



Graduate Theses, Dissertations, and Problem Reports

2010

The use of microbial community fingerprinting as a marker for tracking the source of water: Application to pathogen and groundwater source tracking

Mary Clement
West Virginia University

Follow this and additional works at: <https://researchrepository.wvu.edu/etd>

Recommended Citation

Clement, Mary, "The use of microbial community fingerprinting as a marker for tracking the source of water: Application to pathogen and groundwater source tracking" (2010). *Graduate Theses, Dissertations, and Problem Reports*. 3210.

<https://researchrepository.wvu.edu/etd/3210>

This Thesis is protected by copyright and/or related rights. It has been brought to you by the The Research Repository @ WVU with permission from the rights-holder(s). You are free to use this Thesis in any way that is permitted by the copyright and related rights legislation that applies to your use. For other uses you must obtain permission from the rights-holder(s) directly, unless additional rights are indicated by a Creative Commons license in the record and/ or on the work itself. This Thesis has been accepted for inclusion in WVU Graduate Theses, Dissertations, and Problem Reports collection by an authorized administrator of The Research Repository @ WVU. For more information, please contact researchrepository@mail.wvu.edu.

The use of microbial community fingerprinting as a marker for tracking the source of water: Application to pathogen and groundwater source tracking

Mary Clement

**Thesis submitted to the
College of Engineering and Mineral Resources
at West Virginia University
in partial fulfillment of the requirements
for the degree of**

**Master of Science
in
Civil and Environmental Engineering**

Approved by:

**Benoit Van Aken, Ph.D., Committee Chairperson
Radhey S. Sharma, Ph.D.
Lian-Shin Lin, Ph.D.
Dorothy Vesper, Ph.D.**

**Department of Civil and Environmental Engineering
Morgantown, West Virginia
2010**

Key words: Terminal-restriction fragment length polymorphism, microbial community fingerprinting, microbial source tracking, genomic cloning library

Abstract

The use of microbial community fingerprinting (using T-RFLP) as a marker for tracking source of water: Application to pathogen and groundwater source tracking

Mary Clement

Microbial community structure in groundwater is extremely complex and constitutes a fingerprint of water. Terminal restriction fragment length polymorphism (T-RFLP) analysis is a polymerase chain reaction (PCR)-based fingerprinting method that is commonly used for comparative microbial community analysis, and to track specific bacterial markers in complex environments.

The *Objective of this research* was to propose an innovative approach for tracking the origin of groundwater, microbial source tracking and origin of surface water, based on the specific structure of the microbial community using T-RFLP analysis, which can be used for microbial source tracking.

The first specific aim was to validate the use of T-RFLP profiles of deep and shallow waters in Huntingdon, PA, Great Valley, WV and Berkeley springs, WV, as a marker of the groundwater source. Alternatively, a 16S ribosomal DNA genomic library was constructed to identify bacterial species that can inform about the source of groundwater. It was observed that clustering based on screening a clone library gave more reliable information of the groundwater source than the T-RFLP profiles.

The second specific aim focused on the characterization of the microbial community from different streams along the Left Fork Mud River watershed in Lincoln County, WV. T-RFLP analyses were carried out for determining the source of contaminated water. Results showed that different streams of the Mud River watershed has a specific microbial terminal restriction fragment (TRF) profiles, which helps in determining the origin of water.

The third specific aim was to characterize the entire microbial community in several streams in Corridor-H watersheds, WV using T-RFLP analysis. It was observed that water samples collected from different locations of the same stream carried a very specific signature that can be used for microbial source tracking analysis. The T-RFLP method application on spring waters and Mud River watersheds was based on the preliminary results obtained from the analysis of the microbial community structure of several streams in Corridor-H watersheds.

Our results, thus provide the evidence that natural waters has a 'microbial signature' (T-RFLP profile) that can be used as a marker for pathogens and groundwater source tracking.

Acknowledgements

There are a number of people I would like to sincerely thank for the guidance, assistance, and insightfulness they provided to allow me to complete this research and thesis. Foremost, I would like to thank my advisor and thesis director, Dr. Benoit Van Aken, for his valuable guidance, advice and support. I also wish to thank my committee members: Dr. Radhey Sharma, Dr. Lian-Shin Lin, and Dr. Dorothy Vesper, for their helpful comments and suggestions.

Having a chemical engineering background and performing a molecular biology-based research project resulted in having to learn a substantial amount of concepts and information that were previously unknown to me. I would like to thank Dr. Harris Edenborn (NETL), Dr. Dorothy Vesper (Committee member) and T. Vandivort from WVVRI for providing with data, which has been used in this thesis. Very special thanks to my lab mates Melin and Paola, for their valuable help and suggestions in learning laboratory techniques. I would like to thank Danielle Ellis and Dhiraj from Biology department at WVU for T-RFLP technology support.

I sincerely express my gratitude towards everyone mentioned in these acknowledgements!

Thank you all!

Mary Clement

Table of Contents

List of symbols / Nomenclature	ix
Chapter 1: General Introduction	1
1.1 Introduction.....	1
1.2 Statement of critical regional or state water problem	2
1.3 Objective of the Research	6
1.4 Working Hypothesis	6
1.5 Significance.....	6
1.6 Specific aims	6
2.1 Microbial communities in natural water	8
2.1.1 Microbial ecology in natural waters	8
2.1.2 Bioindicators for water quality	10
2.2 Microbial Source Tracking	10
2.3 Molecular DNA fingerprinting methods.....	12
2.4 Terminal-restriction fragment length polymorphism analysis (T-RFLP).....	13
2.5.1 Disadvantages of the T-RFLP Method	14
Chapter 3: Materials and Methods.....	15
3.1 Water samples.....	15
3.1.1 Spring water	15
3.1.2 Stream water	16
3.2 Membrane filtration bacterial enumeration	17
3.3 DNA Extraction	17
3.4 Microbial community fingerprinting using T-RFLP	18
3.4.1 16S rDNA Amplification.....	19
3.4.2 Agarose Gel Electrophoresis (AGE).....	19
3.4.3 Enzymatic digestion.....	20
3.4.4 Capillary electrophoresis	20
3.4.5 Terminal fragment profile analysis.....	21
3.5 Bacterial Identification by screening of 16S rDNA gene library	22
Chapter 4: Results and discussion.....	24
4.1 Specific aim 1	24
4.1.1 Microbial number and Physico-chemical parameters of spring waters	24
4.1.2 Terminal-restriction fragment length polymorphism (T-RFLP) analysis.....	26

4.1.3 Genotypic identification of bacterial species using genomic library.....	29
4.2 Specific aim 2	33
4.2.1 Indicator Organisms in Left Fork Mud River watershed.....	33
4.2.2 Microbial community in Left Fork Mud River watershed: T-RFLP analysis	35
4.3 Specific aim 3	38
4.3.1 Microbial number and physico-chemical parameters of Corridor-H watershed	38
4.3.2 Microbial community fingerprinting of the Corridor-H watershed using T-RFLP	40
Chapter 5: Conclusion.....	43
5.1 Specific Aim 1	44
5.2 Specific Aim 2	45
5.3 Specific Aim 3	45
5.4 Significance of the results obtained in this thesis	46
5.5 Future work.....	46
References.....	47

List of Figures

Figure 1: Map of the Northern Shenandoah Valley with start dates of county-scale cooperative ground-water assessments	3
Figure 2: Simplified illustration of the metabolic pathways of microorganisms in aquifers.....	9
Figure 3: Schematic representation of some of the methods used for tracking sources of fecal pollution	12
Figure 4: Map of the Shenandoah Valley in Virginia and location of sampling site.....	15
Figure 5: Corridor H alignment and study watersheds: Mill Run, Big Run, Beaver Creek, Patterson Creek, Walnut Bottom Run.....	16
Figure 6: Scheme of a typical T-RFLP process	18
Figure 7: Agarose gel electrophoresis of 16S rDNA amplified by PCR from total DNA extract from spring waters.	20
Figure 8: Profile of a 16S rDNA microbial community in one analyzed sample using T-RFLP.	21
Figure 9: Flowchart summarizing the 'Metagenomic Process'.	22
Figure 10: Bacterial number in each spring, as determined by CFU counting.....	25
Figure 11: Terminal-restriction fragment profiles from 7 springs in Huntingdon (PA), and Great Valley, Berkeley Springs (WV).....	27
Figure 12: Clustering analysis of spring waters based on the microbial community as determined by T-RFLP profiling	28
Figure 13: Bacterial species identified by comparison of their DNA sequences with public genomic databases	31
Figure 14: Clustering analysis of spring waters based on the microbial community as determined by screening a 16S rDNA clone library.....	33
Figure 15: Mean Total Coliform measurement (per 100mL) at each sampling site in the Left Fork Mud River watershed.	34
Figure 16: Mean Fecal Coliform measurement (per 100mL) at each sampling site in the Left Fork Mud River watershed.	34
Figure 17: Mean <i>E.coli</i> measurement (per 100mL) at each sampling site in the Left Fork Mud River watershed.	35

Figure 18: T-RFLP profiles of the bacterial community of different streams from the Left Fork Mud River watershed, Lincoln County (WV). 36

Figure 19: Denderogram obtained from the clustering analysis of T-RFLP profiles of streams from the Mud River watershed. SAS PROC processing using “Ward’s minimum-variance clustering”. 37

Figure 20: Map showing location of the different sites in Beaver Creek, Mill Run and Walnut Bottom Run watersheds located in Preston, Tucker and Hardy counties. 40

Figure 21: T-RFLP profiles of the bacterial community of Beaver Creek watershed, WV. 41

Figure 22: Denderogram obtained from the clustering analysis of T-RFLP profiles of different streams from the Corridor-H watershed. SAS PROC processing using “Ward’s minimum-variance clustering”. 41

List of Tables

Table 1: Physico-chemical characteristics measured along the deep and shallow water sources of the karst groundwater systems	25
Table 2: Bacterial species in spring waters identified by sequencing individual 16S rDNA clones	30
Table 3: Cultivation results of the different streams from the Corridor-H watershed, WV.	38
Table 4: Physico-chemical characteristics measured along the Corridor-H watershed, WV.	39

List of symbols / Nomenclature

AFLP	Amplified fragment length polymorphism
AMD	Acid mine drainage
BPH	Biphenyl Dioxygenase
CDF	Confined disposal facility
cDNA	Complementary deoxyribonucleic acid
CFU	Colony forming unit
DOC	Dissolved organic carbon
LB	Luria Bertani
LH-PCR	Length heterogeneity-polymerase chain reaction
MST	Microbial source tracking
NCBI	National Center for Biotechnology Information
OC	Organic carbon
OTU	Operational taxonomic unit
OD _{260nm}	Optical density at 260 nm
OD _{280nm}	Optical density at 280 nm
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
rDNA	Ribosomal-deoxyribonucleic acid
RDP	Ribosomal database project
RE	Restriction endonuclease
REP-PCR	Repetitive extragenic palindromic
SOC (medium)	Super optimal broth with catabolite repression
TMDL	Total maximum daily load
TOC	Total organic carbon
TRF	Terminal restriction fragments
T-RFLP	Terminal-restriction fragment length polymorphism
USEPA	United States Environmental Protection Agency

Chapter 1: General Introduction

1.1 Introduction

All microbes live in an aqueous environment. Microorganisms, or microbes, are natural and vital members of all aquatic communities, and a direct impact on one species is likely to induce indirect effects on other species (Nico Goldscheider, 2006). The ecology of aquatic environment is complex and our understanding of them is limited. Also, studies have revealed that aquifers are inhabited by a large diversity of microorganisms (Batin, Wille, Sattler, & Psenner, 2001). Man made activities like discharge of noxious substances; into waterbodies are spoiling the aquatic environment to the maximum. Certain microbes, however, when present in excess numbers, pose a threat to human health. In general, fecal indicator bacteria like Total *Coliform* and *Escherichia coli*, are used to assess the microbial quality of water (Lasalde, Guez, & Toranzos, 2005), because they are quite harmless, but their presence in natural water indicates the possible presence of microbial pathogens, which are much more difficult to detect. The standard techniques like cultivation of fecal indicator bacteria, does not identify the source of water contamination. In the US, the need for identifying fecal sources has become a priority to states, territories, and authorized tribes (Santo Domingo, Bambic, Edge, & Wuertz, 2007).

Molecular methods, based on polymerase chain reaction (PCR) are useful both to monitor microbial communities, and to track specific bacterial markers in complex environments (Field, Bernhard, & Brodeur, 2003), which have been gaining popularity among microbial source trackers. The genetic fingerprinting of the microbial community provides information on the presence and relative abundance of the species and thus indicates biodiversity. The terminal-restriction fragment length polymorphism (T-RFLP) analysis, based on amplifying nucleic acids by PCR, enables to compare community DNA amplification profiles derived from different microbial communities (Marsh T. , 1990). The T-RFLP analysis technique introduced by Liu and colleagues in (1997) is highly reproducible and yields a greater number of operational taxonomic units (OTUs) than many other PCR-based fingerprinting methods (Osborn, Moore, & Timmis, 2000). The fluorescent end-labeling of either the 5' or 3' PCR primer allows length separation

and sizing of terminal restriction fragments after hydrolyzing the PCR products with a selected restriction endonuclease (RE), thus producing a molecular “fingerprint” characteristic of the water community analyzed.

Despite a fast development of molecular, PCR-based techniques for source tracking and pathogen detection, no method has been shown to be satisfactory for the reliable determination of pathogen source (Field, Bernhard, & Brodeur, 2003). This research is innovative since it proposes a new approach for microbial source tracking, which uses the entire microbial community profiling as a specific marker for determining the sources of surface water, groundwater and for pathogens tracking. This project was based on the preliminary results obtained from the analysis of microbial community structure in several streams along the Corridor-H watersheds, West Virginia, which showed that single watersheds exhibit very specific T-RFLP profiling that can be used to characterize the source and nature of the microbial contamination.

1.2 Statement of critical regional or state water problem

Sustainability of karst groundwater resources

Groundwater resources from karst aquifers, those formed in soluble rocks play an important role for public water supply in many areas throughout the world (Drew & Hotzl, 1999). (USGS, 2003) have stated that, “In recent years, the Northern Shenandoah Valley of Virginia in West Virginia has been experiencing rapid growth along the I-81 corridor and the eastern margin of the Valley (Figure 1). This growth has the potential to profoundly influence the region's land, water, and biological resources. Regional and local resource managers have real concerns over the region’s ability to sustain future growth. Current resource-management concerns center on the sustainability and vulnerability of the region’s water resources and their ability to provide a reliable long-term water supply”. Increased urbanization and future development in the Appalachian regions require a means for identifying the most sustainable sources. An improved understanding of the complex aquifer systems is required to effectively develop and manage them as sustainable water supplies in the Appalachians. Deep sources may be characterized by brine chemistry (Nativ, 1996) and signals left from deep-karst formation. The hypogenetic process has likely left behind both chemical and microbial signals in the water chemistry

(Doctor, Orndorff, & Plummer, 2008)). Recent studies have indicated the importance of deeper flow systems to the karst water sources but its regional origin and signature are poorly understood. Once that microbial signature is defined, it can be used as a specific marker of deep water input and used to identify the most sustainable water sources.

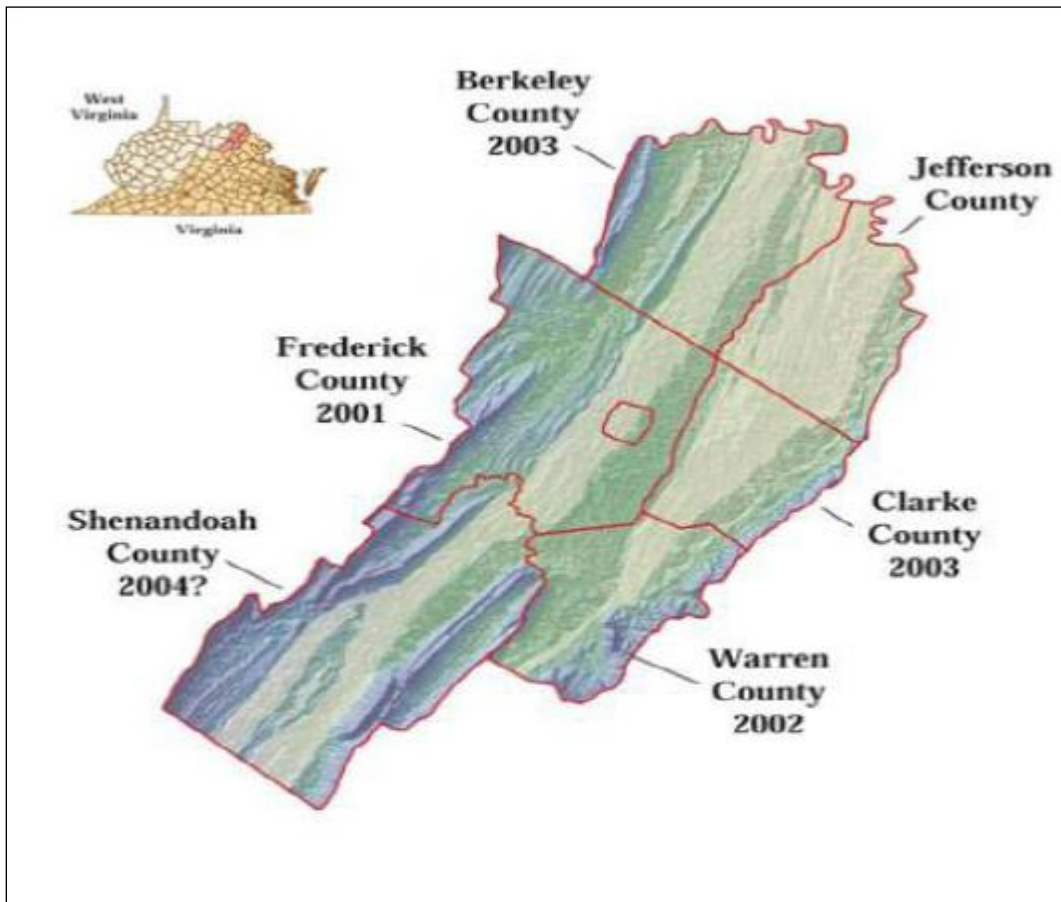


Figure 1: Map of the Northern Shenandoah Valley with start dates of county-scale cooperative ground-water assessments (USGS, 2003).

Contamination of surface water by pathogens in Lincoln County, WV

Fecal pollution of surface water constitutes a critical environmental and public health problem in the United States and worldwide. Also, fecal contamination of natural water negatively impacts fishing and other recreational activities, which is an important asset of the state of West Virginia. West Virginia's Lincoln County is one of the poorest areas in a poor state, and is susceptible to increase the likelihood that untreated wastewater is discharged into streams and rivers. The Left Fork watershed of Mud River in Lincoln County is 6 miles long and has approximately 250 families. In a case study from rural West Virginia, (MacDowell & Basden, 2007) had mentioned that "Tributary samplings since the late 1990's continually have shown higher than acceptable levels of total coliforms and *Escherichia coli*. Alternative wastewater systems have been installed for 28 homes in areas which have highest bacterial contamination, then monitor these areas to demonstrate the intervention's effectiveness in reducing bacterial contamination of surface waters. During the 12 month period ending October 2006, 195 periodic water samples were taken. 61% were over the acceptable *Escherichia coli* limits of 200 parts per 100 millimeters". It is critical that the sources of contamination are identified in order to develop solutions to water quality problems that protect human health.

Streams that are anticipated to be impacted by the construction of Appalachian Corridor-H, WV

In the early 1960s, the Appalachian Regional Commission proposed Appalachian Corridor H, which is part of a network of highways and parkways, to promote economic and social development throughout the Appalachian Region. In West Virginia, Appalachian Corridor H is the last to be completed. Appalachian Corridor H is a divided four lane highway which spans ~232km (144 miles), from Weston, West Virginia, to the Virginia State line, near Wardensville, West Virginia. Watersheds, which runs parallel to the proposed alignment of Corridor H, and many of its major tributaries, currently have low pH, little alkalinity, and high metals loading, due to the impacts of historic and current mining. The potential exists for increases in acid and metals loading as a result of new disturbances of existing mine refuge areas and by the creation of additional sources of water impairment associated with the exposure of new cuts and fills during the construction of Corridor H. While it is necessary to deal with potential acidic drainage resulting from the construction of Corridor H through mine spoils in the creek drainage, it is

equally important to address sources of water impairment originating above the roadway alignment, as such waters will be drained to creek via roadway culverts. Failure to address up-gradient sources of acid dissolved metals, and sediment would result in additional loading to passive treatment processes incorporated into roadway design and construction, which may render such mechanisms ineffective. Under contract with West Virginia Division of Highways, researchers in the Department of Civil and Environmental Engineering at West Virginia University monitor and analyze the water quality of streams that could be affected by the construction of the Appalachian Corridor H.

Thus, it is necessary to validate and optimize an innovative method, which uses the entire microbial community as a ‘fingerprint’ of water. The terminal restriction fragment patterns present in natural water can be used as a marker for tracking the groundwater flows, and pathogens source tracking in surface water. This is done by isolating 16S ribosomal DNA (rDNA) genes from water samples, followed by PCR-amplification and characterizing these 16S rDNA genes by restriction fragment analysis (T-RFLP).

Part of the results presented in this thesis were presented at the Northeastern section (45th annual) and Southeastern section (59th annual) joint meeting for the Geological Society of America (GSA), which took place on March 13-16, 2010 in Baltimore, Maryland.

1.3 Objective of the Research

The *overall objective* of this research is to validate the use of terminal restriction fragment (TRF) profiles present in different spring waters and stream waters as a marker of water source. The DNA pool of a complex microbial community, such as found in most ecosystems, including groundwater and surface water, provides very specific restriction fragment profiles that are used as the fingerprint of a body of water. T-RFLP results were then compared with other geochemical indicators in water, such as elemental analyses.

1.4 Working Hypothesis

The *working hypothesis* of this research is that the specific microbial terminal restriction fragment (TRF) profiles in spring waters and natural streams can be used as a specific signature to track the origin of groundwater, surface water and determine the source of microbial contamination.

1.5 Significance

The significance of this work is that, the results obtained are expected to pave way for more extensive use of complex microbiological profiles as indicators of surface water and groundwater origin. Once the microbial signature is defined, it can be used as a marker to identify the most sustainable water sources.

1.6 Specific aims

In order to achieve the overall objective of this thesis which consists of using the entire microbial community profile, obtained by terminal-restriction fragment length polymorphism (T-RFLP) analysis as a specific signature of deep water flow and stream water, the following *specific aims* were pursued:

Specific aim 1: To test the hypothesis that the specific microbial terminal restriction fragment profiles obtained from different springs in Huntingdon, PA, Great Valley, WV and Berkley springs, WV , can be used as a marker of the groundwater source. Also, the DNA extracted from spring water samples was used to identify bacterial species inhabiting the groundwater, by

screening a 16S rDNA gene library. The correlation between T-RFLP profiles and bacterial species was examined by clustering analysis.

Specific aim 2: To determine the origin of water contamination in the Left Fork Mud River Watershed, Lincoln County, WV, with high *Coliform* and *Escherichia coli* throughout the watershed. In this specific aim, DNA extracted from different streams along the watershed was analyzed using T-RFLP. Cluster analysis of TRF was done to determine the distances between clusters or samples.

Specific aim 3: To establish a proof of concept that TRF profiles of the microbial community structure present in several streams that are anticipated to be impacted by the construction of Corridor-H, WV, can be used as a signature of surface water source. The two previous aims were based on preliminary results obtained from Corridor-H watersheds, WV, showing that single watersheds exhibit very specific T-RFLP profiling that can be used to characterize the source of microbial contamination.

2.1 Microbial communities in natural water

Water forms the basis for all aquatic environments, in which the microorganisms, or microbes, constitutes the foundation of the aquatic abundance. Microbes include bacteria, bacteria-like organisms called archaea, viruses, protozoa, helminthes, and protists. Microbes have evolved that can live in saturated salt solutions at temperatures from below freezing to $>110^{\circ}\text{C}$ and in waters full of toxic substances. Natural water worlds would not be possible without microbes, which are natural and vital members of all aquatic communities. Groundwater originating from deep wells and subterranean springs are virtually free of bacteria due to filtering action of soil, deep sand and rock. However, it may become contaminated when it flows along the channels. Surface water is found in streams, lakes, and shallow wells and the air through which the rain passes may contaminate the water. Other possible sources of microbial contamination of a body of surface water are soil and agricultural run off, farm animals, industrial waste, discharge from sewage treatment plants and storm water run off from urban areas. Changes in the microbial community of water bodies might indicate contamination or other disturbances resulting from human activities. The living conditions (e.g. temperature, pH, water chemistry, flow velocity) in many aquifers are nearly stable. Anaerobic bacteria in natural waters play an important role in the microbial community structure, by increasing the richness of the biological activities (Santegoeds, Damgaard, Zopfi, & Muyzer, 1999).

2.1.1 Microbial ecology in natural waters

Microbial ecology is the relationship of microorganisms with one another and with their environment. Oxygen is a key parameter controlling microbial life. Fresh organic carbon (OC) from the soil and surface waters is the major carbon source for most groundwater bacteria (McMahon & Chapelle, 1991). Some essential elements for microbial growth are present in sufficient quantities in most aquifer (e.g. Ca, Mg, K), while the availability of others might be a limiting factor (e.g. N, S, P, Fe). These elements are essential for their metabolism, which are

released due to mineral weathering of microorganisms (Rogers & Bennett, 2004). Figure 2 illustrates the metabolic interactions between different microorganisms in aquifers. Organisms in aquifers use two main types of metabolic pathways; chemoorganoheterotrophic metabolism and chemolithoautotrophic metabolism (Nico Goldscheider, 2006). Chemoorganoheterotrophic organisms generate energy by the oxidation of organic substances and depend on OC for the biosynthesis of their cellular material. Chemolithoautotrophic organisms use CO_2 as a carbon source for biosynthesis and gain energy from the oxidation of inorganic components. Interaction between microbes depends both on the metabolic pathways and on the groundwater flow direction (Nico Goldscheider, 2006). Consequently, many microorganisms can rapidly adapt to changing environmental conditions due to their metabolic flexibility and, thereby exchanging genetic information (Chapelle, 2001). As groundwater and surface water organisms share complex ecological interactions, a direct impact on one species may induce indirect effects on other species. Thus, microbial ecology of natural waters remains an important and diversified field.

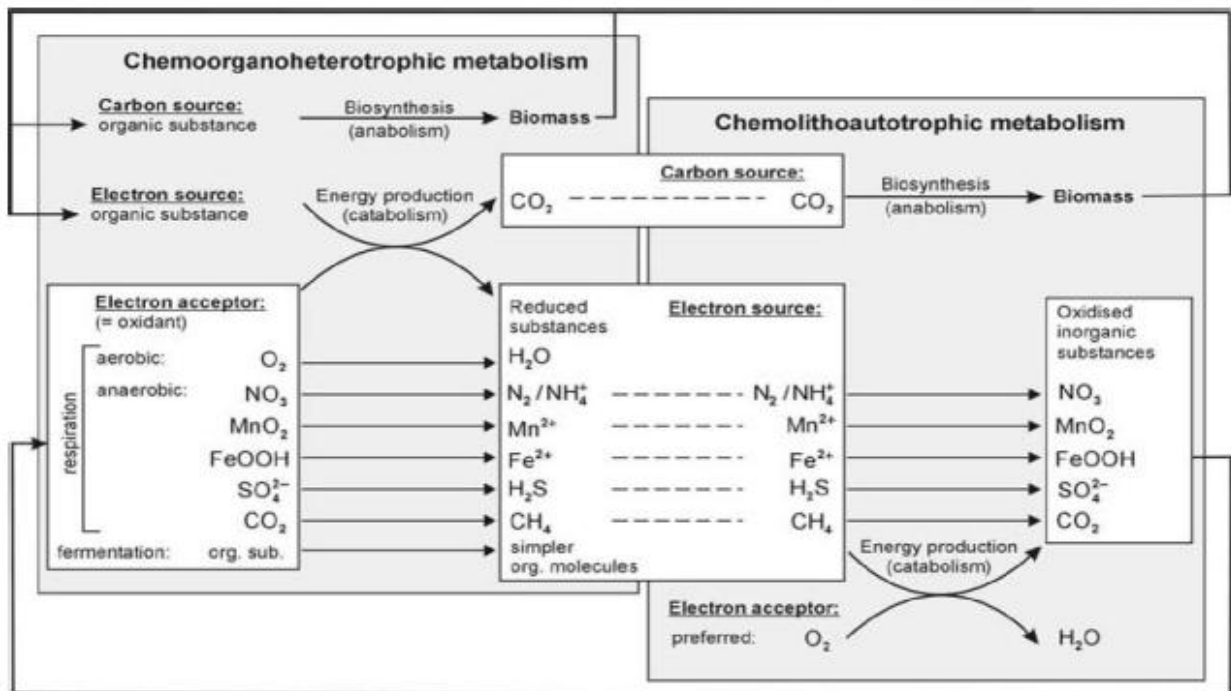


Figure 2: Simplified illustration of the metabolic pathways of microorganisms in aquifers (Nico Goldscheider, 2006)

2.1.2 Bioindicators for water quality

Indicator microorganisms are nonpathogenic and are used to predict the presence of the potential risk associated with pathogenic microbes (Troy M. Scott, 2002). Total and fecal coliforms have been used for many years as indicators for determining the quality of groundwater and surface water. Bacteria of fecal origin, such as *Escherichia coli*, are used extensively as indicators for groundwater quality (Auckenthaler & Huggenberger, 2003). These fecal indicators are found in fecal waste and have similar survival and transport properties as pathogens known to cause disease (Colford 2007). Bacterial indicators have been poor predictors of the presence of viruses and parasites, and are not able to identify the source of the contamination. Thus, fecal indicators have multiple limitations. However, microbial community as a whole can be used as an indicator of surface water and groundwater quality. The genetic fingerprinting of the microbial community provides information on the presence and relative abundance of the bacterial species and indicates the microbial diversity (Nico Goldscheider, 2006). Thereby, the development of microbial source tracking technologies has enhanced the ability of these indicator organisms to be used as tools for tracking sources of fecal pollution (Troy M. Scott 2002), and to identify the origin of surface water and groundwater.

2.2 Microbial Source Tracking

Microbial Source Tracking (MST), also commonly referred to as bacterial source tracking (BST), is a method used to determine the sources of fecal bacteria and establish whether fecal bacteria are being introduced into water bodies through human, wildlife, agricultural, or other wastes. Library-dependent and library-independent methods are the two types of source tracking methods. Library-dependent methods involve creating a larger dataset of particular target indicator bacteria present in waterbodies from a variety of sources. Water samples are then assayed and the results are compared to the identified organisms in the library to determine the origin of the fecal pollution. The disadvantage of library-dependent methods is that it requires a long period of time and resources to collect, identify, analyze, and store all of the organisms that may be present in a given area. This method also returns a large number of false positives and false negatives (Santo Domingo, Bambic, Edge, & Wuertz, 2007). Library-independent methods involve the detection of a marker specific to one species to identify the source of microbial

contamination. Library-independent methods include the use of bacterial genes, and the DNA sequences are first identified and then undergo validation testing to evaluate the specificity of the sequence as unique to a particular species. Water samples are then assayed for the source markers, generally through a non-culture based method (Santo Domingo, Bambic, Edge, & Wuertz, 2007) and DNA amplification through PCR. The library-independent method can use conventional or real-time PCR to detect DNA sequences. These methods return less false positive and false negative results than the library dependent method. However, like all methods for source tracking, library-independent methods do have disadvantages. There are few specific markers available and these markers may not be present in large quantities in the environment (Scott, 2005).

There a number of source-identification methods (Figure 3) that have been developed in response to the need for controlling fecal loading into receiving waters (USEPA, 2005). 99.8% of the microbes present in many environments are not readily culturable (Streit & Schmitz, 2004). However, recent developments in genomics and biotechnology are allowing microbiologists to use culture-independent techniques based on molecular markers to monitor the microbial quality of environmental waters. With the limitations of laborious and costly culture-based techniques, DNA-based molecular methods tend to be more rapid and less expensive; they target a large number of organisms that otherwise would not be quantifiable, because they cannot easily be cultivated in the laboratory. MST researchers are working towards developing molecular methods that are relatively fast, reasonably priced, and do not require host-origin libraries.

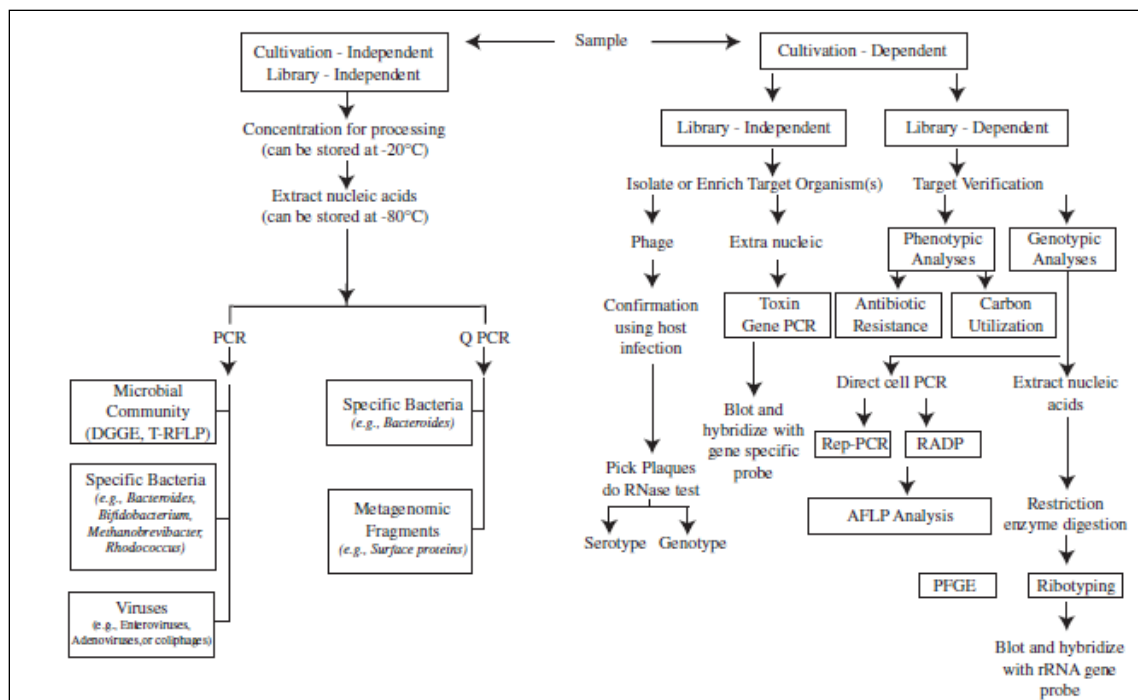


Figure 3: Schematic representation of some of the methods used for tracking sources of fecal pollution (Santo Domingo, Bambic, Edge, & Wuertz, 2007)

2.3 Molecular DNA fingerprinting methods

Molecular (DNA) fingerprinting methods are useful to monitor natural communities of bacteria in a body of water. There are a number of genetic fingerprinting methods that are being developed for discriminating between sources of indicator bacteria in natural waters (Lasalde, Guez, & Toranzos, 2005). The different molecular fingerprinting methods are pulsed-field gel electrophoresis (PFGE), restriction fragment length polymorphism, ribotyping, repetitive extragenic palindromic [REP]-PCR and amplified fragment length polymorphism (AFLP). Length heterogeneity PCR (LH-PCR) and terminal restriction fragment length polymorphism (T-RFLP) analyses (Clement, Kehl, DeBord, & Kitts, 1998) are methods which are used to analyze differences in the lengths of gene fragments and to estimate the relative abundance of each fragment. The data obtained with the molecular fingerprinting approaches are commonly complex, and thus cluster analysis are used to group the isolates by similarities and therefore identify the major source of bacterial contamination (Lasalde, Guez, & Toranzos, 2005). The difficulty of growing strict anaerobes can be circumvented by using molecular methods rather than culture-based methods to detect them.

2.4 Terminal-restriction fragment length polymorphism analysis (T-RFLP)

T-RFLP analysis is a polymerase chain reaction (PCR)-fingerprinting method that is commonly used for comparative microbial community analysis. T-RFLP is best suited for identifying individual strains within similar taxa or species. These polymorphisms allow the DNA to be cut in different locations when digested with a restriction enzyme (Snyder & Champness, 1997). Variations in the sequence of the original DNA fragments will result in the restriction site being different, and thus each different fragment following digestion may represent an individual species (Brown, 2006). These digested fragments are then analyzed using capillary sequencing technology allowing samples contained in 96- or 384- well plates to be sized in an overnight run (Thies, 2007). Variations in the fragment length caused by enzyme digestion are referred to as restriction fragment length polymorphism (RFLP) (Brown, 2006). The T-RFLP analysis technique has gained in popularity in recent years because it is rapid, highly reproducible, and often yields a higher number of operational taxonomic units than other, commonly used PCR-fingerprinting methods (Osborn, Moore, & Timmis, 2000). Statistical approaches have been used to compare T-RFLP fingerprints derived from different communities. The T-RFLP approach has been used successfully to analyze the composition of microbial communities in soil, water, and other environment. (Liu, Marsh, Cheng, & Forney, 1997), first described the T-RFLP method as it is practiced most commonly today. (Marsh T. L., 2005) provided detailed protocols for performing T-RFLP analysis. The growth in the T-RFLP literature has been increasing in recent years, from two referred journal articles in 1997 to 139 during 2006 (Thies, 2007).

The first requisite step for T-RFLP analysis is to extract community DNA or RNA from environmental samples. Purified DNA is quantified and concentrations are standardized between samples before PCR. Selected gene targets are then simplified from the sample DNA by PCR. The PCR reactions are prepared using primers where either the forward or the reverse primer is labeled with a fluorophore like HEX, FAM, or ROX dye chemistries. Fluorophores for primer end-labeling are chosen based on the type of automated sequencer that will be used for fragment sizing, as recommended by the manufacturer. PCR are optimized by passing through a PCR clean-up kit, such as those available from Qiagen (Valencia, CA) or Promega Corp. (Madison, WI). One or several restriction enzymes (REs) are used to hydrolyze PCR products. Amplified

DNA from different organisms that contain different restriction sites will yield terminally labeled fragments of different sizes. These terminal restriction fragment (T-RFs) are then sized on an automated DNA sequencer capillary methods, thus yielding a fingerprint that is characteristic of the community from which the DNA was extracted. The result from automated sequences is in the form of an electropherogram, having peaks representing fluorescently labeled T-RFs. The ABI 3730 capillary sequencer operates GeneMapper v3.5 (Applied Biosystems, Foster City, CA) software for sizing sample fragments by comparing their mobility with that of the size standard.

Once data are processed and fragment lengths assigned, the data set is typically imported into a spreadsheet program, such as Microsoft Excel. The T-RFs for each sample run should be closely examined and the entire run evaluated for the average number of T-RFs detected per sample and the number of T-RFs contained in the various size classes. Fingerprints of complex communities from environmental samples with overall low numbers of terminal fragments are probably compromised and should be rerun.

2.5.1 Disadvantages of the T-RFLP Method

Although T-RFLP has been used extensively, it has many potential pitfalls as well. Some of the pitfalls are while extracting DNA or RNA from environmental samples and analyzed subsequently by PCR (von, Gobel, & Stackebrandt, 1997), such as variation in cell lysis efficiency between different organisms and shearing of DNA during nucleic acid extraction. Sometimes, PCR artifacts, such as chimeras (Wang, Cao, & Cerniglia, 1996) can be generated by amplifying fragmented nucleic acids. However, purification of DNA extracts may resolve problems associated with PCR amplification failure. Loss of DNA from organisms present in low numbers, however, will introduce yet another bias into the analysis. Many sequence types can share the same T-RF length, resulting in overlapping community species that share the same T-RF length, leading to underestimation of community diversity (Thies, 2007).

Chapter 3: Materials and Methods

3.1 Water samples

3.1.1 Spring water

Sampling sites included locations known to discharge deep water and associated shallow-water systems. Water samples were collected from Cold and Warm Springs in Huntingdon, PA, Kilmer Spring, Newman Spring, Jim Smith Farm Spring, Bella Vista Spring, Big Spring and Snodgrass Spring in Martinsburg, WV, and Cacapon State Park and Ladies Spring in Berkeley Springs, WV. Three liters of water samples was filtered in the field through sterile 0.45 μ m filter membrane (Fisher Scientific), using a battery-operated vacuum pump (Soil Measurement Systems, Tucson, AZ). The battery-operated vacuum pump is designed to provide vacuum to suction lysimeters installed in the field. The filters were placed in 1.5mL graduated microtubes (Fisher Scientific) and were stored on ice during transport to the lab.

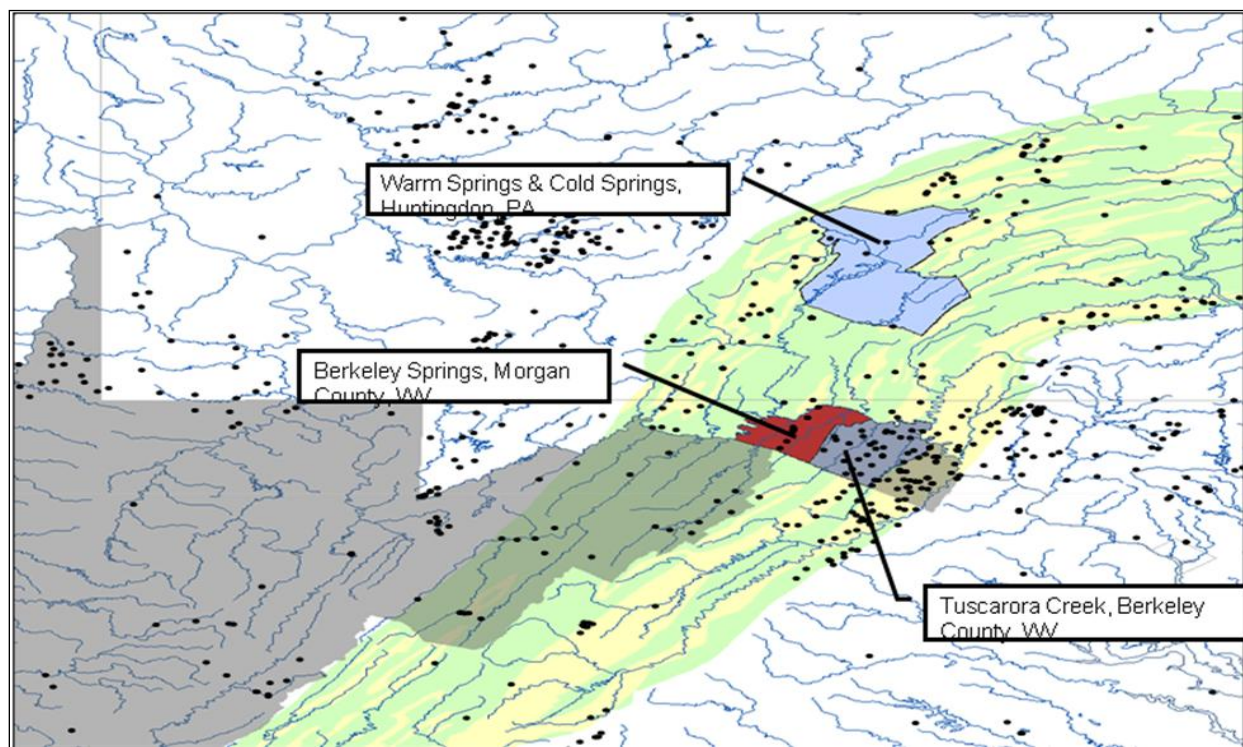


Figure 4: Map of the Shenandoah Valley in Virginia and location of sampling site

3.1.2 Stream water

West Virginia Water Research Institute (WVWRI) staff collected tributary samples from locations previously identified along Left Fork watershed of Mud River in Lincoln County, where wastewater treatment systems have been installed. Different streams include Dianne Adkins (DA), Beecher Adkins (BA), Ken Bailey (KB), Dallas Clay (DC), Willie Clay (WC), Pauley Ferrell (PF) and Delmar White (DW).

Water samples were also collected from different streams that are anticipated to be impacted by the construction of Corridor-H. Samples collected from 5 sites in Beaver Creek (BC) watershed, 6 sites in Mill Run (MR) watershed, and 6 sites in Walnut Bottom Run (WB) watershed, located in Preston, Tucker, and Hardy counties (Figure 5).

All samples were collected in 1-liter containers and were stored on ice during transport to the lab. Upon return to the lab, 500mL of water samples were filtered through sterile 0.2 μ m filter membrane (Fisher Scientific). The filters were placed in 1.5mL microtubes and stored at -80°C.

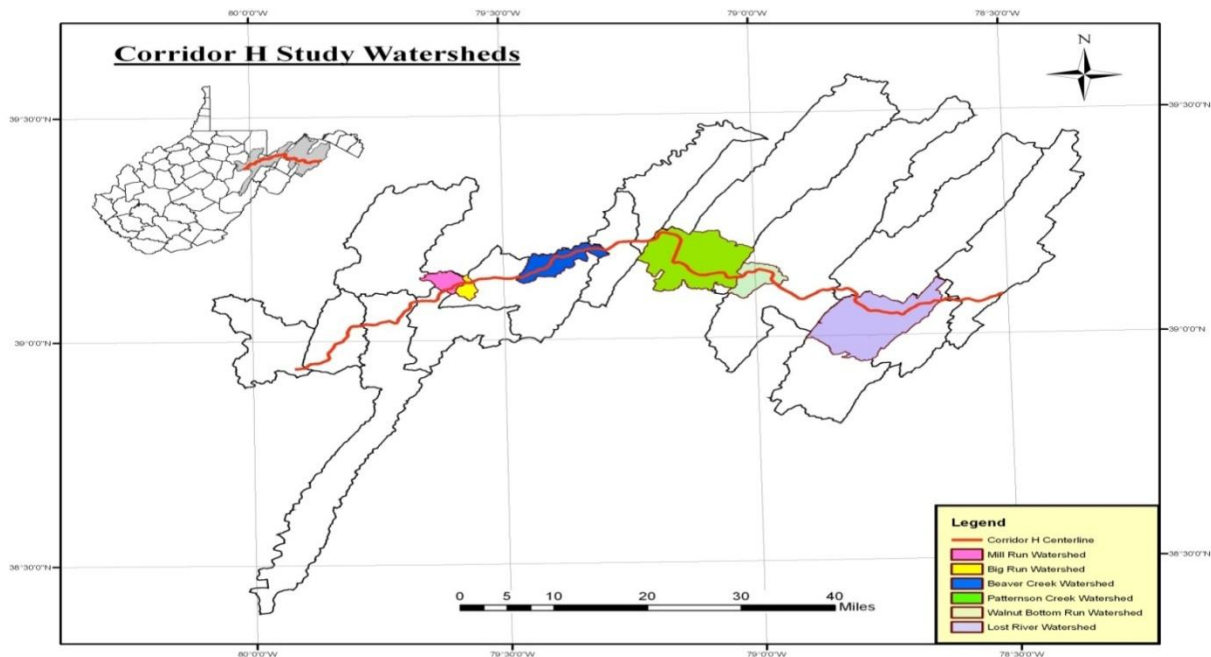


Figure 5: Corridor H alignment and study watersheds: Mill Run, Big Run, Beaver Creek, Patterson Creek, Walnut Bottom Run

3.2 Membrane filtration bacterial enumeration

Indicator organism enumerations were performed according to Standard Methods for the Examination of Water and Wastewater (Eaton, 2005). Specific volumes of water were filtered on sterile 0.45 μ m filters and were placed on an absorbent pad saturated with a specific cultivation medium in individual sterile plastic dishes. Dishes were then incubated in an incubator at a specific temperature for 24 hours. After incubation, colonies were counted visually. Only plates with 20 to 80 colony forming units (CFUs) were considered and at least 3 plates were counted per sample. Bacterial densities are expressed as number of cells per 100mL of water.

Bacterial colony-forming units (CFU) for spring water samples were estimated after two week incubation at 15°C on R2A medium. R2A medium was developed to determine the bacterial count in potable waters during treatment and distribution and has been shown to give significantly higher counts than Plate Count Agar (PCA) or similar high-nutrient media. Since its development in 1979, it has been found to allow the culturing of many other bacteria that will not readily grow on fuller, complex organic media.

3.3 DNA Extraction

Total DNA was directly extracted from the filters using UltraCleanTM Soil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA) following the manufacturer's recommendations. Filters were cut ~ 1-mm width strips and introduced inside micro-bead extraction tubes. Purified DNA was eluted from the silica membrane using two times 50 μ L of sterile elution buffer. The elution buffer is 10mM Tris pH 8 and does not contain salt. Thus, as the elution buffer passes through the silica membrane, DNA is released, and it flows through the membrane, and into the collection tube. The DNA is released because it can only bind to the silica spin filter membrane in the presence of salt. DNA extracts were quantified by the O.D.₂₆₀ using a NanoDrop ND-1000 spectrophotometer (NanoDrop, Wilmington, DE). The quality of the DNA was assessed by the ratio O.D.₂₆₀/O.D.₂₈₀ (all DNA samples showed a ratio higher than 1.9). Purified DNA was stored at -80 °C, until further analysis.

3.4 Microbial community fingerprinting using T-RFLP

The technique involves different steps including DNA extraction, PCR amplification, digestion of purified DNA, and analysis of the T-RFLP patterns (Figure 6). After DNA was isolated from water samples; PCR was conducted to amplify the target region of the 16S rDNA using a fluorescently labeled primer. PCR is a fast, efficient and economically viable method employed in the amplification of specific regions of DNA. It utilizes the specificity of primers to select a target sequence and the heat-stable properties of some DNA polymerases (e.g. *Taq* polymerase) to replicate these sequences at the high temperatures required for PCR (Hartwell et al., 2004). Restriction fragments obtained from digestion was separated by electrophoresis on an ABI (Applied Biosystems, Foster City, CA) DNA sequencer. Finally, the distribution and intensity of taxonomic groups were displayed in a T-RFLP profile. Statistical analysis of the fragments was done for comparing the microbial communities in order to track the source of water.

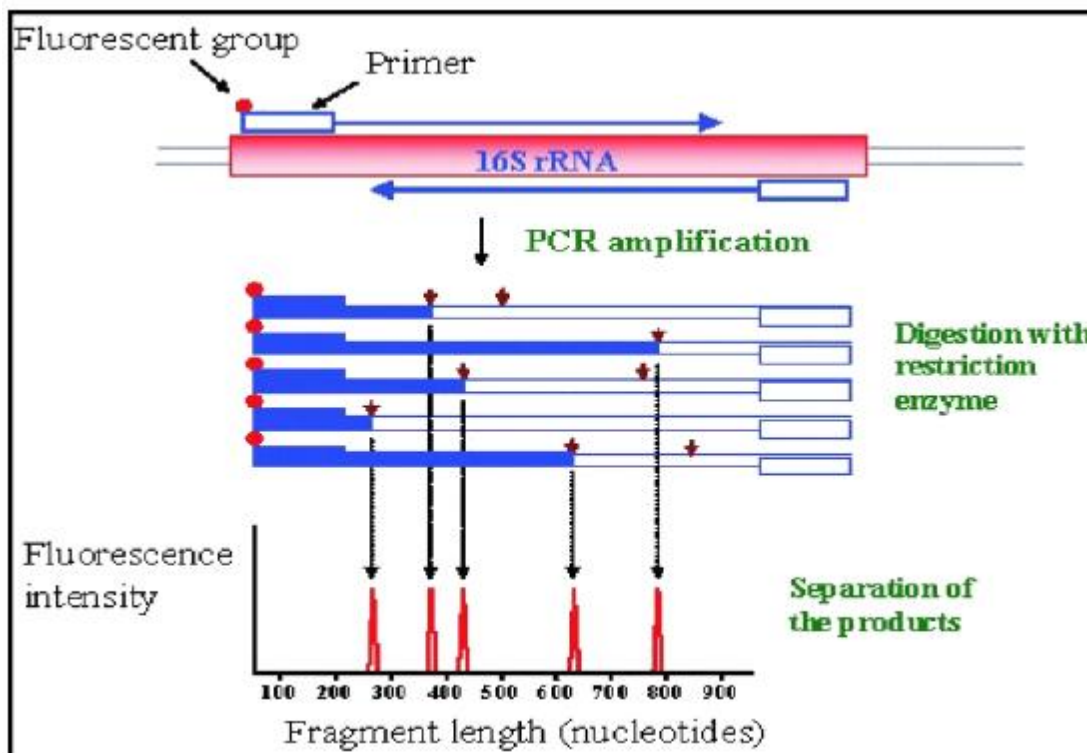


Figure 6: Scheme of a typical T-RFLP process (Source: <http://t-rfpred.sourceforge.net/method> (2007)).

3.4.1 16S rDNA Amplification

PCR amplification of 16S rDNA was performed on a MasterCycler Gradient[®] thermocycler (Eppendorf, Westbury, NY) using HotStarTaq[®] Master Mix Kit (Qiagen, Valencia, CA) and the bacterial universal primers, 27F fluorescently labeled at 5'-end with FAM (5'-/56-FAM/AGAGTTTGATCCTGGCTCAG-3) and 1492R . The DNA was first heated at 95 °C for 15 minutes to denature, or break the double helix into separate strands. The primers then bind, or anneal, to described locations on the single strand of DNA, and the sequence between the primers was filled with the use of an enzyme (Black, 1999). A second heating cycle was performed to denature the strands again at 95 °C for 1 minute, followed by annealing for 1 minute. Followed by, extension at 72 °C for 1.5 min; and final extension at 72 °C for 10 min. PCR products were analyzed by agarose gel electrophoresis (AGE). Three independent PCRs were performed for each sample; PCR products were then pooled and purified using QiaQuick PCR Purification Kit (QIAGEN, Chatsworth, CA). The purified PCR products were stored at -20°C until further analysis.

3.4.2 Agarose Gel Electrophoresis (AGE)

Agarose Gel Electrophoresis (AGE) is a method used to separate, purify and identify fragments of DNA depending upon their individual size, weight and chemical composition. The success and yield of each amplification reaction were determined by electrophoresis on 1.5% agarose gel in TAE buffer with 10µL aliquots of PCR products. The PCR products were quantified by comparing the band intensities to the intensity of 100bp DNA ladder (New England Biolabs, Ipswich, MA). The DNA was visualized in the gel by addition of 5µL ethidium bromide. Separation of fragments was achieved with the application of an electrical current across-linking, porous gel matrix. DNA samples are injected into the dye, and the DNA, being net negatively charged, passes through a gel as the result of the difference in electric potential (Black, 1999). Electrophoresis was done for 1 hour at 100 Volts, 400 mA. The gel was photographed using the gel documentation system (UV light). When performed properly, DNA that has been subjected to ethidium bromide staining becomes visible in the agarose gel as a small band when viewed under ultraviolet light. The location of the band depends on the number of base pairs present in the DNA strand. This is often used simply to determine whether or not PCR has worked. If the expected band is not present or multiple bands exist, something has gone wrong with the original

PCR procedure (Brown, 2006). Standard gel electrophoresis is helpful in estimating the mass of DNA present in a given volume of PCR product, which is important information needed for analyses such as digestion prior to T-RFLP or clone library construction.

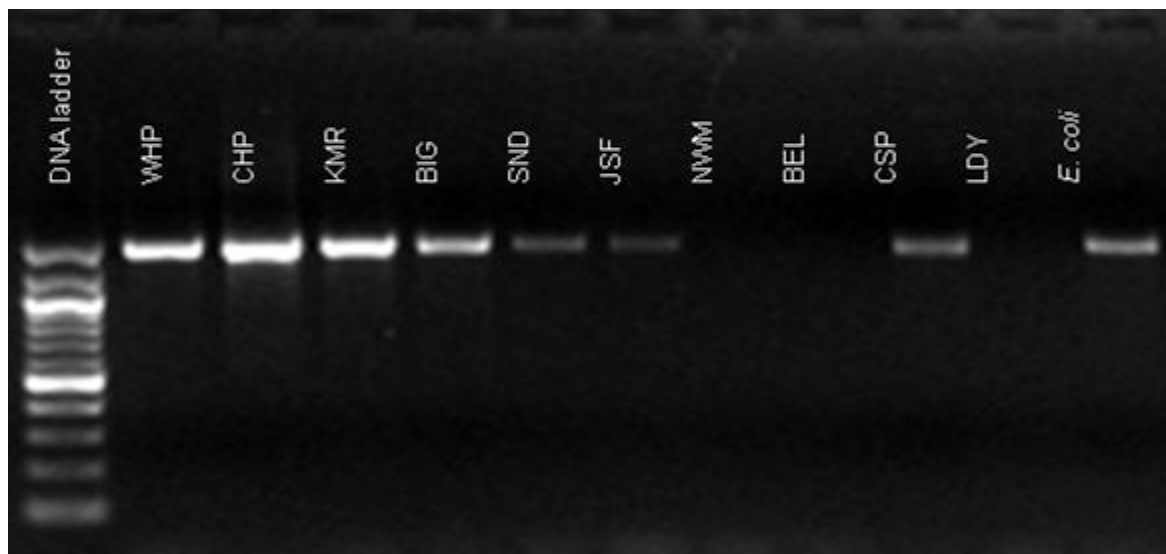


Figure 7: Agarose gel electrophoresis of 16S rDNA amplified by PCR from total DNA extract from spring waters.

3.4.3 Enzymatic digestion

Labeled purified PCR products were digested at 37 °C for 3 hours using restriction enzyme: *RsaI* (recognition site: 5'...GT[^]AC...3') (New England Biolabs, Ipswich, MA). The reaction mixtures with a total volume of 20 μL contained 2 μL of 10 x restriction enzyme buffers, 0.5 μL of the restriction enzyme, 7.5 μL of water, and 10 μL of the purified DNA. The samples were then incubated at 37°C for 4 hours.

3.4.4 Capillary electrophoresis

Digested 16S rDNA products (0.1-0.5 μL) was mixed with 0.5 μL of ROX-1000 DNA Size Standard (Applied Biosystems, Foster City, CA) and 8.5-9 μL of deionized formamide. Aliquots of the mixture (0.5-1 μL) were load in 96-well microplates. For denaturation of the DNA, the tubes were heated in the PCR machine at 95 °C for 3 minutes. The plates were immediately chilled on ice prior to electrophoresis with the automated DNA sequencer and centrifuged to around 2000rpm.

3.4.5 Terminal fragment profile analysis

Finally the fragments were analyzed on a capillary electrophoresis with an ABI 3130 DNA automated sequencer (Applied Biosystems, Foster City, CA). The sizes of the different fluorescently labeled T-RFs were determined by comparison with ROX-1000 DNA Size Standard (Applied Biosystems, Foster City, CA) by using Genescan v2.0.2 and GeneMapper v4.0 software (Applied Biosystems, Foster City, CA). Relative abundances of each microbial species in the samples were estimated by determining the peak areas and the ratios to the total peak area (Wu, Song, Ren, Kan, & Qian, 2004). Terminal fragments smaller than 10bp were discarded to avoid error from internal marker. T-RFLP profiles were obtained for clustering analysis. Figure 8 represents the profile of the species composition in the microbial communities for water samples collected from Dallas Clay stream in WV. The X-axis represents the number of base pairs detected for individual fragment (strain or species), and the Y-axis represents the relative prominence. The prominence is commonly referred to as fluorescence, area, or peak height. Each spike seen on the figure could possibly represent an individual species or strain. It is also possible for one species to be represented by several spikes on the output. Species or strains with larger fluorescent readings are more prominent in the community. Upon completion of the TRFLP analysis, a snapshot of the bacterial groups present in each sampling location was known.

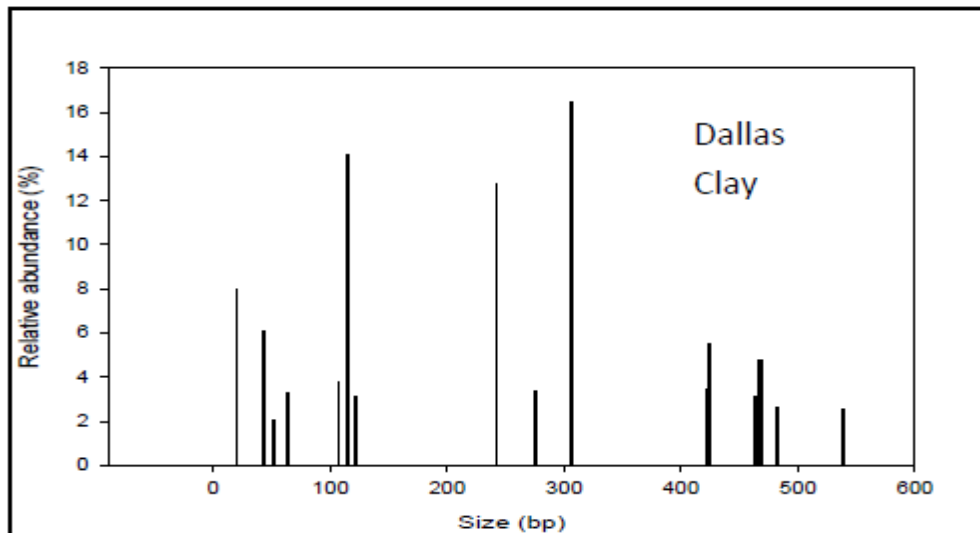


Figure 8: Profile of a 16S rDNA microbial community in one analyzed sample using T-RFLP

3.5 Bacterial Identification by screening of 16S rDNA gene library

To separate individual species and strains from a microbial community, bacterial cloning and construction of 16S rDNA clone libraries can be performed (Figure 10). A clone library is a collection of 16S rDNA that is representative of most of the microorganisms present in a community (Snyder & Champness, 1997).

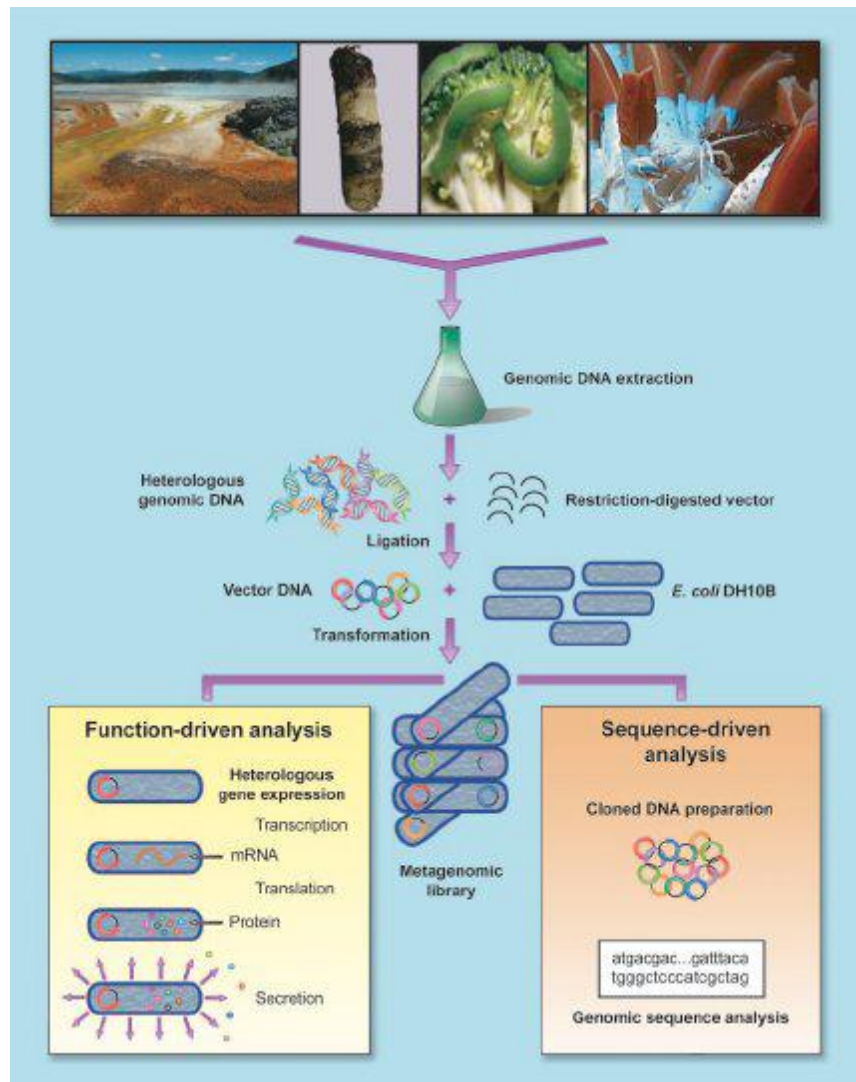


Figure 9: Flowchart summarizing the 'Metagenomic Process' (Handelsman, 2004).

To further characterize the bacterial populations inhabiting the groundwater in each spring, the bacterial diversity was analyzed by cloning PCR-amplified 16S rDNA genes. For each 16S rDNA pool, a clone library was constructed using TOPO TA Cloning[®] Kit (Invitrogen, Carlsbad, CA). DNA isolated from spring water samples was inserted into a plasmid, which acts as a vector. The mixture of DNA and vector (ligation reaction) was then transferred into competent *E. coli* cells. Because of the nature of the vector, only one DNA sequence will end up in each *E. coli* cell. This mixture was chilled on ice for 5 minutes and then exposed to heat shocking at 42 °C during 30 s. The cells were resuspended in super optimal broth with catabolite repression (SOC) medium and incubated for 1 hour at 37°C and 225 rpm. 100 µl of each cell suspension were plated onto LB agar plates containing ampicillin antibiotic, which was used to eliminate most cells that did not receive take up a vector. Each cell grows into a colony of millions of genetically identical cells, also known as clones. Plasmids from sorted colonies were extracted using QIAprep[®] Miniprep Kit (Qiagen, Valencia, CA).

Finally, purified plasmids were sequenced (Davis sequencing, Davis, CA). Sequencing results were edited using BioEdit, v5.0.9.; (www.mbio.ncsu.edu/BioEdit/bioedit.html). Sequences for each individual strain were entered into the classifier using ribosomal database project, RDP10; <http://rdp.cme.msu.edu/classifier/classifier.jsp>. Only sequences from dominant groups were determined. Therefore, this is the mechanism used to separate and isolate individual gene sequences within a mixed bacterial community.

Chapter 4: Results and discussion

4.1 Specific aim 1

Investigation of the hypothesis that Terminal- restriction fragment length polymorphism (T-RFLP) profiles present in deep and shallow spring waters can be used as a marker of groundwater source. In addition, genotypic identification of individual bacterial species inhabiting the groundwater was conducted by screening a 16S rDNA gene library.

4.1.1 Microbial number and Physico-chemical parameters of spring waters

Bacterial count and additional water analyses data were obtained from the spring sampling team. Total bacterial numbers data was obtained from Dr. Harris Edenborn (NETL). Figure 10 shows a relative bacterial numbers in deep and shallow water sources. We observed that, Berkeley Ladies Spring and Bella Vista Spring did not have enough colonies formed. Inside Bella Vista Spring house, the substrate is gravely with little to no other sediment. Whereas, Jim smith farm spring and Newman spring in the Great Valley (Martinsburg, WV) gave a relatively high bacteria count than the other springs. Most groundwater bacteria are viable but non-culturable. Consequently, cultivation techniques give an incomplete picture of the bacterial community, whereas genetic fingerprints of the entire community are obtained using molecular microbiology methods (Nico Goldscheider, 2006).

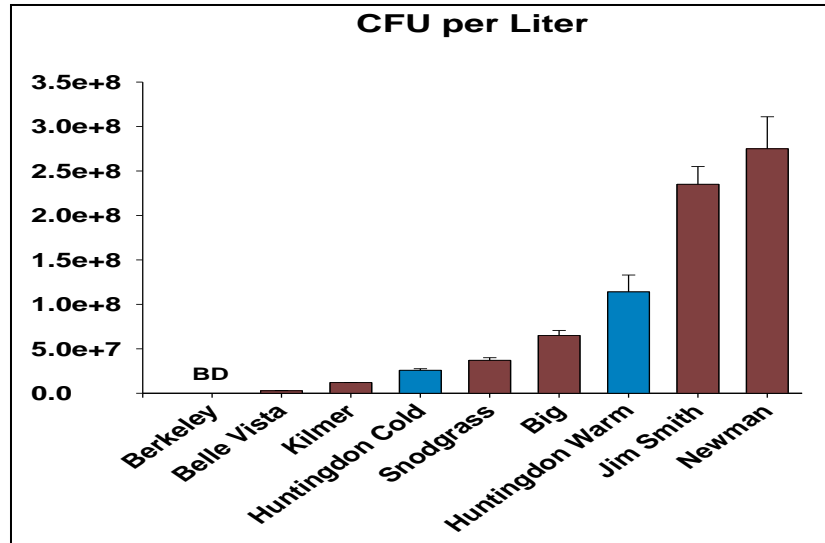


Figure 10: Bacterial number in each spring, as determined by CFU counting (Source: Dr. Harris Edenborn from NETL).

Microorganisms in karst are involved in a variety of geochemical processes, e.g. the dissolution and precipitation of minerals (Culver et al., 2004). Physicochemical parameters of each springs were determined at the time of sampling and the data was obtained from Dr. Dorothy Vesper. Dissolved oxygen (DO), pH, Sulfur (S), Silicon (Si), Nitrate (NO₃) and other parameters were collected.

Site	Deep/ Shallow	Temp	SC (uS/cm)	DO (%)	pH	Alk (mg/L HCO ₃)	S (ug/L)	Si (ug/L)	Ca (ug/L)	K (ug/L)	Mg (ug/L)	NO ₃ (mg/L)
CHP	Deep	10.35	31	78.6	5.11	73.45	2043	2340	2288	520	1128	0.299
WHP	Deep	16.86	163	56.6	6.74	73.45	4730	4108	26169	453	3944	0.13
BEL	Shallow	12.29	574	66.1	7.09	275.28	8933	5367	79757	1908	29338	21.42
JSF	Shallow	12.96	335	56	7.13	147.5	5940	4674	48610	1953	10848	2.69
KMR	Deep	12.42	679	65.3	6.93	295.47	11428	5553	1E+05	2741	21543	19.85
NWM	Shallow	14.08	515	25.5	7.04	251.65	8055	4467	86245	1614	13841	3.79
SND	Shallow	12.82	677	62	6.94	405.54	13608	5074	1E+05	2636	13314	13.05
BIG	Shallow	12.99	664	55.9	6.98	292.05	18425	4458	1E+05	2060	19118	12.29
CSP	Shallow	11.81	319	52.6	7.04	262.83	2258	4200	58838	579	6936	0.29
LDY	Deep	21.7	280	47.1	6.83	133.15	5228	4424	48259	906	4771	0.94

Table 1: Physico-chemical characteristics measured along the deep and shallow water sources of the karst groundwater systems (Data source: Dr. Dorothy Vesper, Dept. of Geology & Geography, WVU).

The dissolved oxygen (DO) and temperature in Cold Spring (CHP), Huntingdon differs from all other springs. Oxygen is the key parameter controlling microbial life (Nico Goldscheider, 2006), and the microbial life changes in the case of low oxygen content in the water. Also, Bella Vista spring (BEL), Big spring (BIG) and Snodgrass spring (SND) in Tuscarora creek, Eastern West Virginia, share high cation concentrations.

4.1.2 Terminal-restriction fragment length polymorphism (T-RFLP) analysis

The process of T-RFLP fingerprinting consists of three major steps: PCR amplification and fluorescence labeling of environmental 16S rDNA genes, digestion of the PCR product with one or more restriction enzymes, and T-RF separation by automated electrophoresis., and the use of the sophisticated electrophoresis system with laser detection provides high resolution and sensitivity for the detection of operational taxonomic units (OTUs) from a complex population. Because of these advantages, the T-RFLP method has been increasingly used for the characterization of bacterial communities in the environment (Clement, Kehl, DeBord, & Kitts, 1998). The 16S rDNA genes can be used to characterize the genetic diversity of microbial communities. Here, the different 16S rDNA genes are separated using terminal restriction fragment length polymorphism (T-RFLP). The result is the genetic fingerprint of the microbial community. This technique makes it possible to compare the microbial community from different spring water samples, and to analyze their similarities using statistical methods. The 16S rDNA extracted from samples was amplified using the universal primers and digested with *RsaI* restriction enzyme. Based on the PCR products analyzed by agarose gel electrophoresis (AGE), Newman spring, Bella Vista spring and Ladies spring did not show any DNA fragments (Figure 7). Comparing agarose gel electrophoresis result with CFU, gives a clear correlation that Bella Vista Spring and Berkeley Ladies Spring have a relatively low bacterial count. So, the TRF profiles for the seven springs were obtained after normalization by the sum of the areas for each group of peaks (Figure 11). The relative abundance obtained is the number of populations represented in the fingerprint of any given community (Liu, Marsh, Cheng, & Forney, 1997). Microbial populations that are not numerically dominant are not represented, because the template DNAs from these populations represents a small fraction of the total community DNA. Consequently, the species diversity of the microbial community is underestimated.

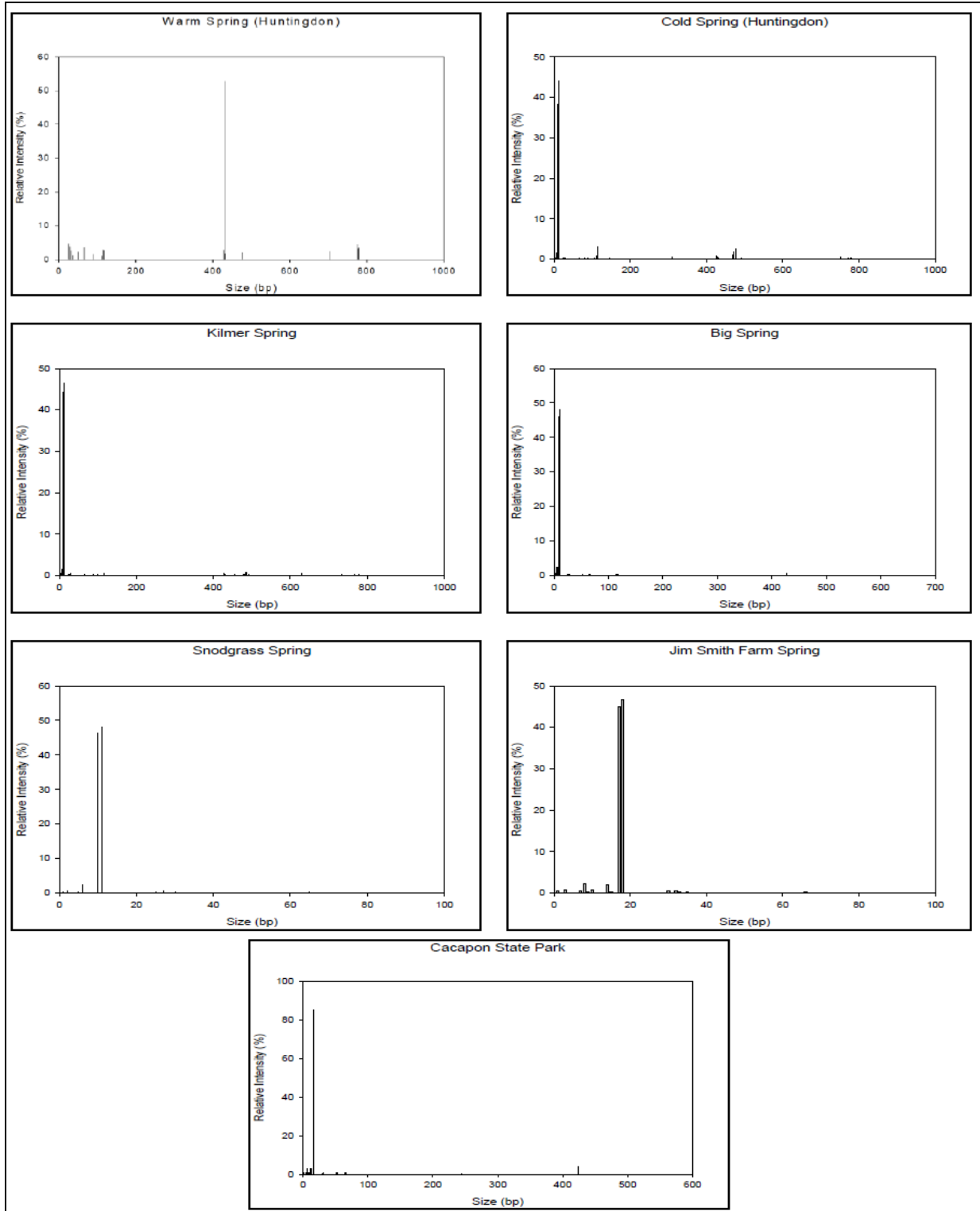


Figure 11: Terminal-restriction fragment profiles from 7 springs in Huntingdon (PA), and Great Valley, Berkeley Springs (WV).

To group samples having similar T-RFLP profiles, statistical methods such as clustering was used (Abdo & Williams, 2005). A cluster analysis was performed using SAS PROC and processed by using the Ward's minimum-variance clustering, which is a hierarchical method designed to optimize the minimum variance within clusters. In order to display the relationships among the clusters, a dendrogram was built (Figure 12). Hieght of each U represents the multidimensional distances between clusters or samples; the distance of the clusters or samples is represented in the length of the horizontal line.

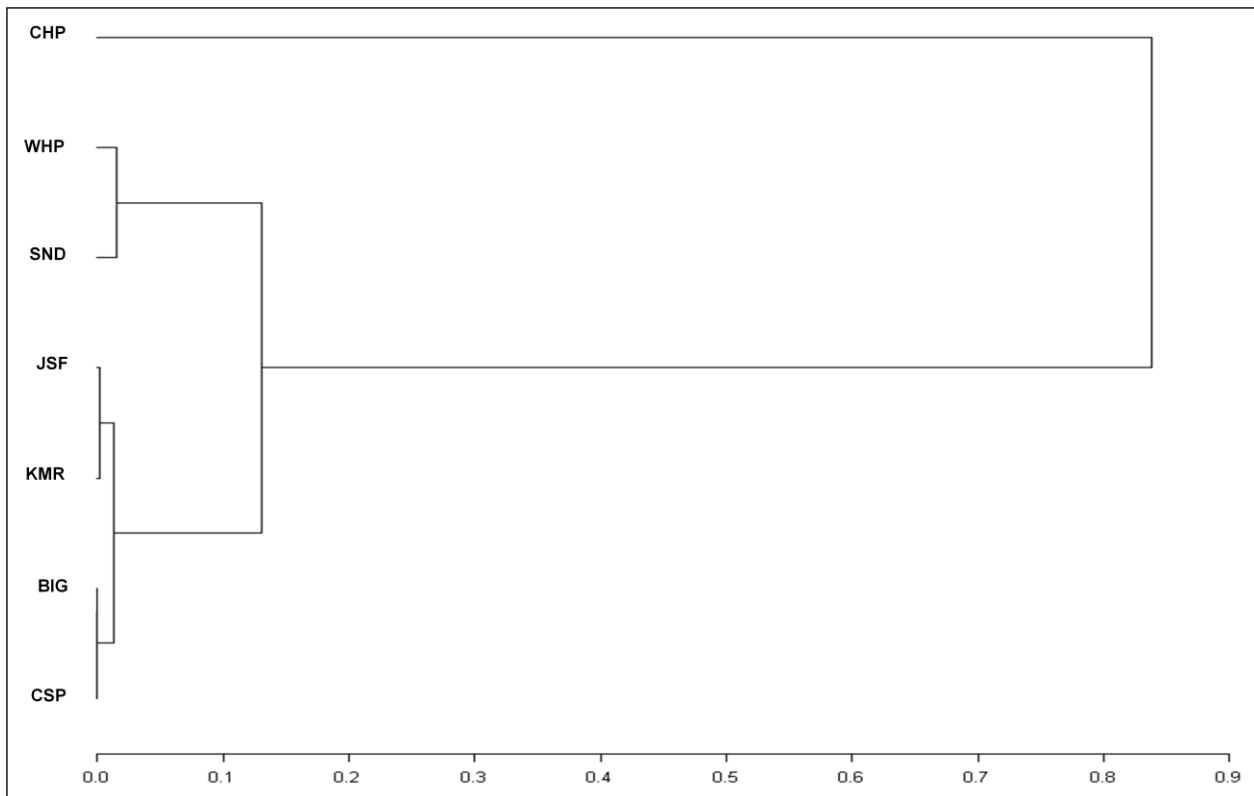


Figure 12: Clustering analysis of spring waters based on the microbial community as determined by T-RFLP profiling (SAS PROC: Ward's minimum-variance clustering).

From the dendrogram (Figure 12), it was observed that Big Spring (BIG) and Snodgrass Spring (SND) seemed not connected to each other, although the two springs are adjacent to each other geographically. On the other hand, Cold Spring (CHP) looks far connected from all other springs. This might be due to lower temperature and DO content in Huntingdon Cold Spring (CHP), which would have caused a variation in flow path due to change of microbial population.

However, except for Huntingdon Cold Spring (CHP), there was no clear relationship between the T-RFLP profiles and physico-chemical characteristics of spring samples. Cluster analysis of the T-RFLP profile did not show any relations between microbial community structure and source of groundwater. This is due to the problems obtained during T-RFLP analysis, which interprets the data.

Technical issues during T-RFLP analysis

There was technical problem with the Capillary Electrophoresis Sequencer at WVU, during analyzing the T-RFLP data. Consequently, the presence of some specific bacterial groups in a mixed population may be underestimated due to these technical problems (Hiraishi, Iwasaki, & Shinjo, 2000).

4.1.3 Genotypic identification of bacterial species using genomic library

In order to determine whether specific bacterial species can inform about the source of groundwater, whether surface or deeper flow, screening of 16S rDNA gene library was conducted. Previously extracted DNA samples from the springs were amplified by PCR with 27F and 1492R primers, purified, ligated with the vector, and transformed into *E.coli* as described previously. Incubation on the ampicilin plates yielded 40 to 50 clones from the spring samples. Plasmid DNA was extracted from 10 clones of each spring samples. Purified plasmids extracted from these clones were digested with the restriction enzyme (RSaI) for three hours at 37°C, and were sent for sequencing to Davis Sequencing, Davis, CA, as described. After sequencing, it was identified a large majority of karst microbial species were characterized as α -, β -, γ - and δ -*proteobacteria*, *Actinobacteria*, *Cyanobacteria*, *Synergistia*, *Sphingobacteria* and *Flavobacteria* are very common environmental bacteria present in water and soil. Figure13 and Table 2, shows bacterial species identified by comparison of their DNA sequences using public genomic databases (www.ncbi.nlm.nih.gov/; <http://rdp.cme.msu.edu/classifier/>).

4.1.3.1 Specific microorganisms as indicators for groundwater quality

Specific microbial species/activities are related to specific physical and chemical properties of groundwater. The composition of the bacterial community at the species level provides information about the microbial metabolic activities in groundwater. Identifying trophic groups in groundwater will help determine the geochemistry and therefore, the origin of water flows in karst systems. Figure 13 and Table 2 shows, some identified bacterial species that may be indicators of activities or chemistries in specific groundwater were *Thermomicrobia* found in Berkeley Ladies Spring (LDY), *Thermodesulfobacteria* in Snodgrass Spring (SND) and Kilmer Spring (KMR), *Aquificae* in Snodgrass Spring (SND), *Halobacteria* in Huntingdon Warm Spring (WHP). Less common bacteria identified in springs that may be used as specific indicators were *Holophage* in Kilmer Spring (KMR), *Opitutae* in Bella Vista Spring (BEL), and *Erysipelotrichi* in Snodgrass (SND) and Kilmer (KMR) springs.

Class	Alpha proteobacteria	Betaproteobacteria	Gammaproteobacteria	Deltaproteobacteria	Cyanobacteria	Actinobacteria	Synergistia	Sphingobacteria	Flavobacteria	Thermomicrobia	Verrucomicrobia	Thermodesulfobacteria	Deinococci	Opitutae	Erysipelotrichi	Clostridia	Chlamydiae	Aquificae	Holophage	Holobacteria
CHP	3	3	2	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
WHP	3	2	1	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	0	1
BEL	1	4	1	0	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0
JSF	5	0	2	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0
KMR	4	0	0	0	0	1	2	0	0	0	0	2	0	1	0	0	0	0	0	0
NWM	1	0	1	0	0	3	1	0	1	0	0	0	0	0	0	1	1	0	0	0
SND	0	3	1	0	0	1	0	1	1	0	0	1	0	0	0	0	0	1	1	0
BIG	0	4	1	0	1	1	0	1	1	0	0	0	1	0	0	0	0	0	0	0
CSP	0	1	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
LDY	3	0	1	0	3	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0

Table 2: Bacterial species in spring waters identified by sequencing individual 16S rDNA clones

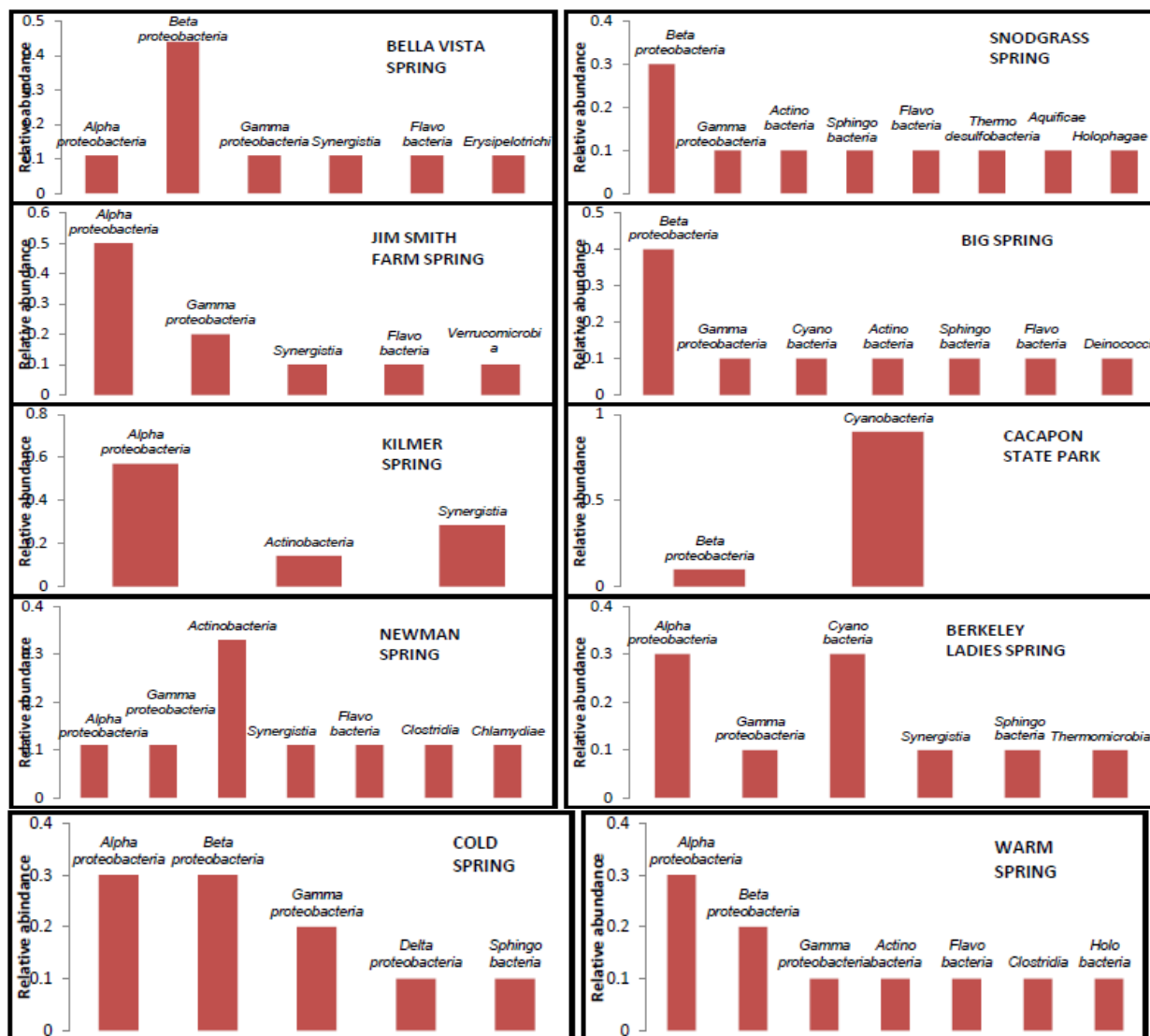


Figure 13: Bacterial species identified by comparison of their DNA sequences with public genomic databases (www.ncbi.nlm.nih.gov/; <http://rdp.cme.msu.edu/classifier/>)

Thermomicrobia: The Thermomicrobia are a group of thermophilic green non-sulfur bacteria.

The genus *Chloroflexus* is a photosynthetic bacterium isolated from hot springs, belonging to the green non-sulfur bacteria. This organism is thermophilic and can grow at an optimum temperature of ~55°C. It uses sulfide as the electron donor.

Thermodesulfobacteria: The Thermodesulfobacteria are a small group of thermophilic sulfate-reducing bacteria. Although not a true hyperthermophile because its growth temperature optimum is only 70°C, *Thermodesulfobacterium* is the most thermophilic of all known sulfate-reducing Bacteria. *Thermodesulfobacterium* is a strict anaerobe, cannot utilize acetate as an electron donor in its energy metabolism and instead uses compounds like lactate, pyruvate, and ethanol, reducing SO_4^{2-} to H_2S .

Aquificae: The Aquificae phylum is a diverse collection of bacteria that live in harsh environmental settings. They have been found in hot springs, sulfur pools, and thermal ocean vents. Members of the genus Aquifex, for example, are productive in water between 85 to 95 °C. They are the dominant members of most terrestrial neutral to alkaline hot springs above 60 °C. They are autotrophs, and are the primary carbon fixers in these environments. They are true bacteria (domain bacteria) as opposed to the other inhabitants of extreme environments, the Archea.

Holobacteria: In taxonomy, the Halobacteria (also Halomebacteria) are a class of the Euryarchaeota, found in water saturated or nearly saturated with salt. They are also called halophiles, though this name is also used for other organisms which live in somewhat less concentrated salt water. They are common in most environments where large amounts of salt, moisture, and organic material are available. Halobacteria can grow aerobically, anaerobically.

Clustering analysis was conducted to determine the similarity between the bacterial species using SAS PROC and processed using Ward's minimum-variance clustering. Figure 14 shows the dendrogram obtained from the clustering analysis of microbial community as determined by 16S rDNA gene library of spring waters. Clustering based on screening a clone library was more meaningful which showed a close association between Big Spring and Snodgrass Spring in the Martinsburg area. But the estimated abundances obtained from the genomic library may not reflect the real percentages of the bacterial groups in groundwater because of the relative small quantity of colonies that were analyzed, resulting in the exclusion of a variety of potentially important bacteria. As a result, there was no correlation between the microbial community structure using T-RFLP analysis and genomic library. This is due to the technical problems experienced while analyzing T-RFLP. However, this can still be used as preliminary result

showing that the microbial community profiles in spring waters can be used as a marker of groundwater source.

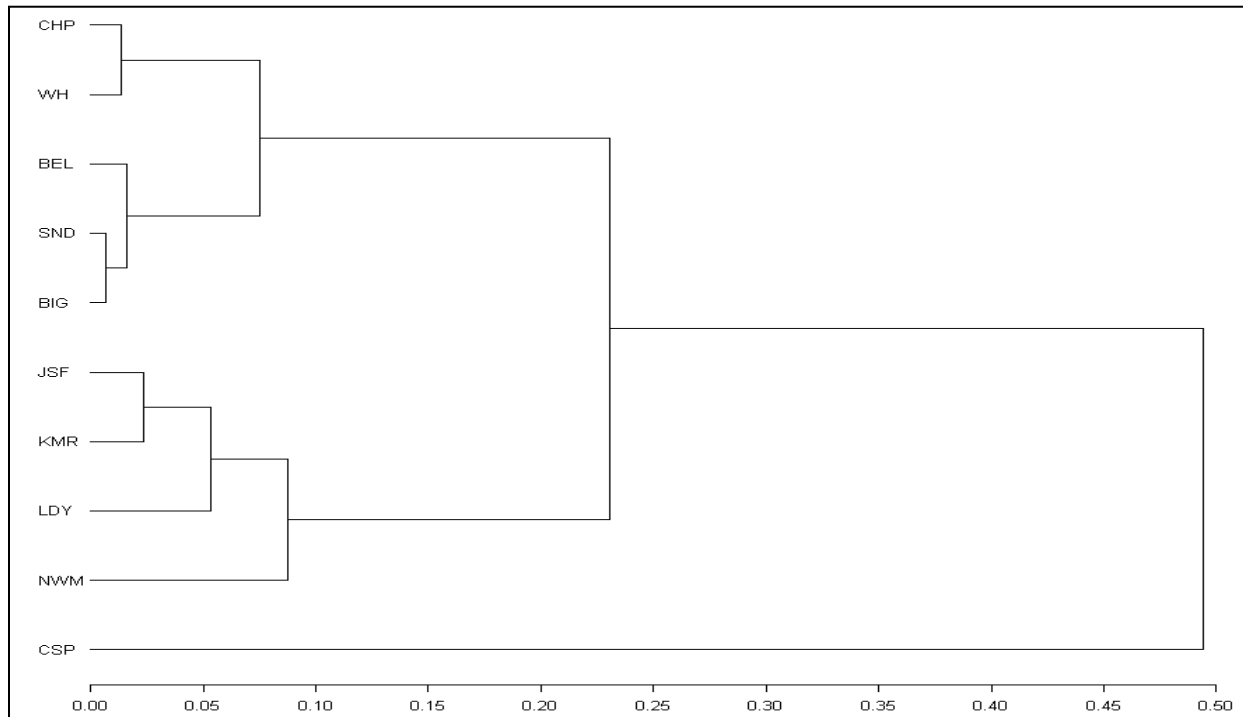


Figure 14: Clustering analysis of spring waters based on the microbial community as determined by screening a 16S rDNA clone library.

4.2 Specific aim 2

To develop a proof of concept that the microbial community structure of surface water can be used as a marker to determine the origin of water contamination, the DNA extracted from different contaminated streams in Left Fork Mud River Watershed, Lincoln County, WV, were analyzed using T-RFLP.

4.2.1 Indicator Organisms in Left Fork Mud River watershed

In coordination with the WVVRI, water samples were collected from the Mud River Watershed in Lincoln County, WV, where wastewater treatment systems have been installed. Different streams were Dianne Adkins, Beecher Adkins, Ken Bailey, Dallas Clay, Pauley Ferrell and Delmar White. The streams have been monitored for indicator organisms using membrane filtration methods and the data obtained from WVVRI staff showed that, there was

periodically high level of total coliforms up to 100,000 cells/100mL (Figure 15), fecal coliforms up to 5,000 cells/10mL (Figure 16), and *Escherichia coli* up to 2,500 cells/100mL (Figure 17).

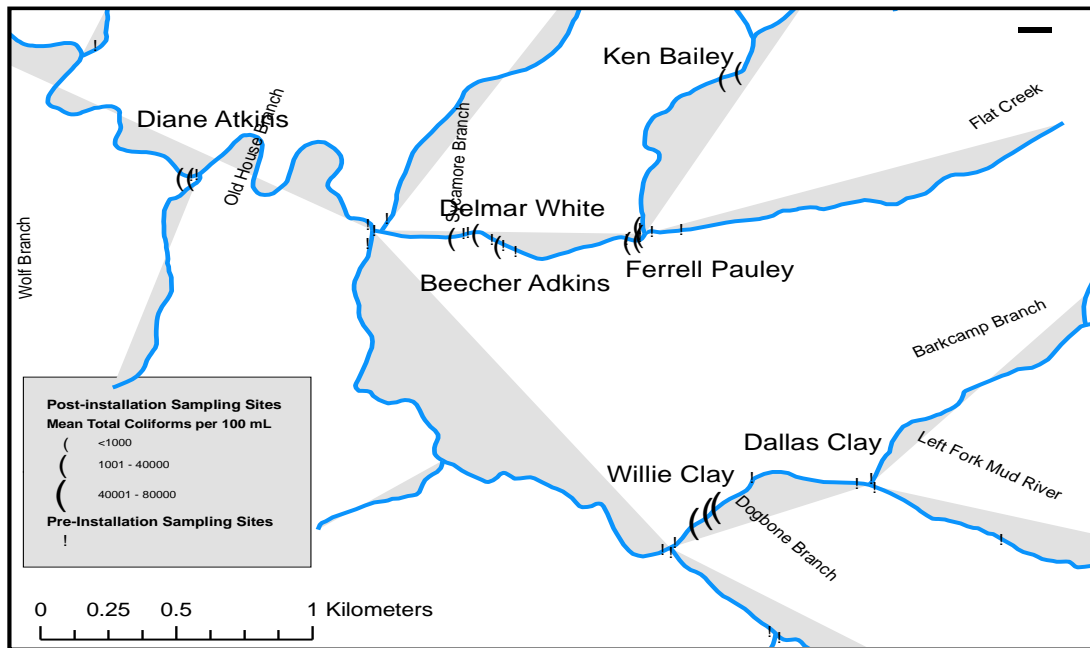


Figure 15: Mean Total Coliform measurement (per 100mL) at each sampling site in the Left Fork Mud River watershed (Vandivort & Fulton, 2008).

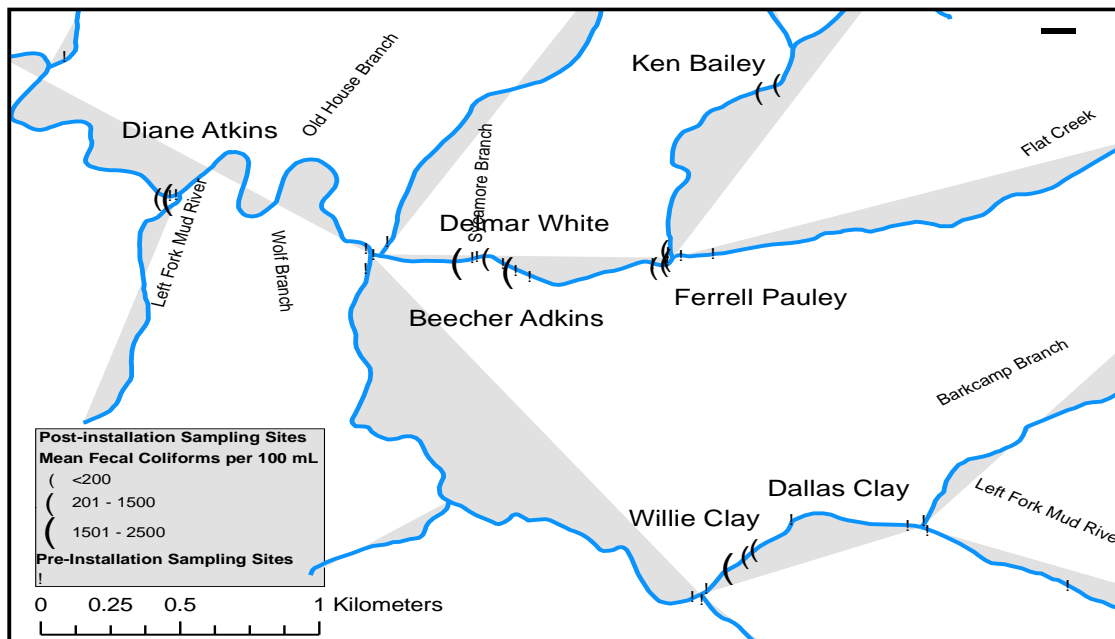


Figure 16: Mean Fecal Coliform measurement (per 100mL) at each sampling site in the Left Fork Mud River watershed (Vandivort & Fulton, 2008).

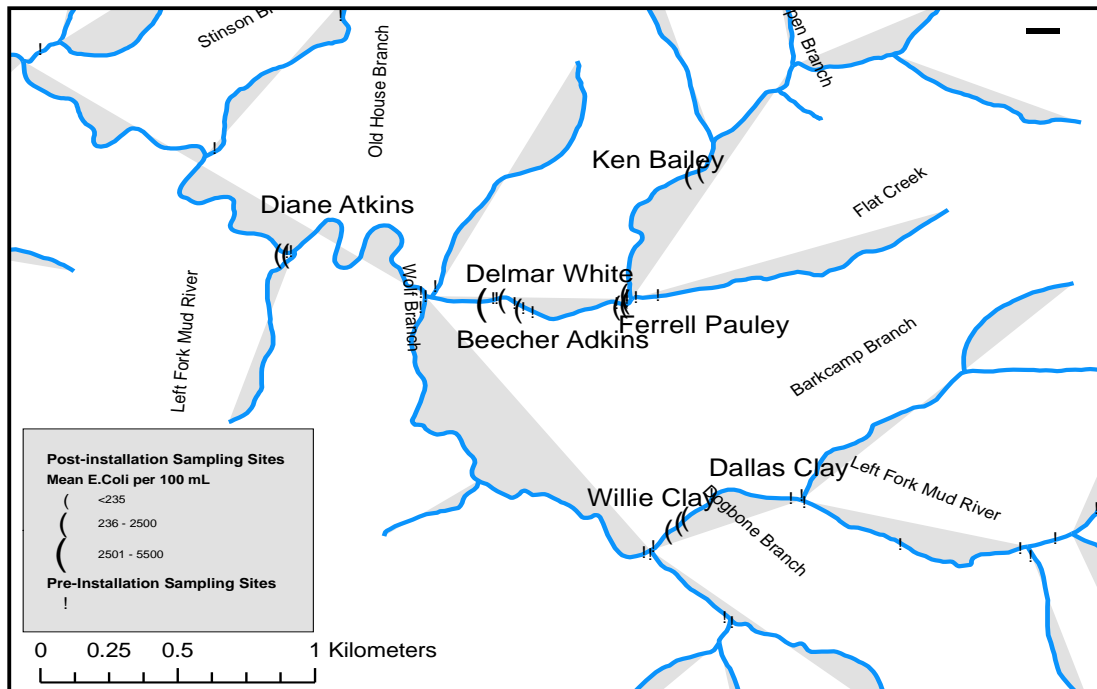


Figure 17: Mean *E.coli* measurement (per 100mL) at each sampling site in the Left Fork Mud River watershed (Vandivort & Fulton, 2008).

Tributary samplings since the late 1990's continually show higher than acceptable levels of Total *Coliforms* and *Escherichia coli*. The Mud River watershed constitutes therefore an adequate site for developing and optimizing T-RFLP method.

4.2.2 Microbial community in Left Fork Mud River watershed: T-RFLP analysis

The 16S rDNA extracted from samples for TRF analysis was amplified with the universal primers 27f fluorescently labeled at 5'-end with FAM and 1492r. PCR amplification was carried as previously described. Three independent PCRs were performed for each sample. Agarose gel electrophoresis (AGE) of 16S rDNA amplified by PCR from different streams in the Left Fork Mud River watershed was run to check the effectiveness of DNA extraction and PCR procedure. PCR products were pooled and purified using QiaQuick PCR purification kit (Qiagen, Valencia, CA). Purified amplifications were digested separately with the restriction enzyme (*RsaI*). Terminal restriction fragment (TRF) sizes were determined using Genescan v2.0.2 and Gene Mapper v4.0 software (Applied Biosystems, Foster City, CA). Finally, the distribution and

intensity of taxonomic groups are displayed in a T-RFLP profile (Figure 18). The peaks represent fluorescently labeled T-RFs detected over time in relation to the size standard. Similarity of microbial communities among the streams was determined by clustering the T-RFLP data using SAS PROC: Ward's minimum-variance clustering.

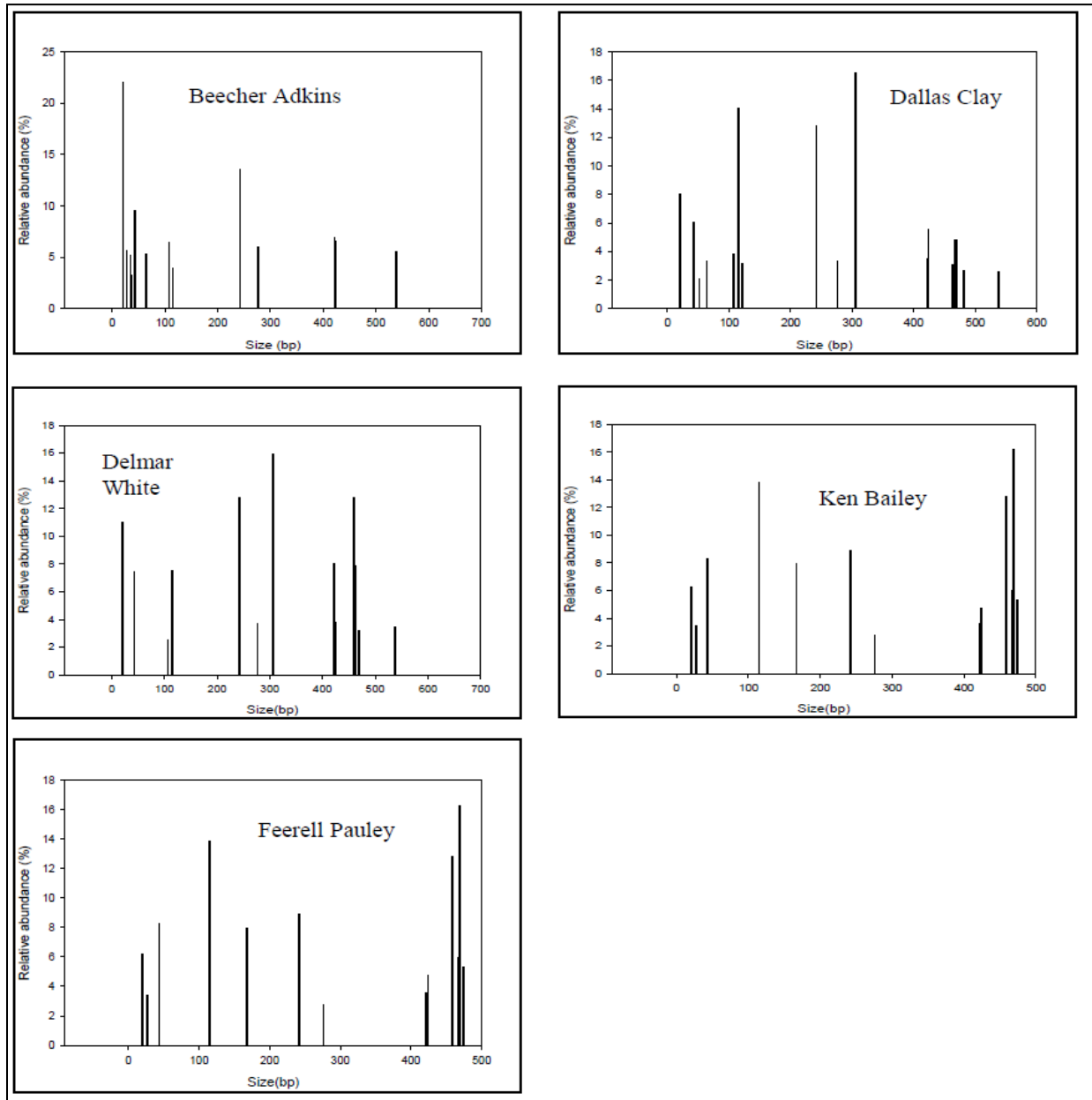


Figure 18: T-RFLP profiles of the bacterial community of different streams from the Left Fork Mud River watershed, Lincoln County (WV).

Figure 19 presents a dendrogram of selected T-RFLP profiles showing that water samples collected from different streams of the Mud River watershed carry a very specific signature that can be used for microbial source tracking analysis. From the dendrogram, it is observed that Delwar White, Ken Bailey and Feerell Pauley streams are closely connected, having similar microbial communities. And, Beecher Adkins and Dallas Clay are connected to each other. Thus, this microbial community structure can be used as a marker for tracking the origin of contaminated water in the Left Fork Mud River Watershed in Lincoln County, WV.

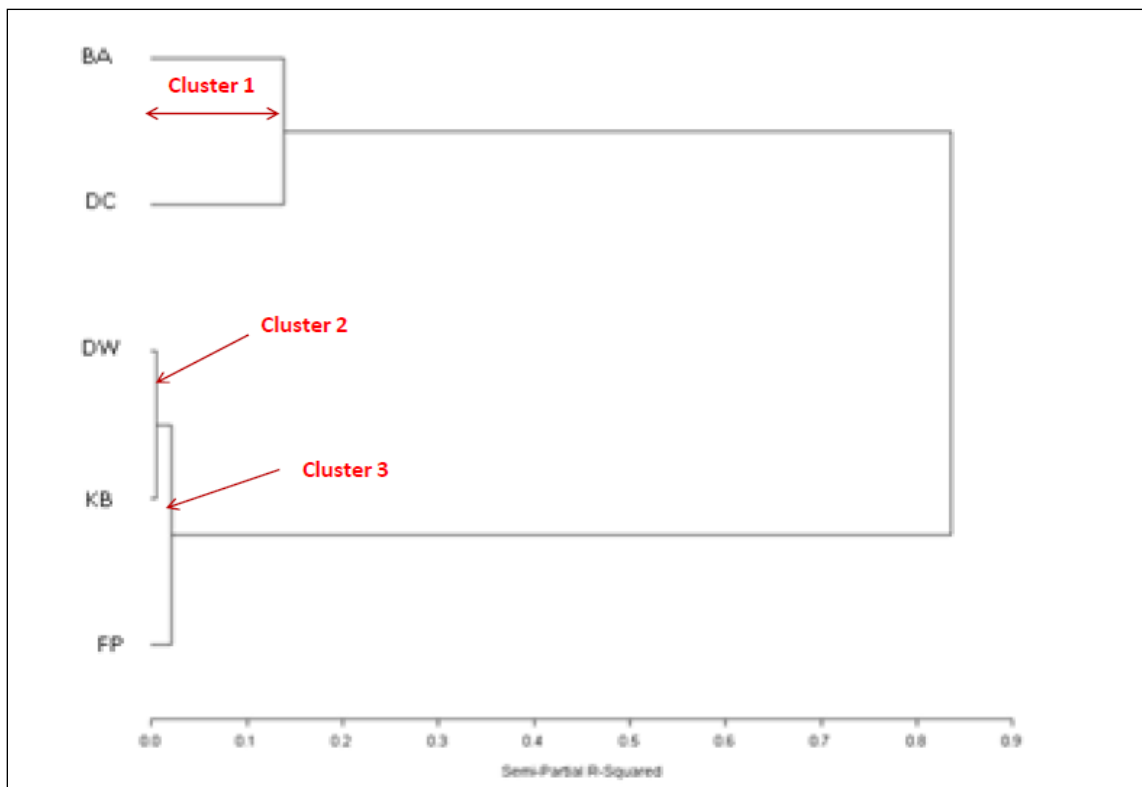


Figure 19: Dendrogram obtained from the clustering analysis of T-RFLP profiles of streams from the Mud River watershed. SAS PROC processing using “Ward’s minimum-variance clustering”.

4.3 Specific aim 3

Investigation of the hypothesis that microbial community structure in several streams along Corridor-H watersheds in WV, provides very specific restriction fragment profiles that can be used as a fingerprint for tracking the source of surface water.

4.3.1 Microbial number and physico-chemical parameters of Corridor-H watershed

Bacterial counts and physico-chemical parameters were obtained from the WV DoH sampling team. Table 3 shows the cultivation results and the relative source of contamination, of different streams in Beaver Creek (BC), Mill Run (MR) and Walnut Bottom Run (WB) watersheds along the Corridor-H, West Virginia. Total Coliforms and Enterococci measurements were performed according to Standard Methods for the Examination of Water and Wastewater.

Site	Source of Contamination	Coliforms CFU/100mL	Enterococci CFU/100mL
WB-1		16	5
WB-2		70	2
WB-3	Pasture	182	ND
WB-4	Pasture	48	ND
WB-5	High pasture	300	12
WB-6	High pasture	2080	15
BC-1		96	ND
BC-3		150	ND
BC-4		196	ND
BC-5		156	14
BC-10		114	ND
BC-12		4	10
MR-2	Residential	44	ND
MR-3	Residential	96	ND
MR-4	Residential	62	2
MR-5		116	ND
MR-6		42	ND
MR-7		70	6

Table 3: Cultivation results of the different streams from the Corridor-H watershed, WV.

Site	Turb. (NTU)	Flow (m ³ /sec)	pH	Sp. Cond. (uS/cm)	TSS (mg/L)	Alkalinity (mg/L CaCO ₃)	Acidity (mg/L CaCO ₃)	Sulfate (mg/L)	Chloride (mg/L)	Iron (mg/L)	Calcium (mg/L)	Magn. (mg/L)	Mang. (mg/L)	Alum. (mg/L)	Nitrate (mg/L)	Amm. (mg/L)
WB-1	4.77	0.003	8.1	312.10	1.25	220.60	14.53	22.30	1.54	0.14	57.20	8.22	0.02	0.08	1.01	0.09
WB-2	1.21	0.017	8	448.27	0.62	242.60	16.60	42.67	1.95	0.07	68.47	10.48	0.02	0.03	0.96	0.06
WB-3	2.12	0.001	7.9	327.10	0.70	179.00	9.89	53.90	3.88	0.07	64.00	6.25	0.02		0.25	0.01
WB-4	24.7	0.039	7.9	483.00	16.00	208.30	10.59	89.50	3.31	0.51	69.40	10.20	0.05	0.24	0.43	0.17
WB-5	8.16	0.023	8.1	430.10	3.65	206.90	8.27	74.20	2.97	0.22	71.60	10.20	0.04		0.44	0.09
WB-6	2.68	0.023	7.7	413.00	1.25	184.70	10.09	86.80	6.76	0.11	67.20	12.20	0.04	0.04	0.12	0.01
BC-1	1.73	0.009	4.8	37.60	0.65	12.20	11.38	8.50	5.37	0.21	2.06	0.32	0.04	0.37	0.11	0.01
BC-3	5.84		6.6	66.20	2.48	21.56	5.69	15.60	4.78	1.40	4.71	1.42	0.09	0.26	0.32	1.17
BC-4	1.55		4.4	392.20	0.50		94.29	181.10	2.44	5.20	22.50	10.08	2.47	3.98	0.10	0.02
BC-5	0.52	0.001	2.9	679.00	6.37		124.16	245.00	2.01	11.40	12.20	17.50	6.20	5.38	0.12	0.07
BC-10	1.53	0.014	3.5	152.60	1.35		36.78	81.60	2.53	1.29	6.10	3.98	0.83	2.16	0.12	0.05
BC-12	1.91	0.008	4.7	11.90	1.33	11.34	6.50	1.70	0.73	0.36	0.65	0.42	0.06	0.21		0.01
MR-2	0.63		7.3	80.90	2.60	110.00	4.98	13.00	2.26	0.01	4.60	2.41	0.01	0.04	0.58	0.06
MR-3	0.45		5.7	43.30	0.15	24.40	3.32	12.90	1.16	0.06	14.30	1.68	0.01		0.50	0.00
MR-4	0.99	0.051	7.5	102.80	0.65	49.24	17.33	12.80	8.46		16.80	2.22	0.01		0.64	0.01
MR-5	0.83		6.7	74.90	0.40	57.90	5.35	12.30	9.96	0.01	8.30	2.48	0.01		0.74	0.01
MR-6	2.33		6.9	90.70	0.20	39.16	3.87	15.80	2.26	0.04	4.90	2.47	0.01		1.00	0.00
MR-7	1.22		7.8	190.20	1.15	71.68	5.35	13.30	16.00	0.04	26.20	2.37			0.69	0.04

High Turbidity/TSS water
Acid mine drainage water

Table 4: Physico-chemical characteristics measured along the Corridor-H watershed, WV.

Watersheds running parallel to the alignment of Corridor-H, WV and many of its major tributaries, have low pH, little alkalinity, and high metals loading, due to the impacts of historic and current mining. And, microorganisms have the capacity to modify the physicochemical conditions of the water either by detoxification or by metabolic exploitation (Bruneel & Duran, 2006).

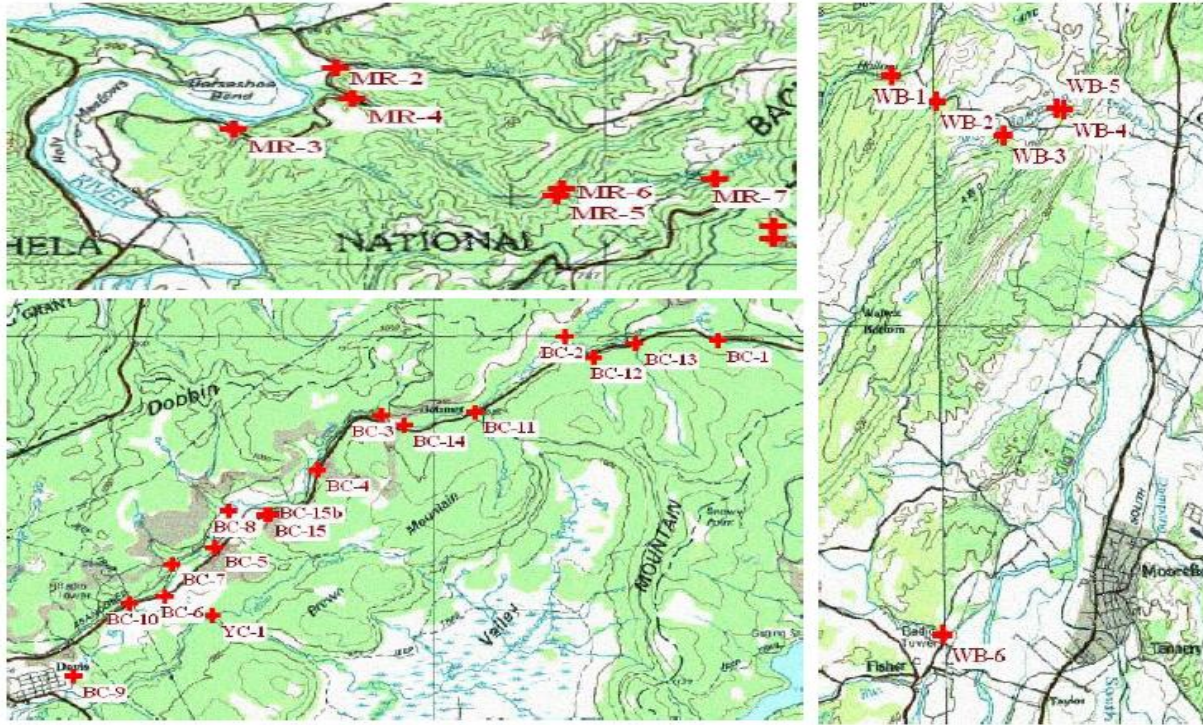


Figure 20: Map showing location of the different sites in Beaver Creek, Mill Run and Walnut Bottom Run watersheds located in Preston, Tucker and Hardy counties.

4.3.2 Microbial community fingerprinting of the Corridor-H watershed using T-RFLP

The process of terminal restriction fragment length polymorphism (T-RFLP) fingerprinting was conducted as described previously. Figure 21 shows an electropherogram of the 5' T-RFLP patterns for Beaver Creek watershed following digestion of PCR products with *RsaI*. Each T-RFLP pattern constitutes a distinct fingerprint of a community based on the size of T-RFs and their intensities. To determine the similarity of the values generated in the T-RFLP experiment, a cluster analysis was performed using SAS PROC and processed by using the Ward's minimum-variance clustering (Figure 22).

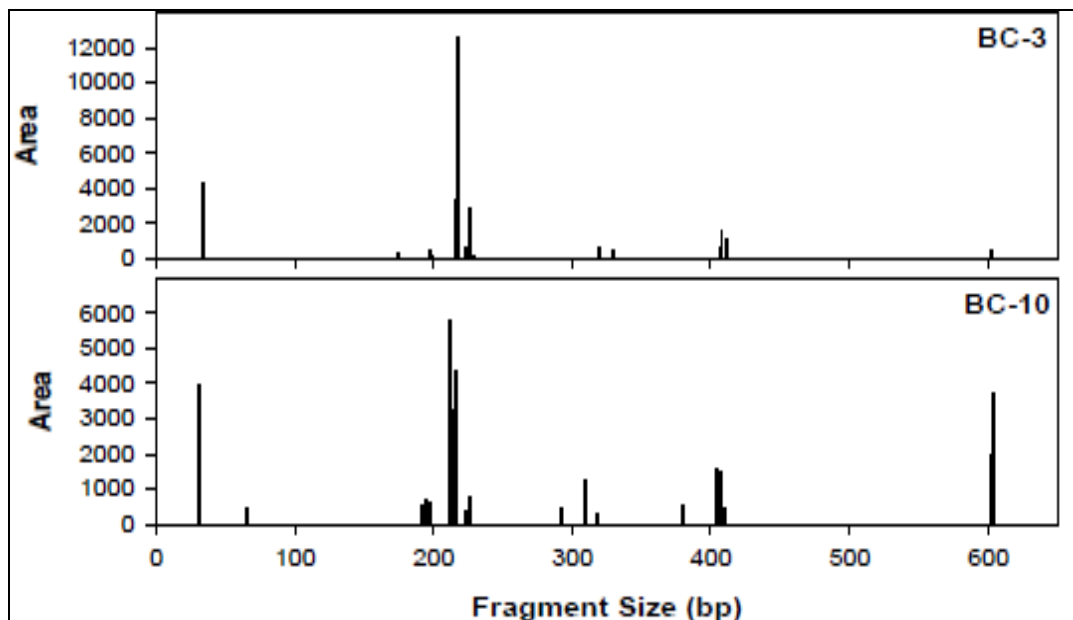


Figure 21: T-RFLP profiles of the bacterial community of Beaver Creek watershed, WV.

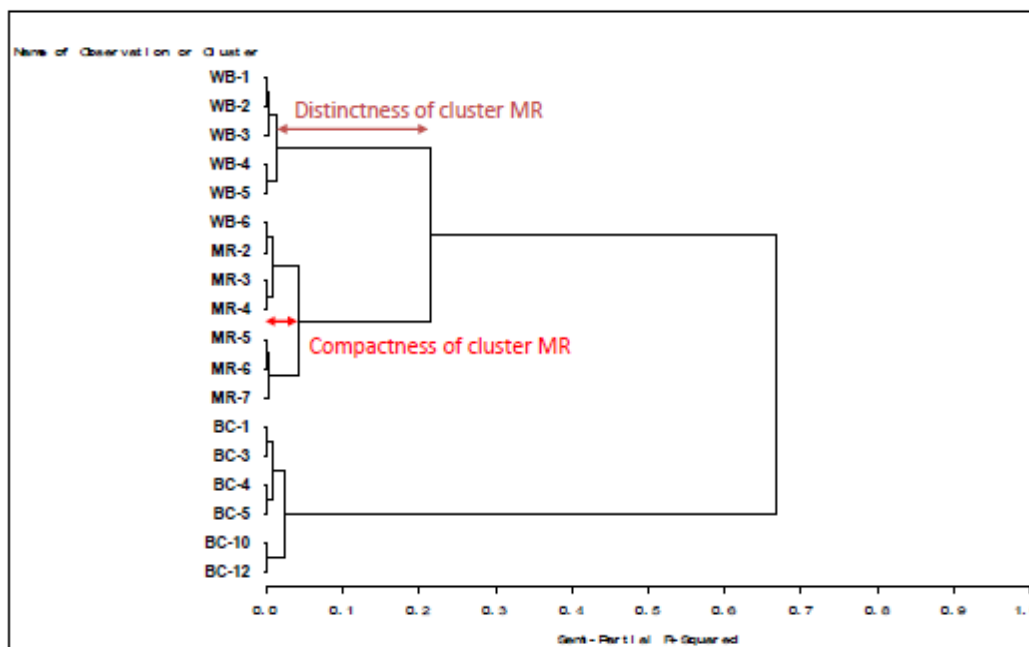


Figure 22: Dendrogram obtained from the clustering analysis of T-RFLP profiles of different streams from the Corridor-H watershed. SAS PROC processing using "Ward's minimum-variance clustering".

The dendrogram shows that single watersheds exhibit very specific T-RFLP profiling that can be used to characterize the source and nature of the microbial contamination. It also gives a clear relation between water samples and their environment, showing that the microbial diversity in Beaver Creek (BC) impacted by acid mine drainage varies significantly from Walnut Bottom Run (WB) and Mill Run (MR). Bacteria in acid mine waters may be useful in AMD bioremediation or that of some other industrial effluents, due to their tolerance of heavy metals (Bruneel & Duran, 2006). Thus, the microbial signature obtained can be used in the application of Microbial Source Tracking.

Chapter 5: Conclusion

The main focus in source tracking in recent years has been on the development of new methods, primarily because many MST methods are cumbersome, and require large sample volumes, unconventional equipment, and significant technical expertise (Santo Domingo, Bambic, Edge, & Wuertz, 2007). Molecular approaches based on polymerase chain reaction (PCR) techniques, have become popular and efficient methods for characterizing and tracking changes in the microbial community structures (Glockner, Fuchs, & Amann, 1999). But, however there is no PCR- based method that has been shown to be satisfactory for the reliable determination of pathogen source (Field, Bernhard, & Brodeur, 2003). Thus, the novelty of this research is that it focused on optimizing an innovative method that uses the entire microbial community profile of natural water for pathogens and groundwater source tracking.

The Objective of this research was to propose an innovative approach for tracking the origin of groundwater, pathogen source tracking and origin of surface water, based on the specific structure of the microbial community using T-RFLP analysis. The hypothesis underlying this research is that, the specific microbial terminal-restriction fragment (T-RF) profiles in spring waters and natural streams can be used to track the origin of groundwater and surface water, and determine the source of microbial contamination. The significance of this work is that, the results obtained are expected to pave way for more extensive use of complex microbiological profiles as indicators of surface water and groundwater origin, which would help to identify the most sustainable water sources and in the elimination of fecal pollution.

The experimental approaches used in this research were based on molecular biology techniques, which included DNA extraction, polymerase chain reaction (PCR), agarose gel electrophoresis (AGE), terminal-restriction fragment length polymorphism (T-RFLP) and screening a 16S rDNA gene library.

5.1 Specific Aim 1

Specific Aim 1 aimed to test the hypothesis that the specific microbial terminal restriction fragment (TRF) profiles present in deep waters and shallow waters, can be used as a marker of the groundwater source. Water samples were collected from Cold and Warm Springs in Huntingdon, PA, Kilmer Spring, Newman Spring, Jim Smith Farm Spring, Bella Vista Spring, Big Spring and Snodgrass Spring in Martinsburg, WV and Cacapon State Park, Ladies Spring in Berkley Springs, WV. The species diversity of the microbial community in these spring waters was determined using Terminal-restriction fragment length polymorphism (T-RFLP) analysis. The result shows that the average terminal restriction fragment (TRF) number was relatively small in almost all the springs, reflecting low bacterial diversity. Clustering analysis of the T-RFLP profiles showed that it was observed that Big Spring (BIG) and Snodgrass Spring (SND) seemed not connected to each other, although the two springs are adjacent to each other geographically. Except for Cold Spring (CHP), there was no clear relationship between physical/chemical characteristics of water and the microbial community structure. This is due to lower temperature and DO content in CHP, which causes a variation in flow path due to change of microbial population.

Furthermore, screening a 16S rDNA gene library was constructed to determine whether specific bacterial species can inform about the source of groundwater. After sequencing, it was seen that some identified bacterial species like *Thermomicrobia*, *Thermodesulfobacteria*, *Aquificae* and *Halobacteria* may be used as indicators of activities or chemistries in specific groundwater. In addition, clustering of the bacterial species showed that Big Spring (BIG) and Snodgrass Spring (SND) are suspected to be connected. However, there were not enough bacteria sequenced to give a statistical significance. Thus, there was no clear correlation between the two methods, T-RFLP and 16S rDNA gene library. This is due to the technical problems experienced during T-RFLP analysis which interprets the data.

5.2 Specific Aim 2

In the specific aim 2, we tried to develop a proof of the concept that the microbial community structure of stream water can be used as a signature to determine the origin of water contamination, using T-RFLP analysis. For this, water samples were collected from five streams in the Left Fork watershed of Mud River in Lincoln County, which has highest bacterial contamination. The results of the microbial community terminal-restriction fragment (TRF) profiles indicated a higher abundance of operational taxonomic units (OTUs) in all of the streams. Clustering of the microbial community structure obtained by T-RFLP analysis showed that, water samples collected from the five streams of the Mud River watershed carried a very specific signature that can be used for microbial source tracking analysis. Delaware White, Ken Bailey and Feerell Pauley, were closely connected having common bacterial communities, which can be used as specific indicators of surface water.

5.3 Specific Aim 3

In Specific Aim 3, we proved the concept that, natural water has a ‘microbial signature’ (T-RFLP) profile that can be used for microbial source tracking. Water samples were collected from different streams along the Corridor-H watershed in West Virginia. While it is necessary to deal with potential acidic drainage resulting from the construction of Corridor-H through mine spoils in the watershed drainage, it is also equally important to address sources of water impairment originating above the roadway alignment. Thus, T-RFLP analysis was conducted in different streams that have been impacted by the construction of Corridor-H. Clustering of these T-RFLP profiles indicated that single watersheds exhibit very specific T-RFLP profiling that can be used to characterize the source and nature of the microbial contamination. The microbial diversity in Beaver Creek (BC) impacted by acid mine drainage varies significantly from Walnut Bottom Run (WB) and Mill Run (MR), due to changes in microbial population. Thereby, indicating a clear relation between the microbial community profile and water parameters.

In summary, the results of this research indicated that microbial community profiles tested as markers of a particular flow or stream of water, and not of a particular type of water. Thus, terminal-restriction fragment length polymorphism (T-RFLP) is an effective method to compare complex communities in natural water, which constitutes a fingerprint of water. The method does produce distinct community signatures that can be used to assess the similarity of different communities.

5.4 Significance of the results obtained in this thesis

The important contributions of T-RFLP method in the field of water analysis is that, the genetic fingerprinting of the microbial community in natural water, provides information on the presence and relative abundance of the species, which can be used as markers for pathogen source tracking and in determining the origin of groundwater. And, the method is rapid, and highly reproducible than other commonly used PCR-fingerprinting methods.

5.5 Future work

However, more research is needed to establish the correlations between different environmental parameters and the microbial community profiles. Screening a 16S rDNA gene library can be constructed to correlate the microbial community profiles obtained by T-RFLP and clone library. Sample collection can be done seasonally and T-RFLP technique can be conducted to observe the variation in microbial diversity.

References

- Abdo, Z., & Williams, C. J. (2005). *Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes*. Society for Applied Microbiology and Blackwell Publishing Ltd.
- Auckenthaler, A., & Huggenberger, P. (2003). Pathogene microorganism in groundwater and drinking water. *Basel* , 184.
- Batin, T. J., Wille, A., Sattler, B., & Psenner, R. (2001). Phylogenetic and functional heterogeneity of sediment biofilms along environmental gradients in a glacial stream. *Applied Environmental Microbiology* , 67: 799-807.
- Black, J. G. (1999). *Microbiology: Principles and Explorations*. John Wiley & Sons, Inc. New York, NY.
- Brown, T. A. (2006). *Gene cloning and DNA analysis: An introduction*. Blackwell Publishing .
- Bruneel, O., & Duran, R. (2006). Diversity of microorganisms in Fe-As-Rich Acid Mine Drainage waters of Carnoules, France. *Applied and Environmental Microbiology* , 72 (1): 551-556.
- Chapelle, F. H. (2001). *Ground water microbiology and geochemistry*. John Wiley and Sons, Inc. New York, NY.
- Clement, B., Kehl, L., DeBord, K., & Kitts, C. (1998). Terminal restriction fragment patterns (TRFPs), a rapid, PCR- based method for the comparison of complex bacterial communities. *Journal of Microbiology* , 31:135-142.
- Doctor, D., Orndorff, W., & Plummer, L. (2008). A proposed hyogenic origin of karst in the Shenandoah Valley, Virginia and West Virginia . *Geological Society of America* .
- Field, K. G., Bernhard, A. E., & Brodeur, T. J. (2003). Molecular approaches to microbiological monitoring: Fecal source detection. *Environmental Monitoring and Assessment* , 81:313-326.
- Glockner, F. O., Fuchs, B. M., & Amann, R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence in situ hybridization. *Applied Environmental Microbiology* , 65: 3721-3726.
- Handelsman, J. (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* , 68(4), 17.

- Hiraishi, A., Iwasaki, M., & Shinjo, H. (2000). Terminal restriction pattern analysis of 16S rRNA genes for the characterization of bacterial communities of activated sludge. *Journal of Bioscience and Bioengineering* , 90 (2): 148-156.
- Lasalde, C., Guez, R. R., & Toranzos, G. A. (2005). Statistical analyses: Possible reasons for unreliability of source tracking efforts. *Applied and Environmental Microbiology* , 71: 4690-4695.
- Liu, W., Marsh, T. L., Cheng, H., & Forney, L. J. (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied Environmental Microbiology* , 63: 4516-4522.
- MacDowell, R., & Basden, T. (2007). Local communities: Equal partners or ignored bystanders? A case study from Rural West Virginia. *USDA-CSREES National Water Conference*.
- Marsh, T. L. (2005). Culture-independent microbial community analysis with terminal restriction fragment length polymorphism. *Methods Enzymology* , 397: 308-329.
- Marsh, T. (1990). Terminal restriction fragment length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. *Current Opinion in Microbiology* , 2:323-327.
- McMahon, P., & Chapelle, F. (1991). Geochemistry of dissolved inorganic carbon in a coastal plain aquifer. 2. Modeling carbon sources, sinks, and C evolution. *Journal of Hydrology* , 127: 109-135.
- Nativ, R. (1996). The brine underlying the Oak Ridge Reservation, Tennessee, USA: Characterization, genesis and environmental implications. *Geochim. Cosmochim. Acta* , 60(5): 787-801.
- Nico Goldscheider, D. H. (2006). Review: Microbial biocenoses in pristine aquifers and an assessment of investigate methods. *Hydrogeology Journal* , 14: 926-941.
- Osborn, A., Moore, E., & Timmis, K. (2000). An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environmental Microbiology* , 2:39-50.
- Rogers, J., & Bennett, P. (2004). Mineral stimulation of subsurface micro-organisms: release of limiting nutrients from silicates. *Chem Geol* , 203(1/2): 91-108.
- Santegoeds, C., Damgaard, L., Zopfi, J., & Muyzer, G. (1999). Distribution of sulfate-reducing and methanogenic bacteris in anaerobic aggregates determined by microsensor and molecular analyses. *Applied Environmental Microbiology* , 65(10): 4618-4629.
- Santo Domingo, J. W., Bambic, D. G., Edge, T. A., & Wuertz, S. (2007). Quo vadis source tracking? Towards a strategic framework for environmental monitoring of fecal pollution. *Water Research* , 41:3539-3552.

- Snyder, L., & Champness, W. (1997). Molecular genetics of bacteria. *American Society of Microbiology*.
- Streit, W., & Schmitz, R. A. (2004). Metagenomics-the key to the uncultured microbes. *Curr Opin Microbiol.*, 7(5), 492-498.
- Thies, J. E. (2007). Soil microbial community analysis using terminal restriction fragment length polymorphisms. *Soil Sci. Soc. Am. J.*, 71: 579-591.
- USEPA, (. S. (2005). Microbial source tracking guide document, EPA-600/R-05-064.
- Vandivort, T., & Fulton, J. (2008). *Watershed bacterial study and decentralized wastewater demonstration, Mud River Watershed, Lincoln County, WV.*
- von, W. F., Gobel, U., & Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: Pitfalls of PCR- based rRNA analysis. *FEMS Microbiology Review*, 21:213-229.
- Wang, R., Cao, W., & Cerniglia, C. (1996). PCR detection and quantification of predominant anaerobic bacteria in human and animal fecal samples. *Applied and Environmental Microbiology*, 62:1242-1247.
- Wu, M., Song, L., Ren, J., Kan, J., & Qian, P. Y. (2004). Assessment of microbial dynamics in the Pearl River Estuary by 16S rRNA terminal restriction fragment analysis. *Continental Shelf Research*, 24: 1925-1934.