

Tracing diffuse sources of fecal contamination through phenotypic and genotypic characterization of *E. coli* isolates

**Final Project Report to Lake Erie Protection Fund
31 July 1999**

INTRODUCTION

A two-year grant, proposed by R.L. Sinsabaugh and L. Glatzer from the University of Toledo, entitled "Tracing diffuse sources of fecal contamination through phenotypic and genotypic characterization of *E. coli* isolates" was awarded by the Lake Erie Protection Fund in September 1996. Funding was received from the Ohio Department of Natural Resources in March 1997.

The impetus for this project was the unusually severe fecal contamination episodes reported for Maumee Bay during the summer of 1995. Swimming advisories were in effect for a total of 69 days at Maumee Bay State Park. Public concern over these advisories resulted in the formation of the Maumee Bay Bacteria Task Force (MBBTF), chaired by Susan Duckworth, head of the City of Toledo Environmental Services Department, which was charged with investigating the problem and suggesting mitigation strategies. The MBBTF included representatives from Maumee Bay State Park, Oregon City Water Department, Ottawa County Public Health Department, Toledo Department of Public Utilities, Toledo Department of Environmental Services, Lucas County Public Health Department, Ohio Department of Health, Ohio Environmental Protection Agency, Toledo Metropolitan Area Council of Governments, Toledo Water Department, and the University of Toledo (UT).

The MBBTF developed a water sampling program for the 1996 season. Sampling began in April and continued through October. Sampling sites included the inland swimming lake and the Lake Erie beach at Maumee Bay State park (MBSP), plus the three ditches (Norden, Curtice, Cousino) that drain the area around MBSP, the Lake Erie plume created by the discharge of cooling water from the Toledo Edison plant, and the outlet of the marsh on the east side of MBSP. Water samples from the beaches and ditches are collected twice per week; sediment samples once per month. Several groups participated in the sampling and in the collection of cognate data, foremost were ODNR, Toledo Public Utilities Department, and Lucas County Public Health Department. The data collected included total coliforms, fecal coliforms, *E. coli*, turbidity, number of bathers, number of birds, current direction, ditch flow, wave height and direction, wind speed and direction, air temperature, water temperature and precipitation.

Researchers from the University of Toledo Lake Erie Center participated in this program by isolating and enumerating *E. coli* bacteria and by collating and analyzing the data collected by the MBBTF. Funding for these efforts was provided by small awards from the Ohio Sea Grant Program, Lake Erie Protection Fund, and the Lucas County Department of Health.

The large grant from the LEPF allowed us to continue and extend the work conducted during 1996. The primary purpose of the work funded by this grant was to analyze *E. coli* for phenotypic and genotypic markers that could be used to identify sources of fecal contamination and to develop a local database for future monitoring. This process involved isolating and identifying cultures of *Escherichia coli* from water

samples, phenotypic characterization of *E. coli* isolates using fatty acid methyl ester (FAME) analysis, and genotypic characterization of *E. coli* isolates by plasmid typing.

The principal outcomes of this project, detailed in the sections that follow, include analysis of approximately 600 *E. coli* isolates from the 1996 sampling program, analysis of approximately 400 isolates from a second sampling program in 1998, establishment of the utility of FAME analysis for resolving *E. coli* strains, estimates of genetic diversity, correlation of fecal contamination episodes with meteorological conditions, partial success in tracing sources of fecal contamination, and training for a Ph.D. student, a M.S. student and several undergraduate researchers. Work in progress that will continue beyond the grant includes screening the archived collection of 1000 *E. coli* isolates for toxin-producing strains, an indicator of pathogenic potential, and fingerprinting *E. coli* isolates by chromosomal restriction patterns. Results from this project have been presented to local agencies and political representatives and at research conferences. Publications for professional journals will come from graduate student theses by Nancy Hatfield (Ph.D.) and Christian Lauber (M.S.).

BACKGROUND

The Enterobacteriaceae comprise a broad and diverse group of bacteria which inhabit the intestinal tract of humans as well as other mammals, birds and reptiles. Some genera of this group contain noted pathogens, such as *Shigella*, *Salmonella* and *Yersinia*, whereas others including *Escherichia coli*, form part of the normal flora of various animals. In addition, some of these bacteria are obligate inhabitants of the host animal, whereas others can proliferate in a wide range of environmental conditions. Because there is significant variation even within species of this group, typing of these bacteria to the species level has not proved especially beneficial in tracking the source of disease outbreaks. Traditionally, clinical isolates that have been identified to the species level through biochemical reactions are further subdivided into strains according to serological response to antibodies.

Ultimately it is the genotypic constitution of the bacteria that proves the relatedness of two separate isolates. In the last fifteen years the development of techniques for genetic analysis have led to more definitive studies tracing the sources of bacteria associated with disease outbreaks in both hospital derived and environmental situations. Of the most widely used genetic analytical techniques are nucleotide sequence analysis of plasmids and ribosomal RNA, sequence analysis of genomic DNA and DNA-DNA hybridization techniques. Sequence analysis of rRNA was used to profile 1291 strains of *Vibrio anguillarum* and *Vibrio ordalii* (Tiainen et al. 1995). Clustering of the strains indicated a genotypic difference between North European and South European strains. Restriction fragment analysis of genomic DNA was recently employed to identify a hot water supply as the source of hospital acquired *Legionella* infections (Luck et al., 1995). Fingerprinting of restriction fragments were utilized by Edberg et al. (1995) to determine the source of a 1992 outbreak of *Enterobacter cloacae* in the New Haven County, Connecticut water distribution system. Comparison of fingerprints of this bacterial species from the water system with those isolated from area hospitals, indicated that the source was the water system itself and did not pose a health threat to the population. DNA-rRNA hybridization were applied to the strain characterization of *Bordetella hilnzii* from poultry and humans (Vandamme et al. 1995.)

In addition to genotypic fingerprinting, there are also a number of biochemical

and/or phenotypic fingerprinting techniques. The advantage of these techniques is that usually they are less expensive, they are more rapid, and for some there are existing databases that facilitate analysis and allowing for comparison between different studies and from one geographic location to another. Included in phenotypic fingerprinting techniques are enzymatic utilization of different growth substrate, occurrence and distribution of flagella or other structures characteristics, and fatty acid methyl ester (FAME) analysis of whole cell extracts. The rationale for these approaches is that the phenotype or physical characteristics of the organism are a true reflection of its genetic content. Biochemical fingerprinting of coliform bacterial population based on utilization of 48 different carbon sources were conducted by Kuhn et al. (1991) to determine cross contamination between 18 different water samples. When using phenotypic characters for fingerprinting, it is a critical requirement that all isolates be grown for a specified amount of time under stringently uniform conditions of pH, temperature, aeration and nutrient substrate. We opted to use fatty acid methyl ester (FAME) analysis of *E. coli* for strain description and identification.

A number of studies confirm that bacterial FAME profiles are correlated with geotypic analysis. Rainey et al. (1994) analyzed 16 reference strains of *Pseudomonas stutzeri* by both fatty acid analysis and genome composition of Zva I and SpeI macrorestriction fragments subjected to pulsed-field electrophoresis, and determined that there was a significant degree of correlation between the genomovar groups and the FAME clusters. Likewise, Birnbaum et al. (1994) analyzed 183 well defined isolates of coagulase-negative *Staphylococci* which were derived from clinical specimens of 5 geographically distant areas. FAME profiles were subjected to cluster analysis and compared to those derived from plasmid profiles, restriction endonuclease and DNA hybridization analyses, and drug susceptibility profiles. They determined that the FAME approach was very effective in predicting the relatedness of strains.

Why fatty acids? The cytoplasmic membrane of the bacterial cell forms the dynamic interface between the cell and the environment. At about 7 nm in thickness, it forms the limiting area of the cell. As in eukaryotes, the membrane is composed of a bilayer of phospholipids. The phospholipids are composed of glycerol esters of straight chain fatty acids. The fatty acids themselves range from 10 to 20 carbons long, the critical length for spanning the membrane. Under varying environmental conditions, the fatty acid profiles change with respect to percentage and identification to the fatty acids present. However, when grown under standard environmental conditions, the profiles are remarkably constant and may be used to finger print bacterial species and strains. This constancy has been utilized by Myron Sasser of Newark Delaware to develop a data base in conjunction with Hewlett Packard which markets the Microbial Identification System. (MIDI)

An analysis of strains of coagulase negative staphylococci by Birnbaum et al (1994) indicated that this system provided a high discriminatory potential. Another advantage is that a large number of isolates (up to 30 per day) can be run by a single technician. The cost of disposable materials is about \$1.50 per sample. In contrast to other methods of typing there is very little disposable waste. Glass, water, and sodium hydroxide are the major waste products.

METHODOLOGY

Isolation: For samples collected during the 1996 program, *E. coli* was isolated by

inoculating water samples into IDEXX 2000 Quanta Tray. For each location, 100 μ l from at least one positive well was streaked onto EMB agar. A single *E. coli* positive colony was picked, grown as a pure culture on L agar and verified as *E. coli* by oxidase, indole reactions and Crystal ID. A total of 585 isolates, selected from samples in which the abundance of *E. coli* exceeded 125 colony forming units per 100 mL, were preserved by freezing in glycerol at -80°C until they were ready to analyze.

By the 1998 sampling season, new products were commercially available that allowed us to simplify the isolation. The simplified procedure allowed us to better coordinate sample collection and isolation, avoiding the backlogs that accumulated during the intensive sampling of 1996.

In 1998, water samples were collected monthly from seven sites (Lake Erie beach, the inland lake, the marina, Curtice Ditch within the Park, Curtice Ditch upstream, the confluence of Curtice Ditch and Wolf Creek, upstream Wolf Creek,). Bird droppings were also analyzed. Water was collected in sterile bottles and kept on ice until processing could begin. *E. coli* was isolated by inoculating one mL of sample onto Petri Film (3 M products) in triplicate. The films were incubated at 37°C for 24 hrs \pm 2 hrs. Gull and goose fecal matter was serially diluted and 1 ml was plated onto Petri Film. From the triplicate plates at each location, ten positive colonies were picked and streaked on L-agar containing MUG (4-methylumbelliferyl - β -D-glucuronic acid) and incubated for 24 h. Fluorescent colonies were transferred to L agar and verified as *E. coli* by oxidase and FAME analysis. A total of 381 isolates were retained. With this procedure the isolation time was reduced to one week as opposed to 3 to 4 weeks in 1996.

Preparation of extracts for FAME analysis: Stock cultures were grown on TSBA for 24 ± 1 hour at $28 \pm 1^{\circ}\text{C}$. A negative control, a calibration standard, and internal duplicates were run with each series. The internal duplicates were of three types: (1) the same extract was run with different batches. (2) The same isolate was cultured and extracted on different dates, (3) Same date/location isolates, identified as clonal from their identical plasmid profile patterns, were used as identity markers. These internal controls permitted an evaluation of the validity of the results, and later, the number of principal components that were necessary to group the clonal isolates adjacent to one another in a cluster analysis.

Statistical Analysis of FAME data: Scatter diagrams and dendrograms based on the first two principal components were generated by the MIDI system. The results failed to cluster identical isolates suggesting that the intrinsic variation was too large for clonal matching when using only two dimensions. Since the MIDI statistical program is proprietary, it was not possible to use it to analyze the information in more depth. Therefore, we decided to import the fatty acid raw data from MIDI into the SAS statistical program for examination.

We did univariate analysis of fatty acid abundance, and frequency, followed by principal component analysis (PCA). Because they were of little classification value, fatty acids that did not occur in at least 2% of the isolates were excluded from the PCA. To establish a similarity criterion for a lipotype designation, we ran a series of cluster analyses, progressively increasing the number of principal components from which similarity indices were calculated. A cluster analysis based on the first 12 principal

components grouped known clonal isolates. Clusters of isolates that with higher dissimilarity coefficients were considered to be distinct lipotypes.

Plasmid Analysis Procedure: *E. coli* isolates were grown on L-agar overnight at 37°C for 16 to 28 hrs prior to plasmid extraction. Plasmid DNA was obtained by using Promega Wizard™ Mini Prep Kit. Putative plasmid extracts were electrophoresed on 0.8% agarose gels along with a molecular mass standard, Promega 2-10 Kb Supercoiled Ladder or BRL Supercoiled Ladder. Polaroid images of the gels were digitized and scrutinized by Quantity-One™ image analysis software. Molecular mass for plasmids for each strain were entered into an Excel spread sheet to facilitate importation into Statistica for cluster analysis. Plasmids of isolates that clustered together were subjected to RFLP to determine relatedness of *E. coli* isolates.

RESULTS

Fifty one different fatty acids were extracted from the 1996 *E. coli* isolates. Of these 14 occurred in only a single isolate. Of the remaining 37, 17 comprised on average less than 0.1% of the total cellular fatty acid extract. The most numerous was 16:0 with a mean abundance of 28.7% of total fatty acid, followed by 16:1(iso I or w7c) at 21.1%, by 18:1 (with double bond at either 7,9 or 12 position), then 17:0 cyclo, 19:0 cyclo w6c and 12:0. Results from the 1998 isolates were similar.

Our statistical analyses indicated a high degree of heterogeneity in *E. coli*. As many as 70 different strains, more accurately lipotypes, of *E. coli* were isolated from the 1996 samples. Of these, most were unique strains that were isolated only once from a single site. A small number of strains, accounting for about half the isolates, appeared to be ubiquitous. Because the ubiquitous lipotypes were also found in bird samples, we concluded that local populations of gulls and geese were the source for at least half the *E. coli* found in bathing areas. The three most abundant lipotypes showed a temporal pattern: one type was common in spring, another in mid-summer and another in late summer.

Of the 585 isolates from 1996, 43% (249) contained at least one plasmid. The average number of plasmids was 2.4 per isolate with a range from 1 to 6. Based on molecular mass alone, 55% of the plasmid bearing isolates were unique (i.e. isolated only once), while the remainder of the isolates matched with at least one other isolate, a pattern similar to that resulting from the FAME analyses. Thirty distinct clusters (genotypes) were observed, many of which contained only 2-3 isolates. Similar plasmid patterns were found in isolates both within and between water sources as well as gull and goose fecal matter. RFLP(Restriction Fragment Length Polymorphism) analysis of plasmids gave no discernable identities for isolates that clustered together based on molecular mass, indicating that genotypic heterogeneity was very high.

RFLP of plasmid DNA was not successful in illustrating the source of fecal contamination at Maumee Bay State Park. The failure of RFLP to indicate a pollution source was not due to the technique but to the high degree of heterogeneity in the 1996 *E. coli* collection. Most of the isolates could not be analyzed by RFLP because of a lack of plasmids (333 isolates) or unique plasmid pattern (138 isolates). The resultant small sample size, 114 isolates, proved to be too small for nonpoint-source pollution identification. However, RFLP worked well to demonstrate the diversity which can exist in environmental populations of *E. coli*. In essence, RFLP indicated all plasmid

bearing strains were unique. RFLP of plasmids was a useful and accurate tool in strain differentiation, but alone was not sufficient for nonpoint -source pollution identification.

Statistical analyses of field observations from 1996, produced a multiple regression model that linked the number of waterborne *E. coli* in the Lake Erie beach area to water temperature and wind direction. The model explained only 37% of the variation in *E. coli* numbers, but is consistent with local observations of the combination of precipitation, water temperature, wind and current that are most likely to create nearshore contamination problems.

Analysis of data from the 1998 sampling season is not yet complete. These will provide a check on our conclusions from the 1996 data and allow us to make a better assessment of the contributions of local bird populations and drainage ditches to the MBSP beach contamination problem. (Surveys by the Lucas County Public Health Department suggest at 40% of the residential septic systems in the drainage area of MBSP may not be functioning properly.)

CONTINUING WORK

Work on this project will continue for at least another year. This work includes:

- Statistical analysis of the 1998 sampling data.
- Correlation of 1996 and 1998 data.
- Screening of archived cultures for potential pathogens:
 - Although the *E. coli* abundance in surface waters is considered to be the best predictor of waterborne disease potential, only a small fraction of bacteria classified as *E. coli* produce toxins associated with pathogenesis. Chris Lauber is in the process of screening our archived cultures to determine the frequency of toxin producing isolates. Attempts to identify virulence factors in *E. coli* from recreational waters have been made in the past (Lang et al., 1994). Many of these studies have focused on *E. coli* from marine and estuarine locations with little attention on *E. coli* from fresh water environments. Christian Lauber's M.S. thesis would be one of few fresh water studies and be the initial study in Lake Erie. The thesis entails probing for virulence factors of enterotoxigenic *E. coli* (ETEC) and enteropathogenic *E. coli* (EPEC). These two broad classes of *E. coli* account for much of the diarrheal disease in the world (Nataro et al., 1998). ETEC elaborate heat stable toxin (ST) and/or heat-labile toxin (LT) that cause a characteristic watery stool. However, EPEC do not produce a toxin but function by altering host epithelial cell ion balance, thus causing diarrhea (Donnenberg et al 1997). As with fimbriae, the types of virulence factors present can implicate possible sources of fecal contamination. This information will allow us to better assess the waterborne disease risk at MBSP.
- Genetic fingerprinting using chromosomal restriction fragment polymorphisms. Because of high interest in the problem of fecal contamination of recreational waters, much research in the past few years has been directed toward identifying markers that can be used to distinguish human-derived *E. coli* strains from animal-derived strains and distinguish pathogenic strains from non-pathogenic strains. We have begun using

pulsed field gel electrophoresis to detect markers associated with the bacterial chromosome. This approach is more technical than the plasmid analyses we have conducted for the past two years, but it has more potential for identifying sources.

- Preparation and defense and theses by Nancy Hatfield and Christian Lauber. These theses will be the basis for manuscripts submitted for publication in professional journals.

Christian Lauber (M.S.) is currently using the 1996 collection to determine the number of pathogenic *E. coli* present in MBSP waters. Attempts to identify virulence factors in *E. coli* from recreational waters have been made in the past (Lang et al., 1994). Many of these studies have focused on *E. coli* from marine and estuarine locations with little attention on *E. coli* from fresh water environments. Christian Lauber's M.S. thesis would be one of few fresh water studies and be the initial study in Lake Erie. The thesis entails probing for virulence factors of enterotoxigenic *E. coli* (ETEC) and enteropathogenic *E. coli* (EPEC). These two broad classes of *E. coli* account for much of the diarrheal disease in the world (Nataro et al., 1998). ETEC elaborate heat stable toxin (ST) and/or heat-labile toxin (LT) that cause a characteristic watery stool. However, EPEC do not produce a toxin but function by altering host epithelial cell ion balance, thus causing diarrhea (Donnenberg et al 1997). As with fimbriae, the types of virulence factors present can implicate possible sources of fecal contamination.

REFERENCES CITED

Birnbaum, D., L. Herwaldt, D.E. Low, M. Noble, M. Pfaller, R. Sherertz and A.W. Chow. 1994. Efficacy of Microbial Identification System for epidemiologic typing of coagulase-negative staphylococci. *Journal of Clinical Microbiology* 32:2113-2119.

Donnenberg MS, Kaper JB and Finaly BB. 1997. Interactions between enteropathogenic *E. coli* and host epithelial cells. *Trends in Microbiology* 5:109-114.

Kuhn, I., G. Allestam, T.A. Stenstrom and R. 1991. Biochemical fingerprinting of water coliform bacteria, a new method for measuring phenotypic diversity and for comparing different bacterial populations. *Applied and Environmental Microbiology* 57:3171-3177.

Lang Al, Tsai Yl, Mayer CL, Patton and Palmer CJ. 1994. Multiplex PCR for detection of heat-labile toxin gene shiga-like toxin I and II genes in *E. coli* isolated from natural waters. *Applied and Environmental Microbiology* 60:3145-3149.

Luck, P.C., J. Kohler, M. Maiwald and J.H. Helbig. 1995. DNA polymorphisms in strains of *Legionella pneumophila* serogroups 3 and 4 detected by macrorestriction analysis and their use for epidemiological investigation of nosocomial legionellosis. *Applied and Environmental Microbiology* 61:2000-2003.

Nataro JP and Kaper JB. 1998. Diarrheagenic *Escherichia coli*. *Clinical Microbiology Reviews*. 11:142-201.

Rainey, P.B., I.P. Thompson and N.J. Palleroni. 1994. Genome and fatty acid analysis of *Pseudomonas stutzeri*. *International Journal of Systematic Bacteriology* 44:54-61.

Tiainen, T., K. Pedersen and J.L. Larsen. 1995. Ribotyping and plasmid profiling of *Vibrio anguillarum* serovar O2 and *Vibrio ordalii*. *Journal of Applied Bacteriology* 79:384-392.

Vandamme, P., J. Hommeez, M. Vancanneyt, M. Monsieurs, B. Hoste, B. Cookson, C.H. Wirsing Von Konig, K. Kersters and P.J. Blackall. 1995. *Bordetella hinzii* sp. nov., isolated from poultry and humans. *International Journal of Systematic Bacteriology* 45:37-45.