Bacterial Source Tracking Analyses to Support Virginia's TMDLs

Non-Shellfish Stations

Prepared by

MapTech, Inc.

in cooperation with

New River Highlands RC&D

for

Virginia Department of Environmental Quality

Contract #9417

October 2005

CONTENTS

FIGURES

TABLES

ACKNOWLEDGEMENTS

Charles Hagedorn, Ph.D., Crop and Soil Environmental Sciences, Virginia Tech (CSES)

MapTech's Environmental Diagnostics Laboratory

Charles Martin, Virginia Department of Environmental Quality (VADEQ)

Jutta Schneider, VADEQ

Roger Stuart, VADEQ

Ram Gupta, VADEQ

Regional VADEQ Offices

Robert Wittman, Virginia Department of Health (VDH)

Regional VDH Offices

Thank you to the many state agency representatives and stakeholders who assisted with sample collection.

MapTech, Inc. of Blacksburg, Virginia conducted this study with funding provided by New River Highlands RC & D (Contract # 9417), made available through a grant from the Virginia Department of Environmental Quality.

1. INTRODUCTION

EPA's document, *Guidance for Water Quality-Based Decisions: The TMDL Process* (USEPA, 1999) states:

> *According to section 303(d) of the Clean Water Act and EPA water quality planning and management regulations, States are required to identify waters that do not meet or are not expected to meet water quality standards even after technology-based or other required controls are in place. The water bodies are considered water quality-limited and require TMDLs.*
>
> *. . . A TMDL, or total maximum daily load, is a tool for implementing State water quality standards and is based on the relationship between pollution sources and in-stream water quality conditions. The TMDL establishes the allowable loadings or other quantifiable parameters for a water body and thereby provides the basis for States to establish water quality-based controls. These controls should provide the pollution reduction necessary for a water body to meet water quality standards.*

The purpose of this project is to use bacterial source tracking (BST) to identify sources of *E. coli* to support the development of *E. coli* TMDLs for impaired segments in Virginia. In fulfilling the state requirement for the development of a TMDL, a systematic process will be utilized to establish the maximum allowable *E. coli* loading for each waterbody to meet the applicable standard, allocate that load among pollutant contributors, and provide a basis for taking actions needed to restore water quality. This report focused on water quality sampling conducted in non-shellfish waters. A companion document will be published later this year to report the results of water quality sampling in shellfish waters. Together, these reports reflect the third year of BST sampling conducted by VADEQ (2004-2005).

Bacterial Source Tracking (BST) methods can be subdivided into three basic groups: Molecular, Biochemical, and Chemical. Molecular (genotype) are typically referred to as "DNA fingerprinting" and are based on the unique genetic makeup of different strains, or

subspecies, of fecal bacteria. Biochemical (phenotype) methods are based on an effect of an organism's genes that actively produce a biochemical substance. The type and quantity of these substances produced is what is actually measured. Chemical methods are based on finding chemical compounds that are associated with human wastewaters, and generally are restricted to determining if sources of pollution are human or not.

Hagedorn's (Hagedorn et al., 1999) Antibiotic Resistance Analysis (ARA) technique was used for this project because it has been demonstrated to be a reliable procedure for confirming the presence of human, livestock, wildlife and pet sources. Compared to DNA fingerprinting, biochemical profiling is much quicker, typically allows for many more isolates to be analyzed (*e.g.*, hundreds per week vs. a few dozen per week for DNA analysis), is more economical, has survived limited court testing, and has undergone rigorous peer review from the scientific community. Additionally, observation of an increased number of isolates allows for an estimate of the relative proportions of the fecal indicator (*e.g.*, *E. coli*) originating from different sources.

2. OBJECTIVES

BST was used to identify sources of *E. coli* as well as the relative percentage contribution from four source groups (*i.e.*, livestock, wildlife, human and pets) to support the development of *E. coli* TMDLs for impairments located throughout Virginia. BST results will be used to improve public awareness of the problem, improve model calibration/validation of *E. coli* densities, and provide a more equitable allocation of loads to source classes.

The specific objectives of the project were to:

1. collect fecal samples from known sources in 22 areas, based on Hydrologic Unit Codes (HUCs),
2. use collected samples to develop a known-source library for each impairment area, and
3. perform bacterial enumerations and BST analyses on whole water samples from impaired segments, using the libraries developed for objective 2.

3. METHODS

Hagedorn's ARA method has been extensively and successfully used by MapTech, and separates fecal sources based on patterns of antibiotic resistance in the *enterococci* or *E. coli*. For this study, *E. coli* was the indicator organism analyzed. The premise of ARA is that fecal bacteria from each source (*e.g.*, human, livestock, wildlife, and pets) will have different resistance patterns to the battery of antibiotics and concentrations used in the analysis. Hagedorn's method for *E. coli* tests each isolate on 28 different combinations of antibiotic type and concentration. Confidence in BST techniques is measured by the level of separation of isolates from known sources, represented as the percentage of isolates that are accurately separated into respective source types (*e.g.*, Average Rate of Correct Classification – ARCC). Additional analyses can be applied to test the specificity of the library. These analyses are discussed further in Section 4 of this document. The ARA method, like other methods (*e.g.*, molecular), requires the collection of source samples from feces of known sources to build a source library. In support of this study, known source samples from the four source classes were collected, analyzed, and entered into known-source libraries.

3.1 Collection of Known Sources

Known source samples were collected in twenty-two HUCs associated with fecal-bacteria impaired waters throughout Virginia (Figure 3.1). In HUCs where known-source samples had not previously been collected to support VADEQ's BST program (newly sampled HUCs), a total of 60 samples were collected in each HUC. In HUCs where known-source samples were previously collected (updated HUCs), a total of 20 samples were collected to update existing libraries. Each set of source samples was distributed evenly between human, livestock, wildlife, and pets (Table 3.1). Specific species within each source category (*e.g.*, deer, raccoon, poultry, beef, etc.) that were selected to represent the sources in each region were identified through field observation, discussion with local stakeholders, and review of available data (*e.g.*, Virginia Agricultural Statistics). From each sample, 8 isolates were analyzed using BST to create a known-source library of 480 isolates for each newly sampled HUC, and to increase known-source libraries by 160 isolates in updated HUCs. To date, approximately 2,844 fecal samples have been collected to support VADEQ's BST program, resulting in over 23,105 isolates

analyzed. In total, 730 fecal samples were collected for this study, resulting in 5,864 isolates analyzed.

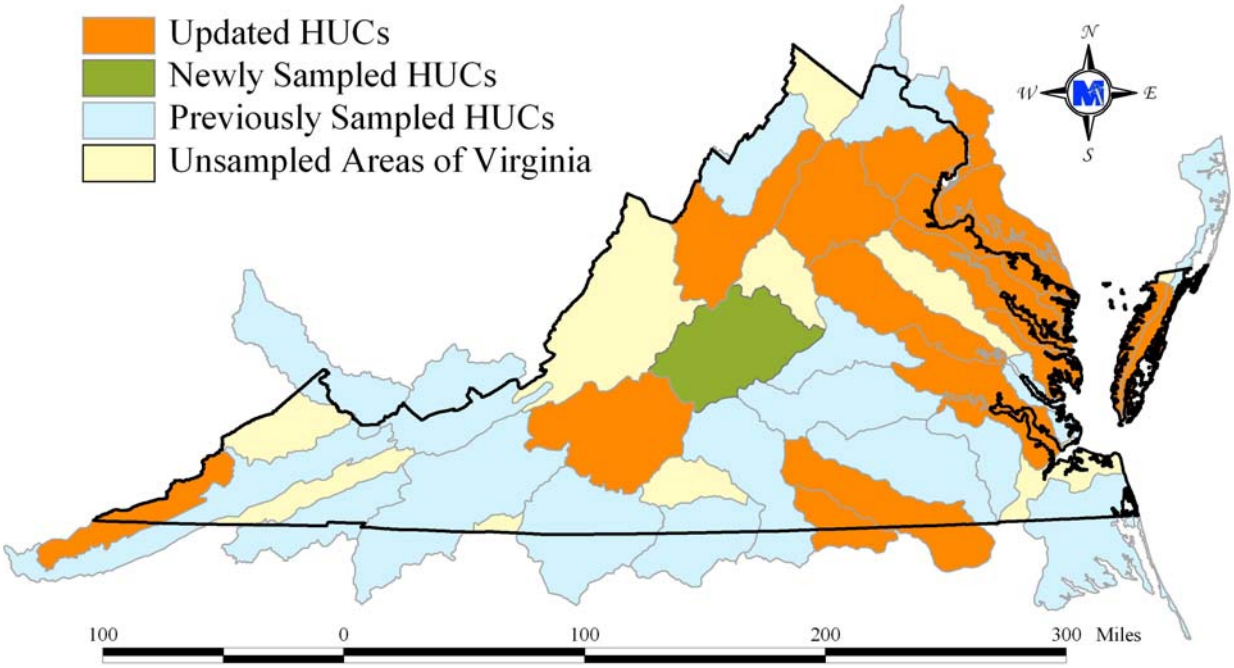

Figure 3.1 Locations of known-source sampling conducted to support this year's and previous years' BST analyses.

Table 3.1 Source samples collected for BST library development.

Source	Source Species	Number of Samples Collected in Newly Sampled HUCs	Additional Samples Collected in Updated HUCs
Human	Septic Systems, Portable Toilets, …	15	5
Livestock	Dairy, Beef, Horse, Sheep, Broilers, Turkeys, Swine, Waste Storage Pits, …	15	5
Wildlife	Deer, Raccoon, Muskrat, Duck, Goose, …	15	5
Pets	Dogs & Cats	15	5
Total		60	20

3.2 Development of Known-Source Libraries

An appropriate known-source library was selected for each of the impairments to complete objective 2. A predictive model was developed from each library using logistic regression. A known-source library must be large enough to prevent an over-specified fit to the library. However, known-source responses to ARA analyses have been observed to vary geographically. The characteristics of this variance have not been well defined, so the regional libraries developed for this study were combined in a stepwise procedure and analyzed to measure the resulting specificity and the predictive accuracy of the combined libraries, as detailed in Section 4 of this document.

3.3 Bacterial Enumerations and BST Analyses

For objective 3, water quality monitoring sites were identified and sampled by the granting agency (Figure 3.2 and Table 3.2). For many sites, the contract began in July 2004. At the conclusion of the study, all sites will have been sampled monthly for one year. Samples were received as whole-water samples (*i.e.*, ambient sampling as presented in Table 3.2). All water samples were analyzed for *E. coli* and fecal coliform. BST was run on bacteria isolated from the whole-water samples. Bacteria were analyzed using Hagedorn's ARA methodology, yielding the percentage of isolates classified as human, livestock, wildlife, and pets. Up to 24 bacterial

isolates were analyzed per sample, limited only by the number of isolates available from the enumeration process.

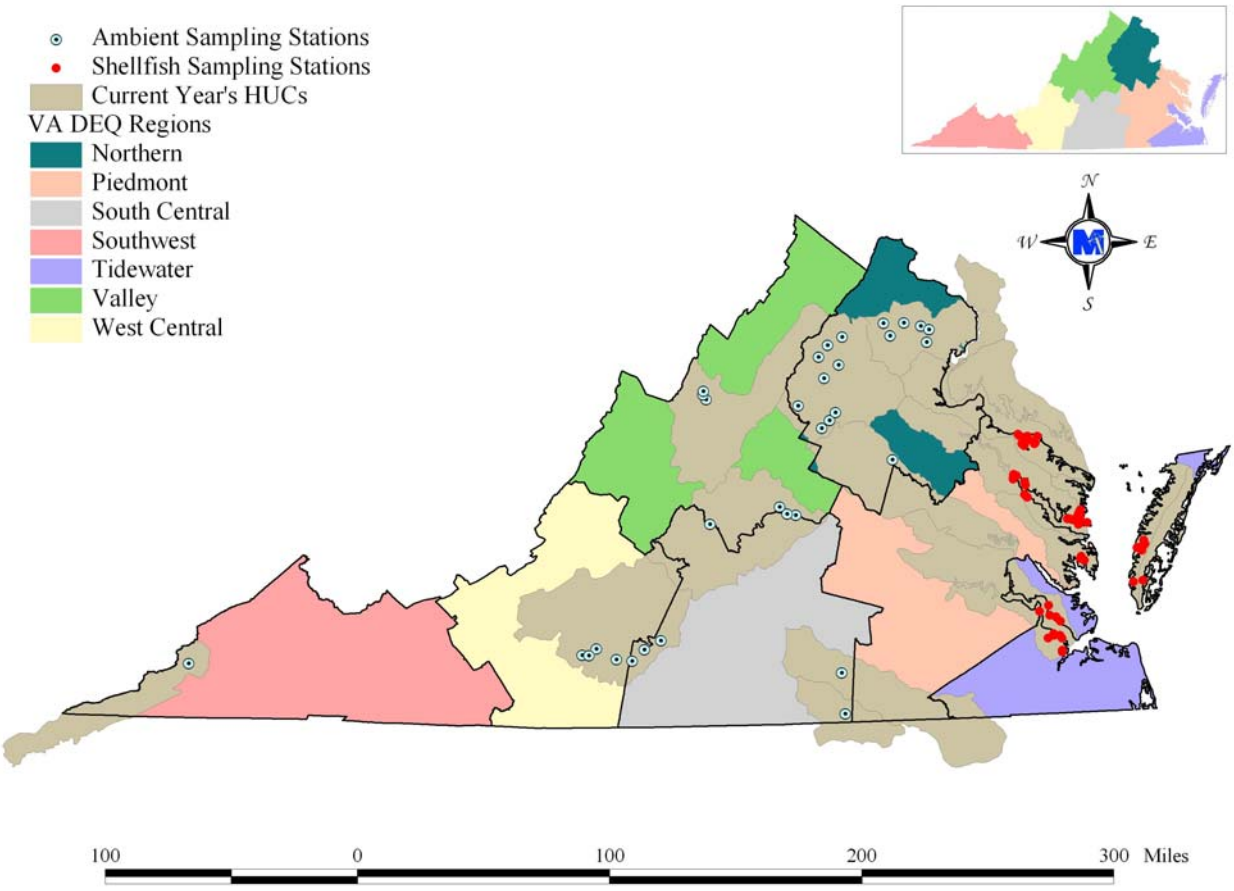

Figure 3.2 Spatial distribution of impaired segments identified by region.

Table 3.2 Distribution of ambient sampling stations addressed in this study.

Waterbody	Hydrologic Unit	BST Stations
Broad Run	A19	1
Bull Run	A23	1
Little Bull Run	A21	1
Occoquan River	A20	1
Popes Head Creek	A23	1
South Run	A19	1
Beaver Creek	B18	2
Union Spring Run	B18	1
Hardware River	H19	1
Little Georgia Creek	H17	1
Piney River	H10	1
Totier Creek	H17	1
Blue Run	E13	1
Hazel River	E04	1
Hughes River	E03	1
Rapidan River	E13	1
Rapidan River	E11	1
Robinson River	E15	1
Rappahannock River	E01	1
Rush River	E05	1
Thornton River	E05	1
Great Creek	L80	1
Old Woman's Creek	L13	1
Pigg River	L18	1
Pigg River	L16	1
Pigg River	L14	2
Story Creek	L14	1
Snow Creek	L17	1
Flat Rock Creek	K03	1
Northeast Creek	F09	1
Chestnut Creek	N06	2

4. KNOWN-SOURCE LIBRARY DEVELOPMENT

As discussed in Section 3, a predictive model was developed from each library (HUC) using logistic regression. Where a previously developed library existed (*i.e.*, updated HUCs), this year's data was combined with the existing data and the updated library was used for further assessment. These regional libraries were combined in a stepwise procedure and analyzed to measure the resulting specificity and the predictive accuracy of the combined libraries. The specificity and predictive accuracy were assessed through three analyses. First, the ARCC was calculated for the library. Second, a randomization test was performed by randomly assigning source categories to samples and assessing the ARCC for the randomized library. Ten randomizations were performed and the results averaged. The expected result of randomization of four source categories is an ARCC of 25%, indicating a completely random result. Greater values for the randomized ARCC indicate a more specified model. Third, a jackknifing routine was conducted, where data from each whole fecal sample were individually withheld during development of the statistical model. The model was then tested for predictive accuracy on the withheld sample. In combining regional libraries, a balance was sought between minimizing the randomized ARCC and maximizing the jackknifed ARCC. Table 4.1 shows the resulting analyses on the finalized libraries, and Table 4.2 shows how the libraries were applied to the analysis of whole-water samples by the HUC in which they were sampled.

Table 4.1 Results of known-source library development.

Known-Source Library	Regional Libraries Included (by HUC)	ARCC (%)	Randomized ARCC (%)	Jackknifed ARCC (%)
2005-01	2070010 + 2070005 + 207008	82	42	73
2005-02	2080103 + 207005 + 2080207	80	37	74
2005-03	3010101 + 2080207	85	40	77
2005-04	2080203 + 2080207	80	38	74
2005-05	2080106 + 2080207	82	39	74
2005-06	3010204 + 3010103 + 2080207	85	38	80
2005-07	3010106 + 2080207	73	37	67
2005-08	2070005 + 6010205	89	41	78

Table 4.2 Known-source libraries associated with HUCs included in this study.

HUC	Known-Source Library
HUC 2070005	2005-08
HUC 2070010	2005-01
HUC 2080103	2005-02
HUC 2080106	2005-05
HUC 2080203	2005-04
HUC 3010101	2005-03
HUC 3010106	2005-07
HUC 3010204	2005-06

5. RESULTS

The results of the water quality analyses for VADEQ's 2004-2005 BST sampling in non-shellfish waters are reported in this section. Fecal coliform enumerations, *E. coli* enumerations, and the results of the BST analyses are reported. The *E. coli* enumerations are reported with the BST results to give an indication of the bacteria concentration at the time of sampling. The proportions reported are formatted to indicate statistical significance (*i.e.*, **BOLD** numbers indicate a statistically significant result). The statistical significance was determined through two tests. The first was based on the sample size. A z-test was used to determine if the proportion was significantly different from zero (alpha = 0.10). During the second test, the rate of false positives was calculated for each source category in each library, and a proportion was not considered significantly different from zero unless it was greater than the false-positive rate plus three standard deviations.

Table 5.53 Bacterial Enumeration for Rush River at Station 3RUS005.66.

Enumerations											
Station ID	Date of Sample	Time of Sample	Lab ID	Lab-In Date	*E. coli* cfu/100ml	Quality	Fecal Coliform cfu/100ml	Quality	Comments	Lab-Out Date	Lab Personnel
3RUS005.66	7/28/2004	10:15	D3578	7/29/2004	400		410			8/31/2004	DM
3RUS005.66	8/30/2004	11:15	D3675	8/31/2004	70		50			9/7/2004	DM
3RUS005.66	9/28/2004	10:30	D3789	9/29/2004	6,000		4,500			10/7/2004	DM
3RUS005.66	10/21/2004	10:50	D3832	10/22/2004	142		480			10/26/2004	DM
3RUS005.66	11/30/2004	10:44	D3956	12/1/2004	164		240			12/7/2004	DM
3RUS005.66	12/16/2004	10:40	D4012	12/17/2004	90		170			1/7/2005	DM
3RUS005.66	2/8/2005	10:22	D4136	2/9/2005	32		60			2/11/2005	DM
3RUS005.66	3/8/2005	11:00	D4200	3/9/2005	790		780			3/16/2005	DM
3RUS005.66	4/13/2005	10:40	D4289	4/14/2005	171		170	B		4/21/2005	DM
3RUS005.66	5/25/2005	10:20	D4427	5/26/2005	90		60	B		5/31/2005	DM
3RUS005.66	6/28/2005	13:30	D4495	6/29/2005	122		80	B		7/5/2005	DM
3RUS005.66	7/26/2005	13:50	D4599	7/27/2005	150	A	460			7/29/2005	DMt

A: Value reported is the mean of two or more determinations.
B: Results based upon colony counts outside the acceptable range.
U: Material was analyzed for, but not detected. Value stored is the limit of detection for the process in use.

Table 5.68 Bacterial Source Tracking for Rush River at Station 3RUS005.66.

VADEQ ID	Date of Sample	Lab ID	HUP ID	Number of Isolates	*E. coli* (cfu/100 ml)	Wildlife	Human	Livestock	Pet
3RUS005.66	7/28/2004	D3578	E05	24	400	8%	4%	**21%**	**67%**
3RUS005.66	8/30/2004	D3675	E05	9	70	0%	0%	**33%**	**67%**
3RUS005.66	9/28/2004	D3789	E05	24	6,000	**17%**	**12%**	**63%**	8%
3RUS005.66	10/21/2004	D3832	E05	24	142	0%	**17%**	**71%**	**12%**
3RUS005.66	11/30/2004	D3956	E05	24	164	**50%**	**29%**	**21%**	0%
3RUS005.66	12/16/2004	D4012	E05	24	90	8%	**46%**	8%	**38%**
3RUS005.66	2/8/2005	D4136	E05	18	32	**33%**	**33%**	**17%**	**17%**
3RUS005.66	3/8/2005	D4200	E05	24	790	**17%**	4%	**62%**	**17%**
3RUS005.66	4/13/2005	D4289	E05	24	171	**54%**	**17%**	4%	**25%**
3RUS005.66	5/25/2005	D4427	E05	24	90	**46%**	8%	**21%**	**25%**
3RUS005.66	6/28/2005	D4495	E05	24	122	**75%**	0%	**21%**	4%
3RUS005.66	7/26/2005	D4599	E05	24	150	**29%**	4%	**21%**	**46%**

BOLD type indicates a statistically significant value.
*NVI - No Viable isolates