MICROBIAL SOURCE TRACKING IN THE UPPER IOWA WATERSHED USING *E. coli* RIBOTYPING

Principal Investigators

Principal Investigator
Mary Skopec, Ph.D.
Adjunct Professor, Dept. of Geography
Iowa Dept. of Natural Resource
Geological Survey
109 Trowbridge Hall
Iowa City, IA 52242-1319
(319) 335-1579

Co-Investigator
Nancy Hall
The University of Iowa
Hygienic Laboratory
102 Oakdale Campus,
#H101 OH
Iowa City, IA 52242-5002
(319) 335-4331

Co-Investigator
Karen Owens
The University of Iowa
Hygienic Laboratory
102 Oakdale Campus,
#H101 OH
Iowa City, IA 52242-5002
(319) 335-4500

AIMS

Iowa's water quality is an important environmental, health, and economic concern as it pertains to recreational use, watershed control, and protection of public drinking water supplies. Monitoring of Iowa's surface waters during the past three years has demonstrated the prevalence of fecal bacteria in surface water resources, but has been unable to definitively identify the major sources of these bacteria. Source identification of bacteria is a crucial piece of information for state officials attempting to isolate and eliminate potential disease vectors. In Iowa, and the Midwest in general, decisions regarding the public health risk from contact with surface water bodies are based on disease vector studies from other parts of the country where large urban centers result in fecal contamination from sanitary sewer malfunctions and overflows. In rural parts of the country, urban sources of contamination most likely do not dominate water resources and therefore may result in a faulty conceptual model regarding the dominant sources of bacteria. For example, in 2001 and 2002, many Iowa beaches were closed or had warning signs posted due to fecal coliform levels that exceeded the U.S. Environmental Protection Agency's (EPA) recommended water quality standard for recreational waters (geometric mean of 200 colony forming units/100ml). None of these closure events were related to a documented sanitary sewer overflow, and instead appear to be related to nonpoint source runoff from agricultural lands. To improve our understanding of nonpoint source disease vectors, methods of tracking bacteria through the environment are needed. This seed grant helped address this need and applied new source-tracking tools to a specific watershed problem on the Upper Iowa River in northeast Iowa.

Materials and Methods

Preparation and Identification of *E. coli* Isolates

Isolates of *E. coli* were obtained directly from the animal feces of cattle, deer, swine, raccoon, sheep and geese within the Upper Iowa Watershed and throughout the state. Isolates of *E. coli* from humans were obtained from stool cultures submitted to University of Iowa Hygienic Laboratory (UHL) for enteric pathogen identification from primarily the study area (Howard and Winneshiek counties) and other volunteers from throughout Iowa. Two *E. coli* isolates were obtained from the feces of each individual animal. Fresh feces or anal surfaces were sampled with a CultureSwab® that contains Stuart's transport media (Becton Dickinson, Sparks, MD). Upon arrival at the laboratory, swabs were aseptically placed in EC broth and incubated in a 44.5ºC waterbath for 24 hours. Inoculum from this broth was streaked onto Levine's EMB plates and incubated at 35ºC for 24 hours. Two individual colonies from each sample were inoculated into specific biochemicals and confirmed as *E. coli.* Those isolates giving the typical biochemical profile for *E. coli* were frozen in small vials at a temperature of -70ºC using inositol calf serum for later processing. *E. coli* from water isolates were enumerated using Standard Methods 9222G (1). Two *E. coli* colonies from each plate were streaked onto Levine's EMB and incubated overnight at 35ºC. Pure isolates of *E. coli* from water were carried through biochemical confirmation and frozen as outlined above. All media used in this procedure was quality control checked (positive, negative and sterility) prior to use.

Ribotyping of *E. coli* Isolates

Isolates of *E. coli* were streaked onto sheep blood agar plates and incubated overnight at 35° C. The next day *E. coli* isolates were harvested from a lawn of growth on the agar using a special calibrated applicator stick. The bacteria were suspended in tubes containing 200µL of sample buffer, heated, enzymatically lysed and placed in the RiboPrinter® Microbial Characterization System (Qualicon, Wilmington, DE). Using the RiboPrinter®, sample DNA was digested with the restriction enzyme *Hin*d III and loaded into a pre-cast agarose gel. Restriction fragments were then separated by electrophoresis and transferred to a nylon membrane through an automated process in the RiboPrinter®. The membrane was then exposed to a series of enzymatic and chemical treatments that ultimately caused the DNA fragments of interest to glow. A low-light camera in the system captured the glowing images while the patterns were stored on the hard disk for analysis by proprietary software. The RiboPrinter® software compared each pattern to others in the *Hin*d III library and assigned a unique alpha-numeric designation to each unique pattern. *Hin*d III single digestion has been shown to be very economical and a suitable method for generating RiboPrint® patterns for the purpose of bacterial source tracking (2).

Analysis and Characterization of *E. coli* Isolates

Normalized pattern data generated by the automated RiboPrinter® Microbial Characterization System was imported into BioNumerics® (Applied Maths) software for statistical analysis (3). Patterns for each animal or human category were grouped into various libraries for comparison and identification purposes: a 5-group library consisting of cattle, deer, human, swine, and geese patterns; a 3-group library consisting of cattle, human, and rest of animals (deer, swine and geese) patterns; and finally a 2-group library consisting of animal (cattle, deer, swine and geese) and human patterns. The method of band analysis for comparison of unknown identifications was the curve-based Pearson correlation coefficient which expressed the degree of linear relationship between two entries. The dendogram type chosen was the unweighted pair group method using arithmetic averages (UPGMA) and optimization and position tolerances were set at 1.5% and 1%, respectively. Quality factors for each identification were also calculated. The Jackknife maximum similarity technique was used for cluster verification which estimated the separation between the defined groups. The rate of correct classification for each animal group comparison was obtained using the Jaccard maximum similarity methods. Discriminate analysis was performed to determine how well the RiboPrint® patterns of the various source group combinations could be distinguished from each other.

Results and Discussion
Group Verification Statistics

A total of 363 *E. coli* isolates were ribotyped. The isolates came from cattle (103 isolates), human (55 isolates), geese (29 isolates), swine (26 isolates), deer (36 isolates), sheep (6 isolates), water (64 isolates), raccoon (4 isolates) and quality control (40 isolates including reruns). After isolate pattern data from fecal samples were imported into

BioNumerics®, obvious outlier patterns within a group based on dendrogram matching were removed leaving the following isolates in the respective groups: cattle (88), deer (35), human (27), geese (26), and swine (24). Each animal group was assigned to a library unit with the exception of sheep and raccoon isolates because the number of isolates was too low (6 and 4 isolates, respectively). To determine the stability of the five-group library, group separation statistics using the Jackknife maximum similarity technique were performed (Appendix A, Table 1). This technique estimates the separation between the defined groups. The percentage of correct identification for each group is on the diagonal, highlighted in bold. The average rate of correct classification for the five defined groups was 69%. The human and cattle groups had the highest correct classification of 85% and 81%, respectively. Discriminant analysis was performed to determine how well the RiboPrint® patterns of the five source groups could be discriminated from each other (Appendix A, Table 1). The multivariate analysis of variance (MANOVA) results demonstrate that only two of the five groups were likely to be drawn from different populations as indicated by discriminants with low Wilkinson's parameter *L* and probability less than 0.05%. The most likely reason for the 5-group comparison not being valid is due to the low number of animal isolates for the geese (26), swine (24), and deer (35) groups compared to the cattle (88) group. Another discriminate analysis was performed using only three groups (cattle, human, and rest of animals) and the MANOVA and group separation statistics are given in Appendix A (Table 2). The low *L* parameters (0.2570 and 0.6871 for discriminants one and two, respectively) and significant probabilities ($p \leq 0.001\%$ and $p = 0.023\%$) demonstrated the validity of this 3-group library. The group separation statistics for this library showed an average rate of correct classification of 81%. This library was then used for the water unknown identifications as described below. The highest average rate of correct classification was achieved when only two groups (human and animal) were used in a comparison (Appendix A, Table 3). Again, the MANOVA results for these two groups showed a valid discriminate analysis (low *L* parameter of 0.3823 and p value of $\leq 0.001\%$).

Water Unknown Identifications

Water isolates from the four locations (27 Silver Creek, 801 Silver Creek, 8 Silver Creek, and 9 Cold Water Creek) in the Upper Iowa Watershed (Appendix B) were compared to the cattle, human and rest of animal library (CHA). Pearson correlation coefficients (maximum similarity) and quality factors were calculated for each water location compared with the CHA library as shown in Tables 4-7 (Appendix C). A quality factor letter of A or B shows a strong confidence of the identification (the unknown fits well in the library group). A letter of C, D, or E shows a poor confidence of the identifications. The criteria used in this study to decide the best association between a water isolate and cattle, human, or rest of animal isolates were similarity coefficients of >90% and a quality factor letter of A or B. Any identification that did not have these values were classified as "unknowns." The identifications for each location with the CHA library are graphically illustrated for the four seasons in Appendix D.

Significance of Study

Currently, *Escherichia coli (E. coli),* the predominant fecal coliform bacterium and the common inhabitant of human and animal intestines, is widely used to assess the quality

of surface water as an indicator of fecal pollution. The presence of fecal coliform bacteria or *E. coli* indicates that disease-producing organisms may be present. However, their presence does not differentiate between human and animal sources of pollution and the potential pathways that exist for pathogens to reach surface water sources. To understand and control fecal contamination problems and to assess human health risks, it is necessary to identify the contamination source. Animal waste can carry various pathogens such as *Salmonella* species, toxigenic *E. coli*, *Cryptosporidium parvum*, and *Giardia lamblia*; while human waste may carry all of the above pathogens plus *Shigella* species, hepatitis A viruses, and noroviruses. DNA identification of sources in watersheds with high levels of fecal indicator bacteria would be beneficial to all those agencies charged with protecting water quality and public health.

Early attempts to differentiate sources of fecal pollution used the ratio of fecal coliforms to fecal streptococci. A ratio of >4.0 would indicate human fecal pollution and a ratio of < 0.7 indicated animal (non-human) pollution. The value of this ratio has been questioned and its use is no longer recommended due to the variable survival rates of fecal streptococci species. Studies have shown that DNA fingerprinting of *E. coli* isolates by a molecular method called ribotyping can be a useful method for identifying human and non-human fecal pollution (2, 4).

The development of an *E. coli* library of DNA "fingerprints" and associated source tracking methodology comes at a critical time in the implementation of the Total Maximum Daily Load (TMDL) provisions of the 1972 Clean Water Act. These provisions require states to determine sources of pollutants and allocate pollutant loads in such a way as to prevent violations of water quality standards. In Iowa, twenty-seven of the 223 impaired waterbodies on Iowa's 2002 303(d) list are listed for fecal coliform contamination. Increased monitoring of Iowa's streams and beaches has resulted in a substantial increase in the number of waters identified as impaired for primary contact purposes. With source tracking information, the TMDL plan will more accurately reflect the input of sources and allocate loads accordingly. Thus, the TMDL plan will be more specific and more achievable.

FUTURE FUNDING

The promising results gained from the CGRER seed grant research helped procure additional funding from the US EPA Section 319 Nonpoint Source Program for bacterial source tracking research in the Lake Darling watershed in Washington, Keokuk, and Jefferson counties. This research includes the ribotyping technology applied in the Upper Iowa Watershed, as well as other source tracking strategies including multiple antibiotic resistance analysis, pathogen analysis, and tracking of sterols, caffeine and cotinine. Identification of bacteria sources at Lake Darling will result in more accurate targeting of best management practices throughout the watershed to address and reduce bacterial contamination. After successfully demonstrating the practical application of ribotyping to differentiate human and nonhuman pollution sources in the Upper Iowa Watershed, it is expected that University of Iowa researchers will have, in addition to the Lake Darling Project, numerous additional opportunities to pursue other funded applications for this ribotyping source tracking tool.

References

1. American Public Health Association, 1998. Standard Methods for the Examination of Water and Waste Water, 20th Ed., Clesceri, Greenberg and Eaton, Editors, American Public Health Association, Washington, D.C.
2. Tseng, C.C., W.T.E. Ting, D. Johnson, G. Thomas, and S. Adams. 2001. Automated Ribotyping of *Escherichia coli* Isolates from Humans and Animals. American Society for Microbiology General Meeting. May 21, 2001. Orlando, Florida.
3. Bionumerics® Manual, Version 2.0, Applied Maths, Kortrijk, Belgium, 2000.
4. 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(7):3123-3147.

Budget

Ribotyping Reagents:	$ 14,668.14
Media and other Supplies:	4,709.56
Student Salary:	622.30
Final Total:	$ 20,000.00
Free Balance	0.00

Acknowledgements

The co-investigators wish to thank the following people that helped with this project:

1. Eric O'Brien, IDNR and UHL – sample collection and manuscript preparation
2. Patrick Henry, Upper Iowa River Project Coordinator and Rick Langel, IDNR - sample collection
3. Alison Houston, UHL - RiboPrinter® and Bionumerics® training and consultation
4. Students Anne Farmer and Mercy Odueyungbo, College of Public Health and Iowa Biosciences Advantage Program, respectively – *E. coli* isolate culture preparation and database management
5. Deb Johnson, Purdue University – Bionumerics® training
6. Lora Friest, Fillmore Co. Minnesota, NRCS – Initial water sample collection and project development
7. Land Owners – Granting permission for fecal sample collection

APPENDIX A

Table 1. Statistics for 5 Defined Groups

	Cluster Verification					**MANOVA report**			
	Swine	Cattle	Deers	Geese	Humans				
Swine	**63.64**	6.76	0.00	12.50	0.00	DISCR_01	EIGV=67.8%	L= 0.0504	p<=0.001%
Cattle	27.27	**81.08**	34.62	31.25	11.11	DISCR_02	EIGV=23.3%	L= 0.2583	p<=0.001%
Deers	0.00	4.05	**65.38**	0.00	0.00	DISCR_03	EIGV= 6.1%	L= 0.6235	p=15.843%
Geese	9.09	4.05	0.00	**50.00**	3.70	DISCR_04	EIGV= 2.8%	L= 0.8556	p=74.748%
Humans	0.00	4.05	0.00	6.25	**85.19**				
ARCC = 69%									

Table 2. Statistics for 3 Defined Groups (Cattle, Humans and Rest of Animals)

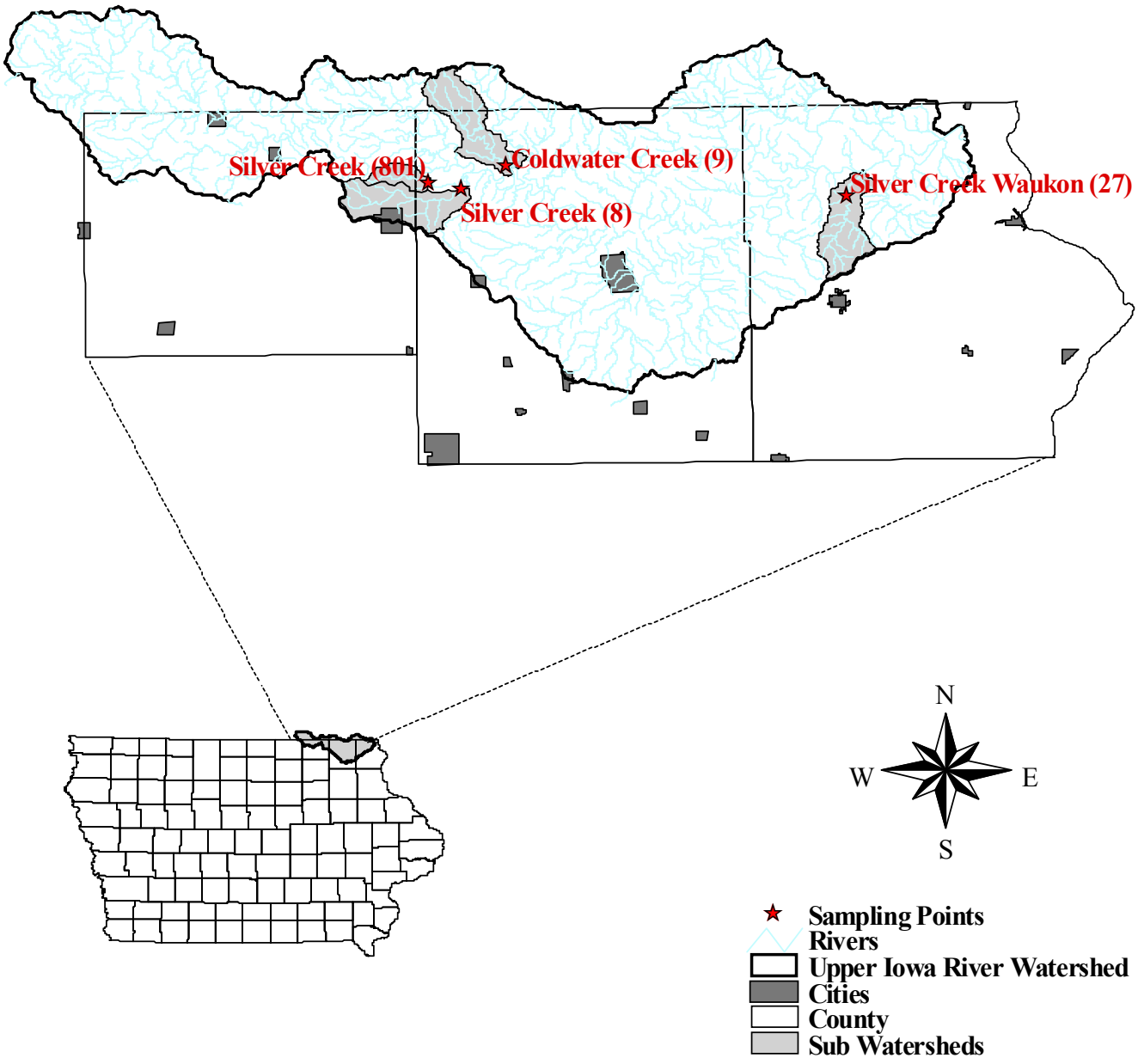
	Cluster Verification			**MANOVA report**			
	Animals	Cattle	Humans				
Animals	**76.56**	14.86	3.70	DISCR_01	EIGV=78.6%	L= 0.2570	p<=0.001%
Cattle	21.88	**81.08**	11.11	DISCR_02	EIGV=21.4%	L= 0.6871	p= 0.023%
Humans	1.56	4.05	**85.19**				
ARCC = 81%							

Table 3. Statistics for 2 Defined Groups (Animals and Humans)

	Cluster Verification		**MANOVA report**			
	Animals	Humans				
Animals	**97.10**	14.81	DISCR_01	EIGV=100.0%	L= 0.3823	p<= 0.001
Humans	2.90	**85.19**				
ARCC = 91%						

Appendix B

Upper Iowa River Sampling Points



Appendix C

Table 4. Location 9 Cold Water Creek Isolates Compared to CAH Library

Lab Number	Season/Year	Pearson Coefficient	Quality Factor	Associated Group
200210890-2	Fall 02	97.4%	B	Human
200210890-1	Fall 02	98.5%	B	Cattle
200208666-2	Fall 02	96.9%	C	Unknown
200208666-3	Fall 02	83.2%	C	Unknown
200211326-2	Fall 02	97.5%	B	Rest of Animals
200211326-1	Fall 02	97.7%	B	Cattle
200303081-1	Spring 03	91.5%	B	Rest of Animals
200303638-1	Spring 03	98.7%	B	Cattle
200303638-2	Spring 03	96.0%	B	Rest of Animals
200305801-2	Summer 03	90.6%	C	Unknown
200305801-1	Summer 03	97.8%	B	Cattle

Table 5. Location 801 Silver Creek Isolates Compared to CAH Library

Lab Number	Season/Year	Pearson Coefficient	Quality Factor	Best Associated Group
200208668-1	Summer 02	97.5%	B	Cattle
200208688-2	Summer 02	93.1%	C	Unknown
200210580-2	Fall 02	98.7%	B	Cattle
200210580-1	Fall 02	94.5%	C	Unknown
2001301825-1	Winter 02	95.3%	C	Unknown
200130825-2	Winter 02	95.3%	C	Unknown
200210892-2	Fall 02	98.1%	B	Cattle
200210892-1	Fall 02	97.4%	B	Cattle
200303640-1	Spring 03	91.9%	B	Rest of Animals
200303640-2	Spring 03	92.2%	B	Cattle
20030447-2	Spring 03	92.0%	B	Rest of Animals
20030447-1	Spring 03	89.8%	C	Unknown
200305805-2	Summer 03	97.5%	B	Rest of Animals
200305805-1	Summer 03	94.5%	B	Rest of Animals

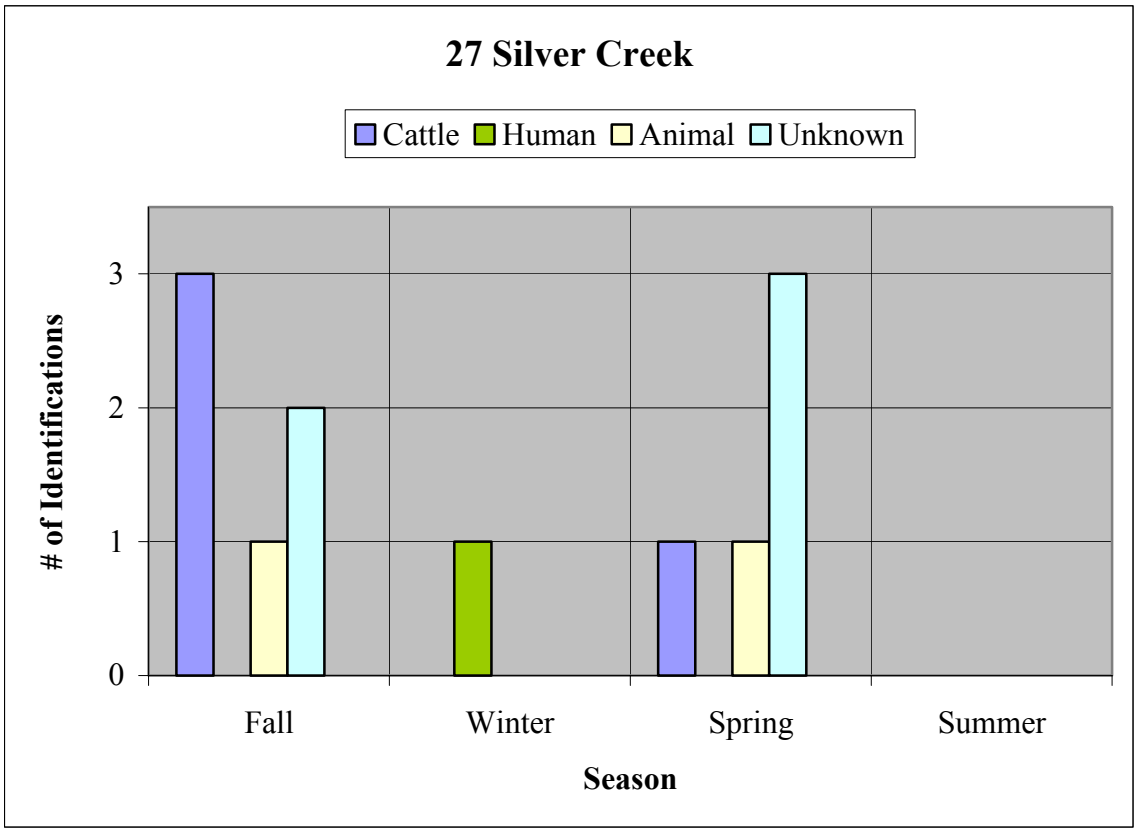
Table 6. Location 27 Silver Creek Isolates Compared to CAH Library

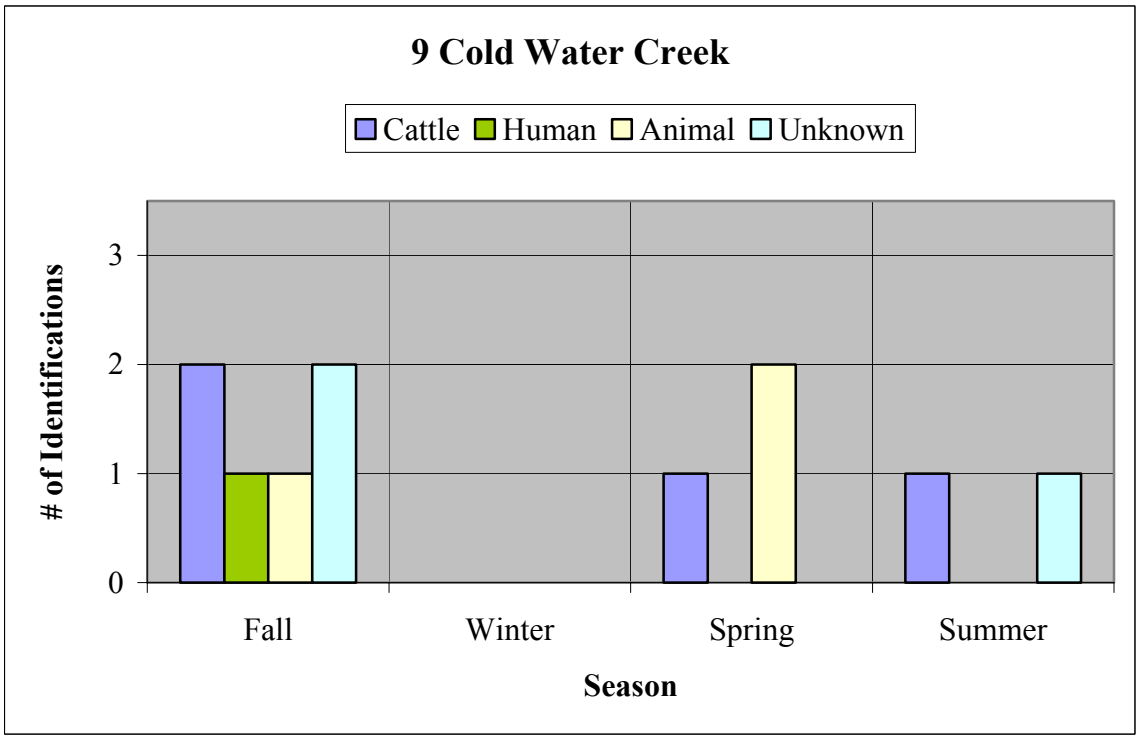
Lab Number	Season/Year	Pearson Coefficient	Highest Quality Factor	Best Associated Group
200209166-1	Fall 02	90.6%	B	Rest of Animals
200209166-2	Fall 02	95.9%	C	Unknown
200210927-2	Fall 02	98.2%	B	Cattle
200210927-1	Fall 02	89.7%	D	Unknown
200209453-2	Fall 02	97.4%	B	Cattle
200209453-1	Fall 02	98.5%	B	Cattle
200301248-2	Winter 02	93.9%	B	Human
200303081-2	Spring 03	97.4%	B	Cattle
200303244-1	Spring 03	96.8%	C	Unknown
200303244-2	Spring 03	95.3%	C	Unknown
200303515-2	Spring 03	96.6%	B	Rest of Animals
200303515-1	Spring 03	94.7%	C	Unknown

Table 7. Location 8 Silver Creek Isolates Compared to CAH Library

Lab Number	Season/Year	Pearson Coefficient	Quality Factor	Best Associated Group
200210889-5	Fall 02	92.9%	B	Rest of Animals
200210889-4	Fall 02	98.4%	B	Rest of Animals
20028665-3	Fall 02	93.8%	C	Unknown
200210889-2	Fall 02	89.7%	C	Unknown
200301822-1	Winter 02	98.7%	B	Rest of Animals
200301822-2	Winter 02	99.0%	B	Cattle
41810253-2	Spring 03	98.0%	B	Cattle
200303080-1	Spring 03	96.8%	B	Rest of Animals
200303637-1	Spring 03	95.1%	C	Unknown
20030637-2	Spring 03	90.0%	D	Unknown
20030444-2	Spring 03	93.2%	B	Rest
20030444-1	Spring 03	91.2%	C	Unknown
200305800-1	Summer 03	95.6%	C	Unknown

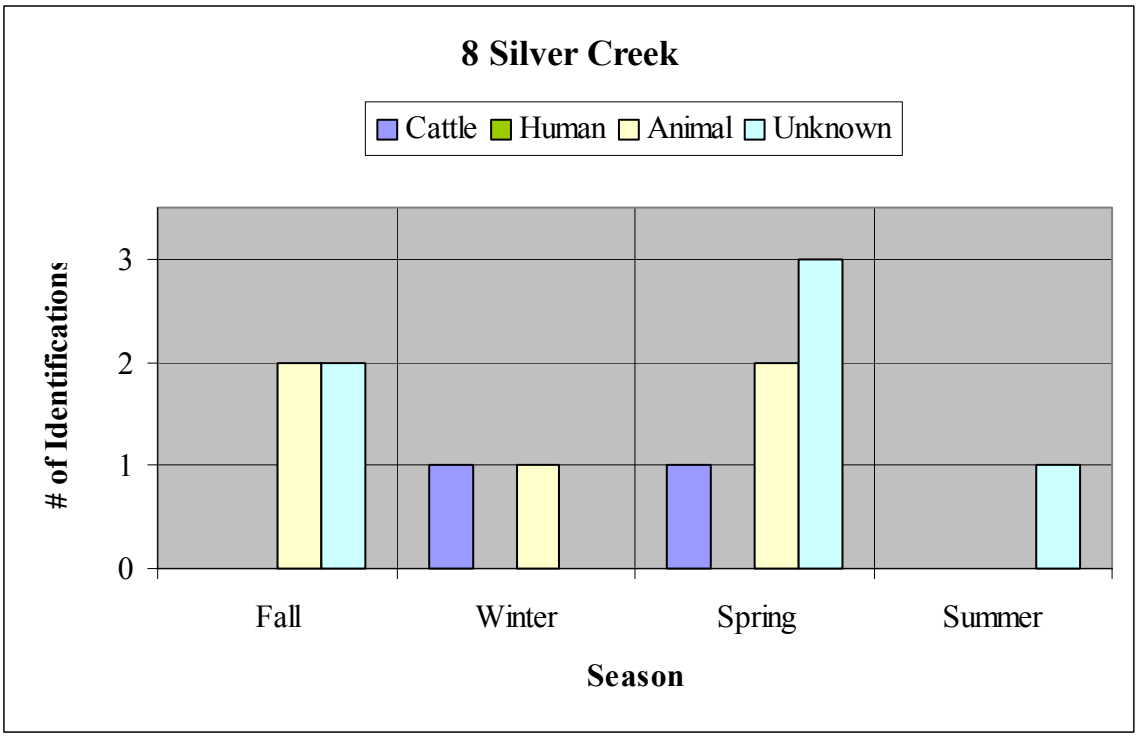
Appendix D

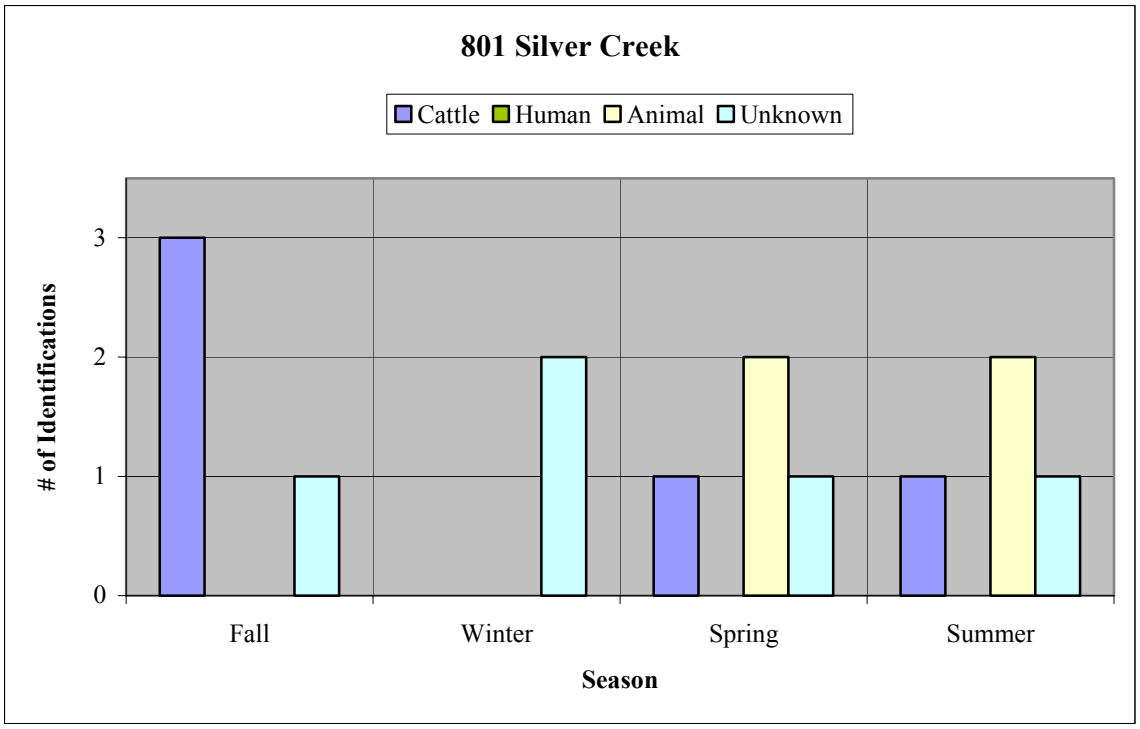

27 Silver Creek
Cattle
Human
Animal
Unknown
# of Identifications
0
1
2
3
Fall
Winter
Spring
Summer
Season


9 Cold Water Creek
Cattle
Human
Animal
Unknown
# of Identifications
0
1
2
3
Fall
Winter
Spring
Summer
Season

(continued)

Appendix D (continued)




801 Silver Creek
Cattle
Human
Animal
Unknown
# of Identifications
0
1
2
3
Fall
Winter
Spring
Summer
Season