Report as of FY2007 for 2006MT98B: "Student fellowship: A genomic and proteomic approach to characterizing natural variation in E. coli: Toward construction of a microbial source tracking database to identify sources of fecal water contamination in the State of Montana"

Publications

Project 2006MT98B has resulted in no reported publications as of FY2007.

Report Follows

Margie Kinnersley
Montana Water Center Student Fellowship Final Report
June 22nd, 2007

Research Project Title: A Genomic and Proteomic Approach to Characterizing Natural Variation in *E. coli:* Toward Construction of a Microbial Source Tracking Database To Identify Sources of Fecal Water Contamination in the State of Montana.

Abstract

The overarching goal of this study is to characterize variation in naturally occurring *E. coli* populations in Western Montana at the genomic, transcriptomic and proteomic levels and to correlate this data with animal host species information to identify animal sources of fecal water contamination. A rep-PCR fingerprint database has been created for the Many Glacier region of Glacier National Park that can be used to roughly classify unknown *E. coli* isolates. Preliminary 2D gel protein profiles for a subset of human, bear and deer *E. coli* show that there are a number of protein composition differences that may be useful for distinguishing human and animal isolates. Microarray-based comparative genome hybridization analysis has also revealed several differences in genome composition that are being investigated as potential biomarkers.

Objectives

The proposed research objectives are as follows:

Objective 1: To collect fecal samples and isolate *E. coli* from humans and several different species of local wild mammal that may be potential nonpoint water contamination contributors.

Objective 2: To characterize genetic and phenotypic differences in *E. coli* populations within individual animals and between host species.

Objective 3: To utilize genotypic and phenotypic information to determine the origin of *E. coli* isolated from contaminated water sources using a relational database.

Project Progress

Objective 1: One hundred and fifty-three *E. coli* strains have been isolated from the feces of ten different species of animal that reside in the Many Glacier area of Glacier National Park, Montana, USA per EPA method 1603. In addition, a reference strain collection (the ECOR collection) consisting of 41 human and 33 animal *E. coli* isolates has been obtained from Michigan State University and 253 human, sheep, goose, cow and pig isolates have been generously donated by Dr. Michael Sadowsky at the University of Minnesota.

Objective 2: Traditional rep-PCR fingerprinting using the BoxA1R primer (genotypic portion of objective 2) has been completed for all Montana isolates. A small database has been constructed for the Many Glacier region of Glacier National Park. Jackknife analysis (a measure of the internal consistency of the database that assesses how often a fingerprint is correctly re-classified into its original group) has been performed. In general, the rate of correct re-

assignment of the fingerprints when each animal species is considered a separate group was somewhat low with an average of 52%. Values ranged from 0% for moose to 100% for coyote. These low values are likely due to the fact that a small number of morphologically distinct colonies were selected from a relatively small number of animals across a wide range of species that all share a common habitat. The overall performance of the database may be improved by the addition of more isolates from each species.

A modified rep-PCR technique in which restriction enzymes were used to create additional fingerprint variation was also applied to a subset of strains from Montana. This new type of fingerprinting has identified a gene region that is widely distributed in human *E. coli* but occurs infrequently in *E. coli* from animal feces. A PCR assay was designed to test for the presence or absence of this marker. Results indicate that 67% of the human isolates from the Many Glacier Hotel raw sewage possess the marker while 95% of the animal strains from the surrounding area do not (Table 1). Eighty-five percent of the human isolates from the ECOR collection are also positive for the marker while only 42% of the Minnesota human isolates and 19% of the Montana clinical isolates score positive. The lower numbers for the clinical and Minnesota strains may be due to differences in how the *E. coli* were initially isolated or may simply reflect geographical variation in genome content. Experiments to determine the effects of isolation technique on this assay are currently underway.

Table 1. Distribution of putative marker identified by rep-PCR in Glacier National Park and ECOR strain collections

	+ for marker	**- for marker**
Glacier human isolates (9 total)	6 **(67%)**	3 **(33%)**
Glacier animal isolates (64 total)	3 **(5%)**	61 **(95%)**
ECOR human isolates/lab strains (39 total)	33 **(85%)**	6 **(15%)**
ECOR animal isolates (33 total)	8 **(24%)**	25 **(76%)**
Minnesota human isolates (48 total)	20 **(42%)**	28 **(58%)**
Minnesota animal isolates (204 total)	43 **(21%)**	161 **(79%)**
Clinical human isolates (16 total)	3 **(19%)**	13 **(81%)**

Twelve isolates with unique fingerprints have been selected for proteomic and microarray analysis. These twelve strains represent three species of host (brown bear, white-tail deer and human) with significantly different digestive system physiology. 2D gel analysis of these strains in triplicate is nearly complete and preliminary analyses have indicated that (1) there exists significant proteomic variation between isolates from different host species and (2)

differences that may be diagnostic for host-species can be identified (Figure 1). Mass spec analysis of all protein spots with host species distribution differences is in progress.

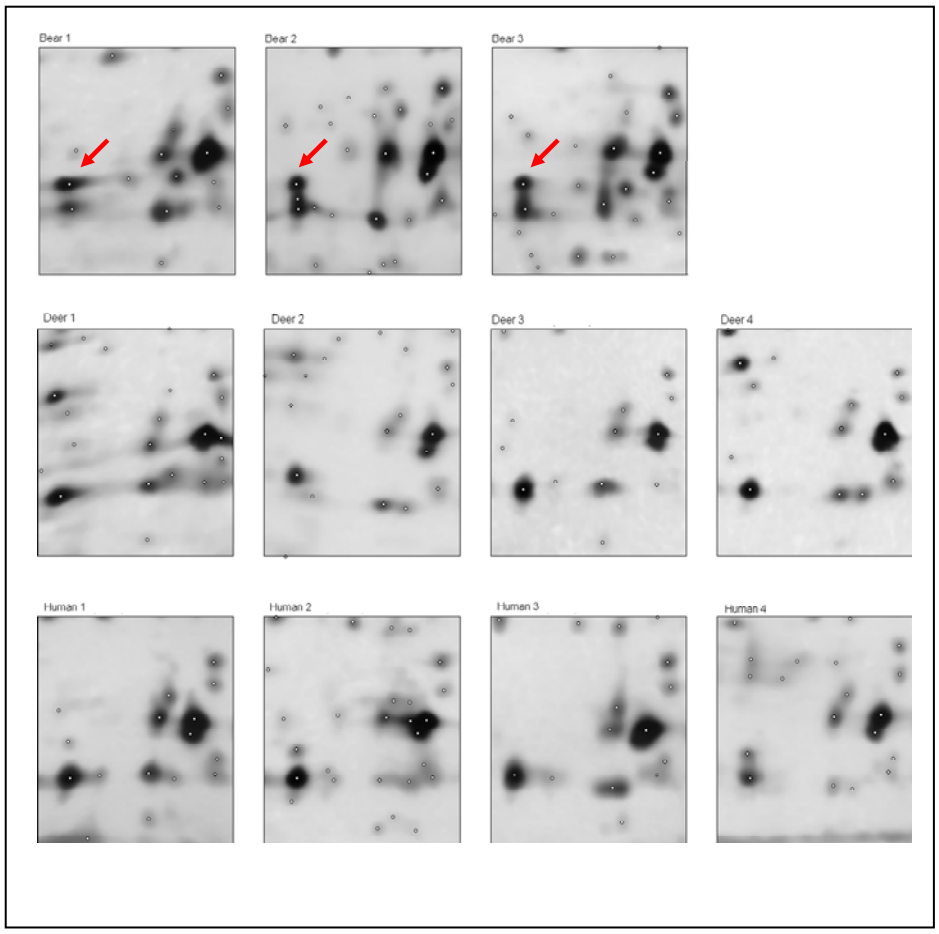

Figure 1. A representative subset of 2D protein expression patterns for three bear, four deer and four human *E. coli* isolates. Each panel represents a small portion of a single 2D gel. The red arrow indicates the position of a protein that may be useful for distinguishing bear samples from human and deer samples.

Microarray-based comparative genome hybridization of these isolates has revealed several gene regions whose presence or absence differs between bear, deer and human *E. col*. Experiments to determine the distribution of these potential markers in the larger strain collection are in progress.

Objective 3. The rep-PCR fingerprint database can currently be used to roughly classify unknown *E. coli* isolates. The completion of objective 2 will increase the reliability and utility of the database.