA for colony-based multiplex PCR. The primers used for PCR amplification of the 14 tetracycline resistance genes were as described by Ng *et al.* (18). The primers were aliquoted into four groups: Group I contained primers for *tetB*, *tetC* and *tetD*; Group II contained *tetA*, *tetE* and *tetG*; Group III contained *tetK*, *tetL*, *tetM*, *tetO*, and *tetS*; and Group IV contained primers for *tetA(P)*, *tetQ* and *tetX*. PCR was done in 96 well plates using a MJ Model PTC100 Thermocycler (Waltham, MA), using the following

conditions as described (18): 5 min initial denature at 94°C, followed by 35 cycles of 94°C for 1 min, 55°C for 1 min and 72°C for 1.5 min. PCR products were separated by gel electrophoresis in 1% (w/v) agarose gels in 1×TAE buffer, stained with ethidium bromide, and visualized under U.V. illumination. The validity of multiplex PCR reactions and product sizes was ascertained by using the following positive control plasmids: pSL18, pRT11,pBR322, pSL106, pSL1504, pJA8122, pAT102, pVB.A15, pJ13, pUOA1, pAT451, pJIR39, pNFD13-2, and pBS5, for *tet* genes A, B, C, D, E, G, K, L, M, O, S, A(P), Q, and X, respectively (18). Sizes of PCR products were determined by comparison to the migration of 100-bp ladder (Gibco, BRL). The identity of all *tet* genes in a representative sample of non-clinical isolates was ascertained by DNA sequencing of PCR products, following extraction from agarose gels. A representative agrose gel of PCR products obtained using primer Group I, amplifying *tetB*, *tetC,* and *tetD*, is shown in Figure 2.

Of the 325 strains analyzed by PCR, 97% contained at least one of 14 (*tetA, tetB, tetC, tetD, tetE, tetG, tetK, tetL, tetM, tetO, tetS, tetAP, tetQ,* and *tetX*) tetracycline resistance determinants. The most common determinants were *tet*(B) (63% of isolates) and *tet*(A) (35% of isolates) (Figure 3). However, *tet*(C), *tet*(D), and *tet*(M) were also found with varying frequencies. The frequencies of *tetA, tetB, tetC,* and *tetD* in the tested isolates (Figure 3) were consistent with those previously reported for lactose-fermenting coliforms using colony hybridization (11). In contrast, Sengeløv and coworkers (22) reported that 71% and 25% of 100 isolates from diseased and healthy pigs, cattle and chickens they tested for five tetracycline resistance determinants contained *tetA* and *tetB*, respectively. None of the tested strains contained genes for *tetE, tetG, tetK, tetL, tetO, tetS, tetAP, tetQ,* or *tetX*. Because in our studies only highly resistant isolates were analyzed by PCR, it is possible that additional resistant genes

were present in the *E. coli* populations, but were non-functional or only provided intermediate or a low-level of resistance.

Isolates from pigs and chickens were the only strains to contain *tetM*, and commonly had more than one tetracycline resistance determinant per strain (Figure 4). *E. coli* from these animals had the greatest number of strains with high MIC values. Over 30% of *E. coli* isolates from turkeys, pigs, and horses contained two Tet resistance determinants, and 4.5% of the pig isolates contained three *tet* genes. However, the presence of more than one resistance determinant did not lead to noticeably higher MIC values. It is possible that strong selection pressures provided by environments containing elevated levels of tetracycline leads to the acquisition of more than one tetracycline gene in a given strain due their prevalence in the environment, rather than a selective advantage. Results of our studies also showed that 22.2% and 1.9% of the isolates contained two and three *tet* genes, respectively. This is in contrast to results from previous studies, in which only 3.5% (16) and 5.4% (22) of isolates had two genes, perhaps due to our use of a larger number and variety of isolates, and the greater number of genes examined.

To our knowledge, this is the first report documenting the presence of the *tetM* gene in *E. coli* (7). Due to the uniqueness of these results, the presence of *tetM* in one of our *E. coli* isolates from pigs was verified by sequencing the PCR product produced using *tetM*-specific primers. Blast analysis (3) indicated that of the 386 bp of high-quality and continuous sequence examined, there was 98% nucleotide sequence identity to the *tetM* gene from *Enterococcus faecalis* (accession number M85225). The *tetM*, which imparts resistance to tetracyclines by encoding a ribosomal protection mechanism, commonly occurs in transposons Tn*916* and Tn*1545*, and is widely dispersed among various gram-positive organisms, but has only rarely

been documented in gram-negative bacteria (19, 21). The presence of *tetM in E. coli* is most likely due to genetic transfer from *Enterococcus*, a common carrier of *tetM* (8). Evidence for this possibility is provided by studies of Poyart, *et al.* (19) who demonstrated the *in vitro* transfer of Tn*916* from *E. faecalis* to *E. coli* (16).

In summary, by examining the frequency and distribution of tetracycline resistance among diverse natural *E. coli* populations present in different animal species, a picture of the selection pressures in the various host animals can be inferred. Those animal hosts that presumably had continuous exposure to tetracycline not only had a greater percentage of tetracycline resistant *E. coli* isolates, but those isolates carried a greater diversity of resistance genes. Moreover, these isolates often had more than one *tet* resistance determinant, and contained a *tet* gene previously thought not to be present in *E. coli*. This suggests that human activity provides environments that select for resistant strains and encourages the transfer of genetic information from unrelated bacterial species. Although this study examined only non-clinical *E. coli* isolates, the prevalence of *tet* resistant genes among these unrelated bacteria, and circumstantial and direct evidence of horizontal gene transfer, suggests that these same resistance determinants may also be present in animal and human pathogens.

ACKNOWLEDGEMENTS

We would like to thank LeeAnn Johnson, John Ferguson, Mary Brown, and Priscilla Dombek for help with the *E. coli* strains. We would also like to thank Dr. Lai-King Ng for the generous gift of positive control strains and cloned *tet* resistance genes.

This study was supported in part from grants from the University of Minnesota Agricultural Experiment Station, and from the Legislative Commission on Minnesota Resources,

through the Environment and Natural Resources Trust Fund and the MN Future Resource Fund (to MJS), and by funding from the Life Sciences Summer Undergraduate Research Program (to AB).

REFERENCES

1. **Aarestrup, F. M., A. M. Seyfarth, H. D. Emborg, K. Pedersen, R. S. Hendriksen, and F. Bager.** 2001. Effect of abolishment of the use of antimicrobial agents for growth promotion on occurrence of antimicrobial resistance in fecal enterococci from food animals in Denmark. Antimicrob. Agents Chemother. **45:**2054-2059.
2. **Akkina, J. E., and R. Johnson.** 1999. Antibiotic Use in U.S. Livestock Production, p. 20-. *In* Antimicrobial Resistance Issues in Animal Agriculture. USDA/APHIS/VS/ CEAHI/CEI, Washington, D.C.
3. **Altschul, S. F., T. L. Madden, A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman.** 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research **25:**3389-3402.
4. **Arzese, A. R., L. Tomasetig, and G.A. Botta.** 2000. Detection of *tetQ* and *ermF* antibiotic resistance genes in *Prevotella* and *Porphyromonas* isolates from clinical specimens and resident microbiota of humans. J. Antimicrob. Chemother. **43:**1523–1524.
5. **Blake, D. P., R. W. Humphry, K. P. Scott, K. Hillman, D. R. Fenlon, and J. C. Low.** 2003. Influence of tetracycline exposure on tetracycline resistance and the carriage of tetracycline resistance genes within commensal *Escherichia coli* populations. J. Appl. Microbiol. **94:**1087-1097.

6. **Boyce, J. M.** 2001. Consequences of inaction: importance of infection control practices. Clin. Infect. Dis.. **33:**S133-137.

7. **Chopra, I., and M. Roberts.** 2001. Tetracycline antibiotics: mode of action, applications, molecular biology, and epidemiology of bacterial resistance. Microbiol. Molec. Biol. Rev. **65:**232-260.

8. **DeFlaun, M. F., and S. B. Levy.** 1989. Genes and their varied hosts. *In:* S. B. Levy and R. V. Miller (eds.), Gene Transfer in the Environment, p. 1-32. McGraw-Hill, New York.

9. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66:**2572-2577.

10. **DuPont, H. L., and J. H. Steele.** 1987. The human health implication of the use of antimicrobial agents in animal feeds. Vet. Quart. **9:**309-320.

11. **Guillaume, G., D. Verbrugge, M. Chasseur-Libotte, W. Moens, and J. Collard.** 2003. PCR typing of tetracycline resistance determinants (Tet A-E) in *Salmonella enterica* serotype Hadar and in the microbial community of activated sludges from hospital and urban wastewater treatment facilities in Belgium. FEMS Microbiol. Ecol. **32:**77-85.

12. **Jones, M. E., J. A. Karlowsky, D. C. Draghi, C. Thornsberry, D. F. Sahm, and D. Nathwani.** 2003. Epidemiology and antibiotic susceptibility of bacteria causing skin and soft tissue infections in the USA and Europe: a guide to appropriate antimicrobial therapy. Int. J. Antimicrob. Agents. **22:**406-419.

13. **Lenart, J., A. A. Andersen, and D. D. Rockey. 2002**. Growth and development of tetracycline-resistant *Chlamydia suis*. Antimicrob. Agents Chemother. **45:**2198-2203.

14. **Levy, S. B.** 2001. Antibiotic resistance: consequences of inaction. Clin. Infect. Dis. **33:**S124-129.

15. **Levy, S. B., G. B. Fitzgerald, and A. B. Macone.** 1976. Changes in intestinal flora of farm personnel after introduction of tetracycline-supplemented feed on a farm. N. Eng. J. Med. **295:**583-588.

16. **Marshall, B., C. Tachibana, and S. B. Levy.** 1983. Frequency of tetracycline resistance determinant classes among lactose-fermenting coliforms. Antimicrob. Agents Chemother. **24:**835-840.

17. **Mathew, A. G., D. B. Arnett, P. Cullen, and P. D. Ebner.** 2003.Characterization of resistance patterns and detection of apramycin resistance genes in *Escherichia coli* isolated from swine exposed to various environmental conditions. Int. J. Food Microbiol. **89:**11-20.

18. **Ng, L. K., I. Martin, M. Alfa, and M. Mulvey.** 2001. Multiplex PCR for the detection of tetracycline resistant genes. Molec. Cell. Probes. **15:**209-215.

19. **Poyart, C., J. Celli, and P. Trieu-Cuot.** 1995. Conjugative transposition of Tn916-related elements from *Enterococcus faecalis* to *Escherichia coli* and *Pseudomonas fluorescens*. Antimicrob. Agents Chemother. **39:**500-506.

20. **Roberts, M. C.** 1996. Tetracycline resistance determinants: mechanisms of action, regulation of expression, genetic mobility, and distribution. FEMS Microbiol Rev **19:**1-24.

21. **Roberts, M. C.** 2003. Tetracycline therapy: update. Clin. Infect. Dis. **36:**462-467.

22. **Sengeløv, G., B. Halling-Sørensen, and F. M. Aarestrup.** 2003. Susceptibility of *Escherichia coli* and *Enterococcus faecium* isolated from pigs and broiler chickens to

tetracycline degradation products and distribution of tetracycline resistance determinants in *E. coli* from food animals. Vet. Microbiol. **95:**91-101.

23. **Smilack, J. D.** 1999. The tetracyclines. Mayo Clin. Proc. **74:**727-729.
24. **U.S. Congress, Office of Technology Assessment.** 1995. Impacts of Antibiotic-Resistant Bacteria, OTA-H-629, GPO 052-003-01446-7. Government Printing Office, Washington, D.C.
25. **Wegener, H. C., F. M. Aarestrup, P. Gerner-Smidt, and F. Bager.** 1999. Transfer of antibiotic resistant bacteria from animals to man. Acta Vet. Scand. Suppl. **92:**51-57.

Table1. *E. coli* isolates used in this study and their animal sources.

Animal source of *E. coli*	No. isolates used for MIC analysis	No. isolates used for multiplex PCR
Cat	46	9
Cow	158	24
Deer	74	1
Turkey	82	30
Duck	70	1
Human	176	30
Sheep	48	15
Goose	122	3
Dog	47	9
Pig	182	131
Horse	66	3
Chicken	151	66
Goat	41	3
Total	1263	325

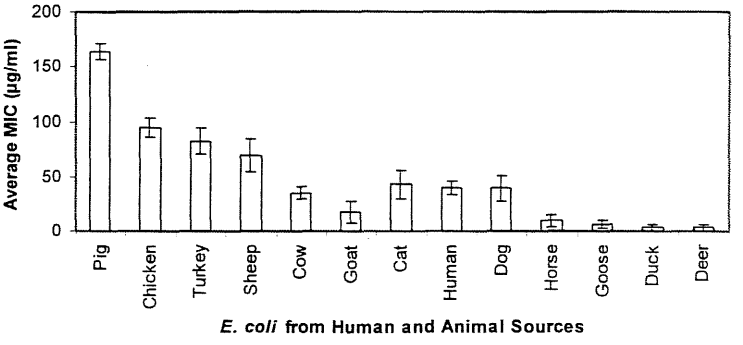
FIGURE LEGENDS

Figure 1. Average minimum inhibitory concentration (MIC) of tetracycline for *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, cats, humans, dogs, horses, geese, ducks, and deer, as determined by the plate dilution method.

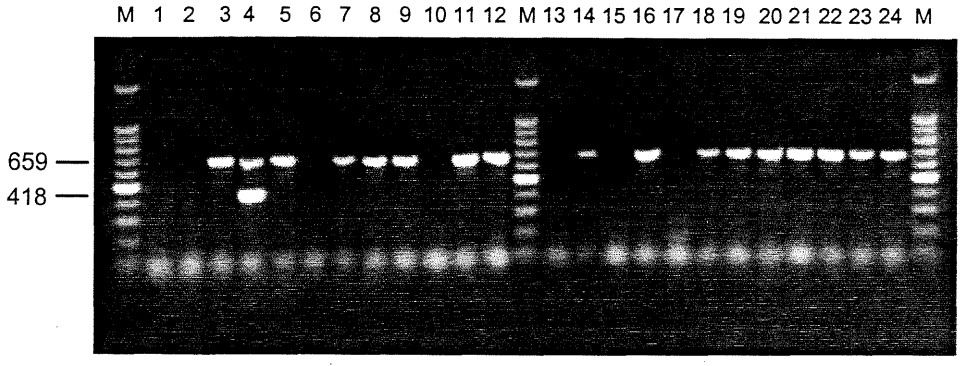
Figure 2. Representative agarose gel of PCR products from non-clinical *E. coli* using primer Group I, containing primers for *tetB*, *tetC*, and *tetD*. Lanes: 1, no template control; 2, *E. coli* H25; 3, *E. coli* H45; 4, *E. coli* H77; 5, *E. coli* P282; 6, *E. coli* P284; 7, *E. coli* P285; 8, *E. coli* P286; 9, *E. coli* P289; 10, *E. coli* P290; 11, *E. coli* P291; 12, *E. coli* P293; 13, *E. coli* P294; 14, *E. coli* P295; 15, *E. coli* P296; 16, *E. coli* P297; 17, *E. coli* P298; 18, *E. coli* P300; 19, *E. coli* P304; 20, *E. coli* P307; 21, *E. coli* P308; 22, *E. coli* P309; 23, *E. coli* P310; and 24, *E. coli* P312. *E. coli* isolate numbers beginning with P and H were isolated from pigs and horses, respectively. Molecular weight markers (100 bp ladder) are in lanes designated M. Sizes of amplicons in base pairs are indicated in the margins.

Figure 3. Frequency of *tetM, tetA, tetD, tetC,* and *tetB* in *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, cats, humans, dogs, horses, geese, ducks, and deer, as determined by colony multiplex PCR. The tetracycline genes *tetE, tetG, tetK, tetL, tetO, tetS, tetA(P), tetQ,* and *tetX* were not found among any of the 325 *E. coli* isolates tested.

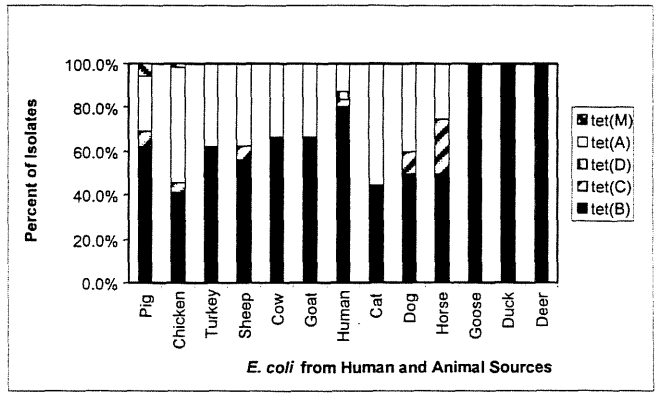
Figure 4. Percent of *E. coli* isolates obtained from pigs, chickens, turkeys, sheep, cows, goats, cats, humans, dogs, horses, geese, ducks, and deer, containing multiple tetracycline resistance genes as determined by multiplex PCR using primers for *tetA, tetB, tetC, tetD, tetE, tetG, tetK, tetL, tetM, tetO, tetS, tetA(P), tetQ,* and *tetX.*



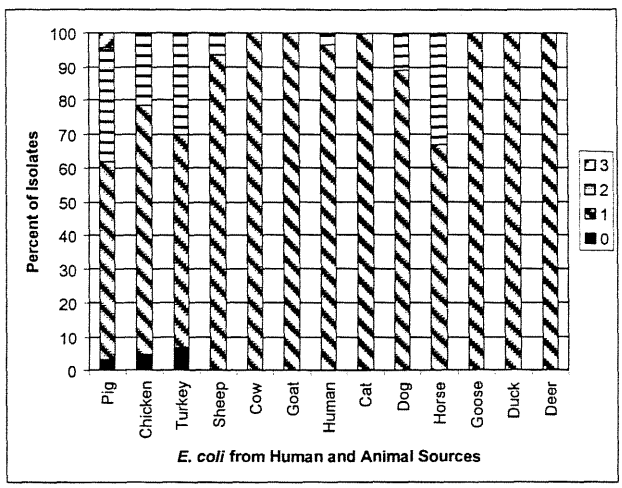
Bryan *et al.* Figure 1



Bryan *et al.* Figure2



Bryan *et al.* Figure 3



Bryan *et al.* Figure 4

Submitted Manuscript – DO NOT CITE

Duplication of Genotypes in DNA Fingerprint Libraries and a High Degree of Genetic Diversity Among Natural Populations of *Escherichia coli* from Different Animals Influences Accuracy of Determining Sources of Fecal Pollution

LeeAnn K. Johnson[†], Mary B. Brown, Ethan A. Carruthers, John A. Ferguson, Priscilla E. Dombek and Michael J. Sadowsky[*]

Department of Soil, Water, and Climate, University of Minnesota, St. Paul, MN 55108

Running Title: Diversity Among Natural *E. coli*

[*]Author to whom correspondence should be addressed:

Michael J. Sadowsky
University of Minnesota
Dept. of Soil, Water, and Climate
1991 Upper Buford Circle, 439 Borlaug Hall
St. Paul, MN 55108
Phone: 612-624-2706
FAX: 612-625-2208
Email: Sadowsky@soils.umn.edu

[†]Current Address: Minnesota Department of Agriculture, 90 West Plato Boulevard, Saint Paul, Minnesota 55107

ABSTRACT

A horizontal, fluorophore-enhanced, rep-PCR DNA fingerprinting technique (HFERP) was developed and evaluated as a means to differentiate human from animal sources of *Escherichia coli*. Box A1R primers and PCR was used to generate 2,466 rep-PCR and 1,531 HFERP DNA fingerprints from *E. coli* strains isolated from fecal material from known human and 12 animal sources: dogs, cats, horses, deer, geese, ducks, chickens, turkeys, cows, pigs, goats, and sheep. HFERP DNA fingerprinting reduced within gel grouping of DNA fingerprints and improved alignment of DNA fingerprints between gels, relative to that achieved using rep-PCR DNA fingerprinting. Jackknife analysis of the complete rep-PCR DNA fingerprint library, done using Pearson's product-moment correlation coefficient, indicated that 69.3-97.1% of animal and human isolates were assigned to the correct source groups, with a 82.2% average rate of correct classification. However, when only unique isolates were examined, isolates from a single animal having a unique DNA fingerprint, Jackknife analysis showed that 44.3–73.8% of the isolates were assigned to the correct source groups, with a 60.5% average rate of correct classification. The percentage of correctly classified isolates were about 15 and 17% greater for rep-PCR and HFERP, respectively, when analyses were done using the curve-based Pearson's product-moment correlation coefficient, rather than the band-based Jaccard algorithm. Rarefaction analysis indicated that despite the relatively large size of the known source database, genetic diversity in *E. coli* was very great, and is most likely accounting for our inability to correctly classify many environmental *E. coli* isolates. Taken together, our data indicates that duplication of genotypes within the DNA fingerprint library, database size, method of statistical analysis, and alignment of band data within and between gels impacts the accuracy of microbial source tracking methods.

INTRODUCTION

Protection of humans from pathogen contamination is dependent on the purity of waters designated for recreation, drinking, and shellfish harvesting. Bacterial pathogens have been listed as the second leading cause of impairment of rivers and streams, and the leading pollutant in estuaries (53). Restoration of impaired waters is currently being accomplished through the development of Total Maximum Daily Loads (TMDLs). Source assessment is an important component of TMDL development in which pollutants are identified and characterized by type, magnitude, and location (54). The implementation of TMDLs has provided one of the driving forces for the development of methods to distinguish between human and animal sources of fecal pollution. Sources of fecal coliform bacteria may include runoff from feedlots and manure-amended agricultural land, wildlife, inadequate septic systems, urban runoff, and sewage discharges. The ability to distinguish between human and animal sources of fecal contamination will be an important assessment tool for the evaluation of possible health risks and for the development of effective control strategies.

Both phenotypic and genotypic methods have been explored as means to study the ecology of fecal bacteria related to host specificity, and determining potential sources of fecal bacteria found in surface water (9,45,48). The mostly widely investigated bacteria for these studies have been *Escherichia coli* and *Enterococcus* sp. strains. Phenotypic approaches that have been explored to date include: fecal coliform/fecal streptococci ratios (10), antibiotic resistance profiles (15,16,21,37,61,62), coliphage typing (23,35), *Bacteroides* phage typing (39,51), and sorbitol-fermenting *Bifidobacterium* (31). In contrast, genotypic approaches including ribotyping (6,7,25,38,44), pulsed field gel electrophoresis (33,36,47), rep-PCR DNA fingerprinting (7,9,33), multilocus enzyme electrophoresis and virulence factors (12), 16S rRNA

analysis (15) and amplified fragment length polymorphism analysis (2,15), and PCR analysis of host specific 16S rDNA fragments from members of the genus *Bifidobacterium* and the *Bacteroides-Prevotella* group (3), have also been investigated as means to determine sources of fecal bacteria.. The use of these methods is based on the hypothesis that specific strains, or a strain's phenotypic or genetic attributes, are related to specific host animals (e.g. bacteria from the intestinal tracts of humans have a greater propensity to associate with humans than other animal species). This hypothesis, however, has only been tested in a limited manner with a minimum number of host animals and a minimal number of bacterial strains.

The majority of phenotypic and genotypic methodologies require the construction of known-source libraries (a host origin database) to differentiate among isolates, which is subsequently used to determine the host origin of unknown environmental isolates (48). However, in most cases, the size of the host origin databases are rather limited, consisting of 35 to about 500 isolates (1,5-7,9,15,18,19,33,34,36,37,44,46,59,60), making broader comparisons to larger populations of *E. coli* and *Enterococcus* in the environment difficult. In addition, temporal and geographic variation in bacterial genotypes within and between animal species (11,18,25,44), multiple strains within a single animal (33), and diet variation within a host animal (19) have been shown to influence the comprehensiveness of known source libraries. Moreover, while microbial source tracking studies done using phenotypic approaches and antibiotic resistance patterns have frequently used large known-source libraries, consisting of about 1000 – 6,000 isolates (5,14,16,21,61-63), many of the strains examined were isolated from the same source material or sample, and thus libraries may be biased due to the presence of multiple replications (clones) of the same bacterial genotype.

The rep-PCR DNA fingerprinting technique uses the polymerase chain reaction and primers based on highly conserved and repetitive nucleotide sequences to amplify specific portions of the microbial genome (24,32,42,50,55,57,58). When the PCR products are separated by agarose gel electrophoresis and visualized following staining with ethidium bromide, the resulting banding patterns produce a "fingerprint" unique to each strain. Bacteria having identical fingerprints are regarded as being the same strain, and those having nearly identical or similar banding patterns are regarded as being genetically related. While rep-PCR has proven to be a valuable tool to identify and track medically and environmentally important microorganisms (8,26,43,55), it has also been recently evaluated for its use as a source-tracking tool (1,7,9,29,33). The rep-PCR DNA fingerprinting technique is relatively quick, easy, and inexpensive to perform, and lends itself to high throughput applications, making it an ideal method for microbial source-tracking studies.

Initial studies done in our laboratory indicated that rep-PCR done with Box A1R primers and *E. coli* yielded a more consistent and complex DNA fingerprints than did studies done using REP primers (9). However, rep-PCR reactions done with Box, ERIC, and REP primers have all been evaluated in microbial source-tracking studies (1,7,9,33). Dombek *et al.* (9) used a minimal data set consisting of about 200 non-unique *E. coli* isolates and reported that 100% of chicken and cow isolates, and between 78-90% of human, goose, duck, pig and sheep isolates were correctly assigned to host source groups using rep-PCR DNA fingerprinting and Box AIR primers. Similarly, Carson *et al.* (7) reported that rep-PCR DNA fingerprinting done using Box A1R produced a 96.6% average rate of correct classification (ARCC) for human and non-human *E. coli* isolates, and McLellan *et al.* (33) reported a 79.3% ARCC for *E. coli* analyzed using rep-PCR done and REP primers.

While all these initial analyses indicated that the rep-PCR technique may be useful for determining animal sources of *E. coli*, these studies were done with relatively small datasets. Moreover, since rep-PCR, and most other genotypic methods, require the construction of libraries of known-source fingerprints, which is labor-intensive and time-consuming, it is very important that the fingerprint database is unbiased, and representative of the diversity of *E. coli* potentially present in animal hosts and in environmental samples. Furthermore, since the database itself can be influenced by many factors (52), including the reproducibility and alignment of DNA fingerprint patterns between and within gels, these variables need to be minimized by using highly standardized protocols and by avoiding known problem conditions.

Binary similarity coefficients are used to analyze presence/absence data (28), and simple banding data obtained from DNA fingerprints can be analyzed using binary coefficients such as Dice or Jaccard band matching algorithms. However, more quantitative algorithms, such as Pearson's product-moment correlation coefficient, can also be applied to complex DNA banding patterns, such as those found using rep-PCR. In this case, fingerprints are analyzed as densitometric curves, taking into account both peak position and height (intensity) (17).

In this study we created a large-scale, known-source rep-PCR DNA fingerprint database from 2,466 *E. coli* isolates obtained from 13 animal sources: cows, pigs, sheep, goats, turkeys, chickens, ducks, geese, deer, horses, dogs, cats, and humans. The database was assembled using a new fingerprinting method, horizontal, fluorophore-enhanced, rep-PCR (HFERP), and the usefulness of this method to differentiate human from animal sources of fecal *E. coli* was evaluated.

MATERIALS AND METHODS

Isolation of *E. coli* from known animal sources. Fecal samples, representing 13 animal source groups, were collected from wild and domesticated animals throughout Minnesota and western Wisconsin. Fresh fecal material was collected from individual animals as previously described (9) by swabbing the rectal or cloacal region using a Culturette7 swab transport system (BD Diagnostic System, Sparks, MD), or by collecting freshly voided feces with a sterile tongue depressor. Fecal samples were placed into sterile Whirl-Pak® bags (Nasco, Fort Atkinson, WI) and kept at 4°C until processed, usually within 6 hr. Fecal material was streaked onto mFC agar plates (Difco, BD Diagnostic Systems, Sparks, MD), and incubated at 44.5°C for 24 hours. Characteristic blue colonies (usually six) from mFC plates were picked and evaluated using selective and differential media as previously described (9). Isolates were used for subsequent studies if growth and color responses on all media were typical for *E. coli*. Isolates giving atypical responses with any test were further screened using API 20E test kits (bioMerieux, Inc., St. Louis, MO). Isolates yielding a "good" to "excellent" *E. coli* identification by the API 20E kit were used for DNA fingerprinting. Three *E. coli* colonies from each individual fecal sample were used for DNA fingerprinting and were stored at –80°C in 50% glycerol.

***E. coli* preparation and rep-PCR conditions**. *E. coli* isolates were streaked onto Plate Count Agar (Difco, BD Diagnostic Systems, Sparks, MD) and grown overnight at 37°C. Colonies were picked with a 1 µl sterile inoculating loop (Fisher Scientific, Pittsburgh, PA), suspended in 100 µl of distilled H_2O in 96-well microtiter plates, and 2 µl of the resulting suspension was used as template for PCR. The rep-PCR fingerprints were obtained using the Box A1R primer: 5'-

CTACGGCAAGG CGACGCTGACG-3', and PCR reactions were done as described previously (9,40,41). PCR was performed using an MJ Research PTC 100 (MJ Research, Waltham, MA) using the protocol specific for this thermocyclers and the Box A1R primer. PCR was initiated with an incubation at 95°C for 2 minutes, followed by 30 cycles, consisting of 94°C for 3 seconds, 92°C for 30 seconds, 50°C for 1 minute, and 65°C for 8 minutes (40). PCR reactions were terminated after an extension at 65°C for 8 min, and stored at 4°C. Reactions that were not used immediately for gel electrophoresis analysis were stored at -20°C.

Electrophoresis was done at 4°C for 17-18 hours at 70V with constant buffer recirculation (9,40). Gels were stained for 20 min in 0.5 μg/ml ethidium bromide prepared in 0.5× TAE buffer. Gel images were captured as TIF files using a FOTO/ Analyst Archiver electronic documentation system (Fotodyne Inc., Hartland, WI).

HFERP studies. Horizontal, fluorophore-enhanced, rep-PCR (HFERP) analyses were performed as follows: *E. coli* colonies were picked with a 1 μl sterile inoculating loop (Fisher Scientific, Pittsburgh, PA), suspended in 100 μl of 0.05 M NaOH in 96-well, low profile, PCR plates (MJ Research, Waltham, MA), heated to 95 °C for 15 min, and centrifuged at 640 RPM for 10 min in a Hermle/Labnet Z383K centrifuge. A 2 μl aliquot of the supernatant in each well was used as template for PCR using the protocol described above for rep-PCR. The primer consisted of a mixture of 0.09 μg of unlabeled Box A1R primer per μl and 0.03 μg of 6-FAM fluorescently labeled Box A1R primer per μl (Integrated DNA Technologies, Coralville, IA). The primer mixture was used at a final concentration of 0.12 μg per 25 μl PCR reaction. A 6.6 μl aliquot of a mixture of 50 μl Genescan-2500 ROX internal lane standard (Applied Biosystems, Foster City, CA) and 200 μl non-migrating loading dye (150 mg Ficoll 400 per ml, and 25 mg blue dextran per ml) was added to each 25 μl PCR reaction prior to loading the PCR reaction

into agarose gels, 12 µl of the resulting mixture was loaded per gel lane. DNA fragments were separated as described for rep-PCR, and HFERP images were captured using a Typhoon 8600 Variable Mode Imager (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) operating in the fluorescence acquisition mode using the following settings: green (532 nm) excitation laser; 610 BP 30 and 526 SP emission filters in the autolink mode with 580 nm beam splitter; normal sensitivity; 200 micron/pixel scan resolution; + 3 mm focal plane; and 800 V power.

Computer-assisted rep-PCR fingerprint analysis. Separated gel images (ROX-stained standards and HFERP banding patterns) were processed using ImageQuant image analysis software (Molecular Dynamics/Amersham Biosciences, Sunnyvale, CA) and converted to 256 gray scale TIF images. Gel images were normalized and analyzed using BioNumerics v.2.5 software (Applied-Maths, Sint-Martens-Latem, Belgium). Rep-PCR gel lanes were normalized using the 1 kb ladder from 298 bp to 5090 bp, as external reference standards, while HFERP gel lanes were normalized using the Genescan-2500 ROX internal lane standard from 287 bp to 14,057 bp. Band matching for rep-PCR DNA fingerprints was accomplished by using the following BioNumerics settings: minimum profiling 5%, gray zone 5%, minimum area 0%, and shoulder sensitivity of 5; while band matching for HFERP DNA fingerprints was done by using 3% minimum profiling, 0% gray zone, 0% minimum area, and 0 shoulder sensitivity. DNA fingerprint similarities were calculated by using either the curve-based cosine or Pearson's product-moment correlation coefficient, with 1% optimization, or the band-based Jaccard coefficient. Dendrograms were generated using the unweighted pair-group method using arithmetic means (UPGMA). The percentages of known-source isolates assigned to their correct source group were calculated by using Jackknife analysis, with maximum similarities.

RESULTS

Evaluation of isolates. Most genotype-based bacterial source tracking methods rely on the construction and use of libraries of known-source fingerprints. Library construction is time consuming and expensive. It is often assumed, however, that isolates from sewage or fecal materials that grow on selective and differential media are bona fide *E. coli* or *Enterococcus* sp. strains. Of the 2,672 *E. coli* strains obtained from known human and animal sources using an array of selective and differential plating media, 219 isolates gave at least one atypical result when examined by routine biochemical screening tests. The biochemical characteristics of these isolates were examined further by using the API 20E system. Results of this analysis indicated the majority of these isolates, 167, were bona fide *E. coli,* while the remainder, 52, could not be confirmed as this bacterium. The latter group was not used in rep-PCR analysis or included in the DNA fingerprint database. This result indicates that it is important to confirm the identity of bacteria used in source tracking libraries, rather than relying solely on growth or reactions on selective/differential plate media.

Influence of duplicate *E. coli* strains on classification of known source library. While we previously described the use of rep-PCR DNA fingerprinting to determine sources of fecal bacteria (9), our initial studies, and many others, used libraries consisting of a relatively small number of samples, some of which were obtained from the same individual animal. Since results from several studies suggest that *E. coli* is genetically diverse and clonal in origin, and that this may influence the usefulness of this bacterium for source tracking studies (11), we evaluated this technology using a large library of *E. coli* obtained from 13 human and animal sources collected throughout Minnesota and Western Wisconsin (Table 1).

2,466 high-quality rep-PCR DNA fingerprints were generated using the Box A1R primer and template DNA from *E. coli* strains obtained from the 13 human and animal sources (Table 1). Jackknife analysis performed on the 2,466 DNA fingerprints from the entire known-source rep-PCR DNA fingerprint database, using Pearson's product-moment correlation coefficient, indicated that 69-97% of animal and human *E. coli* isolates were assigned into correct source groups (Table 2). This corresponds to an 82.2% average rate of correct classification for the 2,466 rep-PCR DNA fingerprints.

However, since identical DNA fingerprints from *E. coli* strains obtained from the same individual most likely represent isolates of clonal origin, and can artificially bias subsequent analyses of strain groupings (e.g. increase the average rate of correct classification) and the fidelity of the database, we eliminated duplicate DNA fingerprints originating from *E. coli* strains obtained from the same individual animal or human. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%.

Of the 2,466 DNA fingerprints analyzed, 1,535 (62%) remained in the "unique" DNA fingerprint library (Table 1). The influence of duplicate DNA fingerprints on the correct classification of library strains is shown in Table 2. When the 1,535 DNA fingerprints from the unique *E. coli* isolates were examined, Jackknife analyses indicated that only 44-74% of the isolates were assigned to the correct source group (Table 2). The average rate of correct classification for these 1,535 unique rep-PCR DNA fingerprints was 60.5%. Taken together, these results indicate that inclusion of duplicate DNA fingerprints in the library can artificially influence strain groupings and increase percentages of strains correctly assigned to source groups.

Influence of library size on usefulness of DNA fingerprint libraries. We also determined whether *E. coli* isolates obtained in this study were sufficient to capture the genetic diversity present within the *E. coli* populations sampled. *E. coli* isolates with rep-PCR DNA fingerprint similarities of 90% or greater (based on cosine coefficient, 1% optimization and UPGMA) were assigned to the same genotype. By this definition, 657 genotypes were identified from the 1,535 unique *E. coli* isolates in the known-source database. The isolates were randomized, and a rarefaction curve was constructed by summing the number of genotypes that accumulated with the successive addition of isolates. Results in Figure 1 show that despite a library size of 1,535 DNA fingerprints, genetic diversity has not been saturated. This is evidenced by the apparent first order relationship between isolate numbers (sampling effort) and accumulation of new genotypes. Moreover, 58.75% of the genotypes from isolated strains, across all animal groups, occurred only once in the database and a limited number occurred multiple times (Figure 2). Consequently, such a library is most likely not optimal for determining sources of unknown fecal bacteria from water, and if used would result in a large proportion of environmental strains not being classified to correct source groups.

HFERP DNA fingerprinting. In our studies we noted that cluster analysis of rep-PCR DNA fingerprint data often produced groupings that were more related to the gels from which they originated, than the host animal from which they were isolated. We hypothesized that within-gel clustering of DNA fingerprints was in part due to intrinsic gel-to-gel variation, differential DNA migration in repeated runs of the same and different PCR samples, and the inability to correct for heat and buffer-induced gel distortion across and between single and multiple gels. Since DNA fingerprint libraries are assembled from many different gels, this could have a major impact on the fidelity of DNA fingerprint libraries and their subsequent use for

tracking sources of unknown fecal bacteria. To overcome these major limitations, we developed and evaluated the use of a horizontal, fluorophore-enhanced, rep-PCR (HFERP) technique as a means to differentiate human from animal sources of fecal bacteria. In this method, alignment, correction, and normalization of fluorescently-labeled, rep-PCR DNA fingerprint bands within and between gels is facilitated by the use of internal ROX-labeled molecular weight markers that are present in each lane. The technique is similar to that previously described for use with a DNA sequencer (56), but instead uses a standard horizontal agarose gel and a dual-wavelength scanner. An example of an unseparated HFERP gel displaying the ROX-labeled internal lane standard and 6-FAM-labeled Box A1R DNA fingerprints is shown in Figure 3A, and the separated gel images are shown in Figures 3B and 3C. Typically, and with our *E. coli* strains, 12 to 20 DNA bands per strain were revealed using the HFERP technique.

To test whether HFERP reduced within-gel groupings of DNA fingerprints, we analyzed DNA fingerprints from 40 *E. coli* strains obtained from dogs on 2 different gels using Pearson's product-moment coefficient. Results of these studies indicated that rep-PCR DNA fingerprints from strains run on the same gel were, on average, 50% (range 29 – 57%) more likely to be grouped together as the same strains analyzed by using the HFERP technique (data not shown). This indicates that HFERP method considerably reduces within gel grouping of DNA fingerprints. In addition, the HFERP method reduced alignment difficulties due to within- and between-gel variation in band migration found with rep-PCR gels (Figure 4).

The repeatability of the rep-PCR and HFERP DNA fingerprinting methods was examined by fingerprinting a single, reference, control *E. coli* strain (pig isolate number 294) that was included on each gel. DNA fingerprints from 29 and 41 repetitions of *E. coli* control pig strain 294, each from a separate gel, were generated by using the rep-PCR and HFERP methods,

respectively. When analyzed using the curve-based Pearson's correlation coefficient, the rep-PCR DNA fingerprints had an average similarity of 88%, whereas the HFERP-derived DNA fingerprints had an average similarity of 92%. Taken together, our results indicate that the HFERP technique has the ability to improve alignment of bands and the repeatability of banding patterns across different gels that are used to create DNA fingerprint libraries. This technology will have application to other DNA fingerprinting methods that rely on the use of PCR primers.

Analysis of HFERP-generated DNA fingerprint library. Of the 1,535 previously selected unique *E. coli* isolates from animals and humans (Table 1), 1,531 were subjected to HFERP DNA fingerprinting using a combination of fluorescently labeled and unlabeled Box A1R PCR primers. Jackknife analyses of HFERP gels done using the curve-based Pearson's correlation coefficient indicated that 38-73% of the isolates were assigned to the correct source group using this technique (Table 3). For the curve-based analysis, the HFERP technique had the lowest percent of correctly classified strain in cases where the numbers of analyzed fingerprints were relatively small (for sheep, horses, and goats). The average rate of correct classification for the unique HFERP-generated DNA fingerprints was 59.9%.

In contrast, Jacknife analyses of HFERP-generated DNA fingerprints done using the band-based Jaccard analysis showed that only 8-56% of the *E. coli* isolates were assigned to the correct source group, with a 43.0% average rate of correct classification. This indicates that for this type of data, the Pearson's product-moment correlation coefficient was superior to Jaccard's band matching algorithm for assigning known isolates to the correct source groups. Interestingly, results in Table 3 also show that despite problems associated with within- and between-gel variation, within-gel grouping of isolates, and repeatability issues, Jacknife analysis of rep-PCR DNA fingerprints, analyzed using Pearson's correlation coefficient, indicated that 48-74% of the

isolates were assigned to the correct source group, a 60.9% average rate of correct classification. Analysis of rep-PCR DNA fingerprint data using the Jaccard band-based method was not as useful in separating *E. coli* isolates into their correct source group as was the curve-based method.

Groupings of fingerprint data. In some instances, it may be sufficient to identify unknown watershed *E. coli* isolates to larger groupings, rather than to individual animal types. To determine if the HFERP-generated DNA fingerprint data from our library of unique *E. coli* isolates grouped well into larger categories, we assembled DNA fingerprints from pets (dogs and cats), domesticated animals (chickens, cows, goats, horses, pigs, sheep, and turkeys), wild-life (deer, ducks, and geese), and humans, and used Jacknife analysis to assess the percent of correctly classified strains. Results in Table 4 show that the HFERP DNA fingerprints, analyzed using Pearson's product-moment correlation coefficient, correctly classified 83.2, 53.8, 71.4, and 59.1% of the isolates into the domesticated, human, wildlife, and pet categories, respectively. The average rate of correct classification for these groups was 74.3%. However, when DNA fingerprints were analyzed using Jaccard's coefficient, the average rate of correct classification was 66.2%. As before, the least precision was found in categories having the smallest number of fingerprints, pets and humans, suggesting that there is an apparent relationship between the number of fingerprints analyzed and the percentage of correctly classified isolates.

In microbial source tracking studies it may often be useful to determine if unknown isolates belong to either animal or human source groups, rather than to more specific categories. Results in Table 5 show that about 94% and 54% of *E. coli* from animals and humans, respectively, were assigned to the correct source groups using HFERP-generated DNA fingerprints and Pearson's correlation coefficient. The average rate of correct classification was

88.2 and 86.1% for analyses done using Pearson's and Jaccard's algorithms, respectively. The lower percentage of correctly classified human isolates may, in part, be due to the smaller size of fingerprints analyzed for this category.

DISCUSSION

The purpose of this study was to develop a large-scale known-source *E. coli* rep-PCR DNA fingerprint library that can be used to identify the source of *E. coli* bacteria isolated from impacted watersheds in Minnesota. The known-source DNA fingerprint library included 2,466 *E. coli* isolates obtained from nearly 1,000 individuals belonging to thirteen source groups: cows, pigs, sheep, goats, turkeys, chickens, ducks, geese, deer, horses, dogs, cats, and humans. Earlier work in our laboratory, examining a much smaller subset of *E. coli* isolates, indicated that the rep-PCR technique had the necessary sensitivity and resolving power to differentiate between strains of fecal coliform bacteria originating from different human and animal sources (9). However, in our earlier studies, and those done by several researchers, the size of the host origin databases were limited, consisting of 35 to about 500 isolates (1,5-7,9,15,18,19,33,34,36,37,46, 59,60). The relatively small size of these libraries do not take into account the tremendous genetic diversity present in *E. coli* (11,20,33) and enterococci, and makes broader comparisons to larger populations of these organisms in the environment difficult.

In our studies reported here we show that increasing the size of the known source library to 2,466 isolates did not necessarily lead to an increase in the ability to correctly assign strains to the correct source group. In fact, the average rate of correct classification decreased 4.2% using the larger library reported here, relative to what was seen using a smaller library in our previous

studies (9). This may in part be due to the uncovering of increased genetic diversity among isolates, increased accumulation of errors due to gel-to-gel variation, or the presence of duplicate genotypes (DNA fingerprints) from the same individual within our original library. Reduction in the percentage of known-source *E. coli* isolates that were correctly classified was especially apparent when our unique library of 1535 *E. coli* isolates was examined. Unique DNA fingerprints were defined as DNA fingerprints from *E. coli* isolates obtained from a single host animal whose similarity coefficients were less than 90%. Since DNA fingerprints from *E. coli* strains obtained from the same individual represent isolates of clonal origin, these duplicate strains (or fingerprints) can artificially bias the average rate of correct classification and the fidelity of the database. Results in Table 2 show that there was a 21.7% reduction in the average rate of correct classification by using the unique DNA fingerprint library, relative to that seen with the complete library. Moreover, the 60.5% average rate of correct classification found with the unique library was less than we previously reported using a smaller library of *E. coli* strains containing duplicate DNA fingerprints from the same individual animal (9), and less than reported by other authors using libraries containing duplicate entries (7,33). More importantly, our results show that failure to remove identical fingerprints from analyses resulted in an overestimation of the ability of the database to assign isolates to their correct source group, perhaps in part due to the clonal composition of *E. coli* populations (11,20,33). Taken together, our results indicate that inclusion of duplicate DNA fingerprints in the library can artificially influence strain groupings and incorrectly increases percentages of strains correctly assigned to source groups.

Results presented here also show that despite our use of an increased number of individuals in our library for DNA fingerprinting, we still failed to capture the genetic diversity

present in *E. coli*. Populations of *E. coli* have been shown to be very diverse (49) and this is evidenced by rarefaction analysis results shown in Figure 1. Despite having a known source library or over 1500 unique isolates, the number of genotypes uncovered by DNA fingerprinting continued to increase at a constant rate. Moreover, across all animal hosts, the majority of these fingerprints occurred only once. For a library to be truly representative it needs to be large enough to capture all the unknowns present in an environmental sample, otherwise strain assignment will most likely be incorrect, or a large number of isolates will be characterized as being unknowns or cosmopolitan. Since the rarefaction curve in Figure 1 has not become asymptotic, our data cannot be used to predict the ultimate size that this library needs to be. However, data presented in Figure 2 indicates that with our current library size, each new isolate added to the library only has a greater than 50% chance of being new. It has been suggested that a library size of 20,000 to 40,000 isolates may be needed to capture all the genetic diversity present in *E. coli* (Mansour Samadpour, personal communication). One suggested strategy to avoid this under-representation problem in large regional or national libraries, is to develop moderate sized libraries for a highly confined geographical region, wherein isolates are only obtained from the animals in the study area. In this way only animals pertinent to the study site, and those likely to have an impact on the targeted watershed, need to be examined in detail

We also report here the development and evaluation of HFERP as an alternative to the standard rep-PCR method. HFERP was shown to reduce gel-to-gel variability and illegitimate clustering of fingerprints within gels. HFERP utilizes a fluorescent-labeled rep-primer (6-FAM-labeled Box A1R) in the PCR reaction, and a size standard set labeled with a second fluorophore (ROX) in each gel lane. Previously, Versalovic, *et al.* (56) and Rademaker, *et al.* (40) reported on the use of fluorophore-enhanced rep-PCR (FERP), whereby polyacrylamide gel

electrophoresis and automated DNA sequencers were used to separate and detect bands generated by the FERP protocol. While the more automated method presented by these authors has some advantages, the increased cost of analyses and the limited dynamic range of fragment size separation on sequencing gels did not make this technique useful in our applications. Thus, in our HFERP studies we separated PCR products using horizontal agarose gel electrophoresis in the same manner as the standard rep-PCR protocol. This allows for the separation of a large range of DNA band sizes using more standard laboratory equipment. Moreover, the presence of a size standard in each lane of the HFERP gel allows for the very accurate normalization of bands within and between gels, which corrects for band-migration variation that occurs during electrophoresis. The result of the normalization process is that fingerprint patterns from different gels can be accurately compared. It should be noted, however, that the intensity of HFERP bands are more variable than those generated by rep-PCR, and that some of the gains achieved by more precise alignment of bands may be offset by more variation in band intensity. We found that this variation in intensity can be overcome by the careful mixing of all reagents in the PCR master mix and greater pipetting precision when loading gels (data not presented). Further improvements to increasing the intensity of HFERP-generated DNA fingerprints may also be obtained by varying the ratio of labeled to unlabeled primer and the final concentration of the primer mixture in PCR reactions. Nevertheless, our results clearly show that HFERP-derived DNA fingerprint bands are more precisely aligned than the rep-PCR bands. In addition, we show that HFERP DNA fingerprints generated by our method reduce within gel groupings of fingerprints, which can have profound ramifications for the assembly of libraries and the analysis of unknown environmental isolates.

A variety of similarity measures exist. Binary similarity coefficients are mostly used to analyze presence/absence data (28) and band- matching data obtained from DNA fingerprints can be analyzed using binary coefficients. However, quantitative similarity coefficients require a measure of relative abundance (27). Quantitative coefficients can be applied to DNA fingerprints when the fingerprints are analyzed as densitometric curves that take into account both peak position and intensity (peak height). Häne, *et al.* (17) demonstrated that for complex DNA fingerprints, such as those produced with the techniques we used here, a curve-based method such as Pearson's product-moment correlation coefficient more reliably identified similar or identical DNA fingerprints than band matching formulas, such as simple matching, Dice, or Jaccard. Results presented here confirm that the curve-based Pearson's product-moment correlation coefficient was superior to the band-based Jaccard algorithm is correctly assigning isolates to the correct source group. Similarly, Louws and co-workers (30) reported that curve-based statistical methods worked best for analysis of complex banding profiles generated by rep-PCR, since comparison of curve data is less dependent on DNA concentration in loaded samples and is relatively insensitive to background differences in gels. More recently, Albert *et al.* (1) performed a statistical evaluation of rep-PCR DNA fingerprint data and reported that *k*-nearest neighbor's classification was similar to Person's product-moment coefficient in its ability to correctly classify fingerprints of 584 *E. coli* isolates.

Further refinements to the Jackknife analysis, including the pooling of source groups into domesticated, human, and wild-life categories, were found to improve the ability to correctly classify isolate to their respective source groups. Over 83, 53, and 71% of domesticated animals, humans, and wild-life animals, respectively, were correctly classified using this approach with the unique DNA fingerprint library analyzed by HFERP. When all animal sources were pooled

into one group, the overall correct classification rate for humans and animals by HFERP was improved to about 94 and 54%, respectively, when analyzed using the curve-based Pearson's correlation coefficient. Accordingly, these results indicated that (1) broader classifications of source groups should be used when appropriate, or (2) a targeted subset of the DNA fingerprint database should be used to more precisely determine sources of fecal pollutants in watersheds where specific source groups are known to be present. The pooling of source groups into a more limited number of categories has previously been shown to increase the average rate of correct classification following discriminant analysis of antibiotic resistance (16,21,62), ribotype (6,7), and rep-PCR DNA fingerprint analyses (7).

In summary, our results suggest that HFERP-generated Box A1R DNA fingerprints of *E. coli* are useful to differentiate between different *E. coli* subtypes of human and animal origin and that this method reduces within gel groupings of DNA fingerprints, and ensures more proper alignment and normalization of fingerprint data. However, our results further indicate that other important issues must also be resolved to more fully understand the potential applications and limitations of this and other library-based microbial source tracking methodologies. Among these are questions concerning the inclusion of identical DNA fingerprints from the same animal in the library, the number of fingerprints that must be included in an *E. coli* known source library to adequately capture the diversity of *E. coli* genotypes that exist among potential host animals, and ultimately, whether *E. coli* exhibits a sufficient level of host specificity to allow unambiguous assignment of unknown environmental *E. coli* to specific host animals.

ACKNOWLEDGEMENTS

This work was supported, in part, by funding from the Legislative Commission on Minnesota Resources through the Environment and Natural Resources Trust Fund and the MN Future Resource Fund, the Metropolitan Council Environmental Services, and The University of Minnesota Agricultural Experiment Station (to M.J.S.).

We would like to thank Priscilla Dombek for DNA fingerprinting assistance, Linda Kinkel and Anita Davelos for help with rarefaction analysis, and Vivek Kapur, Todd Markowski and Bruce Witthun for help with laser image analysis.

REFERENCES

1. **Albert, J. M., J. Munkata-Marr, L. Tenorio, and R. L. Siegrist.** 2003. Statistical evaluation of bacterial source tracking data obtained by rep-PCR DNA fingerprinting of *Escherichia coli.* Environ. Sci. technol. **37:**4554-4560.
2. **Arnold, C., L. Metherell, G. Willshaw, A. Maggs, and J. Stanley.** 1999. Predictive fluorescent amplified-fragment length polymorphism analysis of *Escherichia coli*: high resolution typing method with phylogenetic significance. J. Clin. Microbiol. **37:**1274-1279.
3. **Bernhard, A. E., and K. G. Field.** 2000. Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. Appl. Environ. Microbiol. **66:**1587-1594.

4. **Boualam, M., L. Mathieu, S. Fass, J. Cavard, and D. Gatel.** 2002. Relationship between coliform culturability and organic matter in low nutritive waters. Water Res **36:** 2618-2626.

5. **Burnes, B. S.** 2003. Antibiotic resistance analysis of fecal coliforms to determine fecal pollution sources in a mixed-use watershed. Environ. Monit. Assess. **85:**87-98.

6. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and A. Asfaw.** 2001. Identification of fecal *Escherichia coli* from humans and animals by ribotyping. Appl. Environ. Microbiol. **67:**1503-1507.

7. **Carson, C. A., B. L. Shear, M. R. Ellersieck, and J. D. Schnell.** 2003. Comparison of ribotyping and repetitive extragenic palindromic-PCR for identification of fecal *Escherichia coli* from humans and animals. Appl Environ Microbiol **69:**1836-1839.

8. **de Bruijn, F. J.** 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. Appl. Environ. Microbiol. **58:**2180-2187.

9. **Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky.** 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Appl. Environ. Microbiol. **66:**2572-2577.

10. **Geldreich, E. E., and B. A. Kenner.** 1969. Concepts of fecal streptococci in stream pollution. J. Wat. Poll. Cont. Fed. Suppl. **41:**R336-R332.

11. **Gordon, D. M.** 2001. Geographical structure and host specificity in bacteria and the implications for tracing the source of coliform contamination. Microbiol. **147:**1079-1085.

12. **Gordon, D. M., S. Bauer, and J. R Johnson.** 2002. The genetic structure of *Escherichia coli* populations in primary and secondary habitats. Microbiol. **148:**1513-1522.

13. **Grabow, W. O. K., T. E. Neubrech, C. S. Holtzhausen, and J. Jofre.** 1995. *Bacteroides fragilis* and *Escherichia coli* bacteriophages: excretion by humans and animals. Wat. Sci. Tech. **31:**223-230.

14. **Graves, A. K., A. T. Hagedorn, M. Mahal, A. M. Booth, and R. B. Reneau Jr.** 2002. Antibiotic resistance profiles to determine sources of fecal contamination in a rural Virginia watershed. J. Environ. Qual. **31:**1300-1308.

15. **Guan, S., R. Xu, S. Chen, J. Odumeru, and C. Gyles.** 2002. Development of a procedure for discriminating among *Escherichia coli* isolates from animal and human sources. Appl. Environ. Microbiol. **68:**2690-2698.

16. **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.

17. **Häne, B.G., K. Jäger, and H. Drexler.** 1993. The Pearson product-moment correlation coefficient is better suited for identification of DNA fingerprinting profiles than band matching algorithms. Electrophor. **14:**967-972.

18. **Hartel, P. G., J. D. Summer, J. L. Hill, J. Collins, J. A. Entry, W. and I. Segars.** 2002. Geographic variability of *Escherichia coli* ribotypes from animals in Idaho and Georgia. J. of Environ. Qual. **31:**1273-1278.

19. **Hartel, P. G., J. D. Summer, and W. I. Segars.** 2003. Deer diet affects ribotypo diversity of *Escherichia coli* for bacterial source tracking. Water Res. **37:**3263-3268.

20. **Hartl, D. L., and D. E. Dykhuizen.** 1984. The population genetics of *Escherichia coli.* Annu. Rev. Genet. **18:**31-68.

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

22. **Havelaar, A. H., W. M. Pot-Hogeboom, K. Furuse, R. Pot, and M. P. Horman.** 1990. F-specific RNA bacteriophages and sensitive host strains in faeces and wastewater of human and animal origin. J. Appl. Bacteriol. **69:**30-37.

23. **Hsu, F-C., Y.-S. Carol Shieh, J. van Duin, M. J. Beekwilder, and M. D. Sobsey.** 1995. Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. Appl. and Environ. Microbiol. **61:**3960-3966.

24. **Hulton, C. S. J., C. F. Higgins, and P. M. Sharp.** 1991. ERIC sequences: a novel family of repetitive elements in the genomes of *Escherichia coli, Salmonella typhimurium* and other enterobacteria. Mol. Microbiol. **5:**825-834.

25. **Jenkins, M. B., P. G. Hartel, T. J. Olexa, and J. A. Stuedemann.** 2003. Putative temporal variability of *Escherichia coli* ribotypes from yearling steers. J Environ Qual. **32:**305-309.

26. **Judd, A. K., M. Schneider, M. J. Sadowsky, and F. J. de Bruijn.** 1993. Use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. Appl. Environ. Microbiol. **59:**1702-1708.

27. **Krebs, Charles J**. 1999. Ecological methodology. Benjamin/Cummings, Menlo Park, CA.

28. **Legendre, P., and L. Legendre.** 1998. Numerical Ecology, 2nd English edition. Elsevier Science, Amsterdam, The Netherlands.

29. **Lipman, J. A., A. de Nijs, T. J. G. M. Lam, and W. Gaastra.** 1995. Identification of *Escherichia coli* strain from cows with clinical mastitis by serotyping and DNA polymorphism patterns with REP and ERIC primers. Vet. Microbiol. **43:**13-19.

30. **Louws, F. J, J. L. W. Rademaker, and F. J. de Bruijn.** 1999. The three Ds of PCR-based genomic analysis of phytobacteria: diversity, detection, and disease diagnosis. Ann. Rev. Phytopath. **37:**81-125.

31. **Mara, D. D., and J. I. Oragui.** 1983. Sorbitol-fermenting bifidobacteria as specific indicators of human fecal pollution. J. Appl. Bacteriol. **55:**349-357.

32. **Martin, B., O. Humbert, M. Camara, E. Guenzi, J. Walker, T. Mitchell, P. Andrew, M. Prudhomme, G. Alloing, R. Hakenbeck, D. A. Morrison, G. J. Boulnois, and J.-P. Claverys.** 1992. A highly conserved repeated DNA element located in the chromosome of *Streptococcus pneumoniae*. Nucl. Acids Res. **20:**3479-3483.

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

34. **Nebra, Y., X. Bonjoch, A. R. Blanch.** 2003. Use of *Bifidobacterium dentium* as an indicator of the origin of fecal water pollution. Appl. Environ. Microbiol. **69:**2651-2003.

35. **Osawa, S., K. Furuse, and I. Wantanabe.** 1981. Distribution of ribonucleic acid coliphages in animals. Appl. Environ. Microbiol. **41:**164-168.

36. **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.

37. **Parveen, S., R. L. Murphree, L. Edmiston, C. W. Kasper, K. M. Portier, and M. L. Tamplin.** 1997. Association of multiple-antibiotic-resistance profiles with point and nonpoint sources of *Escherichia coli* in Apalachicola bay. Appl. Environ. Microbiol. **63:**2607-2612.

38. **Parveen, S., K. M. Portier, K. Robinson, L. Edmiston, and M. L. Tamplin.** 1999. Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and nonhuman sources of fecal pollution. Appl. Environ. Microbiol. **65:**3142-3147.

39. **Puig, A., J. Jofre, and R. Araujo.** 1997. Bacteriophages infecting various *Bacteroides fragilis* strains differ in their capacity to distinguish human from animal faecal pollution. Wat. Sci Tech. **35:**359-362.

40. **Rademaker, J. L. W., F. J. Louws, and F. J. deBruijn.** 1998. Characterization of the diversity of ecologically important microbes by rep-PCR genomic fingerprinting, suppl. 3, chapter 3.4.3, p.1-26. *In* A. D. L. Akkermans, J. D. van Elsas and F. J. de Bruijn (ed.), Molecular microbial ecology manual. Luwer Academic Publishers, Dordrecht, The Netherlands.

41. **Rademaker, J. L. W., F. J. Louws, U. Rossbach, P. Vinuesa, and F. J. deBruijn.** 1999. Computer-assisted pattern analysis of molecular fingerprints and database construction, suppl. 4, chapter 7.1.3, p.1-33. *In* A. D. L. Akkermans, J. D. van Elsas and F. J. de Bruijn (ed.), Molecular microbial ecology manual. Kluwer Academic Publishers, Dordrecht, The Netherlands.

42. **Sadowsky, M. J., and H.-G. Hur**. 1998. Use of endogenous repeated sequences to fingerprint bacterial genomic DNA, 399-413. *In* J.R. Lupski, G. Weinstock, and F. J. de Bruijn (ed.), Bacterial genomes: structure and analysis. Chapman & Hall, New York, NY.

43. **Sadowsky, M. J., L. L. Kinkel, J. H. Bowers, and J. L. Schottel.** 1996. Use of repetitive intergenic DNA sequences to classify pathogenic and disease-suppressive *Streptomyces* strains. Appl. Environ. Microbiol. **62:**3489-3493.

44. **Scott, T. M., S. Parveen, K. M. Portier, J. B. Rose, M. L. Tamplin, S. R. Farrah, A. Koo, and J. Lukasik.** 2003. Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. Appl. Environ. Microbiol. **69:**1089-1092.

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

46. **Seurinck, S., W. Verstraete, and S. D. Siciliano.** 2003. Use of 16S-23S rRNA intergenic spacer region PCR and repetitive extragenic palindromic PCR analyses of *Escherichia coli* isolates to identify nonpoint fecal sources. Appl. Environ. Microbiol. **69:**4942-4950.

47. **Simmons, G. M., Jr., S. A. Herbein, and C. M. James.** 1995. Managing nonpoint fecal coliform sources to tidal inlets. Water Res. Update. **100:**64-74.

48. **Simpson, J. M., J. W. Santo Domingo and D. J. Reasoner.** 2003. Microbial source tracking: state of the science. Environ. Sci. Tech. **36:**5280-5288.

49. **Souza, V., M. Rocha, A. Valera, and L. Eguiarte.** 1999. Genetic structure of natural populations of *Escherichia coli* in wild hosts on different continents. **65:**3373-3385.

50. **Stern, M. J., G. F.-L. Ames, N. H. Smith, E. C. Robinson, and C. F. Higgins.** 1984. Repetitive extragenic palindromic sequences: a major component of the bacterial genome. Cell **37:**1015-1026.

51. **Tartera, C. and J. Jofre.** 1987. Bacteriophages active against *Bacteroides fragilis* in sewage-polluted waters. Appl. Environ. Microbiol. **53:**1632-1637.

52. **Tyler, K. D., G. Wang, S. D. Tyler, and W. M. Johnson.** 1997. Factors affecting reliability and reproducibility of amplification-based DNA fingerprinting of representative bacterial pathogens. J. Clin. Microbiol. **35:**339-346.

53. **U.S. Environmental Protection Agency.** 2000. National Water Quality Inventory: 1998 Report to Congress. EPA-841-R-00-001. Office of Water, Washington, D.C.

54. **U.S. Environmental Protection Agency.** 2001. Protocol for Developing Pathogen TMDLs. EPA 841-R-00-002. Office of Water, Washington, DC.

55. **Versalovic, J., F. J. de Bruijn, and J. R. Lupski.** 1998. Repetitive sequence-based PCR (rep-PCR) DNA fingerprinting of bacterial genomes, p. 437-454. *In* F. J. de Bruijn, J. R. Lupski, and G. M. Weinstock (eds.), Bacterial genomes: physical structure and analysis. Chapman & Hall, New York, NY.

56. **Versalovic, J., V. Kapur, T. Koeuth, G. H. Mazurek, T. S. Whittam, J. M. Musser, and J. R. Lupski.** 1995. DNA fingerprinting of pathogenic bacteria by fluorophore-enhanced repetitive sequence-based polymerase chain reaction. Arch. Pathol. Lab. Med. **119:**23-29.

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

58. **Versalovic, J., M. Schneider, F. J. de Bruijn, and J. R. Lupski.** 1994. Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. Methods Mol. Cell. Biol. **5:**25-40.

59. **Wheeler, A. L., P. G. Hartel, D. G. Godfrey, J. L. Hill, and W. I. Segars.** 2002. Potential of *Enterococcus faecalis* as a human fecal indicator for microbial source tracking. J. Environ. Qual. **31:**1286-1293.

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

61. **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.

62. **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. L. Torrijos, and M. E. Zimmerman.** 1999. Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. Appl. Environ. Microbiol. **65:**3483-3486.

63. **Wiggins, B.A., P. W. Cash, W. S. Creamer, S. E. Dart, P. P. Garcia, T. M. Gerecke, J. Han, B. L. Henry, K. B. Hoover, E. L. Johnson, K. C. Jones, J. G. McCarthy, J. A. McDonough, S. A. Mercer, M. J. Noto, H. Park, M. S. Phillips, S. M. Purner, B. M. Smith, E. N. Stevens, and A. K. Varner.** 2003. Use of antibiotic resistance analysis for representativeness testing of multiwatershed libraries. Appl. Environ. Microbiol. **69:**3399-3405.

Table 1. Animal source groups and rep-PCR DNA fingerprints generated from *E. coli* isolates.

Animal Source Group	Individuals Sampled	Total Fingerprints	Unique Fingerprints[a]
Cat	37	108	48
Chicken	86	231	144
Cow	115	299	191
Deer	64	179	96
Dog	71	196	106
Duck	42	122	81
Goat	36	104	42
Goose	73	200	135
Horse	44	114	79
Human	197	307	211
Pig	111	303	215
Sheep	37	101	61
Turkey	69	202	126
Total	**982**	**2,466**	**1,535**

[a]Identical *E. coli* genotypes from each individual animal were removed.

Table 2. Percentage of known-source rep-PCR DNA fingerprints assigned to the correct source group by Jackknife analysis[a].

Animal Source	All Fingerprints (n=2,466)	Unique Fingerprints (n=1,535)
	Percent Correctly Classified Isolates	
Pets[b]	91.8 (279)[d]	61.7 (95)
Chicken	81.4 (188)	59.7 (86)
Cow	79.6 (238)	55.0 (105)
Deer	85.5 (145)	55.2 (53)
Waterfowl[c]	81.4 (262)	66.2 (143)
Goat	97.1 (101)	66.7 (28)
Horse	69.3 (79)	44.3 (35)
Human	78.3 (240)	59.2 (125)
Pig	77.9 (236)	63.7 (137)
Sheep	79.0 (80)	47.5 (29)
Turkey	88.6 (179)	73.8 (93)
Overall	82.2 (2,027)	60.5 (929)

[a]Done using Pearson's product moment correlation coefficient with 1% optimization and maximum similarities.
[b]Pet group consists of cats and dogs
[c]Waterfowl group consists of ducks and geese.
[d]Values in parentheses are number of isolates correctly classified.

Table 3. Unique *E. coli* isolates correctly classified into source groups by rep-PCR and HFERP DNA fingerprinting methods.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]			
		rep-PCR		HFERP	
		Pearson	Jaccard	Pearson	Jaccard
Pets[b]	154	61.7 (95)[d]	45.5 (70)	59.1 (91)	44.8 (69)
Chicken	144	59.7 (86)	38.9 (56)	63.2 (91)	31.9 (46)
Cow	189	55.0 (104)	47.6 (90)	62.0 (117)	48.2 (91)
Deer	96	55.2 (53)	36.5 (35)	62.2 (60)	42.6 (41)
Waterfowl[c]	216	66.2 (150)	52.8 (114)	70.4 (152)	56.5 (122)
Goat	42	66.7 (27)	59.5 (25)	47.6 (20)	42.9 (18)
Horse	78	44.3 (35)	34.2(27)	52.6 (41)	32.1 (25)
Human	210	59.2 (124)	47.4(100)	53.8 (113)	45.2 (95)
Pig	215	63.7 (137)	43.7 (94)	54.4 (117)	36.3 (78)
Sheep	61	7.5 (29)	39.3 (24)	37.7 (23)	8.2 (5)
Turkey	126	73.8 (93)	52.4 (66)	73.0 (92)	54.8 (69)
Overall	1,531	60.9 (933)	45.8 (701)	59.9 (917)	43.0 (659)

[a]Based on Jackknife analysis with 1% optimization and maximum similarities using curve-based (Pearson's product moment correlation coefficient) or band-based (Jaccard's coefficient) similarity calculations.
[b]Pet group consists of cats and dogs.
[c]Waterfowl group consists of ducks and geese.
[d]Values in parentheses are number of isolates correctly classified.

Table 4. Percentage of *E. coli* isolates correctly classified into domestic, human and wildlife source groups by using the HFERP DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]	
		Pearson	Jaccard
Domesticated[b]	855	83.2 (711)[e]	77.5 (663)
Human	210	53.8 (113)	45.2 (95)
Wildlife[c]	312	71.4 (223)	59.6 (186)
Pets[d]	154	59.1 (91)	44.8 (69)
Overall	1,531	74.3 (1,138)	66.2 (1,013)

[a]Done using Jackknife analysis with 1% optimization and maximum similarities using curve-based Pearson's product-moment correlation coefficient and band-based Jaccard similarity calculations.
[b]Domesticated group includes, chickens, cows, goats, horses, pigs, sheep and turkeys.
[c]Wildlife group includes deer, ducks and geese.
[d]Pet group includes dos and cats.
[e]Values in parentheses are number of isolates correctly classified.

Table 5. Percentage of *E. coli* isolates correctly classified into human and animal source groups by using the HFERP DNA fingerprinting method.

Source group	Number of DNA Fingerprints	Percent Correctly Classified[a]	
		Pearson	Jaccard
Animal	1321	93.7 (1,237)[b]	92.6 (1,223)
Human	210	53.8 (113)	45.2 (95)
Overall	1,531	88.2 (1,350)	86.1 (1,318)

[a]Done using Jackknife analysis with 1% optimization and maximum similarities using curve-based Pearson's product moment correlation coefficient and band-based Jaccard's similarity calculations.
[b]Values in parentheses are number of isolates correctly classified.

Figure Legends

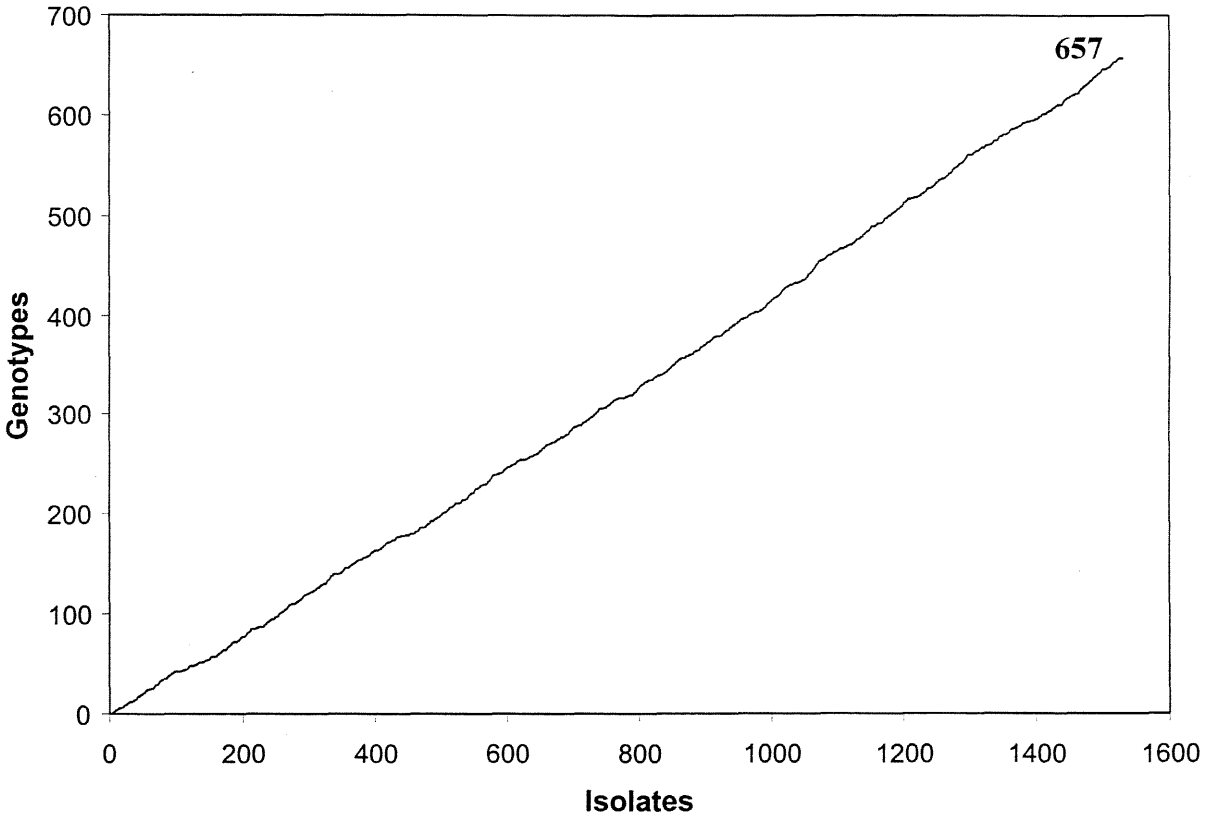
Figure 1. Accumulation curve of genotypes from *E. coli* isolates. Of 1,535 unique *E. coli* isolates in the known-source database with rep-PCR DNA fingerprint similarities of 90% or greater (based on the cosine coefficient), 657 genotypes were identified. The isolates were randomized, and a rarefaction curve was constructed by summing the number of genotypes that accumulated with the successive addition of isolates.

Figure 2. Frequency of occurrence of genotypes among rep-PCR DNA fingerprints from unique *E. coli* isolates. Analysis was limited to the 657 genotypes identified among the 1,535 unique *E. coli* isolates with rep-PCR DNA fingerprint similarities of 90% or greater.

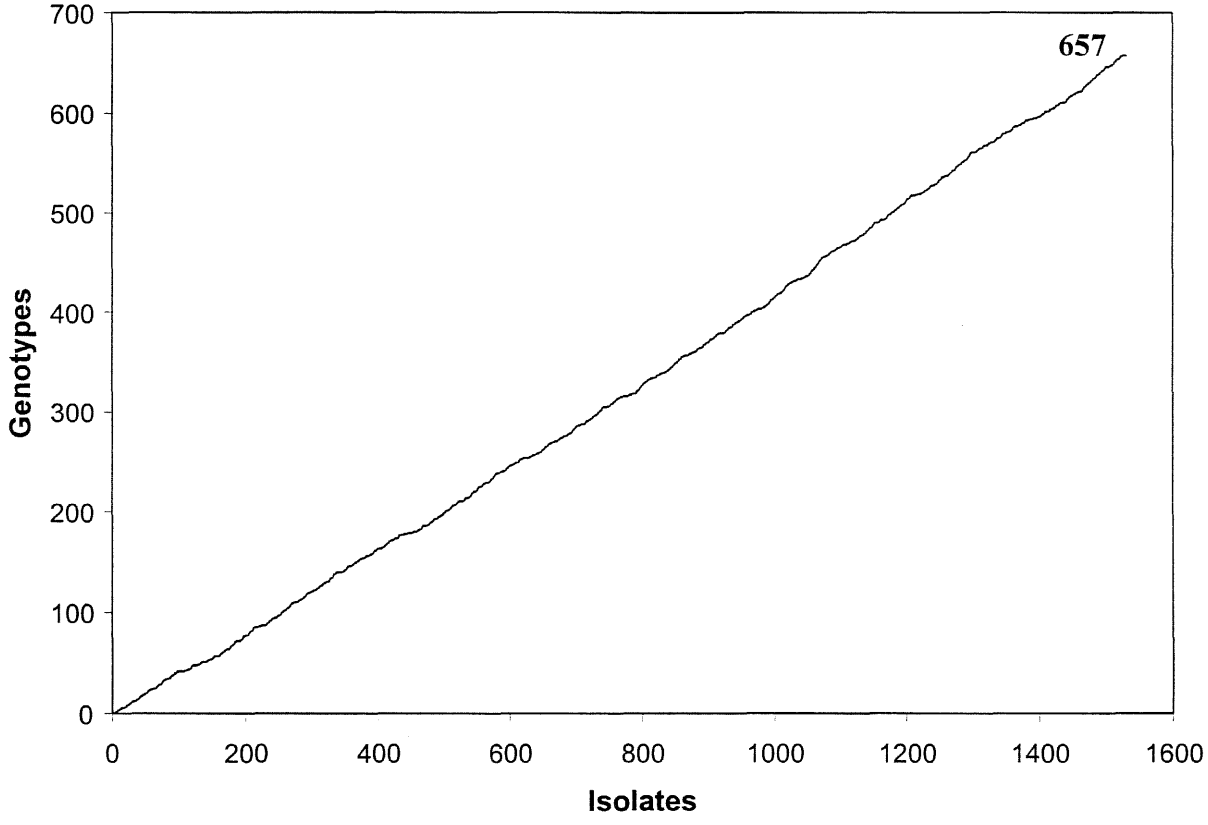
Figure 3. Representative examples of HFERP DNA fingerprint images. Genomic DNAs from 24 *E. coli* strains were subjected to HFERP DNA fingerprint analysis using a mixture of unlabeled Box A1R and 6-FAM fluorescently labeled Box A1R primers. Each lane contained Genescan-2500 ROX internal lane standards and HFERP DNA fingerprints. The combined, dual colored, HFERP image (A) was captured using a Typhoon Imager and two emission filters. Values in margin are in base pairs. Individual images of the HFERP DNA fingerprints (B) and Genescan-2500 ROX internal lane standard (C) were acquired using one filter at a time.

Figure 4. Comparison of DNA fingerprint patterns of a reference *E. coli* strain generated using rep-PCR and HFERP. (A) rep-PCR DNA fingerprint patterns were assembled from 29 individual

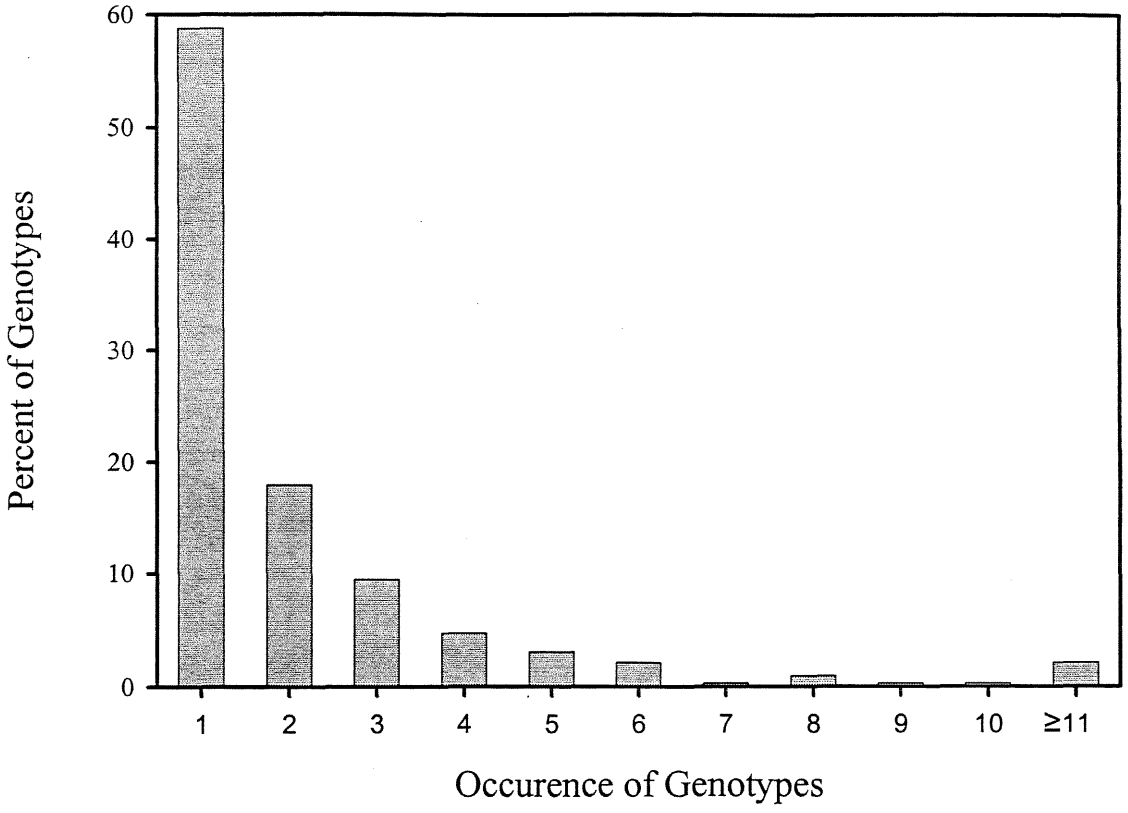
PCR reactions, each of which was run on a separate agarose gel. Fingerprints were generated using *E. coli* isolate P294 as template DNA and the Box A1R primer. (B) HFERP DNA fingerprint patterns were assembled from 29 individual PCR reactions each, of which was run on a separate agarose gel. Fingerprints were generated using *E. coli* isolate P294 as template DNA and a mixture of unlabeled Box A1R and 6-FAM fluorescently labeled Box A1R primers. Bands were aligned using Genescan-2500 ROX internal standards, which were present in each lane. Similarities were determined using the cosine algorithm of Bionumerics and dendrograms were generated using the unweighted pair-group method using arithmetic means (UPGMA).



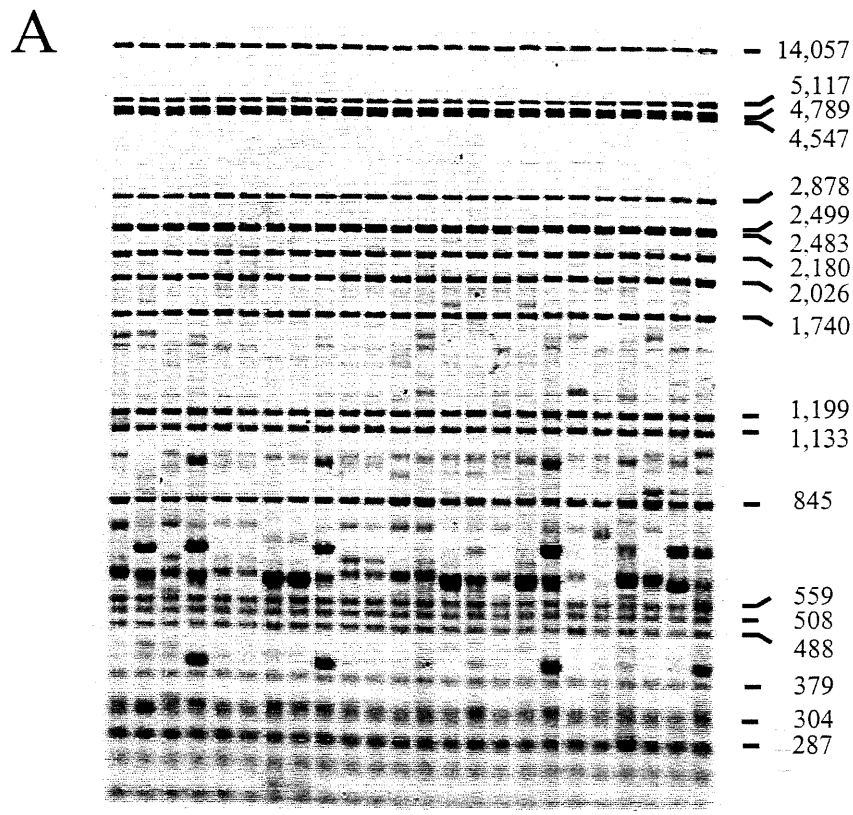
Johnson *et al.* – Figure 1



Johnson *et al.* – Figure 1

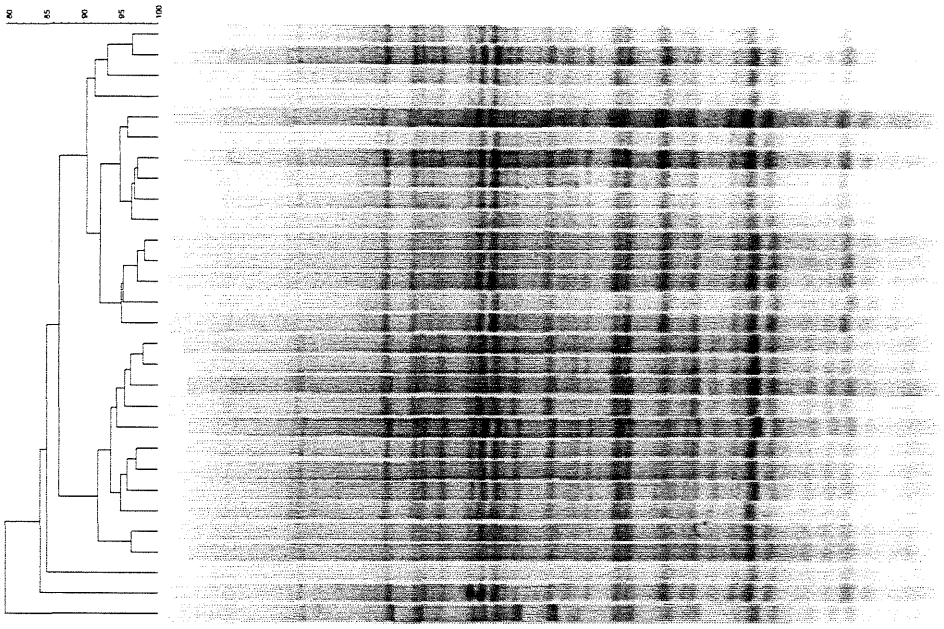


Johnson *et al.* – Figure 2

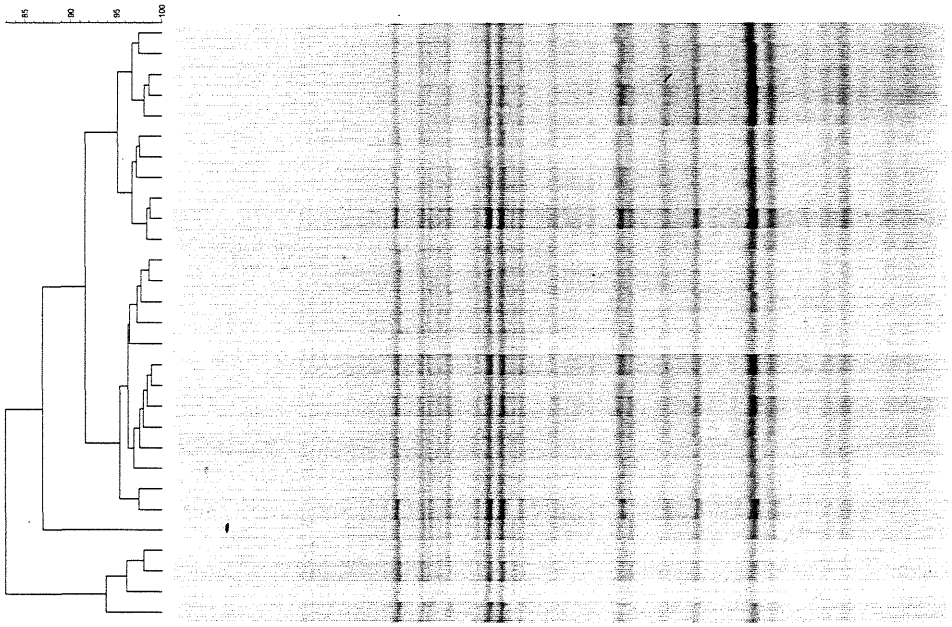


Johnson *et al.* – Figure 3

A



B



Johnson *et al.* – Figure 4

ATTACHMENT A

Project Title: Determination of Fecal Pollution Sources in Minnesota Watersheds

Project Number: IR13

LCMR Recommended Funding: $275,000

Attachment A Deliverable Products and Related Budget												
2001 LCMR Project Biennial Budget[A]	Objective/Result											
	Result 1 Budget:	Result 1 Current invoice:	Result 1 Balance:	Result 2 Budget:	Result 2 Current Invoice:	Result 2 Balance:	Result 3 Budget:	Result 3 Current Invoice:	Result 3 Balance:	PROJECT TOTAL:		
Budget Item	Acquire E. coli Bacteria			Generate DNA Fingerprints			Dissemination Activities			BUDGET TOTAL:	CURRENT INVOICE TOTAL:	BALANCE TOTAL:
Wages, salaries & benefits												
Senior Scientist	40,100	40,064	36	57,169	57,073	96	28,288	28,288	0	125,557	125,425	132
Junior Scientist	28,000	27,899	101	60,470	60,470	0				88,470	88,369	101
Student Workers (2@15 hours/week)	11,900	11,879	21	3,983	3,754	229				15,883	15,633	250
Contracts										0		
Professional/technical (Metropolitan Council Environmental Services for fecal coliform analysis)	4,800	4,196	604							4,800	4,196	604
Printing and Publication Costs							1,000	0	1,000	1,000	0	1,000
Laboratory Supplies	14,400	14,378	22	13,631	13,502	129				28,031	27,880	151
Office Supplies	500	457	43	1,000	416	584				1,500	873	627
Local automobile mileage paid	1,500	1,480	20				264	264	0	1,764	1,744	20
Other travel expenses in Minnesota (lodging and Meals)	0	0	0				200	0	200	200	0	200
Office equipment & computers				2,500	2,411	89				2,500	2,411	89
Software				4,800	4,699	101	495	349	651	5,295	5,048	247
COLUMN TOTAL	$101,200	$100,354	$846	$143,553	$142,325	$1,228	$30,247	$28,901	$1,851	$275,000	$271,581	$3,419

[A] Dollar amounts are estimates of balances on 12/31/03