University of South Florida

DIGITAL COMMONS
@ UNIVERSITY OF SOUTH FLORIDA

Digital Commons @ University of South Florida

USF Tampa Graduate Theses and Dissertations
USF Graduate Theses and Dissertations

4-1-2003

EVALUATION OF AUTOMATED RIBOTYPING AS A TOOL FOR BACTERIAL SOURCE TRACKING IN AQUATIC ENVIRONMENTS

Tracy L. Idocks
University of South Florida

Follow this and additional works at: https://digitalcommons.usf.edu/etd

Part of the Marine Biology Commons

Scholar Commons Citation
Idocks, Tracy L., "EVALUATION OF AUTOMATED RIBOTYPING AS A TOOL FOR BACTERIAL SOURCE TRACKING IN AQUATIC ENVIRONMENTS" (2003). *USF Tampa Graduate Theses and Dissertations.*
https://digitalcommons.usf.edu/etd/9053

This Thesis is brought to you for free and open access by the USF Graduate Theses and Dissertations at Digital Commons @ University of South Florida. It has been accepted for inclusion in USF Tampa Graduate Theses and Dissertations by an authorized administrator of Digital Commons @ University of South Florida. For more information, please contact digitalcommons@usf.edu.

EVALUATION OF AUTOMATED RIBOTYPING AS A TOOL FOR BACTERIAL SOURCE TRACKING IN AQUATIC ENVIRONMENTS

by

TRACY L. IDOCKS

A thesis submitted in partial fulfillment
of the requirements of the degree of
Master of Science
College of Marine Science
University of South Florida

Major Professor: Joan B. Rose, PhD.
Valerie J. Harwood, Ph.D.
John T. Lisle, Ph.D.

Date of Approval:
April 1, 2003

Keywords: Indicator bacteria, *E. coli*, Fecal Contamination, Watershed, Tampa Bay

© Copyright 2003, Tracy L. Idocks

Table of Contents

List of Tables ii

List of Figures iii

Abstract iv

Chapter One: Review of Related Literature 1
Introduction 1
Microbial Water Quality Monitoring and Bacterial Indicators 1
Bacterial Source Tracking 5
Methods for Bacterial Source Tracking 5
F+ coliphage serotyping 6
O-serotyping of *E. coli* 7
Rep-PCR 7
Antibiotic Resistance Analysis 8
Ribotyping 9
Automated Ribotyping 11
E. coli Ribotype Library for the Tampa Bay Area 13
Specific Research Objectives 14

Chapter Two: Methods and Materials 15
Sample Preparation 16
RiboPrinter® Methods 16
Data Analysis Software 18
Determination of Percent Similarities for the RiboPrinter® and BioNumerics® 19
Terminology 22

Chapter Three: RiboPrinter® Results 23
Library Analysis 23
Proficiency Isolates 26
Human or Non-Human Library 27

Chapter Four: Results from BioNumerics® 29
Library Analysis 29
Jackknife Analysis 32
Proficiency Isolates 33
Human or Non-Human Library 34
Individual Holdout Analyses 34

Chapter Five: Comparison of RiboPrinter® and BioNumerics® Software 38

Chapter Six: Discussion and Conclusions 43
The RiboPrinter® Microbial Characterization System 43
RiboPrinter® Software and BioNumerics® Software Results for Library 43
BioNumerics® Additional Library Analyses 50
Conclusions 52

References 55

List of Tables

Table 1	Results based on 95% mean similarity within source categories, as determined by the RiboPrinter® system.	24
Table 2	Number of Ribogroups formed from single sources and cosmopolitan source patterns based on the RiboPrinter® system.	25
Table 3	Results of Proficiency Isolate Ribogroup Assignment based on the four source category library.	27
Table 4	Number of ribogroups formed from single sources and cosmopolitan source patterns based on the RiboPrinter® system.	28
Table 5	Results of proficiency isolate ribogroup assignment based on the three source category library.	28
Table 6	Results from BioNumerics® showing number of ribogroups found by source.	29
Table 7	Number of ribogroups formed from single sources and cosmopolitan source ribogroups based on the BioNumerics® cluster analysis.	30
Table 8	Results of Jackknife analysis on the entire library of four sources.	33
Table 9	Results of proficiency isolate ribogroup assignment by BioNumerics®.	34
Table 10	Number of ribogroups formed from single sources and cosmopolitan source ribogroups based on the BioNumerics® software.	34

List of Figures

Figure 1	Summary of isolate sampling for the creation of the library.	20
Figure 2	Gel image of a batch run with only control strain. Band size of the marker is displayed on the right.	21
Figure 3	Percentages of Ribogroups per source category.	25
Figure 4	Distribution of number of ribogroups per individual based on 95% mean similarity as determined by the RiboPrinter® System.	26
Figure 5	Percentages of ribogroups per source category.	31
Figure 6	Distribution of number of ribogroups per individual based BioNumerics® results.	32
Figure 7	Percentages of isolates that classified correctly during individual holdout analyses.	36
Figure 8	Results of correct classification for holdout analysis for wastewater samples.	36
Figure 9	Classification of isolates within an individual sample.	37
Figure 10	Gel image from the RiboPrinter® of eight dog isolates.	39
Figure 11	Gel image from the RiboPrinter® of eight cow isolates.	39
Figure 12	Percentage of ribogroups per source for the RiboPrinter® and BioNumerics®.	40
Figure 13	Percentage of ribogroups based on single source and cosmopolitan source ribogroups between the RiboPrinter® and BioNumerics®.	41
Figure 14	Percentage of ribogroups when library is human, non-human or cosmopolitan between the RiboPrinter® and BioNumerics®.	42

Evaluation of Automated Ribotyping as a Tool for Bacterial Source Tracking

in Aquatic Environments

Tracy L. Idocks

ABSTRACT

An *Escherichia coli* ribotype profile library was created for the Tampa Bay watershed for application to bacterial source tracking of human or non-human fecal contamination to the bay and surrounding rivers. Bacterial source tracking uses a variety of methods to identify the source of fecal contamination, not just the quantity of the indicator bacteria present, allowing implementation of effective management practices. The 700 confirmed *E. coli* isolate library was equally distributed among four sources: human, bird, dog and cow. The library was tested for applicability for source tracking within the Tampa Bay watershed. The DuPont Qualicon RiboPrinter® Microbial Characterization System, a fully automated ribotyping instrument, was used to generate this library. The ribotype patterns were analyzed by software packages from the RiboPrinter® and BioNumerics®. Proficiency isolates were used to test the accuracy of the library. The RiboPrinter® defined 166 ribogroups, whereas, BioNumerics® defined 234 ribogroups for the same isolates. While the numbers of ribogroups differed, percent classifications in the four source categories were similar when normalized data were compared between software packages. The RiboPrinter® found dog, bird, and cow

isolates represented in 34, 38, 29 and 31% of the ribogroups, respectively. BioNumerics® found dog isolates, bird, human and cow isolates in 32, 35, 31 and 30% of the ribogroups, respectively. The RiboPrinter® and BioNumerics® software respectively assigned 58% and 32%, of the proficiency isolates to their correct source categories. The average rate of correct classification, as determined using BioNumerics® and Jackknife analysis, ranged from 72.57% for human isolates to 77.14% for cow isolates. Conversely, the individual holdout analyses showed low rates of correct classification, with most individuals less than 50% correctly classified, of individuals (a single sample containing multiple isolates). The individual holdout analyses and the proficiency isolate data indicate low correct classification rates(less than 60%), suggesting, the results of the Jackknife analysis over-estimates the rates of correct classification. Currently, this library does not offer the discrimination needed for most bacterial source tracking applications and a large library containing the diversity of isolates will be needed. The 700 isolates and the ribogroups that they form provide foundation for future work.

Chapter One: Review of Relevant Literature

Introduction

Water resources and aquatic health are vital to the success of coastal communities. Florida's water resources are important economically, recreationally, and aesthetically. Beach and recreational water closures due to microbial contamination have a direct impact on the utilization of Florida's aquatic environments by tourists and residents alike. Understanding the quantity and source of microorganisms, associated with pollution, in a watershed is the first step in correcting the problem of microbial water pollution, thus providing a cleaner and safer environment for all to enjoy.

Microbial Water Quality Monitoring and Bacterial Indicators

Microbial water quality monitoring is fundamental to the management of recreational and drinking water. Understanding the potential risk to human health is necessary to determine safe uses (drinking, bathing, fishing or shellfish harvesting) of the water source in question. The potential risk from water pollution has been historically based upon the presence and abundance of indicator bacteria know as coliform bacteria. The characteristics of ideal bacterial indicators are summarized for the marine environment by Griffin *et. al.* (2001). Ideal indicators are non-pathogenic microorganisms that occur alongside pathogens, survive longer than pathogens, and have some direct correlation to the number of pathogens, are more resistant to disinfection than pathogens, cannot grow in the environment and are easy to isolate. These are key

characteristics for identifying and using indicator organisms. Many of the indicators currently used as measures of water quality fail to live up to one or more of the ideal indicator criteria.

Commonly used indicators of fecal water pollution are total coliforms and fecal coliforms (Wolf 1972) and *Escherichia coli.* (Clesceri *et.al.* 1998). *Clostridium perfringens* (Bisson and Cabelli 1979), *Enterococcus spp.* (Levin *et. al* 1975) and coliphage (Griffin *et. al.* 2001) have been suggested as alternative indicators for water quality monitoring.

The total coliform group, which encompasses the genera *Enterobacter, Citrobacter, Escherichia* and *Klebsiella*, is a collection of Gram negative rod shaped bacteria that ferment lactose with the formation of gas and acid within 48 hours at 35 ^{o}C (Clesceri *et.al.* 1998). The detection of these bacteria in water can be accomplished through membrane filtration, incubation at 35 ^{o}C on mENDO media. Pink colonies with a green metallic sheen are counted after 24-48 hrs (Wolf 1972). Fecal coliforms, a subgroup of total coliform group include the genera *Escherichia* and many species of *Klebsiella* and are defined as bacteria that ferment lactose with the formation of acid and gas at 44.5 ^{o}C within 48 hours. One of the standard methods for detection of fecal coliform bacteria in water is membrane filtration, incubation at 44.5 ^{o}C on mFC media and counting blue colonies after 24-48 hours (Clesceri *et.al.* 1998). *E. coli*, a specific fecal coliform, can be identified by its ability to cleave 4-methylum-belliferyl-β-D-glucuronide by the enzyme β-glucuronidase. The product is detected by fluorescence (under UV light) of *E. coli* colonies after incubation on EC-MUG media for 24-48 hours

at 37 °C (Clesceri *et.al.* 1998). These bacterial indicators are recognized standards for monitoring ambient water quality by the USEPA Federal Water Pollution Control Act.

Total and fecal coliforms have been used as the primary bacterial indicators of fecal pollution since coliforms are found in the intestines of most warm-blooded animals (Clesceri *et.al.* 1998). These bacteria are found in both human and animal feces so they are not capable of identifying the specific source of pollution. The use of the coliform indicator groups, especially the fecal coliforms, has come under question due to increased stability and longevity in the sediments in tropical settings (Fujioka and Shizumura 1985, Hazen and Toranzos 1990 and Roll & Fujioka 1997). Samples may have a large number of coliform (both total and fecal coliforms) bacteria present, but the fecal pollution event may have been long since removed.

The United States Environmental Protection Agency (USEPA) has recommended the use of *Enterococcus* spp. for water quality monitoring (EPA 821-R-97-004) Enterococci are a subgroup of the fecal streptococci, which are Gram positive coccoid bacteria commonly associated with the gastrointestinal tract of warm blooded animals (Geldrich 1969). The current USEPA standard for measuring enterococci is Method 1600 which involves membrane filtration, incubation at 41 °C on mEI media, and colonies that have a blue halo after 24 hours are counted as enterococci (EPA 821-R-97-004). However, enterococci may be found in sediments where human fecal inputs are not present (Hardina and Fujioka 1991).

Coliphage, which are viruses that infect *E. coli*, have also been suggested by the EPA as a measure of fecal input into water. USEPA method 1602 uses a single agar overlay method for the enumeration of viruses in water samples (EPA 821-R-01-029).

Another alternative indicator is *Clostridium perfringens,* which is an anaerobic spore forming bacterium that is detected and counted by membrane filtration and the use of a specialized media called mCP agar (Bisson & Cabelli 1979).

While the indicator concept has been criticized for accuracy, it is still the standard that is used for many water quality studies. In a study of Little Sarasota Bay, FL in 2001, Lipp *et.al*. determined that the use of water quality indicators indicators, including fecal coliforms, enterococci, *Clostridium perfringens* and coliphage, with the addition of cluster analysis could identify regions of the bay that could cause a potentially high risk to human health. This study was conducted and samples were collected at 11 sites between May and September 1996. Repeated sampling over a four month period revealed the inputs to the system, not just the residual bacteria that may remain in the sediments.

Fecal coliforms, enterococci, *Clostridium perfringens* and coliphage, were also used to determine the seasonal fluctuations of bacterial abundance in Charlotte Harbor when studied over a twelve month period (Lipp *et. al.* 2001). Fecal coliform levels were greatest in August and December to February and were positively correlated to rainfall. *C. perfringens* was positively correlated to water column turbidity and showed a peak in abundance in March. Enterococci were also positively correlated to rainfall. The enterococci levels were the greatest between December and February. Coliphage were most prevalent in water samples in December and showed statistical relationships to rainfall and river flow entering the bay. In this study, septic tanks were the suspected source of fecal inputs; however no definitive determination could be made.

Bacterial indicators and coliphage have been criticized for their use in the marine environment due to lack of correlation with feces, pathogens and human health risk

(Griffin *et.al* 2001). One of the greatest issues with the use of indicators is the inability to discriminate between human and animal inputs to a water body.

Bacterial Source Tracking

Bacterial indicators have been primarily used for determining the impact of fecal pollution, yet still leave a central question unanswered; "Where or what are the bacteria coming from?". In most watersheds, multiple sources are found including septic tanks, runoff from farms, and wildlife, but the exact source impacting the water is not often identified with traditional methods. Thus new methods are needed to determine the source of indicator bacteria in the water so the best management practices can be implemented. Bacterial source tracking (BST) is a new approach that holds promise for identifying the sources of fecal pollution in water systems. Some ideal characteristics for BST methods are that the method is reproducible, works on all target organisms, and has high discriminatory power (Olive 1999).

Methods for Bacterial Source Tracking

Numerous methods have been used to determine the sources of microbial pollution. The earliest methodology attempted to use the ratio of fecal coliforms to fecal streptococci. A ratio of >4 was considered a human source and a ratio of <0.7 was considered an animal source (Geldrich *et. al.* 1969). . This did not prove to be a reliable test since the two bacterial groups had different survival rates in the environment (Feachem 1975 and Pourcher 1991). Various *Enterococcus* spp. were found to persist longer in the environment and overestimated the contribution of non-human sources of

fecal pollution. Currently, the most commonly used methods for BST are Antibiotic Resistance Analysis (ARA)(Wiggins 1996) and Ribotyping (Parveen 1999). Studies have also shown that F+ coliphage (Hsu *et. al.* 1995), O-serotyping (Parveen 2001) and repetitive DNA sequences (Bruijn 1992) may also be useful source tracking methods. Different microorganisms have been used for source tracking, including: fecal coliforms (Whitlock 2002), *E. coli (*Dombek *et. al.* 2000), the fecal streptococi group (Wiggins 1996), *Bifidobacterium* and the *Bacteroides- Prevotella* group (Bernhard *et. al.* 2003 and Bernhard & Field 2000).

F+ coliphage serotyping

F+ coliphage are viruses that infect *E. coli* cells via pili which are attachment structures on the surface of susceptible *E. coli* cells. There are four distinct serotypes categorized as I, II, III and IV. Source tracking using F+ coliphage is based on these serotypes (Hsu *et. al.* 1995). F+ coliphage Serotypes II and III are predominately of human origin, while serotypes I and IV are predominately of animal origin. In a study in Homssassa Springs, FL, Griffin *et. al.* (2000) used total coliforms, fecal coliforms, enterococci, *C. perfringens* and F+ coliphage to monitor the spring and determine the source of the high level microbial pollution. This study found the coliphage to be animal serotypes, I and IV, and the authors concluded that the fecal contamination was from animals in the Homossasa Wildlife Park or from indigenous animals and not from the septic tanks in the surrounding area. While this method was successful in determining human versus non-human sources it does not distinguish between types of animals. In a study performed in South Africa and Spain, Schaper *et.al.* (2002) found that serotypes II

and III were associated with human sewage, but human samples also contained serotypes I and IV. Animal samples also contained all four serotypes, with the majority of the F+ coliphage being serotypes I and IV. This study found the assignment of serotypes to specific human or animal sources to be statistical significant. However, the distinction between serotypes may not be as definite as previously thought, as there was overlap between the serotypes and their expected animal sources.

O-serotyping of *E. coli*

O-serotyping is based on the presence or absence of somatic (O) antigenic determinants on *E. coli*. This method was presented by Parveen *et.al.* (2001) as a method of BST. One hundred and four known source isolates (53 from human sources and 51 from non-human sources) that had been previously collected from the Apalachicola National Estuarine Research Reserve were used in this study. In addition to O-serotyping, pulse field gel electrophoresis (PFGE) and fatty acid methyl ester (FAME) were also used. The investigators showed no source discrimination from PFGE or FAME. However, O-serotyping showed 77% of the isolates were successfully serotyped. Human and non-human sources had distinct serotypes, suggesting that this may be a reliable method of differentiating between sources of fecal contamination.

Rep-PCR

Repetitive intergenic DNA sequences amplified by polymerase chain reaction (PCR), has been used to determine strain differences in many bacterial species including *Rhizobium meliloti* (Bruijn1992), *Bradyhizobium japonicum* (Judd *et.al* 1993),

Streptomyces strains (Sadowsky *et. al.* 1996) and *E. coli* (Dombek *et.al.* 2000). The method involves polymerase chain reaction (PCR) amplification of variable strain specific DNA regions. These sequences are found in non-translated regions of the DNA and many copies are found throughout the chromosomal DNA. The PCR products are separated on a gel and the pattern of bands creates a "fingerprint" for each bacterium (Stern *et.al.*1984). This method is sensitive enough to discriminate between closely related bacterial strains (Sadowsky *et.al* 1996). The *E. coli* library used by Dombek *et.al.* (2000) included, human swab isolates (29 from 14 individuals), geese swab isolates (21 from 8 individuals), duck fecal isolates (23 from 10 individuals), chicken fecal isolates (20 from 10 individuals), pig fecal isolates (21 from 9 individuals), sheep fecal isolates (19 from 10 individuals) and cow fecal isolates (21 from 12 individuals). This study used BOXA1R primers (5' CTACGGCAAGGCGACGCTGACG 3'), which yielded bands approximately 0.25kb to 2.3kb. Dombek *et. al.* (2000) found 78-90% correct classifications of source groups.

Antibiotic Resistance Analysis

Antibiotic Resistance Analysis (ARA) has been widely used in BST for both fecal coliforms (Whitlock *et.al.* 2002 and Harwood *et.al.* 2000) and the fecal streptococci group (Hagedorn *et. al.* 1999, Harwood *et.al.* 2000, Wiggins *et.al.* 1999 & Wiggins 1996). Introduced by Wiggins (1996), ARA is based on the bacterial growth on a suite of antibiotics of varying antibiotic concentrations. Patterns of antibiotic resistance or susceptibility of known source isolates are compared to unknown source isolates (i.e. ones collected from water) in order to identify a source of the unknown isolates. The

major drawback to this method is the need to create a large reference library of known source isolates to compare environmental or unknown isolates. The generation of a reference library is time consuming, since most libraries have hundreds of isolates from numerous sources. While this method can be labor intensive, it is not without merit. In a study by Hagedorn *et.al* (1999), ARA was performed on fecal streptococci isolates from a Virginia watershed that had high levels of fecal indicator bacteria. The library for this study contained 1398 beef cattle fecal isolates, 728 dairy cattle fecal isolates, 824 chicken fecal isolates, 1245 deer fecal isolates, 1284 waterfowl fecal isolates and 1579 isolates from human wastewater. The average correct classification rates (ARCC) for this study ranged from 84 to 90% correct classification of isolates. The ARA results showed that the contamination source was beef cattle. Once the cattle were restricted from access to the stream, the level of fecal indicator bacteria decreased and the percentage of isolates identified as being from cattle decreased by 45%.

Ribotyping

Ribotyping has been used in many studies for source tracking using the *Enterococcus* spp. (Brisse *et.al.* 2002 and Turlak *et. al.* 2001), *Staphylococcus aureus* (Barbour *et.al.* 1994), *Listeria monocytogenes* (Wiedmann *et.al.* 1996) and *E. coli* (Carson *et.al.* 2001, Parveen *et.al.* 1999, Scott *et. al.* 2003 and Tseng *et.al.*2001). Ribotyping is also a library-based method of BST. The use of a library, due to diversity in *E. coli*, over a large geographical area has been questioned in most BST studies. Based on a study by Scott *et. al.* (2003) libraries need to be created for each individual watershed being evaluated and is only useful for time geographical area (Carson *et. al*

2001). The ribotyping method involves the collection, digestion and separation of DNA and detection of the resulting bands. The banding patterns are then compared statistically to obtain similarities between patterns that can be used to identify unknown bacterial isolates collected from a water sample. In most cases the enzymatic digestion of *E. coli* is performed using *Hin*dIII.

Parveen *et.al.* (1999) successfully demonstrated that ribotyping and discrimnant analysis were useful in determining sources of microbial pollution. This study analyzed 84 human source isolates, from sewage treatment plant effluents, that were from a previous ARA study, 95 non-human source isolates from the same previous ARA study, 30 new human feces isolates and 29 new isolates collected from wildlife feces. Discriminant analysis of the isolates (subset of the previous ARA study isolates) showed correct classifications of 97% and 67% for non-human source and human source isolates respectively. The newly isolated human feces average rate of correct classification was 67 %, while the animal feces correctly classified 100%.

In a study by Carson *et. al.* (2001) correct classification for human and non-human sources was 95.0% and 99.2% respectively (ARCC=97.1%). This Missouri library consisted of fecal samples from the following animals: beef and dairy cattle (39 isolates from 24 individuals), pigs (44 isolates from 30 individuals), horses (37 isolates from 10 individuals), dogs (29 isolates from 15 individuals), geese (49 isolates from 24 individuals) and composite samples from chickens (23 isolates), turkeys (26 isolates) and human swab samples (40 isolates from 15 individuals). When all eight sources were classified separately, the ARCC decreased to 73.4% from the 97.1% ARCC of the human or non-human library. This decrease is due to overlap found within the non-human

categories which caused better classification when the library was analyzed with human and non-human categories only (Caron 2001).

Automated Ribotyping

Most BST methods have a common caveat; they are time consuming to perform. Results are achieved after numerous days and many hands on hours of work. Prior to analyzing unknowns many methods require, that a large reference library be created. It takes multiple days for the initial isolation of the bacteria followed by many days of processing, depending on the method employed and months to generate of a large database of known source isolates for comparison with unknown source isolates. Removing the time element could make BST a much less formidable task.

Dupont Qualicon (Wilmington, DE) has marketed a possible solution to the time issue involved in performing ribotyping. The Dupont Qualicon RiboPrinter® Microbial Characterization System, hereafter referred to as the RiboPrinter® is a fully automated ribotyping system that performs the same methods as the bench top ribotyping methods, but yields results in approximately eight hours, rather than several days (Bruce 1996). The instrument has been used by the food industry to track bacterial contamination throughout processing (Bruce 1996 and Wiedmann 1997). The RiboPrinter® has also been used to identify pathogenic strains of *E. coli* (Bruce 1997). While the system was originally designed for industrial and clinical applications, it has been shown to be useful in environmental BST (Tseng 2001). Tseng *et. al.* (2001) showed that the RiboPrinter® system could be used to discriminate sources of *E. coli* from a group of 160 isolates collected from four sources: human (40 isolates from 40 individuals), cow(39 isolates

from 39 individuals), horse (41 isolates from 41 individuals) and geese (40 isolates from 40 individuals) with an using *Hin*dIII for digestion. The ARCC was calculated to be 94% using DICE, Jaccard, Jeffrey's and Ochiai similarity coefficients separately. Increased profile discrimination was accomplished in this study by using *Hin*d III as compared to *Cla* I, *Eco* RI, *Mlu* I and *Pvu* II. While the RiboPrinter® does not routinely use *Hin*dIII (it uses *EcoR*I) as the restriction enzyme, the instrument's protocol can be modified to accommodate this alternate enzyme. The instrument has also been used for tracking vancomycin-resistant *Enterococcus faecium* by double enzyme digest, using *Ase*I and *BamH*I (Brisse *et. al.* 2002 & Turlak *et. al.* 2001).

The use of *E. coli* as an ideal indicator and as a target for BST has come under question. *E. coli*, and most of the fecal coliform group, has been shown to regrow in the environment (Desmarais *et.al.* 2002 & Solo-Gabriele *et.al.* 2000). This may make determining a rather recent pollution event difficult. The clonal distribution with the species *E. coli* has also shown to be problematic for BST (Spratt & Maiden 1999). In this study a clonal distribution was defined as a group of isolates that had little or no recombination of chromosomal DNA in its evolutionary history. These clonal populations of bacteria were characterized by low levels of sequence diversity (Levin 1981).

Gordon (2001) discusses the characteristics of an ideal target for coliform pollution tracking. The characteristics are geographical differences in clonal strains, host specificity, no difference between clones in primary and secondary habitats and stable clones through time. In a series of publications Gordon (Gordon 2001, Gordon & Lee 1999, Gordon *et. al.* 1998 and Gordon 1997) shows that *E. coli* may not live up to these

characteristics. Geographical differences in house mice populations did not show distinct differences in clonal strains of *E. coli* isolates collected from house mice (Gordon 1997). The special differences in the two populations studied only accounted for 2% of the genetic diversity. Patterns formed by *E. coli* collected from Australian mammals were not found to be mutually exclusive when compared across host categories. Gordon *et.al.* (2002) found that the clones isolated from the environment (secondary source) were not always the same as isolated from the septic tank source of the isolates. This suggests that the *E. coli* isolates collected from the environment may not share the same clones as the original source. The clonal structure on a temporal scale has been shown to change (Gordon *et. al.* 1998). Isolates collected from house mice over a year long period differed with each sample. Some clonal isolates were found in all sampling events, some were intermittent. This series of studies appears to argue that *E. coli* is not an acceptable target for indicator monitoring or for BST. In a separate study by Kariuki *et. al.* (1999) *E. coli* isolates were shown to be distinct when compared between chickens and children living in close contact. Kariuki *et. al.* found clonality within the two sources, but very little between the two sources. This study suggests that *E. coli* may still be a valuable target organism for BST.

E. coli Ribotype Library for the Tampa Bay Watershed

Tampa Bay is an economically and recreationally important estuary in west-central Florida. The bay covers nearly 400 square miles and the associated watershed is 2200 square miles. Within the bay's watershed are extensive, highly urbanized areas, wastewater treatment plants, active phosphate plants and agricultural areas. This shallow

estuary discharges into the Gulf of Mexico and has four major rivers flowing into the bay: Hillsborough River (average flow rate 449 ft^3/sec), Manatee River (average flow rate 110 ft^3/sec), Little Manatee River (average flow rate 40 ft^3/sec) and Alafia River (average flow rate 336 ft^3/sec). The northern region of the bay is home to an active port used for industry and cruise lines. Understanding the sources of non-point microbial pollution to Tampa Bay will facilitate the design of effective best management practices for clean-up or maintenance of the bay's waters.

The purpose of this study was to use *E. coli* isolates from various sources, including dogs, birds, cattle and humans, which could inhabit the Tampa Bay watershed and impact the bay's waters to create a library from these isolates and determine the library's potential application to tracking sources of bacterial water pollution in Tampa Bay.

Specific Research Objectives

- Evaluate the use of the Dupont Qualicon RiboPrinter® for use in the generation of a Tampa Bay regional library using *E. coli* as the study organism
- Compare ribotype data analysis using two statistical software packages (RiboPrinter® software and BioNumerics® software)
- Analyze the library for efficiency as a four source library (human, dog, cow and bird) and as a human or non-human library

Chapter Two: Methods and Materials

The library for Tampa Bay region included 700 *E. coli* isolates, 175 isolates from each of the following four sources: human (175 isolates), dogs (175 isolates), cattle (175 isolates) and birds (175 isolates). The sampling scheme is shown in Figure 1. The 175 human isolates were collected from four wastewater influent samples (130 isolates) and nine human anal swab samples (45 isolates) and were considered one group, termed human, for the analyses. Twenty-five isolates were collected from one of the wastewater samples and 35 isolates were collected for each of the three remaining wastewater samples. The dog, cattle and bird isolates were all collected from feces. Bird isolates were collected from a variety of wild birds including ducks, geese, pigeons and anhinga. No poultry isolates were included in this library. The 175 cow isolates represent both beef cow and dairy cow, but will be considered as one group, termed cow, for the analyses. Five isolates were collected from each of the fecal samples. After the library was collected and analyzed, the stability of the library was challenged. This was performed by analyzing 20 new known isolates from dogs, cattle and human sources and 18 isolates from birds (different from the ones already in the library) against the library to determine an average correct classification rate for the library. The library was also analyzed as a human or non-human library. This analysis was performed as the discrimination between human and non-human categories can be sufficient in some applications (i.e. determining septic vs. wildlife influence).

Sample Preparation

The wastewater, swabs and feces were collected by the laboratory staffs of Dr. V. J. Harwood Biology Department, University of South Florida and by the laboratory staffs of Dr. J. Rose, College of Marine Science, University of South Florida. Fecal matter was collected using sterile swabs. Swabs were placed in sterile tubes containing a 1ml of phosphate buffered saline and transported to the laboratory on ice. Each sample swab was spread onto mFC agar (Difco Laboratories) for the isolation of fecal coliforms the same day it was collected (Clesceri *et.al*. 1998). After 24 hours in a 44.5 ^{o}C water bath, typical fecal coliform colonies were aseptically transferred to EC/MUG media (Difco Laboratories). After 24 hour incubation at 37 ^{o}C, the cultures were exposed to UV light. Cultures that fluoresced blue were considered MUG positive. *E. coli* positive cultures were transferred and re-isolated on tryptic soy agar plates. The isolated colonies were then tested for a negative oxidase reaction. A subset (10%) of the isolates was verified as being *E. coli* using the API 20E system (BioMerieux, France). Isolates were preserved as a 50/50 mixture of 24 hour liquid culture and 7% working solution dimethyl sulfoxide at -70 ^{o}C.

RiboPrinter® Methods

The RiboPrinter® can process eight isolates simultaneously. Each *E. coli* isolate was prepared by picking a colony from a tryptic soy agar plate and suspending it in the buffer solution supplied with the RiboPrinter®. The samples were then heated at 80°C for 10 minutes and inserted into the RiboPrinter®. The instrument performs the following

steps during the eight hour sample processing: DNA preparation, separation and transfer, membrane processing and detection.

In the initial step of DNA preparation the cells are lysed, causing the release of the chromosomal DNA. The DNA is then digested in 100,000 U/mL *Hin*dIII for 20 minutes at 37 ^{O}C. The DNA fragments, created by the enzymatic digestion are loaded into a 0.8% agarose gel. A molecular weight marker is also added to the gel and the gel is run for two and a half hours. The resulting bands are transferred to a nylon membrane. The membrane is then hybridized to an *E.coli* 16S-rRNA probe. The membrane is washed and prepared for the final analysis. An image of the membrane with the chemically-labeled RNA is captured by the RiboPrinter's internal camera. The resulting banding pattern is then processed through proprietary analyses with the RiboPrinter® and placed into a ribogroup based on its similarities to others members of that group within the internal library.

Two major types of ribogroups can be formed. A single source category ribogroup contains isolates from only one source category. In a four source library these single source categories would be dog, bird, cow or human. In a two source category library the single source category ribogroups would contain either human or non-human isolates. A cosmopolitan ribogroup can also be formed and contains isolates from more than one source category group. In a four source library this can be any combination of isolates from two, three or four sources (i.e. dog and bird or dog, bird and cow, etc). In a human or non-human library the cosmopolitan groups will contain isolates from both human and non-human sources.

In order to test the accuracy of the library, 78 proficiency isolates were used: 20 isolates from dogs, 20 from humans, 20 from cattle and 18 isolates from birds. The isolates were collected from different individuals than were used in the library with the exception of the wastewater isolates. The wastewater isolates were selected from the same wastewater samples, but was not included in the library. These proficiency isolates were *E. coli* isolates from known sources that were not included in the creation of the library. The ribogroups formed by these proficiency isolates were compared to the ribogroups formed by the library isolates.

Data Analysis Software

The ribotype patterns created using the RiboPrinter® were also input into BioNumerics® software (Applied Maths Austin, TX) as a secondary analysis tool. Bands were identified using the auto band selection option with additional bands added in manually if not included in the auto band select. This software allows the user to characterize banding patterns using numerous similarity coefficients, including DICE, Peasron, and Cosine. For this study the DICE similarity coefficient was used to determine similarity between banding patterns. The analysis compares band positions at maximum similarity to determine relatedness.

BioNumerics also has programs that allow the user to perform more intra-library analyses, such as Jackknife and holdout analyses. This software was used to perform Jackknife (maximum similarity) analyses on the library as a four source library (dog, bird, cow or human) as well as a human or non human library. This analysis takes out one isolate from the library and puts it back in as an unknown and is then repeated for all

isolates in the library. The percentage of isolates that are correctly assigned to their original source category provides the user with the average rate of correct classification, and the rate of misclassifications of isolates into other source categories.

Individual holdout analyses were also performed. In these analyses all of the *E. coli* isolates obtained from one individual hosts were removed from the library and then classified as if they were unknowns. The use of two independent data analysis software packages (RiboPrinter® software and BioNumerics®) were compared and used to determine the library's applicability in BST.

Determination of Percent Similarities for the RiboPrinter® and BioNumerics®

Even in the same isolate slight differences will be seen when run multiple times. In order to determine the appropriate percent similarity settings, a control was used. This strain was an American Type Culture Collection *E. coli* strain 9637 that was run multiple times on the RiboPrinter®. The control was run 21 times, which spanned 14 gels. A gel image of an all control strain gel is shown in Figure 2. Seven of the eight times this strain was run the isolate was assigned to a single ribogroup based on the 95% mean similarity used by the RiboPrinter®. These same seven banding patterns were placed into one ribogroup by BioNumerics® at 90% similarity based on the dendogram constructed using the DICE similarity coefficient. Based on these findings 95% mean similarity was the cutoff for the RiboPrinter® analysis and 90% similarity as determined by DICE similarity coefficient was used for the BioNumerics® analysis.

Figure 1. Summary of isolate sampling for the creation of the library.

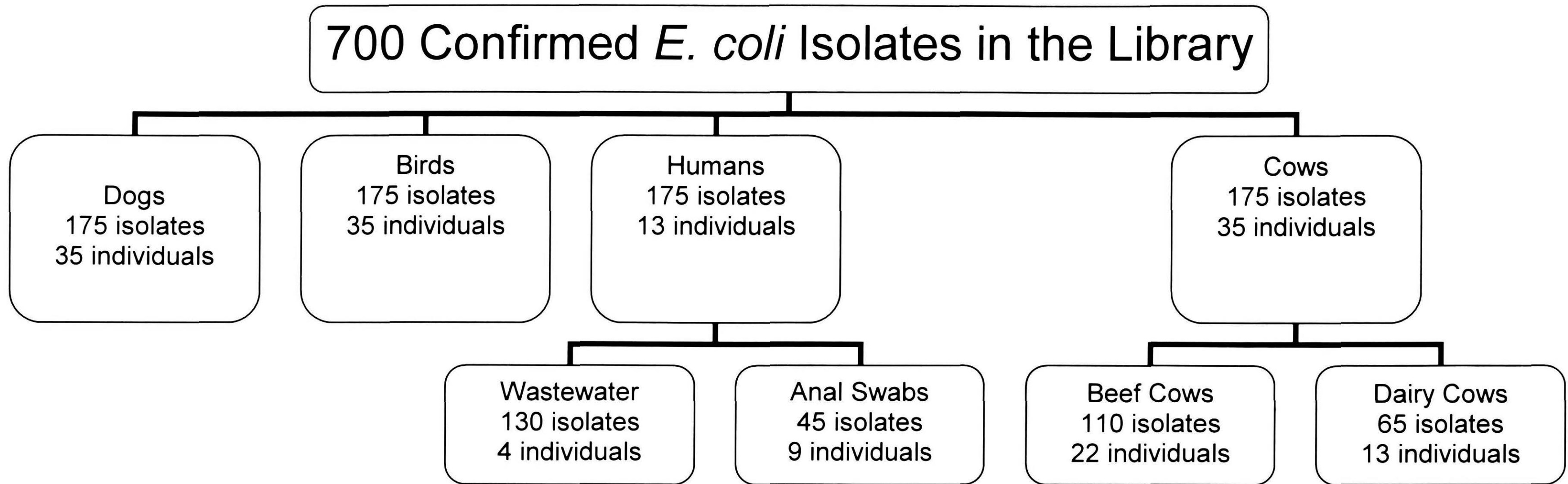
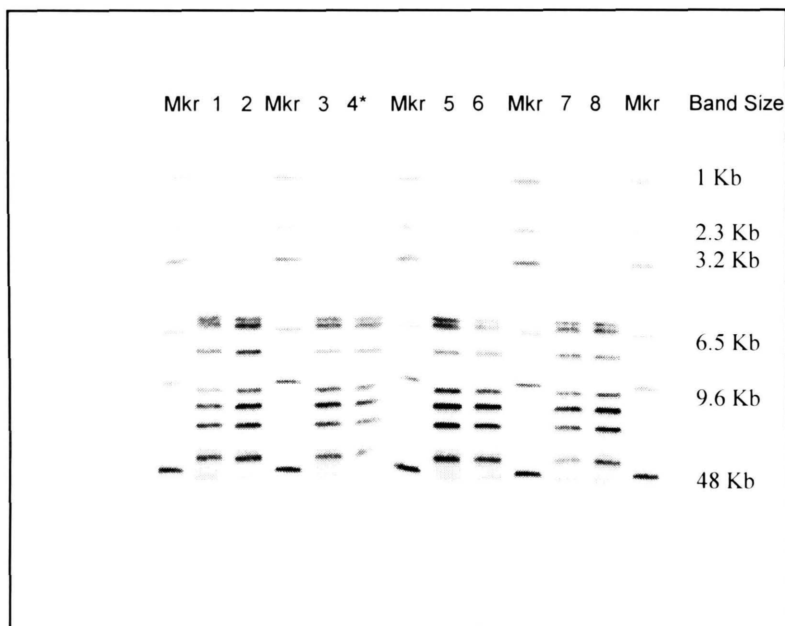


Figure 2. Gel image of a batch run with only control strain. Band size of the marker is displayed on the right. Mkr=marker lane. Numbered lanes are control strain lanes.



* Lane 4 was not used in the analysis due to a problem with the gel in lane 4 and the adjacent lane

Terminology

- Control Strain- An American Type Culture Collection(ATCC) strain that was ribotyped numerous times and used as a standard for determining parameters used on library isolates

- Cosmopolitan Group- A collection of isolates that form one ribogroup or BioNumerics® pattern that represent more than one source category

- Individual- a collection that contains five isolates from one dog, one bird, one cow or one human anal swab. When individual is used for wastewater, the number of isolates increases to 25 and 35 per sample

- Individual Holdout Analysis- An analysis performed using BioNumerics®. One individual is removed from the library and put back in as an unknown. Similar to Jackknife analysis.

- Isolate- A single pure culture of *E. coli* collected from fecal or wastewater sample

- Jackknife Analysis- An analysis performed by BioNumerics®. One isolate is removed from the library at a time and classified as if it was an unknown. The resulting table shows percentage of time the isolate gets placed into each of the defined source categories.

- Proficiency Isolate- A known source *E. coli* isolate that was collected in the same manner as the library isolates, but was not used in the creation of the library. These isolates were used as "unknown" isolates to test how well the library could identify them.

- Ribogroup- A collection of isolates sharing the same ribotype pattern

- Ribotype Profile- Pattern of bands created by the RiboPrinter® System. These bands are the basis for all analysis

- Source Category- Organism from which an isolate was collected (Dog, Bird, Human or Cow)

- Single Source Category- A category which contains only isolates that belong to one of the four source

Chapter Three: RiboPrinter® Results

Library Analysis

The library developed and used in this study contained 700 *E. coli* isolates from four sources: human, dog, bird and cattle. Based on the RiboPrinter®'s group assignment, 166 different groups were formed from the 700 isolate library. The summary of isolates and ribogroups is shown in Table 1. This table includes, for each source, the number of ribogroups containing any number of isolates from that source. The ribogroups may be single source category ribogroups, which contain only one source category (i.e. dog) or they may be cosmopolitan ribogroups, containing more than one source category (i.e. dog & bird or dog, bird and cow). Since this table shows all ribogroups for each individual source there will be overlap seen in the total number of ribogroups represented. Ribogroups contained between 1 and 36 isolates. The range of the number of isolates per ribogroup is also shown in Table 1. Human isolates (both wastewater and anal swabs) were represented in 48 different ribogroups. The wastewater isolates were found in 47 different ribogroups while the human anal swab isolates only divided into ten ribogroups. Nine of the human swab ribogroups overlapped with the wastewater ribogroups. Dog isolates were placed into 56 different ribogroups. Birds had the greatest diversity of all sources with 63 ribogroups containing bird isolates. Cow isolates (both beef and dairy) were found in 52 different ribogroups. Dairy cow isolates

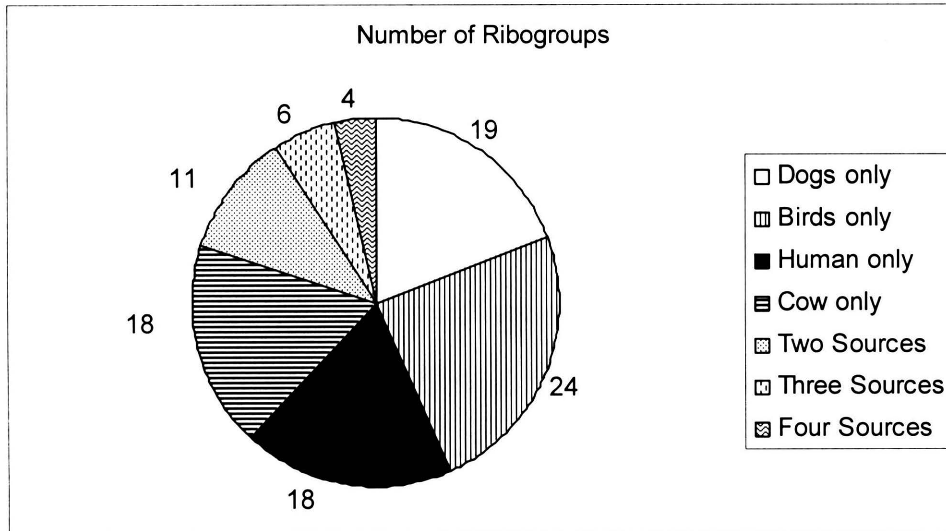
were dispersed among 34 ribogroups while the beef cow isolates were placed into 28 ribogroups.

Table 1. Results based on 95% mean similarity within source categories, as determined by the RiboPrinter® system.

Primary Source of Bacteria	Subgroup of Primary Source	Total Number of Isolates	Number of Ribogroups	Range in the Number of Isolates per Ribogroup
Dog	N/A	175	56	1-18
Bird	N/A	175	63	1-24
Human	**Wastewater**	130	47	1-19
	Swabs	45	10	1-14
	Total Human	175	48	1-20
Cow	**Beef**	110	28	1-25
	Dairy	65	34	1-9
	Total Cow	175	52	1-36

Groups formed by the RiboPrinter® were not always single source category groups. The distribution of ribogroups among source categories is shown in Table 2. The range of the number of isolates per ribogroup is also shown. Of the 166 ribogroups formed by the instrument, 133 ribogroups were single source category groups. The remaining 33 ribogroups were cosmopolitan groups. Eighteen of these cosmopolitan ribogroups contained isolates from two sources, nine ribogroups contained three sources and six ribogroups contained all four source categories. The percentages of ribogroups in each category are shown in Figure 3.

Figure 3. Percentages of Ribogroups per source category.



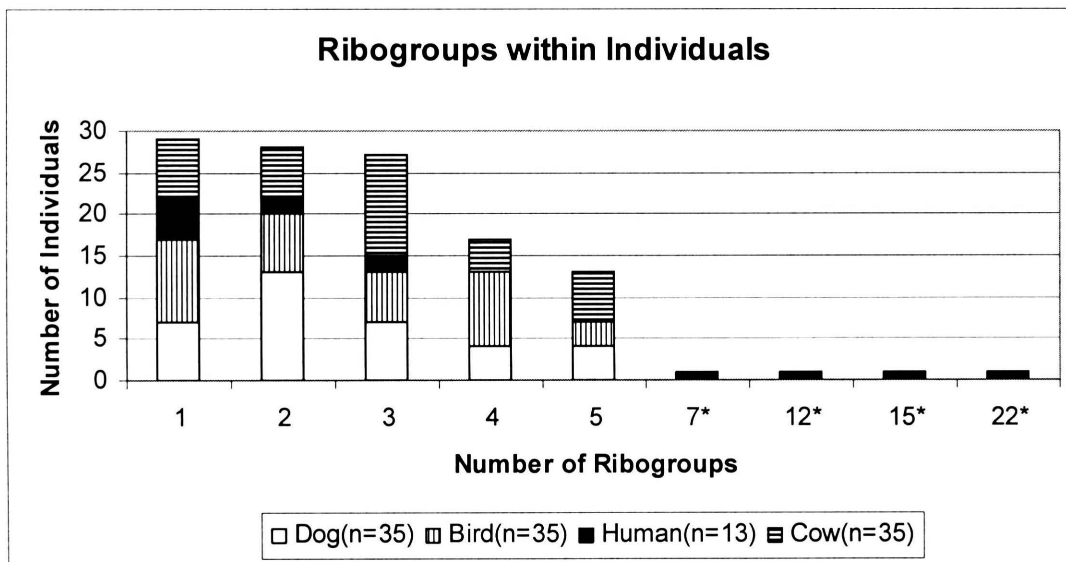
*Numbers shown are percentages

Table 2. Number of Ribogroups formed from single sources and cosmopolitan source patterns based on the RiboPrinter® system.

Source (s)	Number of Ribogroups	Range of Number of Isolates per Ribogroup
Dogs only	32	1-8
Birds only	40	1-18
Human only	31	1-7
Cow only	30	1-10
Dog & Bird	5	3-8
Dog & Cow	3	4-7
Dog & Human	2	4-14
Human & Bird	3	3-24
Human & Cow	2	3 & 4
Cow & Bird	3	4-7
Dog, Bird, & Cow	5	11-15
Dog, Human, & Cow	3	13-23
Human, Bird, Cow	1	8
All 4 Sources	6	14-82
Total	166	

The diversity of ribogroups associated within individuals is shown in Figure 4. All four source categories contained individuals with that had isolates assigned to different ribogroups. One ribogroup per individual represents clonality and five or more ribogroups per individual shows diversity within the individual sample. The human isolates had the widest variation with the human swab samples falling into one, two or three ribogroups per individual versus the wastewater samples (denoted by the *) that were distributed out among 7, 12, 15, and 22 different ribogroups. Birds had the largest number of individuals (ten) that were clonal.

Figure 4. Distribution of number of ribogroups per individual based on 95% mean similarity as determined by the RiboPrinter® System.



*Denotes composite wastewater samples. (n=25-30 per individual sample)

Proficiency Isolates

The summary of proficiency isolate assignments is shown in Table 3. The only single source ribogroup that dog proficiency isolates were associated with was the human ribogroups (10%). Bird proficiency isolates fell into bird only ribogroups (6%) and human only ribogroups (6%). Human proficiency isolates were found in human only

(10%) and bird only (5%) ribogroups. Cow proficiency isolates were not classified into any single source ribogroups from the library. Many (36%) of the proficiency isolates, formed new ribogroups and the isolates could not be assigned to any of the library ribogroups. The majority of the proficiency isolates (37%) were assigned to cosmopolitan ribogroups that contained two or three sources. Some (17%) proficiency isolates were placed into ribogroups containing all four sources.

Table 3. Results of Proficiency Isolate Ribogroup Assignment based on the four source category library.

Ribogroup Assignment \ Proficiency Isolates	**Dog (n=20)**	**Bird (n=18)**	**Human (n=20)**	**Cow (n=20)**
Dog	0	0	0	0
Bird	0	6	5	0
Human	10	6	10	0
Cow	0	0	0	0
Dog & Bird	0	22	0	0
Dog & Cow	25	0	0	0
Dog & Human	15	0	5	0
Human & Bird	0	6	0	0
Cow & Bird	0	0	0	5
Dog, Bird & Cow	15	16	0	10
Dog, Human & Cow	5	6	10	5
Human, Cow & Bird	0	0	0	5
Human, Cow, Bird & Dog	10	16	30	15
New	20	22	40	60
Total	100	100	100	100

*Numbers are shown as percentage of proficiency isolates belonging to each group.

Human or Non-Human Library

The library was also evaluated as a three source (human, non-human or cosmopolitan) library. The results of the ribogroups assignments are in Table 4. Human isolates formed 31 ribogroups with one to seven isolates per ribogroup. The non-human category contained the majority of the ribogroups (118) with the fewest isolates per

ribogroup, one to five. The greatest number of isolates per ribogroup was found in the cosmopolitan ribogroups, which contain human and non-human isolates. These 17 ribogroups had between three and 82 isolates per ribogroup. The cosmopolitan groups, 17% of all ribogroups, could not be classified as human or non-human.

Table 4. Number of ribogroups formed from single sources and cosmopolitan source patterns based on the RiboPrinter® system.

Source	Number of Ribogroups	Number of Isolates per Ribogroups
Human	31	1-7
Non-Human	118	1-5
Cosmopolitan	17	3-82

After the library was divided into the three source category groups the proficiency isolates were re-analyzed. The correct assignment of the human isolates remained the same at 10% (Table 6). The number of correct assignments for the non-human isolates increased from 6% (birds only in Table 3) to 33%. Many of the proficiency isolates were matched with the cosmopolitan ribogroups (which contained human and any non-human sources). As seen previously, the number of new ribogroups created by these isolates was large and accounted for 34% of the ribogroup assignments for non-humans and 40% for the human isolates for the proficiency isolates.

Table 5. Results of proficiency isolate ribogroup assignment based on the three source category library.

	Human	Non-human
Human	10	5
Non-human	5	33
Cosmopolitan	45	28
New	40	34
Total	100	100

*Numbers are shown as percentage of proficiency isolates belonging to each group.

Chapter Four: Results from BioNumerics®

Library Analysis

The 700 isolates used in this library formed 234 ribogroups based on the cluster analysis results from BioNumerics® (Table 6). Seventy three ribogroups were found in human isolates. The wastewater isolates (n=130) formed 63 ribogroups while the swab isolates (n=45) formed 15 ribogroups for the nine individuals sampled. Dog isolates (n=175) formed 74 ribogroups. The bird isolates (n=175) had the greatest number of ribogroups at 81. Cow isolates (n=175) had 71 ribogroups with the largest number of isolates per group being 25. Dairy and beef cow isolates showed a similar number of ribogroups, 38 and 42 respectively.

Table 6. Results from BioNumerics® showing number of ribogroups found by source.

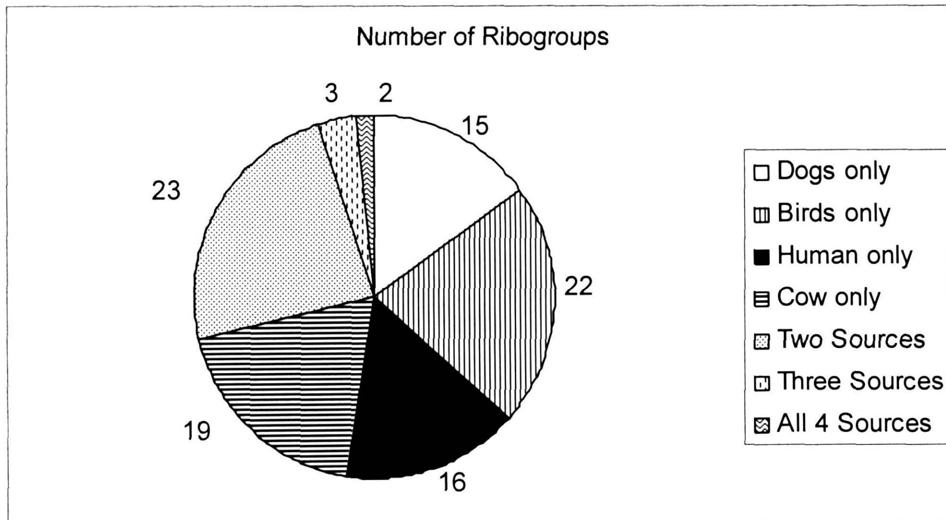
Primary Source of Bacteria	Subgroup of Primary Source	Total Number of Isolates	Number of Ribogroups	Range of Number of Isolates per Ribogroup
Dog	N/A	175	74	1-13
Bird	N/A	175	81	1-16
Human	**Wastewater**	130	63	1-9
	Swabs	45	15	1-14
	Total Human	175	73	1-16
Cow	**Beef**	110	42	1-18
	Dairy	65	38	1-8
	Total Cow	175	71	1-25

Based on the cluster analysis and the calculated 90% similarity cut-off, 167 of the ribogroups formed by BioNumerics® were single source category ribogroups (Table 7). The dog category had 35 ribogroups that contained one to eight isolates per group. The birds had 51 ribogroups that contained one to nine isolates per pattern. The 37 human only ribogroups had the fewest individuals per pattern (one to five) of any of the source categories. Cow isolates formed 44 ribogroups with one to six isolates per pattern. Only four ribogroups contained all four source categories, which also had the largest numbers of isolates per group, seven to 65. Of the remaining ribogroups, 55 had only two source categories per pattern, with less than 15 isolates per pattern. Eight ribogroups contained three source categories. The percentage of isolates per category is shown in Figure 5.

Table 7. Number of ribogroups formed from single sources and cosmopolitan source ribogroups based on the BioNumerics® cluster analysis.

Source (s)	**Number of Ribogroups**	**Range of Number of Isolates per Ribogroup**
Dogs only	35	1-8
Birds only	51	1-9
Human only	37	1-5
Cow only	44	1-6
Dog & Bird	8	2-10
Dog & Cow	15	2-5
Dog & Human	10	2-15
Human & Bird	5	2-6
Human & Cow	6	2-6
Cow & Bird	11	2-7
Dog, Bird, & Cow	1	12
Dog, Human, & Cow	2	6-7
Human, Bird, Cow	3	3-11
Human, Bird & Dog	2	21 & 34
All 4 Sources	4	7-65
Total	234	

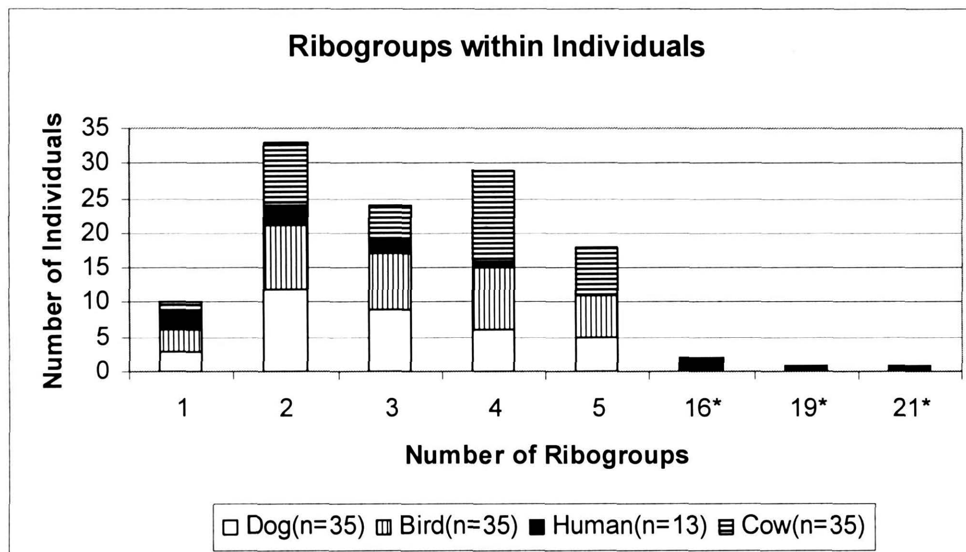
Figure 5. Percentages of ribogroups per source category.



*Numbers shown as percentages.

The diversity of ribogroups found within individuals is shown in Figure 6. Human swab isolates showed individuals with one, two, three and four ribogroups in a single individual. Two wastewater samples showed 16 ribogroups, while the other two wastewater samples showed 19 and 21 ribogroups per sample. Birds were evenly distributed across the two, three, four and five ribogroups per individual. Cattle showed two or four ribogroups per individual more frequently than they did any other number of ribogroups. Diversity of ribogroups in dogs peaked at two ribogroups per individual and decreased as the number of ribogroups per individual increased.

Figure 6. Distribution of number of ribogroups per individual based BioNumerics® results



Jackknife Analysis

The results of Jackknife analysis of the library revealed correct classification rates of 74.9% for bird isolates, 77.2% for cow isolates, 72.6% for human isolates and 75.4% for dog isolates (Table 8). For all four categories there were misclassifications, meaning some percentage of the isolates did not return to their original source category. Every source had some percentage of isolates that did not return to the original correct source category, which were 25.1% for bird isolates, 22.8% for cow isolates, 27.4% for human isolates and 24.6% for dog isolates. No incorrect classification rate exceeded 12% for the other source categories.

Table 8. Results of Jackknife analysis on the entire library of four sources.

	Bird	**Cow**	**Human**	**Dog**
Bird	**74.9**	9.1	11.4	6.9
Cow	10.3	**77.2**	10.3	6.3
Human	6.3	9.1	**72.6**	11.4
Dog	8.5	4.6	5.7	**75.4**
Total	100	100	100	100

*Numbers shown as percentages.

Proficiency Isolates

The same 78 proficiency isolates (20 from dogs, humans and cattle and 18 from birds) used with the RiboPrinter® analysis were also used to test the library using BioNumerics®. The results of the pattern assignments are shown in Table 9. The BioNumerics® library function places each proficiency isolate into a source category group based on what pattern an isolate clusters closest to. Therefore, there is no creation of new group categories, as seen previously in the RiboPrinter® analysis. No dog proficiency isolates, or any other source category isolates, were identified as dog. They were often (80%) misidentified as being from a cow source. Only 6% of the bird isolates were identified as bird, while the remaining isolates were identified as cow. Twenty-five percent of the human proficiency isolates were correctly, but frequently misidentified as bird (15%) and cow (60%). Ninety-five percent of the cow isolates were correctly identified, but 5% of the isolates were misclassified as being from a bird source.

Table 9. Results of proficiency isolate ribogroup assignment by BioNumerics®.

Source	Dog (n=20)	Bird (n=18)	Human (n=20)	Cow (n=20)
Dog	0	0	0	0
Bird	15	6	15	5
Human	5	0	25	0
Cow	80	94	60	95
Total	100	100	100	100

*Numbers are shown as percentage of test isolates belonging to each group.

Human or Non-Human Library

The library was re-analyzed as a human or non-human source library (Table 10). Human isolates still formed 37 ribogroups with only one to five individuals per pattern. The non-human ribogroups included 165 ribogroups with less than 12 isolates per pattern. The 32 cosmopolitan ribogroups contained between two and 65 isolates per pattern.

Table 10. Number of ribogroups formed from single sources and cosmopolitan source ribogroups based on the BioNumerics® software.

Source	Number of Ribogroups	Number of Isolates per Pattern
Human	37	1-5
Non Human	165	1-12
Cosmopolitan	32	2-65

Jackknife analysis was performed using human and non-human as the categories. This resulted in a rate of correct classification for human of 75% and non-human of 93.70%. With only two groups, the misclassification rate for human isolates was high at 25%. The rate of misclassification for non-human was very low 6.3%.

Individual Holdout Analyses

The percentages of isolates per individual that classified correctly during holdout analysis are shown in Figure 7. In this figure 0% means that none of the isolates from the individual were correctly classified. An individual represented as 100% correct classification had five out of five isolates classifying correctly. These results show that most of the dogs (n=32) had no isolates that were identified as dogs during the hold out analysis. The number of individual birds that had a portion of the five isolates correctly classified as birds was greater than was found in dogs, but the majority of the individuals had no isolates classify as birds. Four individual human swabs had all five isolates classify as human. No individual swab had all five isolates misclassified. Cattle had the best classification percentages among the individuals. All 35 individual cattle had greater than 60% of the five isolate correctly classify as cow. Human wastewater samples are shown in Figure 8. Wastewater individuals contained 25 and 35 isolates per individual, instead of five as used in other individuals, and did not classify into 20% increments as shown in Figure 6. Wastewater C had the lowest value of correct classification for all wastewater samples (6%). The largest correct classification rate was found in wastewater sample B (43%).

The classifications of isolates within individual host animals are shown in Figure 9 (A, B, C & D). These graphs do not show correct classification percentages; they demonstrate the distribution of source assignments for the entire source category. For example, if all of the isolates from an individual dog were classified as human source, those isolates would be part of the "human" component of the pie chart. If an individual dog had isolates classifying into multiple sources, three isolates to cow and two isolates into cow, this would be represented by the "dog and cow" component of the pie chart.

Isolates from individual dogs, birds and humans were most commonly misclassified as cow at 51%, 66% and 36% respectively. Cow isolates were most often misclassified as birds (20%).Isolates from individual humans and birds did not misclassify in the dogs.

Figure 7. Percentages of isolates that classified correctly during individual holdout analyses.

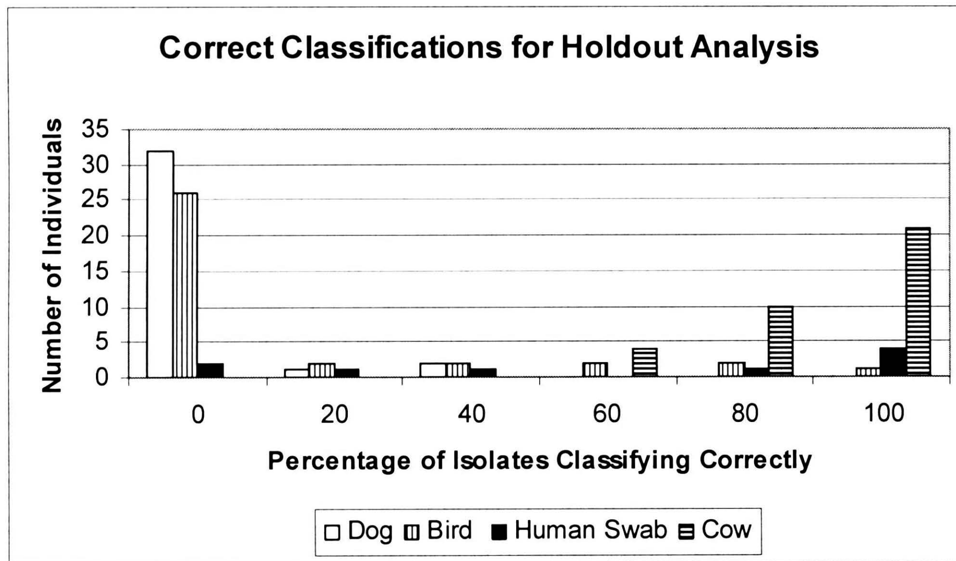


Figure 8. Results of correct classification for holdout analysis for wastewater samples.

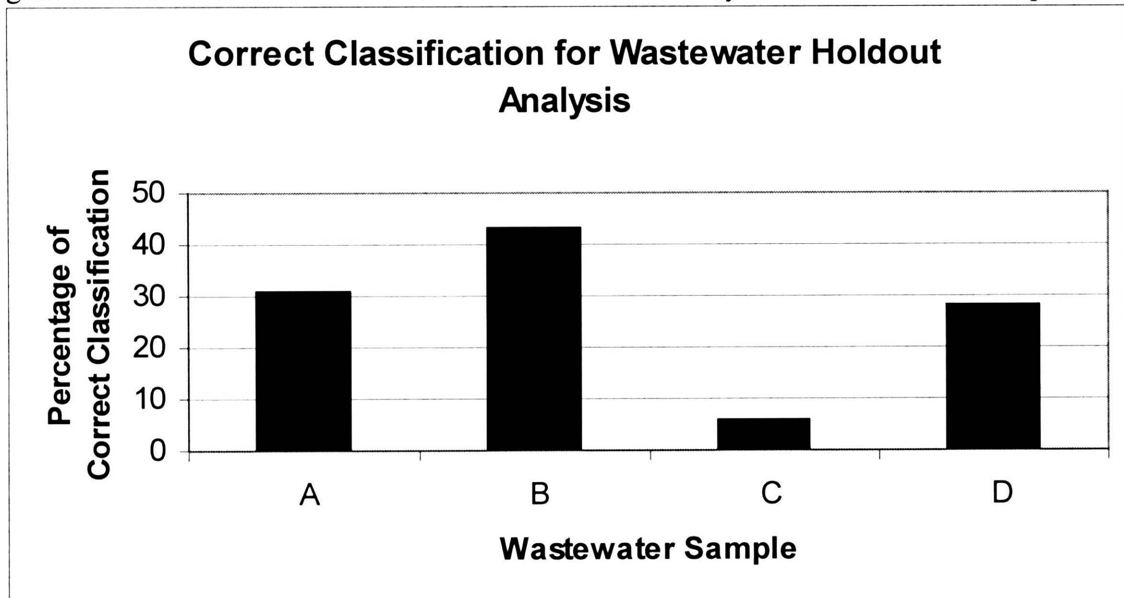
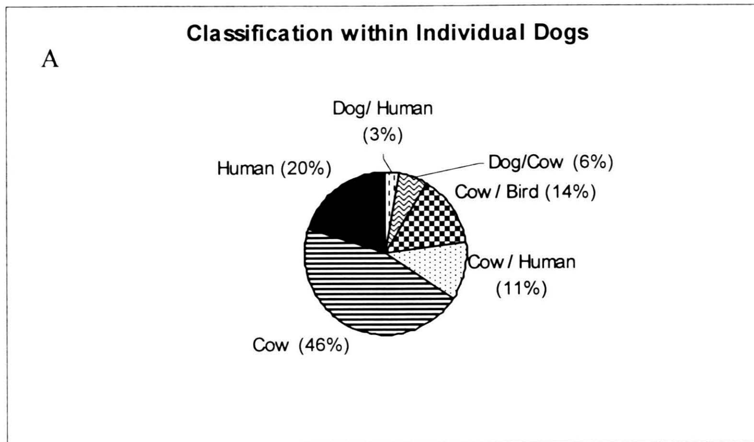
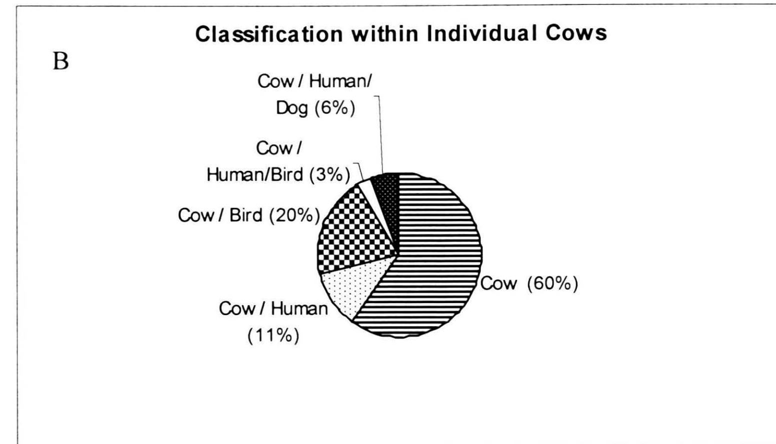
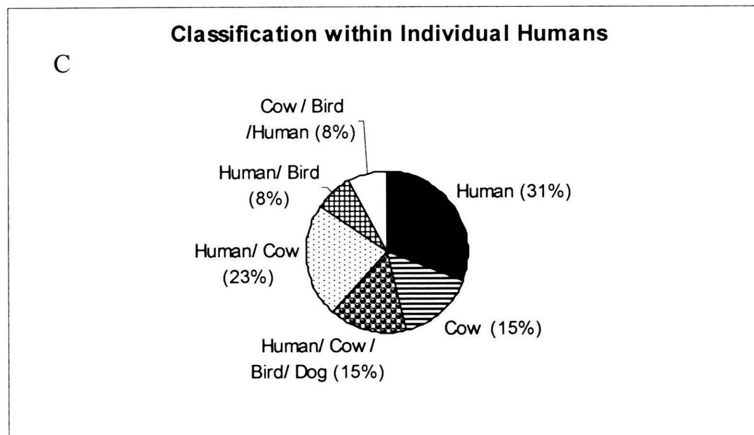
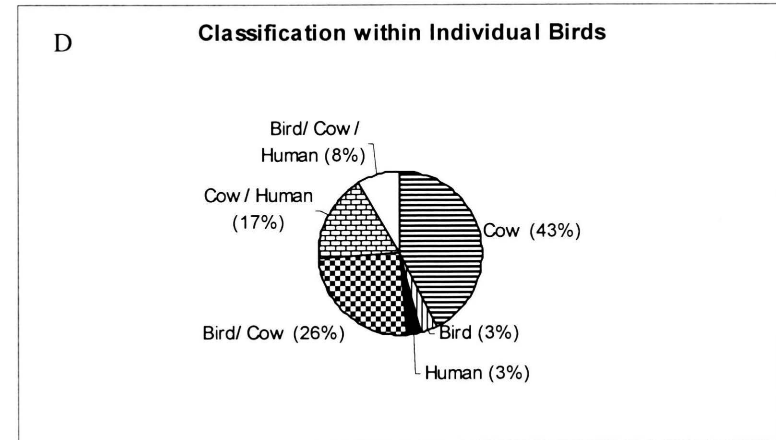


Figure 9. Classification of isolates within an individual sample.









Chapter Five: Comparison of RiboPrinter® and BioNumerics® Software

Both BioNumerics® software and the RiboPrinter® software were used to analyze the same ribotype profiles, but the differences in the statistical methods sometimes assigns the same group of isolates to different source categories. The gel in Figure 10 shows five isolates from one dog (lanes2-6), one isolate from a second dog (lane 1) and two isolates from a third dog (lanes 7-8). The RiboPrinter® assigned the eight isolates to four ribogroups. Lanes 3-6 were one group, lane 1 represents a group, lane 2 represents a group and lanes 7-8 represent the fourth group. BioNumerics® assigned the eight isolates to five groups. Lane 1 represents a group, lane 2 represents a group, lanes 7-8 represent a group, lane 3 represents a group and lanes 4-6 represent the fifth group. Even though lanes 3-6 appear to be the same ribotype profile BioNumerics® did not classify them as one group. Lanes 1, 2, 7 and 8 appear to be different ribotype profiles and were classified as different groups by both software analyses.

In contrast to the previous gel, the gel image shown in Figure 11 shows both software packages assigning the identical number of ribogroups to the eight cow isolates. Patterns in lanes 1-2 are from one individual, lanes 3-7 from a different individual and lane 8 from a third individual cow. Four ribogroups are identified by both the RiboPrinter® and BioNumerics®. Lane 1 represents a ribogroup, lanes 2 and 4 represent a ribogroup, lanes 3 & 5-7 represent a ribogroup and lane 8 represents the fourth ribogroup.

Figure 10. Gel image from the RiboPrinter® of eight dog isolates. Mkr= marker lane. Numbered lanes are sample lanes.

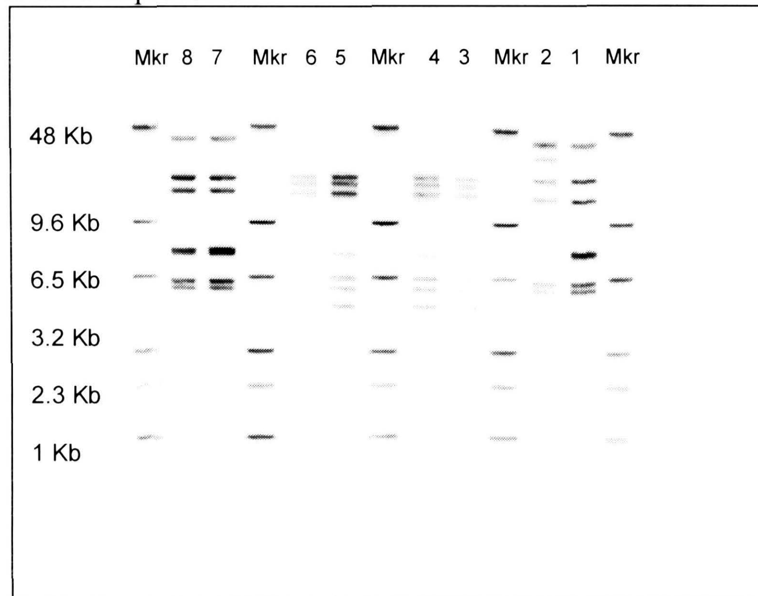
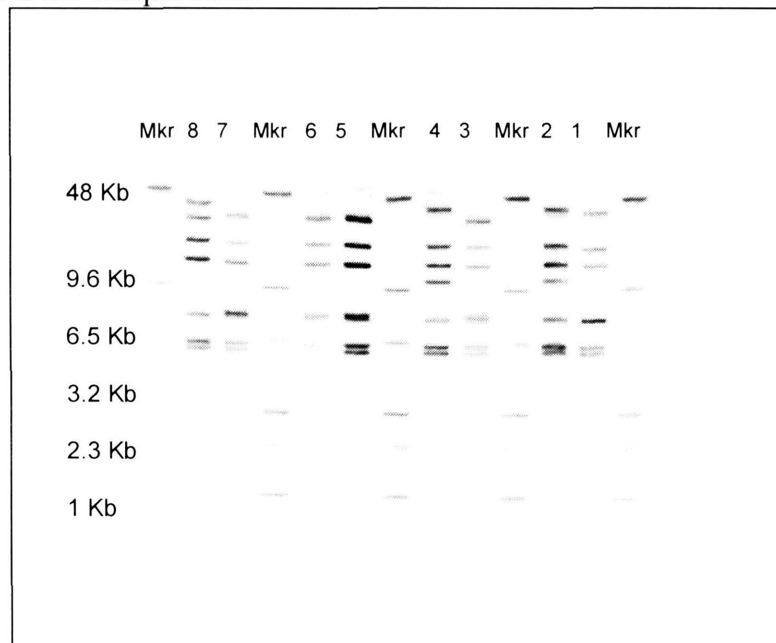


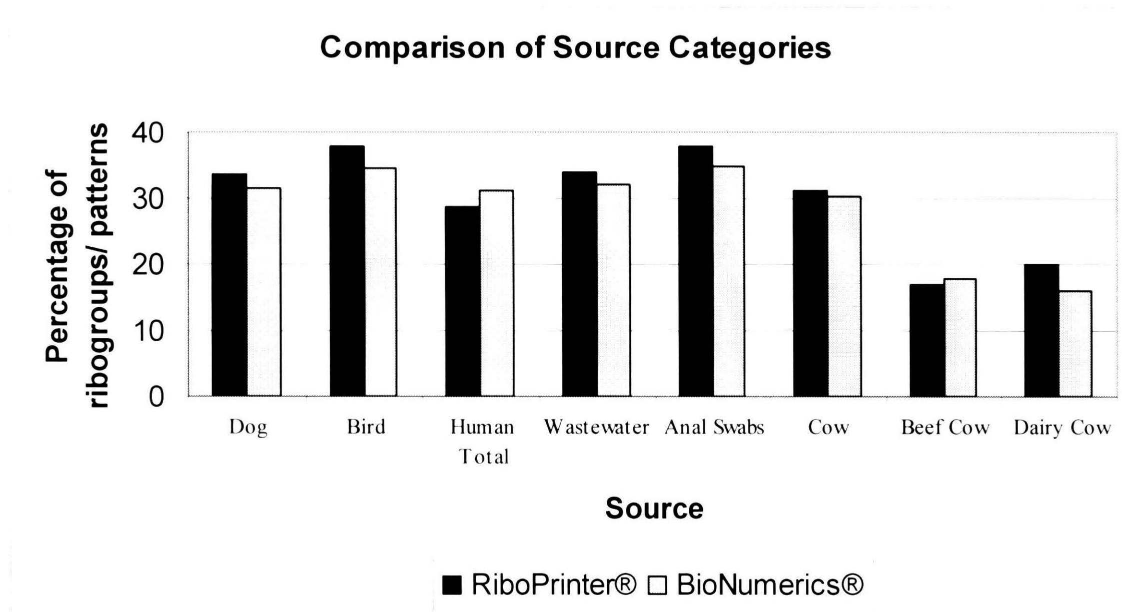
Figure 11. Gel image from the RiboPrinter® of eight cow isolates. Mkr= marker lane. Numbered lanes are sample lanes.



The values in the Tables 1 and 6 were converted to normalize the RiboPrinter® and BioNumerics® analyses (Figure 12). This was done for the RiboPrinter® data by

dividing the number of ribogroups by the total number of ribogroups formed (n=166) and for the BioNumerics® data dividing by the number of ribogroups formed (n=234). Table shows the range of percent values. Sources were between 0 and 4% difference when comparing between the RiboPrinter® and BioNumerics® data. One difference in the percentage of groups per source category is in the cow data. Using the RiboPrinter® software, beef cow isolates occupied 17% of the ribogroups and dairy cow isolates 20%. BioNumerics® placed beef cow isolates into 18% of the ribogroups and dairy cow isolates into 16% of the groups.

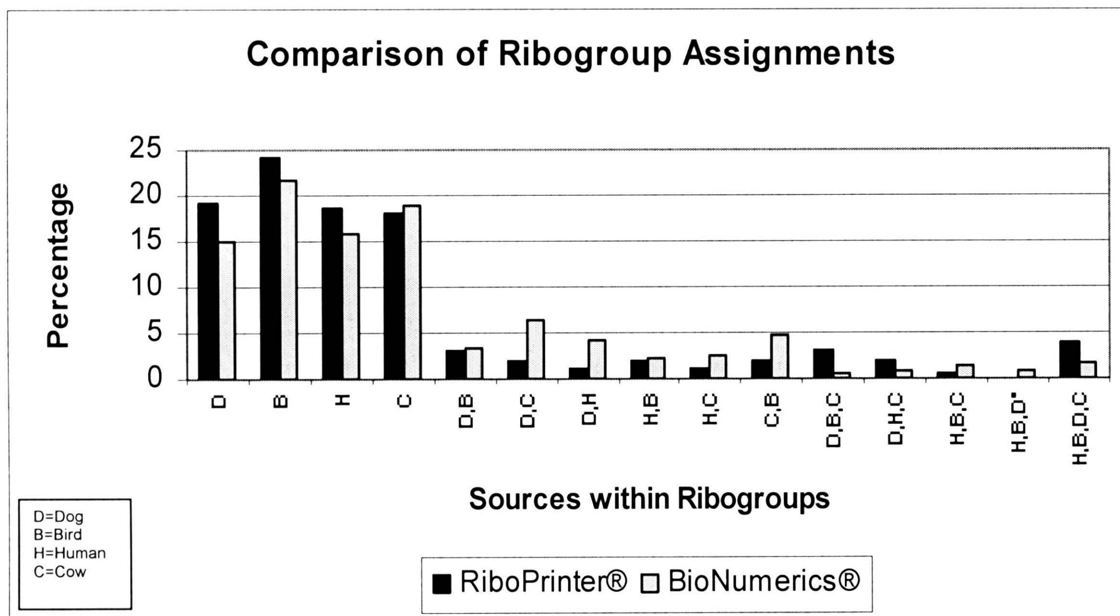
Figure 12. Percentage of ribogroups per source for the RiboPrinter® and BioNumerics®.



The source or sources belonging to the groups formed by the two software packages were examined. Again the numbers were normalized for direct comparison as previously discussed. The results are summarized in Figure 13. The RiboPrinter® software produced ribogroups containing 14 different source combinations while the BioNumerics® software generated 15 types of source combinations in the ribogroups.

These groups contained one, two, three or four source categories. Based on the Riboprinter® software, 80% of the different source combinations were single source categories, while 72% of the different combinations were single source categories using BioNumerics®. Two sources represented 23% of the ribogroups generated by BioNumerics®. Five percent of the groups contained more than two source categories per group. The RiboPrinter® results showed 11% of the ribogroups containing two source categories and 9% of the groups containing more than two source categories.

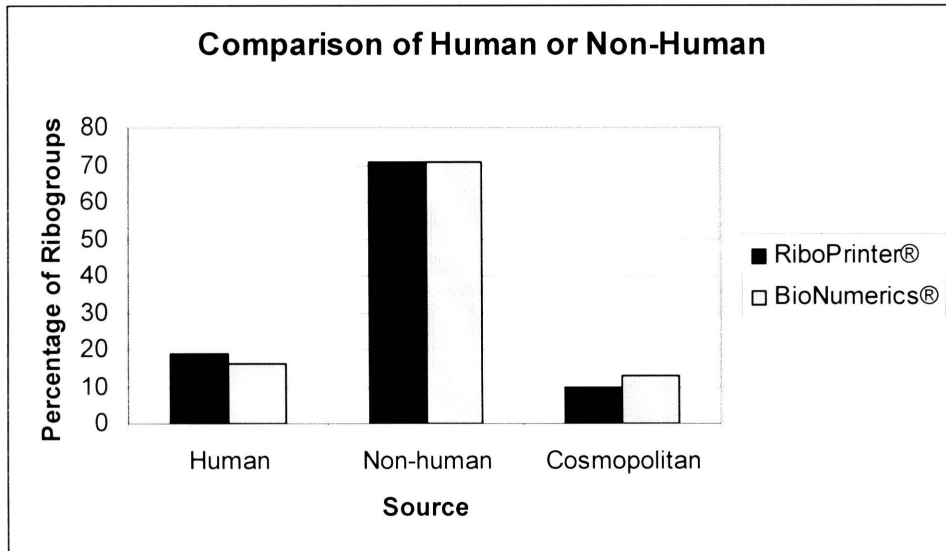
Figure 13. Percentage of ribogroups based on single source and cosmopolitan source ribogroups between the RiboPrinter® and BioNumerics®.



The library was also analyzed as a human or non-human library. Following the normalization of the data from Tables 4 and 10, as previously described, the RiboPrinter® assigned 19% of the ribogroups to the human category. Non-human ribogroups represented 71% of the total ribogroups and 10% of the ribogroups were

cosmopolitan, containing both human and non-human (Figure 14). The BioNumerics® software recognized 16% of the ribogroups as exclusively human and exclusively non-human ribogroups were 71% of the total. Cosmopolitan ribogroups accounted for 13% of the total ribogroups based on BioNumerics®. As was seen in the four source category library the percentage of groupings appears to be similar between the source categories.

Table 14. Percentage of ribogroups when library is human, non-human or cosmopolitan between the RiboPrinter® and BioNumerics®.



Chapter Six: Discussion and Conclusions

The RiboPrinter® Microbial Characterization System

The RiboPrinter® Microbial Characterization System was successfully used in this study to create ribotype patterns for 700 *E. coli* isolates. *Hin*d III was used in the digestions based on Tseng *et.al.* (2001) and Parveen *et. al.* (1999) comparison of multiple enzymes. The instrument allows for a standardization of the ribotype method, since few steps are completed outside the instrument. While numerous laboratories may use the same protocol, slight differences in technique or personal judgment may lead to differences in the resulting data.

A significant difference between the bench top and automated ribotyping is the amount of time needed to run isolates. The RiboPrinter® can process 32 isolates in a normal eight hour day, or 160 isolates per week (Bruce 1996). The same through put of isolates was found in this study. Bench top ribotyping can accommodate an average of 34 isolates per week by a single technician. This increase in throughput for the RiboPrinter® translates into quicker results or creation of a larger library in the same amount of time. The greater expense that comes with the use of the RiboPrinter® (the instrument alone is >$100,000) versus bench top methods may be offset due to the decreased labor hours required to process isolates.

RiboPrinter® Software and BioNumerics® Software Results for Library

The same 700 *E. coli* ribotypes were analyzed using both the RiboPrinter® software and BioNumerics® software. From the gel images shown in Figures 8 and 9 it is found that the software packages do not always group the isolates in the same fashion. Even though by eye, the patterns appeared the same in Figure 10 lanes 3-6, BioNumerics® placed the isolate in lane 3 in a different group than lanes 4-6. This may be due to the band identification system used in BioNumerics®. The band identification allows the user to auto select bands and or manually select bands, offering some external bias, by either the software or the person running the analysis, adding bands that are not truly present or by missing bands that are present. The percent classification that was used to determine groups likely accounted for some of the biases introduced in the band identification procedure. After testing the control isolates, 90% were determined to be a reasonable value, since all isolates formed one ribogroup. Most isolates that appeared to be different by visual analysis were considered different by BioNumerics®. The default setting of the RiboPrinter® is 95% mean similarity and based on the analysis of the control isolates no data supported manually decreasing this value.

The RiboPrinter® software classified the isolates into 166 distinct ribogroups while BioNumerics® separated the isolates into 234 different ribogroups. One would expect BioNumerics® to have fewer ribogroups, due to the decreased stringency in the determination of different groups. This was not the case. One explanation could be the difference in the algorithms used to calculate similarity.

Both software packages revealed a large percentage of ribogroups containing only one source category. The single source ribogroups are the most beneficial in BST. These

data are similar to the findings made by Kariuki *et.al.* (1999) that identified the majority of susceptibility profiles belong to only one source.

Upon comparison of the normalized values for the number of ribogroups per source and the distribution of source categories per ribogroup little difference (<4%) was found between the software packages. It is interesting that individual isolates are sometimes assigned to different ribogroups, but the overall distribution of sources within the ribogroups and composition of those ribogroups differs only slightly between the RiboPrinter® software and BioNumerics®. Tseng *et. al.* (2001) used the RiboPrinter® in conjunction with the BioNumerics® software, similar to this study, but did not compare the two software packages. Since the majority of applications of the RiboPrinter® have been for identification there are no studies with which to compare these findings.

Cosmopolitan ribogroups can be a problem in the assignment of an unknown isolate to a single source category. These profiles may be broadly distributed profiles that offer no discriminatory abilities. While these cosmopolitan groups may not be able to offer an exact source of contamination, they may be useful in determining what the source most likely is not. For example, unknown isolates from a water sample consistently go into a dog and bird cosmopolitan ribogroup, and then one could infer that cattle and humans may not be impacting the water body.

The profiles created with *Hin*dIII may show a lack of discrimination needed for BST. The combination of *Hin*dIII and a second digestion may reveal profiles that are distinct to each source category when ribotype profiles are combined into one composite

profile. Studies have suggested the inclusion of a second enzyme (Carson 1999 & Scott 2003) and specifically the use of *Pvu*II (Tseng 2001).

A major difference between the assignments of isolates to ribogroups can be seen when the distribution of ribogroups within an individual is shown (Figures 3 and 4). Using the RiboPrinter® software, there are approximately the same numbers of individuals having one, two, or three ribogroups within an individual. This represents a more clonal or less diverse population of *E. coli*. In the BioNumerics® analysis, the majority of the individuals had two or four ribogroups per individual, representing a more diverse population of *E. coli* in the library. The BioNumerics® data are consistent with Gordon & Lee (1999), showing that there is as much diversity within individuals as there is between source categories.

In order to test the library quality in this study, proficiency isolates were used. As described previously, these were known source isolates that were not part of the library. The use of proficiency isolates has not been shown in previous BST studies. The theory behind proficiency isolates is to use isolates that are not included in the library in order to test the library. The two software packages treat these isolates differently. The RiboPrinter® assigns a ribogroup to every isolate based on 95% mean similarity. If an isolate does not match another ribogroup by at least 95%, the isolate will be assigned to a new ribogroup. This allows for the proficiency isolates to not be characterized as any know source category ribogroup. These new categories created by the RiboPrinter® do not offer source discrimination needed for BST. They do offer an insight into how representative the isolates included in the library are of the true diversity seen in the isolates from each source. Using the RiboPrinter® software, only 4% of the proficiency

isolates were correctly categorized as a single source ribogroup (one bird isolate and two human isolates). These numbers are quite low, but when correct classification is expanded to include being placed in any ribogroup containing that source category, the percentage of isolates that were placed in the correct ribogroup increased dramatically to 58%. Using this expanded group (single source and correct cosmopolitan) may over estimate the actual correct classification rate of the proficiency isolates.

In this study, the library was not found to be very representative of source diversity. Only 40% of the cow isolates and 60% of the human isolates were represented in the library. Birds and dogs were represented well in comparison, approximately 90% for each source. This suggests that in general the library was unrepresentative of the natural diversity.

In BioNumerics®, all isolates were assigned to a pattern based on the library function in the program. This means that every proficiency isolate was given a source category and was not placed into a new pattern, even though the confidence the proficiency isolate assignment was low. BioNumerics® correctly identified 32% of the proficiency isolates. No dog proficiency isolate were correctly classified. The cow isolates were correctly classified 95% of the time by BioNumerics®, but never classified into a cow only pattern by the RiboPrinter®. Cattle were also the source category that the RiboPrinter analysis created the largest number of new ribogroups. Isolates from non-cow sources were most often misclassified as cow. The rates of incorrect classification for the proficiency isolates, with the exception of cow isolates, were low at 8%. Ninety-five percent of cow proficiency isolates were correctly classified. Again

these low rates of correct classification suggest that the library may not represent the source population very well.

The major difference was the two methods of analysis. The RiboPrinter® assigns a new ribogroup if an isolate doesn't fit into current ribogroups. From a clinical aspect, identification of unknown strains is important. For BST, a definitive answer as to the source category is what is sought. In order for the RiboPrinter® software to be applicable to environmental source tracking, a very large library needs to be used. This large library would hopefully be able to decrease the percentage of new ribogroups formed by the introduction of unknowns. BioNumerics® assigns every isolate to a pattern with a confidence in that assignment. While this method may allow for isolates to be misgrouped, the confidence percentage can be used to determine how much that assignment is supported.

There were low rates of correct classification of proficiency isolates, 58% for the RiboPrinter® and 32% BioNumerics®. With both software packages having less than 60% correct classifications of the proficiency isolates, it is difficult to determine which program is better for use in analysis. The two software packages, by two different companies, placed isolates into similar ribogroups and show similar diversity within the ribogroups. This supports the diversity within sources and within ribogroups is actually present in the library and not an artifact of analysis. There have not been studies that address differences between two analysis software packages used on the same library. For BST the standard method for testing the accuracy of any library has been the ARCC, usually calculated using a Jackknife analysis. This being the case, the use of BioNumerics® for the future analysis of this library is suggested.

The library used in this study does show trends that suggest its possible application to BST for Tampa Bay. In order for the current library to offer the discrimination needed more isolates are required to supplement the current 700 isolates from the four sources. An exact number of isolates that would be needed to supplement the current is unknown, since the ideal library size has not been determined. Isolates would need to be added in equal numbers across source categories and the library frequently reanalyzed. The library could be considered the optimal size when no new ribogroups are formed by the RiboPrinter® or when an acceptable rate of correct classification is achieved for the proficiency isolates. No value for correct classification rate has been determined, but ARCC of greater than 65% are considered sufficient in most BST studies.

Since the current 700 isolate library was limited as a four source library, a second round of testing as human or non-human was performed. As was shown when the library was treated as a four source library, the two software packages group isolates similarly when treated as a human or non-human source library. In a human or non-human library, cosmopolitan groups offer no discrimination between sources of unknown isolates. The library appeared to offer more discriminating power when only comparing human or non-human sources.

When the proficiency isolates were compared to this new three source library based on the RiboPrinter® analysis, the correct classification rates for non-human isolates improved. While the correct classification rates for the proficiency isolates has increased by this broadened library, the large number of cosmopolitan and new ribogroups is a problem for BST. A library based on human or non-human categories is

only applicable in a situation where a result of the non specific non-human source is sufficient to solve the source of the pollution. An example would be a small water body that is contaminated and managers are trying to determine if the source is the wildlife or damaged sewer lines.

BioNumerics® Additional Library Analyses

In addition to the analyses presented in comparison with the RiboPrinter® system, BioNumerics® has the ability to perform additional analyses on the library. The Jackknife analysis of the library offers a simple test of the library that can reveal much about composition. The results of Jackknife analysis give the average rate of correct classification (ARCC) for the library. This analysis also shows where the misclassifications of the isolates are located.

When the library was analyzed as a four source library the ARCC was 75%. The rates of correct classification for the four sources were 75% for birds, 77% for cattle, 73% for humans and 75% for dogs. These values are approximately three times higher than the probability of placing an isolate into the proper source category by chance (25%). The source of misclassification for any given source category is spread out among the remaining three sources. Birds and humans were most often misclassified as cow. Dogs were most often misclassified as human. Cattle, birds, and human isolates were least likely to be misclassified as dogs.

In a study by Carson *et. al.* (2001), the ARCC was 73.6% for a 267 isolate eight source category library. Since there were numerous source categories the range of rates of correct classification for any given source was 48.7% to 95.7%. It has been suggested

that smaller libraries have higher ARCC (Whitlock 2002), but this does not appear to be the case here. The ARCC for this study appears to be similar to the Carson *et.al* (2001) study.

The individual holdout analyses results suggest that the ARCC created by Jackknife analysis may be too generous for this library. The individual holdout analysis is basically an extension of the Jackknife analysis concept, removing an individual (multiple isolates) instead of a single isolate. Cow appeared to be the only source category that had a high rate of correct classification in the holdout analyses. There was no appreciable difference between cattle and dogs, birds or humans in the Jackknife results. Using the Jackknife analysis as the standard for determining the potential accuracy of a library may not be valid and other methods of internal library accuracy determination need to be found.

Both the Jackknife analysis and the individual holdout analyses are useful in predicting the accuracy that the library represents. These analyses are only performed using the isolates included in the library and may not be able to correctly predict the accuracy once unknowns are introduced. This is the reason proficiency isolates should be used to address how well a library will categorize unknown isolates. The proficiency isolates are treated as unknowns, are not included in the library and may therefore show how well the library will actually classify unknowns from a water sample.

BioNumerics® can classify isolates as human or non-human without having to use cosmopolitan categories as is needed with the RiboPrinter®. When the library is examined as human or non-human in BioNumerics, the increase in ARCC is large. The ARCC goes from 75% as a four source category library to 89% in a human or non-human

library. This increase is explained by the increased overlap between non-human source isolates versus human and non-human source isolates. Non-humans are classified correctly 93.70% of the time and humans are classified correctly 75%. Human isolate rates of correct classification did not change since the source, human, was not altered.

In the study by Carson *et.al.* the 267 isolate library that was created and treated as an eight source category library was also re-evaluated as a human or non-human library. The ARCC for this study increased from 73.6%, eight sources, to 97.1%, two sources. In a study by Parveen *et.al.* (1999) the ARCC for a human or non-human library was 82% based on a 238 isolate library. The ARCC of 89% found in this study is well within the range of published ARCC for similar studies using human or non-human as the source categories.

Another possible way to increase the discriminatory ability of this library would be to use composite ribotype profiles from digestion with two different enzymes. Tseng *et. al.* (2001) tested the use of *Eco* RI/*Pvu* II and *Hin*dIII/*Pvu*II for source tracking. The *Hin*dIII/*Pvu*II combination showed better discrimination, but due to the added time and expense only *Hin*dII was used in the study. In spite of the additional cost of a double enzyme digestion, this should be considered as a possible improvement to any library.

Conclusions

The RiboPrinter® Microbial Characterization System is an asset to the methodologies of ribotyping. The decreased labor and time to obtain results can benefit scientists in performing BST. This is accomplished by being able to get results about a source of pollution quicker or by being able to process more samples in the same time frame (allowing for the large libraries needed for source tracking).

The two software programs shared similarities as the number of ribogroups per source or sources was very similar when compared across both software packages. The average correct classification of the proficiency isolates was slightly better for the RiboPrinter® (58%) than BioNumerics® (51%). Although the RiboPrinter® was slightly better at identifying proficiency isolates, the software creates new ribogroups for anything not matching at 95% mean similarity. In a well sampled library, the creation of new groups would decrease. The creation of new ribogroups can be useful in determining how representative the library is of the actual diversity in the source population. The method BioNumerics® applied would be more useful on a library that may or may not be under sampled. By assigning every isolate to a source, BioNumerics® gives the best match for an isolate, rather than assigning the isolate to a new group. BioNumerics® gives a percentage confidence with each isolate assignment. BioNumerics® also offers more tools for research and analysis of the library, including Jackknife and individual holdout analyses. BioNumerics® appears to offer more options when using a small library, such as the one generated by this study.

Both software packages do show one thing, the library used in this study is not representative of the diversity found in *E. coli*. The poor rates of correct classification of the proficiency isolates support this. The 700 isolates and the ribogroups that they form provide foundation for future work. In order to increase the average rates of correct classification for the library, more isolates need to be added and the library retested frequently to determine the appropriate library size. To avoid further misclassification, no new source categories should be added until the library has been tested further.

In order to create a library that will be useful in Tampa Bay a large number of *E. coli* isolates will need to be collected from all areas surrounding the Bay, including but not limited to St. Petersburg, Sarasota and its surrounding towns and Tampa. Since the bay has multiple uses for agriculture, industry and recreation a variety of isolates from all of these uses should be included.

References

Barbour, W.M., Cover, C., Gannon, J., Cole, E., Bruce, J. & Webster, J. (1994) The use of ribotyping to track bacterial strains: determination of the origin of *Staphylococcus aureus* isolated from laboratory animals. In *Proceedings of the 94th American Society for Microbiology General Meeting*, Las Vegas, Nevada.

Bernhard, A.E. , Goyard, T., Simonich, M.T. & Field, K.G. (2003) Application of a rapid method for identifying fecal pollution sources in a multi-use estuary. *Water Research*. 37, 909-913.

Bernhard, A.E. & Field, K.G. (2000) Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Applied Environmental Microbiology*. 66, 1587-1594.

Bisson, J.W. & Cabelli, V.J. (1979) Membrane filter enumeration method for *Clostridium perfringins*. *Applied and Environmental Microbiology*. 37, 55-66.

Brisse, S., Fussing, V., Ridwan, B., Verhoef, J. & Willems, R. J. L. (2002) Automated ribotyping of vancomysin-resistant *Enterococcus faecium* isolates. *Journal of Clinical Microbiology*. 40, 1977-1984.

Bruce, J. 1996. Automated system rapidly identifies and characterizes microorganisms in food. *Food Technology*. January 1996, 77-81.

Bruce, J.L., Cole, E.M. & Wilson, R.A. (1997) Automated ribotyping, a standardized method for monitoring current and emerging populations of *Escherichia coli.* In *Proceedings from the 33rd Joint Conference U.S. Cooperative Medical Science Program Cholera and Related Diarrheal Diseases*, St. Petersburg, Florida.

Brujin, 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. *Applied Environmental Microbiology*. 58, 2180-2187.

Carson, C.A., Shear, B.L., Ellersiek, M.R. & Asfaw, A. (2001) Identification of fecal *Escherichia coli* from humans and animals by ribotyping. *Applied Environmental Microbiology*. 67, 1503-1507.

Clesceri, L.S., Greenberg, A.E., Eaton, A.D., eds. (1998) Standards for the Examination of Water and Wastewater, 20th ed. American Public Health Association. Washington, D.C.: American Public Health Association

Desmarais, T.R., Solo-Gabr Iele, H.M., and Palmer, C.J. (2002) Influence of soil on fecal indicator organisms in a tidally influenced subtropical environment. *Applied Environmental Microbiology*. 68, 1165-1172.

Dombek, P.E., Johnson, L.K., Zimmerly, S.T. & Sadowsky, M.J. (2000) Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. *Applied Environmental Microbiology*. 66, 2572-2577.

Feachem, R. (1975) An improved role for fecal coliform to fecal streptococci ratios in the differentiation between human and non-human pollution sources. *Water Research*. 9, 689-690.

Fujioka, R.S. amd Shizumura, L.K. (1985) *Clostridium perfringens*: a reliable indicator of stream water quality. *Journal of Water Pollution Control Federation*. 578, 986-992.

Geldrich, E.E. and Kenner, B.A. (1969) Concepts of fecal streptococci in stream pollution. *Journal of Water Pollution Control Federation*. 41, R336-R352.

Griffin, D.W., Lipp, E.K., McLaughlin, M.R. & Rose, J.B. (2001) Marine recreation and public health microbiology: quest for the ideal indicator. *Bioscience*. 51, 817-825.

Griffin, D.W., Stokes, R., Rose, J.B. & Paul, J.H. (2000) Bacterial indicator occurrence and the use of an F+ specific RNA coliphage assay to identify fecal sources in Homosassa Springs, Florida. *Microbial Ecology*. 39, 56-64.

Gordon, D.M., Bauer, S. and Johnson, J.R. (2002) The genetics structure of *Escherichia coli* populations in primary and secondary habitats. *Microbiology*. 148, 1513-1522.

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

Gordon, D.M. and Lee, J. (1999) The genetic structure of enteric bacteria from australian mammals. *Microbiology*. 145, 2673-2682.

Gordon, D.M., Riley, M.A. and Pinou, T. (1998) Temporal changes in the frequency of colicinogeny in *Escherichia coli* from house mice. *Microbiology*. 144, 2233-2240.

Gordon, D.M. (1997) The genetic structure of *Escherichia coli* populations in feral house mice. *Microbiology*. 143, 2039-2046.

Hagedorn, C., Robinson, S.L., Filtz, J.R., Grubbs, S.M., Angier, T.A. & Reneau, R.B. (1999) Determining sources of fecal pollution in a rural Virginia watershed with antibiotic resistance patterns in fecal streptococci. *Applied Environmental Microbiology*. 65, 5522-5531.

Hardina, C.M. & Fujioka, R.S. (1991) Soil: the environmental source of *Escherichia coli* and enterococci in Hawaii's streams. *Environmental Toxicology and Water Quality*. 6, 185-195.

Harwood, V.J., Whitlock, J., and Withington, V. (2000) Classification of antibiotic resistance patterns of indicator bacteria by discriminate analysis: use in predicting the source of fecal contamination in subtropical waters. *Applied Environmental Microbiology*. 66, 3698-3704.

Hsu, F., Shieh, C., Duin, J., Beekwilder, M.J. & Sobsy, M.D. (1995) Genotyping male-specific RNA coliphages by hybridization with oligonucleotide probes. *Applied Environmental Microbiology*. 61, 3960-3966.

Hazen, T.C. & Toranzos, G.A. (1990) Tropical Source Water. In McPeters, G.A. ed.Drinking Water Microbiology. New York:Springer-Verlag.

Judd, A.K., Scheider, M., Sadowsky, M.J. & Bruijn, F.J. (1993) Use of repetitive sequences and the polymerase chain reaction technique to classify genetically related *Bradyrhizobium japonicum* serocluster 123 strains. *Applied Environmental Microbiology*. 59, 1702-1708.

Kariuki, S., Gilks, C., Kimari, J., Obanda, A., Muyodi, J., Waiyaki, P. & Hart, C.A. (1999) Genotype analysis of *Escherichia coli* strains isolated from children and chickens living in close contact. *Applied Environmental Microbiology*. 65, 472-476.

Levin, B.R. (1981) Peroidic selection of infectious gene exchange and the genetic structure of *E. coli* populations. *Genetics*. 99, 1-23.

Levin, M.A., Fischer, J.R. and Cabelli, V.J. (1975) Membrane filtration technique for enumeration of enterococci in marine waters. *Applied Environmental Microbiology*. 30, 66-71.

Lipp, E.K., Farrah, S.R. & Rose, J.B. (2001). Assessment and impact of microbial fecal pollution and human enteric pathogens in a coastal community. *Marine Pollution Bulletin*. 42, 286-293.

Lipp, E.K., Kurz, R., Vincent, R., Rodriguez-Palacios, C., Farrah, S.R. & Rose, J.B. (2001) The effects of seasonally variability and weather on microbial fecal pollution and enteric pathogens in a subtropical estuary. *Estuaries*. 24, 266-276.

Olive, D. M. & Bean, P. (1999) Principles and applications of methods for DNA-based typing of microbial organisms. *Journal of Clinical Microbiology*. 37, 1661-1669.

Parveen, S., Hodge, N.C., Stall, R.E., Farrah, S.R. & Tamplin, M.L. (2001). Phenotypic and genotypic characterization of human and non-human *Escherichia coli*. *Water Research*. 35, 379-386.

Parveen, S., Portier, K.M., Robinson, K., Edmiston, L. & Tamplin, M.L. (1999) Discriminant analysis of ribotype profiles of *Escherichia coli* for differentiating human and non-human sources of fecal pollution. *Applied Environmental Microbiology*. 65, 3142-3147.

Pourcher, A.M., Devriese, L.A., Hernandez, J.F., & Delattre, J.M. (1991) Enumeration by a miniaturized method of *E. coli, Streptococcus bovis,* and enterococci as indicators of the origin of faecal pollution in waters. *Journal of Applied Bacteriology*. 70, 525-530.

Roll, B.M. & Fujioka, R.S. (1997). Sources of feacal indicator bacteria in a brackish, tropical stream and their impact on recreational water quality. *Water Science and Technology*. 35, 179-186.

Sadowsky, M.J., Kinkel, L.L., Bowers, J.H. & Schottel, J.L. (1996) Use of repetitive intergenic DNA sequences to classify pathogenic and disease-suppressive *Streptomyces* strains. *Applied Environmental Microbiology*. 62, 3489-3493.

Schaper, M., Jofre, J., Uys, M., & Grabow, W.O.K. (2002) Distribution of genotypes of F-specific RNA bacteriophages in human and non-human sources of faecal pollution in South Africa and Spain. *Journal of Applied Microbiology*. 92, 657-667.

Scott, T.M., Parveen, S., Portier, K.M, Rose, J.B., Tamplin, M.L., Farrah, S.R., Koo, A. & Lukasik, J. (2003) Geographical variation in ribotype profiles of *Escherichia coli* isolates from humans, swine, poultry, beef, and dairy cattle in Florida. *Applied Environmental Microbiology*. 69, 1089-1092.

Solo-Gabriele, H.M., Wolfert, M.A., Desmarais, T.R. and Palmer, C.J. (2000) Sources of *Escherichia coli* in coastal subtropical environments. *Applied Environmental Microbiology*. 66, 230-237.

Spratt, B.G and Maiden, M.C. (1999) Bacterial population genetics, evolution and epidemiology. *Philosophical Transactions: BioSciences (The Royal Society)*. 354, 701:710.

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

Tseng, C.C., Ting, W-T.E., Johnson, D., Thomas, G. & Adams, S. (2001) Automated ribotyping of *Escherichia coli* isolates from humans and animals. *Proceedings of the American Society for Microbiology 101st General Meeting*, Orlando Florida.

Turlak, A. Cole, E.M., Brinton, B., Eutopius, L., Samore, M. & Carroll, K.C. (2001) Comparison of double restriction enzyme digest procedure using the Qualicon RiboPrinter to pulsed field gel electrophoresis (PFGE) for strain characterization of vancomycin-resistant *Enterococcuc facium* (VRE). *Proceedings of the American Society for Microbiology 101st General Meeting*, Orlando Florida.

United States Environmental Protection Agency. 2001. Method 1602: Male specific (F+) and somatic coliphage in water by single agar overlay procedure. Office of Water. EPA 821-R-01-029.

United States Environmental Protection Agency. 2001. Method 1600: Membrane filtration test method for enterococci in water. Office of Water. EPA 821-R-97-004.

Whitlock, J.E., Jones, D.T. & Harwood, V.J. (2002) Identification of the sources of fecal coliforms in a urban watershed using antibiotic resistance analysis. Water Research 36, 4273-4282.

Wiedmann, M., Bruce, J.L., Knorr, R., Bodis, M., Cole, E.M., McDowell, C.I., McDonough, P.L. & Batt, C.A. (1996) Ribotype diversity of *Listeria monocytogenes* strains associated with outbreaks of Listeriosis in ruminants. *Journal of Clinical Microbiology*. 34, 1086-1090.

Wiggins, B.A., Andrews, R.W., Conway, R.A., Corr, C.L., Dobratz, E.J., Dougherty, D.P., Eppard, J.R., Knupp, S.R., Limjoco, M.C., Mettenburg, J.M., Rinehardt, J.M., Sonsino, J., Torrijos, R.L. & Zimmerman, M.E. (1999) Use of antibiotic resistance analysis to identify nonpoint sources of fecal pollution. *Applied Environmental Microbiology*. 65, 3483-3486.

Wiggins, B. (1996) Discriminant analysis of antibiotic resistance patterns in fecal streptococci, a method to differentiate human and animal sources of fecal pollution in natural waters. *Applied Environmental Microbiology*. 62, 3997-4002.

Wolf, J.W. (1972) The coliform count as a measure of water quality. In Mitchel, R. ed. Water Pollution Microbiology. New York: Wiley Interscience.