Molecular Analysis of Environmental *Escherichia coli* Using Repetitive DNA Sequences and the PCR primer BOXA1R to Differentiate Between Human and Animal Contamination Source

Trent Elliott and Christopher Kitts
Environmental Biotechnology Institute, Cal Poly, San Luis Obispo, CA 93407

Total coliform levels were monitored in San Luis Creek over a five-month period to determine contamination level of the recreational waters. From these frequent samplings, environmental *Escherichia coli* strains were isolated, denoting fecal contamination. The use of repetitive DNA sequences and the BOXA1R primer was used as a fingerprinting technique for each of the *E. coli* strains involved. Our study revealed that the BOXA1R primer was effective in clustering various *E. coli* strains of known origin according to organism source. Using a dendogram constructed by gel comparison software and a scatter plot of principle component values one and two, differentiation between human and animal *E. coli* strains was evident. Analytical methods successfully clustered the human source group apart from most of the animals and all of the environmental isolates. Thus, it was concluded that humans were not the source of contamination. The compilation of all results suggests that the use of rep-PCR performed with the BOXA1R primer is a viable option for determining fecal contamination sources.

Bacterial contamination of the nations' recreational waters with microbes is becoming an increasing public health concern and environmental problem. The concern with recreational water contamination is believed by many to be inadequately addressed by current water safety standards and research. The focus on water pollution problems has long been on chemical pollutants such as to overshadow the significant risks posed by microbial contamination. Recent reports by the Centers for Disease Control (CDC) document nearly 900,000 cases of a variety of illnesses occurring relative to waterborne microbial infections (Chang, 1997). Evidence suggests that a substantial amount of contamination results from disposing of waste into waterways and allowing the waste to filter into groundwater. Current American Society for Microbiology estimates that millions of septic tanks foster as much as 175 billion gallons of wastewater which may contaminate ground and surface waters with numerous pathogens, including viruses (1997). Additionally, it is known that farming cattle can pollute creeks with bovine fecal contamination harboring pathogenic strains of *Escherichia coli* like O157: H7 and a variety of other parasitic offenders.

According to the CDC, water is one of the most frequent vehicles for dissemination of microbial diseases, “responsible for the development of epidemics or widespread occurrences of disease in many people in short periods of time”(Howard, 1998). Contamination of water by fecal waste post water treatment may result in outbreaks of typhoid fever, shigellosis, and salmonellosis, gastroenteritis, giardiasis, dysentery, cholera and infectious hepatitis (1998).

Recreational hazards associated with microbial contaminated waters have been directly attributed to various ailments from such water contact. Recent epidemiological studies by the World Health Organization show a variety of acquired health problems due to recreational contact with contaminated water. These ailments include gastrointestinal disorders, various skin, and ear nose and throat infections (WHO, 1998). Of these, the most common complaint is that of gastrointestinal disturbance ranging from mild to severe. Reports reveal the most common and severe infections are in young children. Additionally the World Health Organization presents numerous acute febrile episodes as biologically plausible to recreational contact with contaminated water (1998).

In order to determine the amount of fecal contamination in a water supply, it is necessary to monitor the levels of coliform bacteria. Coliform bacteria are from the Enterobacteriaceae family, which are present as normal microbial flora of the gastrointestinal tract of a variety of organisms. Of the fecal coliform indicator organisms, *E. coli* is the most commonly used standard for microbial assessment of water contamination levels. Once high levels of fecal contamination have been discovered, determining the source of the fecal pollution is required before effective control measures can be instigated.

In this paper, we monitor the coliform levels from San Luis Creek in California over the period of five months. The weekly most probable number of coliform bacteria present is compared to environmentally accepted standard of 200 colony forming units (CFU) per 100 mL of water. Those MPN tubes that showed positive growth for coliform bacteria were then analyzed further to isolate *Escherichia coli* specifically. Following isolation, the utilization of repetitive DNA sequences and the polymerase chain reaction (PCR) method for molecular analysis of each isolate in order to determine possible organism sources of contamination. Genetically stable motifs that differ only in their copy number and chromosomal locations are fairly consistent throughout multiple species of bacteria, therefore these regions make them desirable targets for strain differentiation (Tyler, 1997). Collectively these elements have been referred to as rep-PCR. In rep-PCR DNA fingerprinting, PCR amplification of the DNA between adjacent repetitive extragenic elements is used to obtain strain specific DNA fingerprints, which can be easily analyzed with pattern recognition computer software (Dombek, 2000).

Materials and Methods

Most probable number of total coliform levels. Samples were typically taken during the morning or afternoon hours over a five-month period between September 1999 and February 2000. Attempts were made to consistently collect samples on a weekly basis, however, over certain time periods, collections varied in periodicity. All samples were collected in a like manner using sterile plastic Whirly bags. Samples were taken in the shallow, flowing waters of San Luis Creek. The bags, once opened, were submerged partially into the stream and water was allowed to flow into the opened container. The filling of the bags allowed one to two inches of air above the collected water. The sample collecting location was always upstream relative to the investigator. Two samples were collected and placed within a small cooler. Samples were immediately transported to the lab and typically processed within ten to fifteen-minutes. Upon arrival to the lab the creek samples promptly underwent a series of ten-fold dilutions with sterile 10mL dilution blanks through 10^{-3}. Those dilution blanks with final creek water concentrations of 0.1mL, 0.01mL, and 0.001mL were used to inoculate a five tube MPN series of 1% lactose broth according to the protocol of the American Public Health Association (APHA, 1985). After an eighteen to twenty-four hour incubation period at 37°C, tubes showing both acid and gas production denoted a positive result. Each of the MPN series were analyzed and observations were recorded.

Health Department *Escherichia coli* strains. During 1999, the San Luis Obispo County Health Department provided the Environmental Biotechnology Institute with a collection of *Escherichia coli* isolates that had been obtained and isolated from known organism sources. The *E. coli* strains were from several different humans, a couple of horses, a dog, cat, elephant seal, harbor seal, and sea lion. These strains were used as the standard that the environmental isolates were compared against to determine probable origin relatedness.

***E. coli* isolation and DNA extraction** . With an inoculation loop, samples from the least concentrated tubes that exhibited both acid and gas production were plated onto Eosine Methylene Blue (EMB) Agar using a three quadrant streak for isolation. Growth on EMB that is dark green to black colonies accompanied by a metallic green sheen across the surface of the media is a presumptive identification for *E. coli*. Following another incubation at 37°C, colony isolates, suspected to be *E. coli*, were each transferred to a separate Tryptic Soy Agar (TSA) plate. Each isolate then was subjected to the Api 20E identification system for *Enterbacteriaceae* and other gram-negative rods from bioMérieux. Only those colony isolates that were confirmed via this battery of twenty biochemical tests were used in subsequent analysis. DNA was extracted from each of the confirmed *E. coli* isolates using the UltraClean Soil DNA Isolation Kit from Mo Bio (Solana Beach, CA) with minor alterations to the given protocol. (Changes included: 1. Two loopfuls of colonies from

young 24 hour culture, 5. Bead beater settings of 4 m/s for 10 seconds, 6. One-minute centrifuge time, 7. 500_L of supernatant transferred, 17. Elute with PCR water rather than TE buffer.) Quantification of the DNA extraction was accomplished using a SPECTRAmax UV spectrophotometer at A_{260}. DNA extract was stored in sterile PCR water at -80°C until they were used.

PCR conditions and gel electrophoresis. The PCR was performed according to the conditions described by Rademaker and De Bruijn (Rademaker, 1997). Using the BOX A1R primer, consisting of the sequence (5'-CTACGGCAAGGCGACGCTGACG-3'), the PCR amplification of regions between primer specific sites were obtained, again, using the techniques outlined in Rademaker (1997). Each PCR was performed using Perkin Elmer GeneAmp PCR System 9600. The reaction was initiated with 2 minutes at 95°C followed by 30 cycles of 3 seconds at 94°C, 30 seconds at 92°C, one minute at 40°C, and 8 minutes at 65°C. The termination of the reaction requires an extension of 8 minutes at 65°C followed by a hold at 4°C until the samples can be taken from the machine. Both open and closed control wells were utilized to ensure that no other DNA contaminated the reaction set up. To confirm that the PCR protocol was successful, a 1.5% agarose gel was used to verify the presence of DNA from each sample as well as the absence of DNA in both the open and closed controls. The separation of various DNA fragments was achieved through the use of a Novex (Carlsbad, CA) 4%-20% TBE acrylamide gel. Each gel is precast and has ten wells, capable of holding 25_L. Into the first, fourth, and tenth wells was

loaded 5_L of a 1 Kb standard ladder. Into the remaining seven wells was placed 2_L of DNA, 1_L of 5X loading dye, and 2_L of TE Buffer. Each gel electrophoresis was performed in a vertical position, using 95 volts for approximately two hours. Final gels were stained with ethidium bromide, destained with deionized water, and photographed on a UV viewing plate.

Computer assisted analysis of DNA fingerprint. Gel images were loaded into GelCompar (BioSystematica, Tavistock, Devon, United Kingdom) analytical software. In this program, the images were normalized, bands were identified, and band patterns were compared. From the information obtained through GelCompar, an agglomerate hierarchical cluster analysis was performed to produce a relatedness dendogram4, somewhat similar in appearance to phylogenic tree. Statistical analysis of each band was performed through S-Plus 4.5 (MathSoft Inc., Cambridge, MA) software and clustering of gel pattern relatedness was organized using the first and second principle component values.

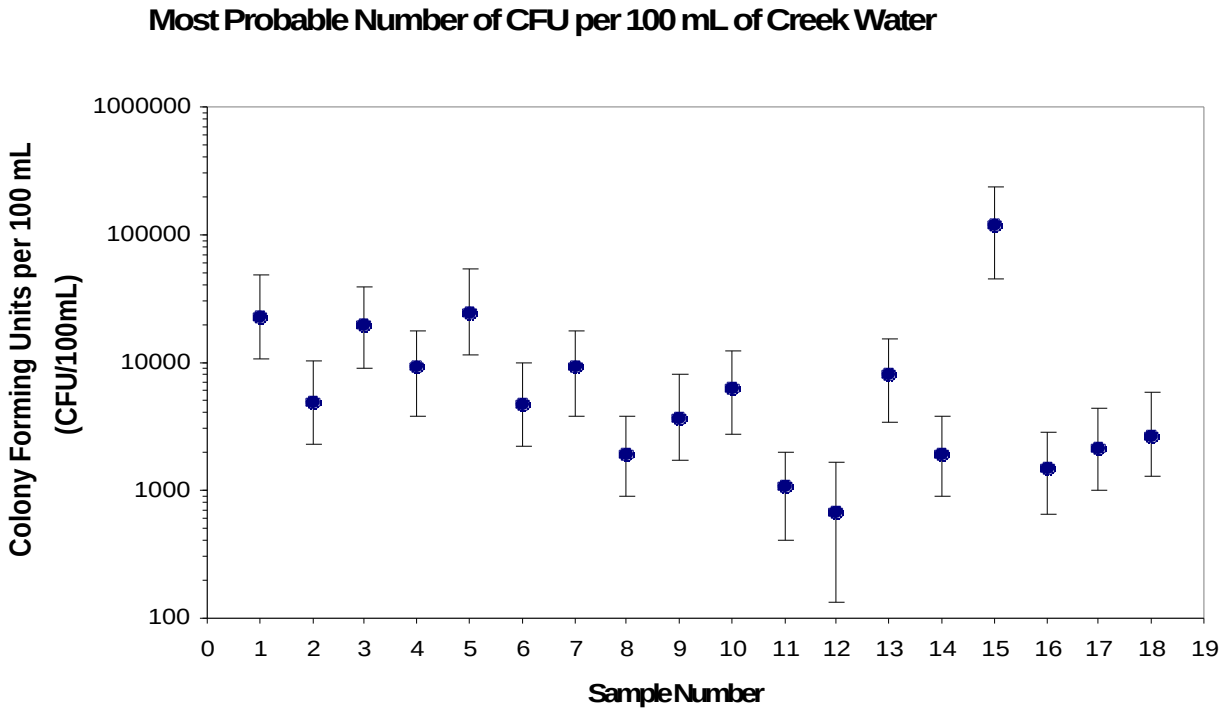

Figure 1: Most Probable Number values for total coliform levels from September 1999 through February 2000

Results and Discussion

Coliform levels in San Luis Creek . Over the five-month period that tests were performed to monitor MPN values, levels consistently exceeded environmentally accepted standards. Environmental standards established are 200

Colony forming units per 100 mL of water (EPA, 2000). Figure 1 displays the total coliform level determinations over the course of the study with a 95% confidence level. From the graph it is clearly evident that levels consistently and in some instances grossly exceeded the levels that have been deemed appropriate and acceptable. During samplings that followed a rain, there were extreme spikes in the coliform count, which is to be expected. The next step in this project focused on the possibility of determining the organism source for each of the *E. coli* strains isolated.

Analysis of DNA fingerprint . The PCR product that is loaded into the acrylamide gel undergoes constant voltage over a period of a couple hours. This process separates the amplified DNA fragments according to size. After the allotted time in the electrophoresis chamber, staining with ethidium bromide, and de-staining with DI water, distinct bands are present. This band pattern is referred to as the DNA fingerprint. Figure 2 shows a common DNA fingerprint as a result of this process.

Each of the individual lanes entered into the GelCompar program were normalized into straight columns. Individually, these normalized lanes were studied and band assignments were given. Often, debris from the surface of the UV viewing table appeared on the gel image as dark spot. The computer program initially assigned these dark regions as a DNA band. The manual filtering of each lane was necessary before

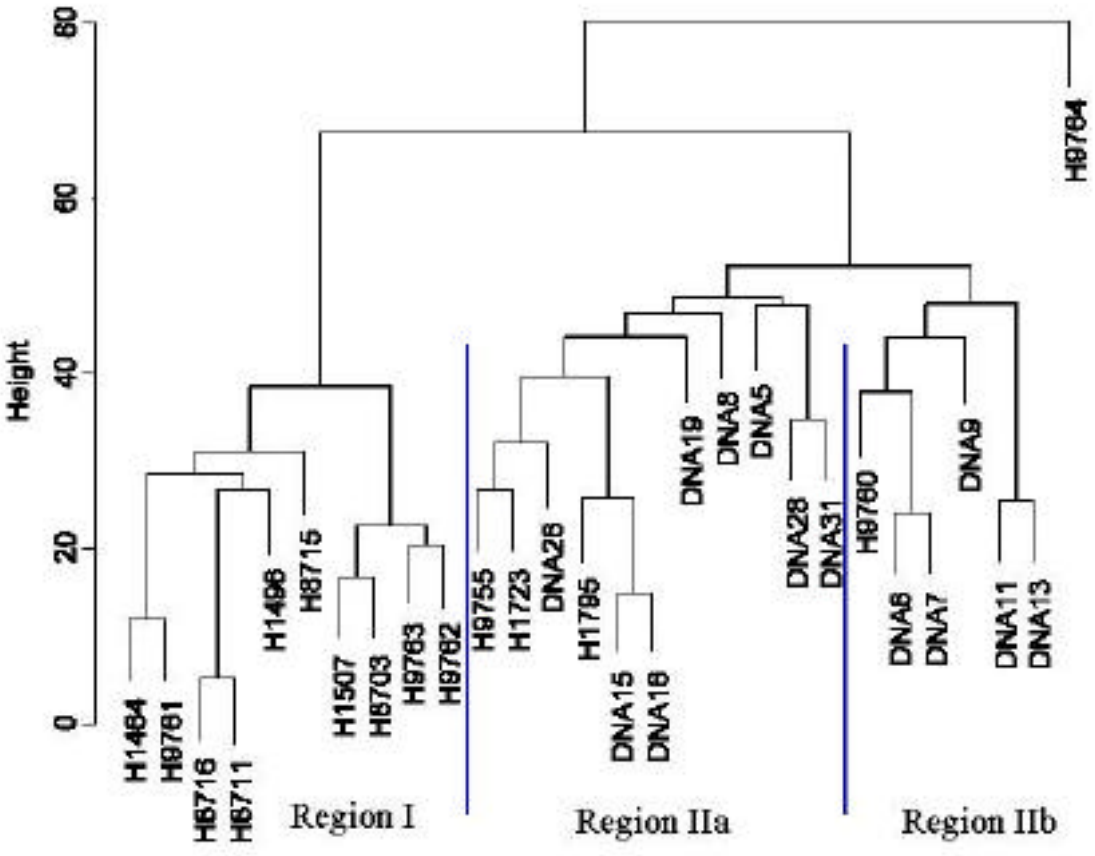

Figure 3: Dendogram constructed from GelCompar analysis. Bands with similar fingerprint patterns are clustered in proximity to each other on the tree.

further analysis was possible. In order to correct the smile in the gel that resulted from a warped

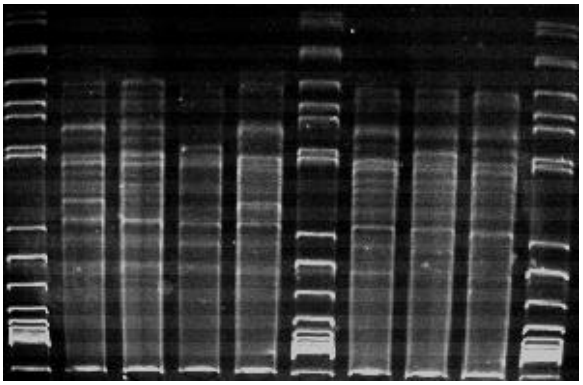

Figure 2: Repetitive PCR final products undergo electrophoresis on acrylamide gel. The resulting DNA fingerprint is generated and the band patterns are compared. This gel represents the band patterns from seven environmental *Escherichia coli* isolates.

electric field during electrophoresis, the three separate lanes of 1Kb Standard Ladders were needed. From those standard lanes, bands in each lane that corresponded to the same DNA fragment were aligned.

The band patterns in every lane were compared to each of the others and the computer program grouped them according to similarity. From the similarities a dendogram was produced that showed the relatedness of each *E. coli* strain. It is from this relatedness and natural clustering that inferences regarding possible organism sources can be obtained. Figure 3 shows the resultant dendogram from this set of *E. coli*. This method of clustering clearly depicts three individual regions of relatedness. Clustered in Region I are solely health department strains including the dog, cat, and harbor seal, however, the majority of which are human in origin. Region IIa contains the horses, the elephant seal, and the majority of our environmental isolates. Grouped together in Region IIb are the remainder of the environmental isolates along with the sea lion and a stray human strain. A comparison of the dendogram with the graphical depiction of the clustering according to principle components one and two (Figure 4) further illustrates the tendency of grouping by organism source.

The most prominent band that separated out the humans, dog, cat, and harbor seal was located at position 224 on the gel. This is the only band that was in common with each of the organisms that were mentioned previously. The horses, which separated out quite nicely, share bands at positions 139, 200, and 269. There are no apparent similarities between many of the other strains.

Many of the environmental isolates clustered together on the graph, however, they were sufficiently removed from Region I to conclude with a great deal of assurance that they were not from humans, dogs, or cats. In fact, the majority of the environmental isolates were positioned on the graph such as to suggest that the carnivorous animals were not the likely source of the contamination. Other possible sources of contamination are ducks, pigeons, and bats. Each of these animals is part of the creek ecology and may be significant contributors to the coliform count in the creek water.It was concluded that the BOX primer was able to cluster each of the health department strains of know organism source out rather efficiently. However, similar complications that have existed in previous experiments using ERIC primers and REP primers still persist. It is not understood what differences exist that create the separation of the various *Escherichia coli* strains from different species of organism source. One possible explanation is that the variety of *E. coli* that has inhabited the gastrointestinal tract of any particular organism is dependent in part to that organism's diet. Similar trends have been noted in clustering analysis resulting from each of the separate primer sequences. It appears that the carnivorous animals such as humans, dogs, and cats will group in one area of the graph while grazing animals like cows and horses will be in different regions of the graph. This theory is only a possible explanation, and no other evidence is given to support this claim.

Unfortunately there were no known samples that had been obtained from cows, therefore it is not possible to know exactly where *E. coli* from this particular organism would cluster. The fact that the Cal Poly dairy and hundreds of acres of grazing land are located upstream from the testing site, it is very likely that many of the environmental *E. coli* isolates are from cattle.

Acknowledgments

I'd like to thank the Environmental Biotechnology Institute at Cal Poly for the use of the lab equipment and supplies. Also, a special thank you to Dr. Christopher Kitts for this time, patients, and dedication to helping this project reach completion.

Works Cited

APHA, et. al., 1985. Standard Methods for the Examination of Water and Wastewater 16th ed. American Public Health Association, Washington, DC. p. 870.

Chang, C. 1997. Arch water in the news: Microbes endanger U.S. Waters. American Society for Microbiology. May; p. 12-17.

Dombek, P. E., L. K. Johnson, S. T. Zimmerley, and M. J. Sadowsky. 2000. Use of repetitive DNA sequences and the PCR to differentiate *Escherichia coli* isolates from human and animal sources. Applied and Environmental Microbiology. Vol. 66, No. 6; p. 2572-2577.

Environmental Protection Agency. 2000. Bacterial Water Quality Standards for Recreational Waters (Fresh and Marine). Region 9, State - California, Class - Central Coast Region. (Site accessed June 2000). http://www.epa.gov/OST/beaches/local/sumtable.html

Howard, Barbara J.; et. al. Clinical and Pathogenic Microbiology 2nd ed. Mosby-Year Book, Inc. St. Louis, MO. p. 11.

Rademaker, J.L.W., and F.J. de Bruijin. 1997. Characterization and classification of microbes by rep-PCR genomic fingerprinting and computer-assisted pattern analysis, p. 151-171. In G. Cactano-Anollés and P. M. Gresshoff (ed.), DNA markers: protocols, applications, and overviews. J. Wiley and Sons, New York, N.Y.

Tyler, K.D., G. Wang, S.D. Tyler, and W.M. Johnson. 1997. Factors affecting Reliability and Reproducibility of Amplification-Based DNA Fingerprinting of Representative Bacterial Pathogens. Journal of Clinical Microbiology. Vol. 35, No. 2; p.339-346.

World Health Organization. 1998. Guidelines for Safe Recreational-Water Environments: Coastal and Fresh Waters. Geneva;70.