2012

A Baseline Study of Chemical Parameters and Microbial Diversity of Two Streams in the Ten Mile Creek Watershed in Southwestern Pennsylvania

Jennifer Rutter

A BASELINE STUDY OF CHEMICAL PARAMETERS AND MICROBIAL DIVERSITY OF TWO STREAMS IN THE TEN MILE CREEK WATERSHED IN SOUTHWESTERN PENNSYLVANIA

A Thesis

Submitted to the Bayer School of Natural and Environmental Sciences

Duquesne University

In partial fulfillment of the requirements for

the degree of Master of Science

By

Jennifer Kathleen Rutter

August 2012

A BASELINE STUDY OF CHEMICAL PARAMETERS AND MICROBIAL DIVERSITY OF TWO STREAMS IN THE TEN MILE CREEK WATERSHED IN SOUTHWESTERN PENNSYLVANIA

By

Jennifer Kathleen Rutter

Approved June 21, 2012

_____	_____
Dr. John F. Stolz Professor of Biology (Committee Chair)	Dr. Brady Porter Associate Professor of Biology (Committee Member)
_____	_____
Mr. Edward Schroth CERE (Committee Member)	Dr. John F. Stolz Director, Center for Environmental Research and Education (CERE)
_____	
Dr. David Seybert Dean, Bayer School of Natural and Environmental Science	

ABSTRACT

A BASELINE STUDY OF CHEMICAL PARAMETERS AND MICROBIAL DIVERSITY OF TWO STREAMS IN THE TEN MILE CREEK WATERSHED IN SOUTHWESTERN PENNSYLVANIA

By

Jennifer Kathleen Rutter

August 2012

Dissertation supervised by Dr. John F. Stolz

As drilling for natural gas in the Marcellus Shale becomes increasingly prevalent, both human safety and environmental concerns have arisen. The aim of this study was to assemble an environmental baseline to make an accurate assessment of its possible impacts. Water samples and chemical parameters were collected from Bates Fork, a stream with Marcellus Shale drilling activity and Fonner Run, a sister stream with no drilling activity, on a monthly basis beginning in the summer of 2010. Bacterial DNA was also extracted from water collected at each site and then amplified using primers for the variable ribosomal internal transcribed spacer (ITS) region of 16S and 23S rRNA. The bacterial ribosomal ITS has both sequence and length variability, which can be used

to approximate species abundance and diversity along both streams. The conserved 16S rRNA gene was also amplified, and sequenced to taxonomically categorize organisms. To date, it has been observed that Bates Fork has consistently higher conductivity and chloride levels than Fonner Run. The bacterial diversity was found to be similar between the two streams. As a baseline study, these results show that there have not been any detectable impacts to Bates Fork thus far, but the stream should continue to be monitored.

ACKNOWLEDGEMENTS

This work was supported by the Heinz Foundation. I am grateful to Robert Reiter for the use of his RISA diagram, Dr. John Stolz, and J Lucas Eastham. Helpful comments were provided by Mark Evans during editing.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Background

1.1 Unconventional Shale Gas: Horizontal Drilling & Hydraulic Fracturing

As the United States tries to wean itself off foreign oil, many believe that natural gas is the next logical fuel to turn to because it is cleaner burning than other fossil fuels (Kargbo et al. 2010). The combustion of natural gas emits considerably lower levels of carbon dioxide and sulfur dioxide than does the combustion of coal or oil (US EIA 2012). It is because of this that natural gas has been called the "bridge to renewable energy" (GE 2012). Currently, ninety percent of the natural gas used in the United States is produced domestically (US EIA 2012). There are multiple, sometimes overlapping, shale formations in the United States that contain significant accumulations of natural gas (Figure 1). Several shale plays, such as the Marcellus, are so large that estimation of technically recoverable resources is highly uncertain, with updates being made as drilling and production continue (US EIA 2012).

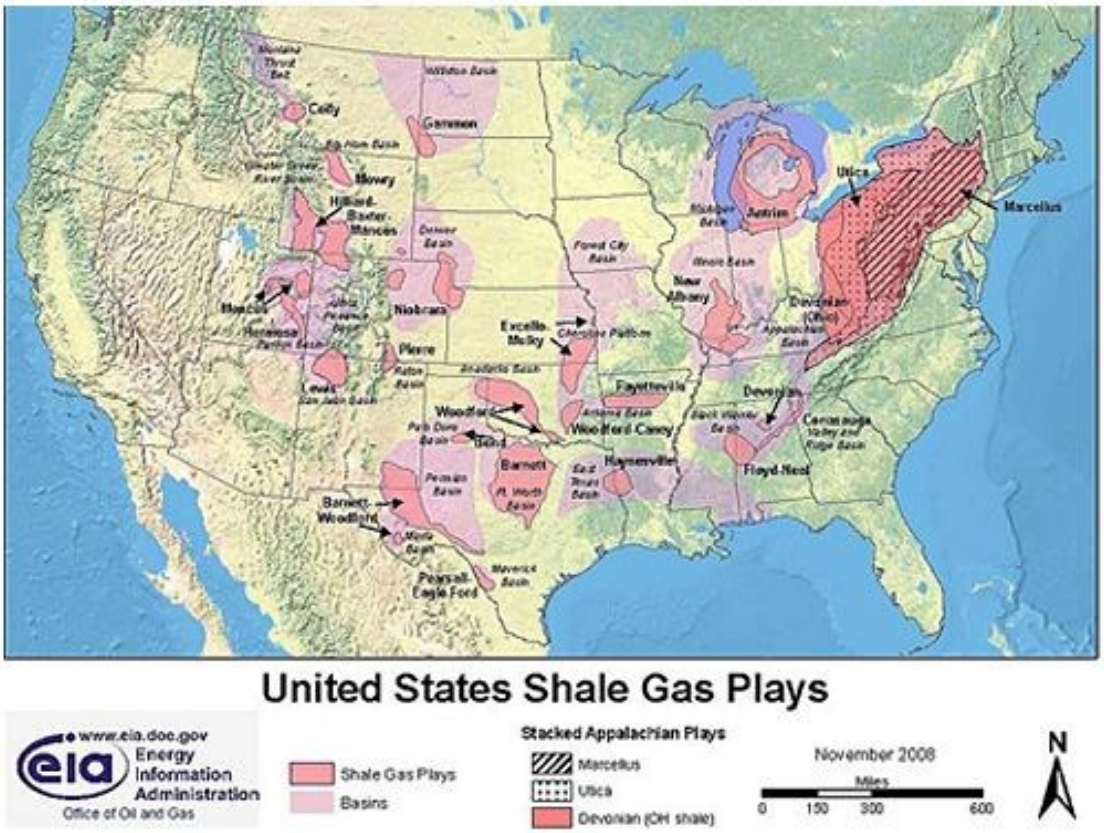


Figure 1. Shale gas plays in the lower 48 states (US EIA 2012).

Pennsylvania has been called "the Saudi Arabia of natural gas" due to its vast shale gas reservoirs with its Devonian Age Marcellus Shale, one of the largest sources of domestic natural gas discovered in the United States thus far (Kargbo et al. 2010). The Marcellus Shale is a highly organic black shale that was deposited when a shallow continental seaway existed in the area that now currently extends from west central New York into northern and western Pennsylvania, western West Virginia, and eastern Ohio (Arthur et al. 2008, Kargbo et al. 2010). Drilling for this shale gas involves horizontal drilling and hydraulic fracturing, which is different from conventional natural gas drilling (Arthur et al, 2008, Kargbo et al. 2010). Conventional gas drilling uses vertical wells, which is, overall, less economical and would require up to 16 separate well pads,

compared to horizontal drilling which would use only one multi-well pad to penetrate the ground several times from one surface location (Arthur et al. 2008).

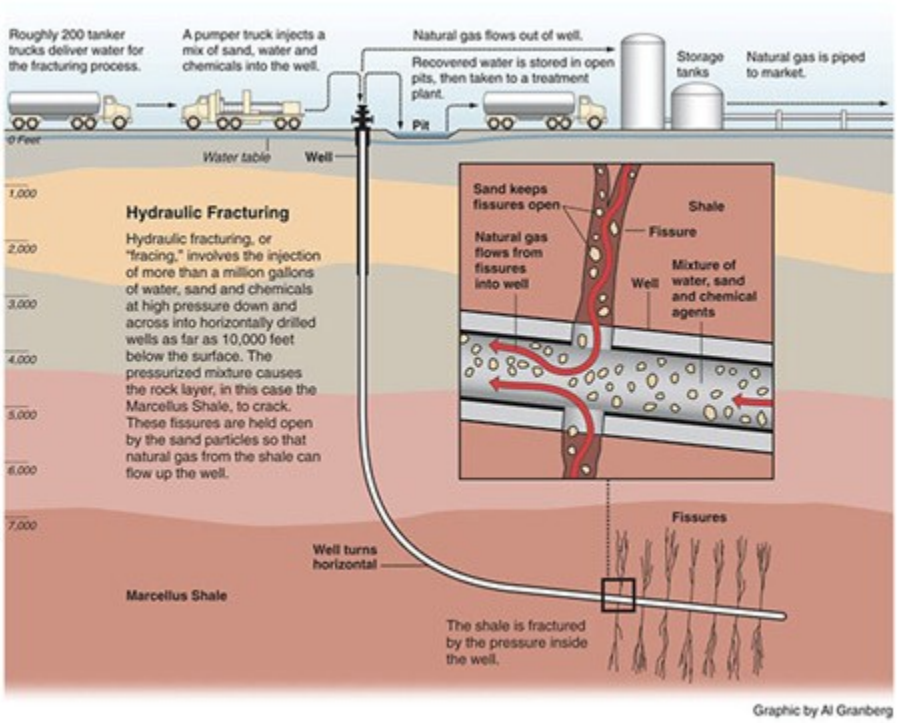


Figure 2. A diagram of how horizontal drilling and hydraulic fracturing occurs in a shale formation (US EIA 2012).

In addition, as indicated by the name of this technique, hydraulic fracturing uses copious amounts of water, about 3-5 million gallons per well, usually taken from local waterways (Kargbo et al. 2010, US EIA 2012). The water is mixed with biocides, lubricants, corrosion inhibitors, and sand, and then pumped at a very high pressure into the well to fracture the shale and release the natural gas (Figures 2 and 3) (Kargbo et al. 2010). This water-chemical mixture is called 'fracking fluid', and although it can be reused several times, it quickly becomes very contaminated and incorporates saline brines from the Marcellus (Kargbo et al. 2010). Up to 70% of the injected water flows back to the surface and contains high levels of total dissolved solids, radionucleotides, and other constituents (NETL 2012). The flow-back and produced water from hydraulic fracturing

activities is contained in fluid capture systems to minimize their exposure to the environment and reduce the possibility for spills to occur (Arthur et al. 2008).

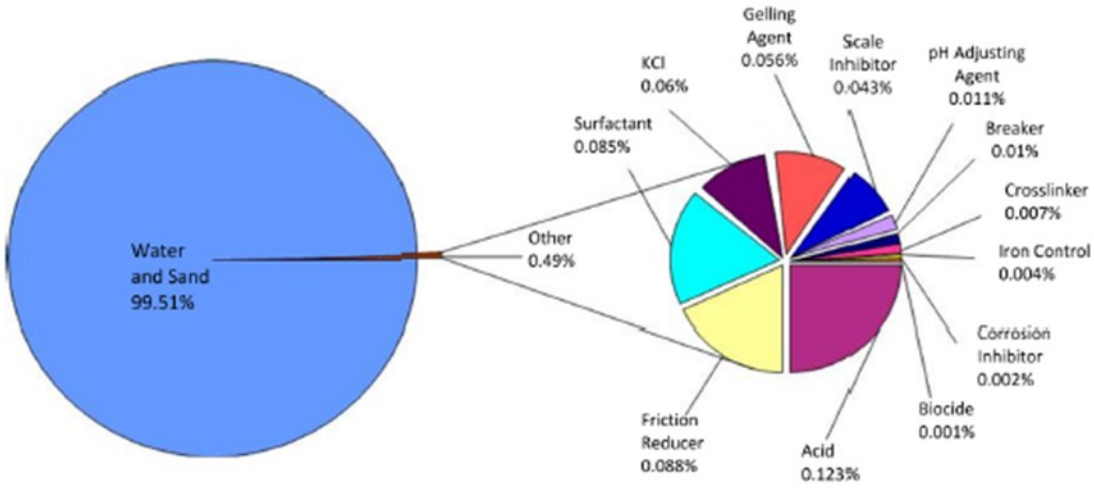


Figure 3. Volumetric composition of a hydraulic fracturing fluid (Ground Water Protection Council and ALL Consulting 2009)

1.2 Potential Environmental Impacts

Through anecdotal evidence, it is thought that hydraulic fracturing can have adverse environmental impacts, specifically ground and surface water contamination, increased volatile organic carbon in the air, and habitat destruction (Kargbo et al. 2010). An increase in drilling activity has resulted in limited, but increasing groundwater and surface water contamination caused by releases of the 'frack fluid' via spills, leaks, faulty well construction, or other exposure pathways (Harrison 1983, US EIA 2012). Any such releases can contaminate surrounding areas. In addition, high-salinity groundwater is often co-produced with the extraction of subsurface natural gas (Goodfellow et al. 2000). Freshwater aquifers can become contaminated by brines or oil from underlying formations, various chemicals used in the drilling process, and the natural gas itself (Harrison 1983). Surface activities on natural gas well sites can lead to water

contamination through storage slush pits, aerosols that are blown off during fracturing or servicing operations, or brines spread on the road for dust control or deicing (Harrison 1983).

1.3 Chemical and Environmental Sampling

Since horizontal drilling and hydraulic fracturing is a relatively new technique in this region, there is currently no environmental baseline to make an accurate assessment of the possible impacts of fracking fluid contamination (Kargbo et al. 2010). There are several ways to assess potential impacts in aquatic systems, including both chemical and biological measurements. Chemical data is important to evaluate water quality because it can give a snapshot of the current conditions and provide numerical information. Specifically, total dissolved solids (TDS), conductivity, and salinity are often measured to determine the ionic load in freshwater (Goodfellow et al. 2000). Goodfellow et al. (2000) cites the extraction of subsurface natural gas as a source of high salinity groundwater and elevated TDS levels as high as 1.7 to 200‰.

One important water quality indicator is the concentration of salts, or salinity (Eldridge et al. 2010). Salts are inorganic and comprised of both cations and anions (Eldridge et al. 2010). Although salts are essential to all organisms, excess amounts can be very damaging. Aquatic organisms have a wide range of salt tolerances, and can vary depending on which specific cation is involved (Eldridge et al. 2010). Salt can enter the environment from natural sources, such as the natural weathering of bedrock or in the oceans, as well as wet and dry deposition from the atmosphere (Eldridge et al. 2010). Salt is also widely used by humans, with NaCl being the most common man-made and mined that is often used as a de-icing agent or water softener (Eldridge et al. 2010).

Increases in salt concentrations in freshwater ecosystems can have a devastating effect on the biota by resulting in shifts in community structures, limiting biodiversity, and causing acute and chronic effects at different life stages (Eldridge et al. 2010; Weber-Scannell and Duffy 2007). A decline in salt-sensitive species can ultimately result in a change in community structure and function of the aquatic ecosystem (Eldridge et al. 2010). It has been reported that salinity and aquatic biodiversity are inversely related in lentic systems (Weber-Scannell & Duffy 2007). Others have suggested that if the conductivity of water flowing into a freshwater system is higher than 2,000 µS/cm, the concentration of TDS can be high enough to negatively affect the flora and fauna (Goodfellow et al. 2000). The conductivity, or specific conductance, of water is the measure of how well it can conduct an electric current, increasing with increasing levels of TDS (Kimmel and Argent 2009). The chemical measurement of TDS does not differentiate among ions and includes the concentration of inorganic salts, organic matter, and other dissolved materials in water (Weber-Scannell and Duffy 2007). Algae productivity and nitrogen fixation by cyanobacteria decline at TDS concentrations greater than 1,400 mg/L (Weber-Scannell and Duffy 2007), with levels of 400 mg/L adversely affecting diverse fish populations (Kimmel and Argent 2009). However, the overall toxicity of TDS to aquatic life depends on the both the concentration and combination of the ions present in the water (Kimmel and Argent 2009).

Water quality can also be assessed through biomonitioring, a technique that is the integrated result of many chemical, biological, and physical factors over time and provides a top-down evaluation of the biological community performance as a sum of the various system processes (OHEPA 2010). Biological indicator organisms are valued for

their response to changes in a system and are often more cost and time effective than complex chemical and physical analyses (Lemke et al. 1997). All organisms respond differently to human-induced environmental stress, with the disturbance characteristics affecting each in a different way (Merkley et al. 2004). Inferences can be made from the presence/absence, condition, relative abundance, and community structure of biological indicators, providing an index of overall ecosystem health (Lear et al. 2009).

One biological indicator that can be used is bacterial community structure, a measure that theoretically encompasses all taxa and can be used to assess broad-scale watershed influences on aquatic ecosystems (Lear et al. 2009). Bacteria may be one of the most sensitive indicators of ecosystem impacts because they have the highest surface area to volume ratio of all organisms, having only a thin membrane and cell wall as a boundary between them and the environment (Merkley et al 2004). The role that bacteria play in aquatic ecosystems is a very important one because they perform essential chemical reactions, including the transformation and demineralization of nutrients and the biodegradation of allochthonous detritus and other substances (Araya et al. 2003). The community structure and relative proportion of bacteria is influenced by physiochemical changes in the environment, as well as changes caused by the metabolism of the organisms themselves (Wintzingerode et al. 1997). Given that 'fracking fluid' can introduce a variety of organics, metals, and increase the TDS, a drilling-impacted stream would likely experience a shift in microbial community composition and activity (NETL 2012). According to Lear et al. (2009), the analysis of bacterial community structure can provide a sensitive measure of the extent of ecosystem

degradation or impact, and may be a valuable tool for monitoring the success of ecological restoration of aquatic ecosystems.

Bacteria are very small, difficult to identify under a microscope, and often difficult to culture, so other tools are needed to aid in the determination of microbial species richness and species identification. Polymerase chain reaction (PCR) is an essential tool in microbiology, amplifying selected DNA sequences and thereby allowing for the discovery of diversity and the establishment of phylogeny in studied communities (Brown et al. 2005, Wintzingerode et al. 1997). PCR techniques also facilitate the detection of microorganisms in a variety of environmental sources and can provide insight into the overall microbial diversity of a sample (Baker et al. 2003, Blackwood et al. 2005). This culture-independent tool has allowed for the amplification of segments of rRNA from bacteria, providing important insight into their community structure (Baker et al. 2003). When using PCR, however, it is assumed that the gene sequences present in the environment are complimentary to the "universal" primers used for amplification (Baker et al. 2003). Regions of the bacterial 16S rRNA range in their variability, from totally conserved to only present in some organisms; with only 10% of the bases being totally conserved and most of them not adjacent to each other, there are no continuous conserved regions for highly effective universal priming (Baker et al. 2003, Garcia-Martinez et al. 1999). Primers designed for any 16S rRNA region cannot fully match all bacterial gene sequences; those designed to be less specific are often also complimentary to eukaryote rRNA sequences (Baker et al. 2003, Wanatabe et al. 2001).

Lemke et al. (1997) used bacteria as biomonitors; they studied two watersheds, one relatively undisturbed and the other with significant point and non-point sources.

The authors measured stream chemical characteristics over time, including water temperature, flow, and pH. Abundances of bacteria in each of the watersheds were determined using colony hybridization with species-specific rRNA probes. Lemke et al. (1997) looked at three different bacterial species and found that *Acinetobacter calcoaceticus* was the only species that was found in significantly higher concentrations in the disturbed watershed, indicating that it has a high tolerance for polluted conditions.

A relatively recent addition to the PCR-based fingerprinting approaches is that of ribosomal intergenic spacer analysis (RISA), which involves the use of universal primers to amplify the internal transcribed spacer (ITS) located between the 16S and 23S ribosomal genes in addition to the 3' region of the 16S ribosomal RNA operon (Figure 4) (Benlloch 2000, Brown et al. 2005, Fisher and Triplett 1999, Leuko et al. 2006). The heterogeneity of the length and composition of tRNAs in the ITS region of bacterial species is important (Leuko et al. 2006). RISA provides a rapid assessment of diversity compared to the creation of a 16S r RNA gene clone library, as well as greater taxonomic resolution which can lead to very complex community profiles (Leuko et al. 2006, Malik et al. 2007). The intergenic spacer region has both sequence and length variability (140-1530 bp) depending on the species, and thus can reflect the composition of the community present in analyzed samples, as well as changes in community structure over time (Benlloch 2000, Garcia-Martinez et al. 1999, Malik et al. 2007).

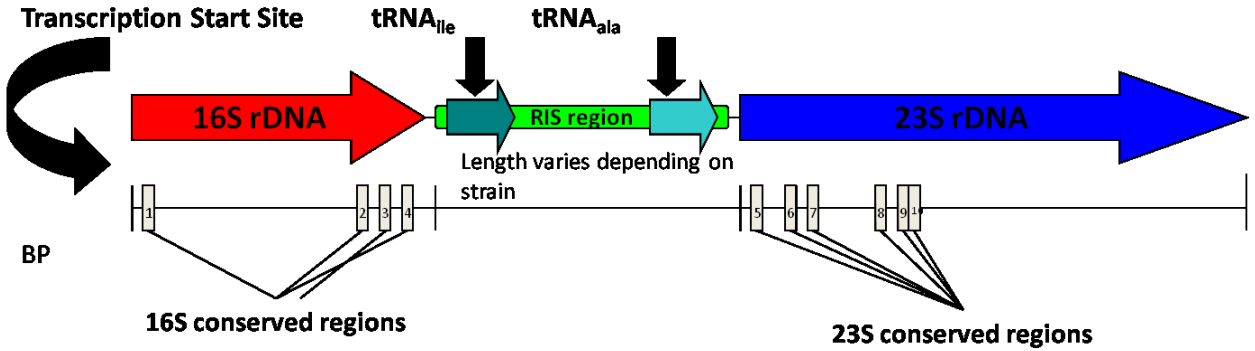


Figure 4. The intergenic spacer region with both the 16S and 23S regions shown (Reiter 2012).

Another molecular technique that has recently been developed is that of the detection of dye-labeled nucleic acids via fluorescence reporting; this method is a valuable tool that is used in automated ribosomal intergenic spacer analysis (ARISA) (Behlke et al. 2005). Similar to RISA, ARISA amplifies the ITS region of the 16S-23S rRNA operon, but instead uses a fluorescent-tagged forward primer and an automated electrophoresis system (Fisher and Triplett 1999). Brown et al. (2005) notes that the ITS in free-living, uncultured bacterioplankton was observed to be highly variable in length between different clades and to be highly conserved within clades. This suggests that ARISA can provide a community-specific profile, with each band length corresponding to at least one organism in the original assemblage (Fisher and Triplett 1999). ARISA is also a rapid and effective method for assessing microbial community diversity and richness, but it has some limitations (Kovacs et al. 2010).

Fisher and Triplett (1999) were the first to publish the ARISA method, and in doing so, were successful in expressing microbial diversity though 16S-23S intergenic spacer lengths. Their study site included several lakes in the Northern Highland Lake District in Northern Wisconsin. The microbial intergenic spacer region was cloned and sequenced, and was found to range in length from 143-1529 bp, most of which were within 150-600 bp; gram-positive bacteria may tend to have shorter spacers because of their lack of tRNAs in that region. The authors ran their samples in triplicate and looked at variability in relative abundance of fragments for the same series of peaks by analyzing the percentage of the total fluorescence. The study found that level of similarity between replicates increased when looking at those peaks greater than 50 fluorescence units, with it getting even better when looking at peaks greater than 100 fluorescence units.

Brown et al. (2005) used both 16S-ITS rRNA clone libraries and ARISA to show marine microbial diversity in the water off the coast of San Pedro, California. The authors obtained a total of 512 clones, but used only 248 due to the high similarity of many of the lengths. Ninety-two percent of the ARISA peaks were matched with corresponding lengths from the clone library, suggesting that ARISA resolution is near the species level. This study was performed over 42 months, revealing shifts in community structure and ecological niches for specific organisms over time. The authors used two relatively sensitive microbial detection techniques to successfully observe temporal patterns of presence and abundance.

Lear et al. (2009) used ARISA as a high-throughput, low-cost molecular analysis method to assess the difference in predominant bacterial community structures in streams affected by acid mine drainage. By using multivariate analyses, it was shown that there was strong clustering of the five samples collected from each stream, with significant differences in ARISA data among streams, but not within streams; the ARISA data strongly separated the data according to pH (Lear et al. 2009). It was also observed that the bacterial communities in the acidic streams were less diverse than the control streams; pH was determined to be the most important determinant of the observed differences in bacterial community variability.

Chapter 2 Specific Aims and Hypotheses

By using two representative streams, this study aims to provide an environmental baseline of chemical and bacterial community composition to assess possible impacts of Marcellus Shale activity on a lotic system.

Hypotheses:

1. Chemical data will show that the stream with drilling in its watershed has higher salinity and bromide levels than the control stream that has no record of drilling in its watershed because of the potential for leaks and spills of fracking and flowback fluid.
2. The stream with drilling activity in its watershed has a less diverse bacterial community than the adjacent control stream.

The specific aims of this study were to 1) determine a chemical baseline for each stream and identify any apparent trends; 2) establish the microbial richness for each stream using the ribosomal intergenic spacer (RISA); and 3) ascertain which species of bacteria are living in each stream by sequencing the 16S rRNA gene.

Chapter 3 Materials / Methods

These specific aims required an array of methods, including DNA extraction, molecular work, and bioinformatics. The experimental methods for each specific aim fell into the two following categories:

1. Sample Acquisition and Chemical Analysis
2. Molecular Profiling

3.1 Sample Acquisition and Chemical Analysis

3.1.1 Sample Acquisition

The representative stream system chosen for the study was the Ten Mile Creek watershed, located in Greene County in southwestern Pennsylvania, and includes Fonner Run, the control stream, and Bates Fork, a stream that has new drilling operations in its watershed. Both streams and their watersheds are illustrated in Figures 5 and 6, as well as their respective sampling sites. Fonner Run is a small third order stream with varied habitats located near SR 4029. Bates Fork is a larger third order stream with highly diverse habitats and about three times the discharge of Fonner Run, most likely due to its larger drainage area and input of several first order streams. Marcellus shale drilling operations was observed to be very active in the area during the scouting trip in July 2010, with multiple functioning and soon-to-be well pads upstream from the sampling site (Figures 5 and 6). Beginning in August 2010, the stream system was visited on a monthly basis, except in the winter months of November 2010 through March 2011, when the streams were frozen over. The winter of 2011-2012 was much milder and therefore water samples were taken throughout those months. Water samples were

collected in one liter French square glass bottles to be taken back to the lab; one liter of water was collected at each location for the first year, then four liters being collected for each location thereafter due to the realization that more DNA would be needed for optimal amplification. The one liter jars were transported on ice and stored at 4°C until they were processed. Several other samples were taken for analyses and comparison to the study streams and are listed in Table 1, including three samples near the Ten Mile Run / Monongahela confluence, two samples from coal mine effluent, one from Dunkard Creek, which experienced a fish-kill in the past year resulting from an algal bloom of Golden Algae, and Blacklick Creek, the site of the outflow of a POTW brine treatment facility.

Table 1. Sample List

Sample Name	Location and Date Sampled
FR_7/10	*Fonner Run (July 2010)*
FR_9/10	*Fonner Run (September 2010)*
FR_5/11	*Fonner Run (May 2011)*
FR_6/11	*Fonner Run (June 2011)*
FR_7/11	*Fonner Run (July 2011)*
FR_8/11	*Fonner Run (August 2011)*
BF_9/10	*Bates Fork (September 2010)*
BF_5/11	*Bates Fork (May 2011)*
BF_6/11	*Bates Fork (June 2011)*
BF_7/11	*Bates Fork (July 2011)*
BF_8/11	*Bates Fork (August 2011)*
TM1_10/10	*Ten Mile Run (October 2010)*
TM2_10/10	*Ten Mile/Monongahela Confluence (October 2010)*
TM3_10/10	*Monongahela River (October 2010)*
WC_6/11	*Whiteley Creek – Receives Contaminated Coal Mine Effluent (June 2011)*
BS_9/11	*Bleeder Shaft #5 - Coal Mine Effluent (September 2011)*
DC_5/11	*Dunkard Creek (May 2011)*
BL_8/11	*Blacklick Creek – Site of Josephine POTW Brine Treatment Plant (August 2011)*

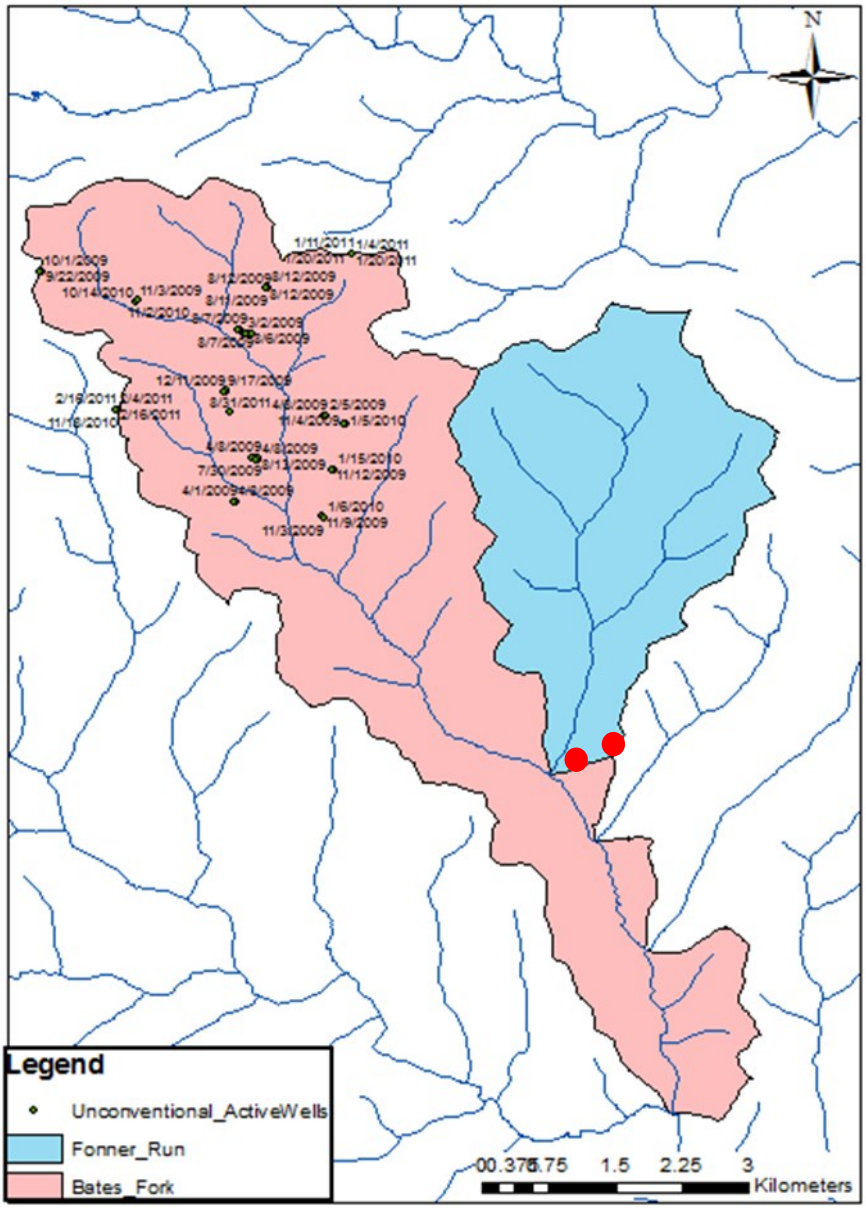


Figure 5. A map of Bates Fork (watershed in pink) and Fonner Run (watershed in blue), indicating both sampling locations (red dots) and Marcellus activity (green dots), with permit dates displayed for each well; created July 4, 2012.

Figure 6. A map of the sampling sites on Bates Fork (red star) and Fonner Run (blue star); created using Google Maps on July 3, 2012.

3.1.2 YSI Multimeter

A YSI-Pro Plus multimeter (YSI Incorporated, Yellow Springs, OH) was taken to each stream sampling event and used to obtain in situ measurements of water temperature, pH, dissolved oxygen, specific conductivity, and chloride. The YSI multi-meter provides snapshot of the stream water conditions, allowing for comparisons temporally and seasonally.

3.1.3 Ion Chromatography: Anion Analysis

In the lab, the water was further analyzed for nitrates, sulfates, and chlorides using ion chromatography, a method approved by the EPA to perform compliance monitoring of common inorganic ions in drinking water (Jackson 2006). To perform the analysis, a Dionex ICS-1100 equipped with an AS-22 anion exchange column, conductivity cell, and UV/VIS detector was used. Separation through the column is accomplished using a 4.5 mM Sodium Carbonate, 1.4 mM Sodium Bicarbonate eluent. Conductivity from the eluent is suppressed before conductivity cell measurements with an ASRS-300 suppressor (Dionex). In order to quantitate anion concentrations, 5-point standards used. Stock 1000 ppm standard solutions (Sigma-Aldrich) were diluted to various concentrations by mass and volumetrically. Standards included fluoride, arsenite, chloride, nitrite, bromide, nitrate, phosphate, sulfate, and arsenate. Any unidentified chromatogram peaks have to be identified through literature searches or further standards. Samples were run for 20 minutes at 0.25 ml/min to allow proper separation and elution of all anions. Concentrations from peak height were determined from the five-point calibration curves. The retention times were: chloride, 4.4 minutes; bromide, 6.1 minutes;

nitrate, 7.01 minutes; and sulfate, 11.23 minutes. The minimum detection level of bromide was 0.05 ppm (500ppb).

3.2 Molecular Profiling

3.2.1 DNA Extraction

To determine the molecular profile of the samples, the DNA was extracted from the stream water using a modified Phenol-Chloroform-IAA extraction protocol. To pellet the cells, the one-liter water samples were centrifuged at 7,000 rpm for 10 minutes using the Coulter Ultra High Speed Centrifuge with the JLA 8.1 rotor. The water was discarded while the pellet was transferred to a 50 mL tube. The 50 mL tubes were centrifuged at 7,000 rpm for 10 minutes. The supernatant was discarded and the pellet was resuspended by vortexing in 10 mL of Lysis Buffer (50 mM EDTA, 100 mM NaCl). The tubes were spun again at 7,000 rpm for 10 minutes, then the supernatant was discarded and the pellet resuspended by vortexing in 5 mL Lysis Buffer. An aliquot of 600 µL of 100 mg/mL lysozyme in Lysis Buffer, 400 µL SDS, and 30 µL of 20 mg/mL proteinase K were added to the tubes and mixed thoroughly. The tubes were incubated for 1 hour at 37°C. In a fume hood, an equal volume of tris-buffered phenol/chloroform/IAA (25:24:1) was added to the tubes and they were mixed thoroughly then spun at 10,000 rpm for 10 minutes. The aqueous supernatant was removed and transferred to a fresh 50 mL tube, being careful not to disturb the interphase layer. The phenol/chloroform/IAA was added again and the step was repeated. The step was again repeated using only chloroform/IAA. The aqueous supernatant was then transferred in 1 mL increments to fresh 1.5 mL tubes (as many as necessary), then 20 µL of 5M NaCl and 600 µL isopropanol was added to each tube to precipitate the nucleic

acids. The tubes were then inverted several times to mix until a stringy, white DNA precipitate was seen. The DNA was pelleted by centrifuging the tubes at 12,000 rpm at room temperature for 5 minutes. The supernatant was carefully discarded then the pellet was washed with 1 mL ice cold 70% ethanol. The sample was spun at 12,000 rpm for 10 minutes. The supernatant was carefully removed and then the pellet was dried in a speed vacuum on medium heat until no liquid remained, or about 20 minutes. The dried pellet was redissolved in 20 µL sterile DI water, and then recombined with like samples, ending with about 100 µL each. An Invitrogen Qubit® fluorometer (Grand Island, NY, USA) was used to quantify nucleic acid concentration in the samples.

3.2.2 Ribosomal Intergenic Spacer Analysis (RISA) and Automated RISA (ARISA)

The entire 16S-23S spacer was amplified using primers designed by Cardinale et al. (2004) (Table 2). PCR reactions were performed in triplicate with 50-100 ng of the extracted DNA into with 12.5 µL of 2x Phoenix Master Mix (Phenix Research Products, Candler, NC), 25 pmol of each forward and reverse primer (6-FAM florescent tag on the forward primer), and nuclease-free water to a total reaction volume of 25 µL. The reactions were run on a Techgene thermocycler (Techne Incorporated, Princeton, NJ, USA) under the following conditions: initial denaturation at 95°C for 5 mins, followed by 45 cycles of denaturation at 95°C for 15 seconds, annealing at 55°C for 15 seconds, extension at 72°C for 45 seconds, and final extension at 72°C for 10 minutes.

Internal spacer polymorphism was investigated by loading 20 µL of the PCR product into 10% TBE polyacrylamide gels and then run in a Mini-PROTEAN gel apparatus (Biorad, Hercules, CA, USA) for 6 hours at 75 V. The gels were then stained with ethidium bromide (0.5 ug/mL) for 45 minutes and rinsed for 1 hour with dIH_2O. All

gels were examined and photographed under ultraviolet light. Each lane represented one environmental sample and each band (or peak in an automatic sequencer) roughly represents one species in the bacterial community; the band patterns can then be analyzed for the degree of similarity (Garcia-Martinez et al. 1999).

To perform ARISA, 1 µL of the PCR product, along with 0.5 µL of LIZ 1200 internal size standard (Applied Biosystems, Foster City, CA, USA) and 8 µL of deionized formamide were combined in a well on a 96-well plate; the mixture was then denatured at 95°C for 5 min followed by 2 min on ice. The samples were run under gene scan mode for fragment analysis on an 3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) to obtain precise ITS lengths for each sample, with the data being expressed numerically and visually in Applied Biosystems Peak Scanner software (Foster City, CA, USA). An electropherogram, or a series of peaks which represent the size of fragment lengths, is generated by the Peak Scanner program; the peak sizes are determined in reference to the internal size standard LIZ1200, and the peak heights are determined by the intensity of the fluorescence of each length.

3.2.3 ARISA Analysis

The fragment size, height, and area for each sample was exported to Microsoft Excel and to reduce background noise of the genetic analyzer, all peaks less than 100 fluorescent units were excluded from the data. Additionally, fragment lengths less than 200 base pairs were excluded. The file was then saved as an MS-DOS text file. A label file for the samples was created in Excel and also saved as a text file. The data was then inputted into the T-REX software developed by Culman et al. (2009) to identify the true peaks over the noise by determining a baseline and aligning the terminal restriction

fragments in the samples. The triplicate results were binned and averaged by peak height, peak area, and peak presence/absence; Kovacs et al. (2010) suggest binning the fragment lengths as a data processing step in the ARISA analysis to minimize the bias caused by the estimation of the lengths by the DNA sequencer. The binning process groups fragments of similar sizes that may be the products of multiple strains belonging to the same species. The refined and reorganized data was then exported in a tab-delimited data matrix format and submitted into the PermutMatrix software, which results in the output of heat maps and a hierarchical clustering tree, using McQuitty's Method, which represents the structure of the similarities of the samples (Caraux and Pinloche 2004).

To compare the ARISA peak data, a multivariate analysis was performed called principal component analysis (PCA). The matrix tables for peak height, peak area and binary (peak presence/absence) that were used for heatmap generation were also used for the PCA; PCA helps reveal common profiles between the samples by simplifying the data. PCA converts data with several dimensions to a set of linear variables called principal components that represent the variability between the samples. PCA plots display the data in such a way that the first coordinate, or x-axis, represents the greatest variance by any projection of the data, while the second coordinate, or y-axis, represents the second greatest variance.

In addition, two diversity indices were calculated for both Fonner Run and Bates Fork, including the Shannon-Weiner Diversity Index and the Simpson Diversity Index, using the equations $H = -\text{sum}(P_i \log_2[P_i])$ and $D = \text{sum}(P_i^2)$, respectively (REWHC 2000). Each ARISA peak was assumed to be a distinct species. The Shannon-Weiner and Simpson Diversity Indices take into account both richness and proportion of each

species (REWHC 2000). Evenness was also determined using the equation: E = H/log(S); this calculation takes into account the richness, or total number of species, and measures how similar the abundances of each species are (REWHC 2000).

3.2.4 RISA Analysis

The digital images of RISA run out on 10% TBE polyacrylamide gels were uploaded to Quantity-One software (BioRad, Hercules, CA, USA) for fragment analysis. The lanes were manually identified and the bands were identified automatically and manually with one 100-basepair ladder as a standard. The bands were matched manually with a 1% matching tolerance by comparing them with bands from all of the other lanes for possible matches. Banding similarity was calculated and then dendograms were constructed with Quantity-One software (Bio-Rad, Hercules, CA, USA) using the weighted pair group method (WPGMA) and unweighted pair group method (UPGMA).

3.2.5 16s Clone Library Construction

16s Gene Amplification

The primers used were considered to be universal bacterial primers for the 16S rRNA gene sequence (Table 3). The DNA extracted from the environmental samples was amplified using the 8F and 1492R 16S primer set (Edwards et al. 1989, Reysenbach et al. 1994). An aliquot of 100 ng of DNA was added to a PCR reaction with 25 þM of each primer, 12.5 µL 2x Phoenix Taq Master Mix (Phenix Research Products, Candler, NC), and brought up to 25µL with nanopure water. The reaction was run on a Techgene thermocycler (Techne incorporated, Princeton, NJ, USA) under the following conditions: initial denaturation at 95°C for 5 mins, followed by 45 cycles of denaturation at 95°C for 15 seconds, annealing at 55°C for 15 seconds, extension at 72°C for 45 seconds, and final

extension at 72°C for 10 minutes. Negative controls were also run using just nanopure water as a template.

Ligation and Transformation

Following the amplification of the 16S gene, the PCR products were cleaned using the Promega SV PCR Clean-up system (Madison, WI, USA). The product was then ligated into the Invitrogen pCR4®-TOPO vector (Grand Island, NY, USA) then transformed into One Shot® TOP10 Chemically Competent *E. coli* following the Invitrogen protocol. First, a water bath was equilibrated to 42°C while a vial of S.O.C. Medium was warmed to room temperature. LB agar plates containing 100 µg/mL ampicillin were warmed in a 37°C incubator for 30 minutes. One vial of One Shot® chemically competent *E. coli* was thawed on ice for each transformation. 2 µL of the DNA was added into a vial of One Shot® cells and mixed gently by tapping the vial. The vials were incubated on ice for 30 minutes, then heat-shocked for 30 seconds at 42°C. The vials were immediately transferred back to the ice and let to cool for 2 minutes. 250 µL of room temperature S.O.C. Medium was added to each vial, and then the vials were capped tightly and shaken horizontally at 37°C for 1 hour. Volumes of 20 and 50 µL from each vial were spread on separate pre-warmed selective plates and allowed to incubate overnight at 37°C.

Colony PCR

After incubating overnight, 10-20 colonies were harvested from each plate using a sterile toothpick, and then resuspended in 50 uL DI water. The colonies were lysed in a Techne thermocycler (Techne Incorporated, Princeton, NJ, USA) at 95°C for 5 minutes

to denature the DNA. The colony PCR consisted of 10 µL of the clone lysate, 12.5 µL 2x Phoenix Taq Master Mix (Phenix Research Products, Candler, NC), 3.2 þmol of each forward and reverse M13 primers, and nuclease-free water for a total volume of 25 µL. The PCR reaction was run in a Techne thermocycler (Techne Incorporated, Princeton, NJ, USA) under the following conditions: initial denaturation at 95°C for 5 mins, followed by 30 cycles of denaturation at 95°C for 30 seconds, annealing at 55°C for 30 seconds , extension at 72°C for 2 minutes , and final extension at 72°C for 10 minutes.

The PCR product was purified using Sephadex G50 and Princeton Separation CENTRI-SEP columns (Adelphia, NJ, USA). To prepare the columns, 800 µL of Sephadex G50 slurry (1g fine grade G50/16 mL highly purified water) was added to each column. The columns were let to drain and then spun at 750xg for 2 minutes. The column was transferred to a 1.5 mL tube. The entire sample was then loaded into the column and spun at 750xg for 2 minutes. The column was then discarded and the sample was stored at 4°C until the next step. An Invitrogen Qubit® fluorometer (Grand Island, NY, USA) was used to quantify nucleic acid concentration in the samples and 10 uL of the PCR product was run on a 2% agarose gel at 100V for 1 hour to assure correct product length.

Sequencing Reaction

The BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) was used. In a reaction with a total volume of 20 µL, the following were added: 10-20 ng of the cleaned PCR product, 3 µL of the BigDye Ready Reaction Mix, 2 µL of the BigDye Sequencing Buffer, 3.2 þM of the forward or reverse primer (8F or U534R), and nuclease-free water to bring up to volume. The reaction was run on a Techne

thermocycler (Techne Incorporated, Princeton, NJ, USA) under the following conditions: incubate at 95°C for 1 minute , then incubate at 95°C for 30 seconds (ramp 1.0°C/sec), 50°C for 5 seconds(ramp 1.0°C/sec), 60°C for 4 minutes (ramp 1.0°C/sec) and cycle 75 times and then hold at 10°C.

The sequencing product was then cleaned to remove the dye terminators using the Sephadex G50 protocol as described above. The purified DNA sample flow-through was lyophilized until dry at medium heat in the speed vacuum. The sample was then resuspended in 20 μL deionized formamide and transferred to a 96-well plate. The plate was then put in a Techne thermocycler (Techne Incorporated, Princeton, NJ, USA) and denatured at 95°C for 2 minutes and quick-cooled on ice. The plate was then analyzed on an ABI-3130 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

Sequencing and Phylogenetic Analysis

The raw sequence data from the Avant-3130 Genetic Analyzer was initially processed in GeneMapper software (Applied Biosystems, Foster City, CA, USA), and then visualized using the freeware Chromas (Technelysium Pty Ltd, Brisbane QLD, Australia). Low quality sequences displayed numerous ambiguous N bases and were not included in the subsequent analyses. The better quality sequences were edited in DNASTAR SeqMan (Madison, WI, USA) to eliminate the TOPO4 primer binding site.

The forward and reverse sequences for each clone were aligned and trimmed using freeware ClustalX (Conway Institute UCD, Dublin, Ireland). The Mega 4 software (Tamura et al. 2007) was used to create a bootstrapped, neighbor-joining, mid-point rooted phylogenetic tree of the aligned 16S sequences; known sequences were used in the tree for reference and the tree was bootstrapped with 5000 replicates.

Table 2. ARISA Primer Sets

Primer	Gene Target	Sequence 5' -> 3'	PCR Conditions	T_m (°C)	Reference
ITSF	16s	(6FAM)-GTC GTA ACA AGG TAG CCG TA	55 ^{0}C , 45 cycles	54	(Cardinale et al. 2004)
ITSReub	23s	GCC AAG GCA TCC ACC		54	(Cardinale et al. 2004)

Table 3. 16s Sequencing Primer Sets

Primer	Gene Target	Sequence 5' -> 3'	PCR Conditions	T_m (°C)	Reference
8F	Bacteria 16s	AGA GTT TGA TCC TGG CTC AG	55 ^{0}C , 30 cycles	52	(Edwards et al. 1989)
1492R	Universal 16s	GGT TAC CTT GTT ACG ACT T		49	(Reysenbach et al. 1994)
U534R	Universal 16s	GWA TTA CCG CGG CKG CTG	55 ^{0}C , 30 cycles	54	(Baker et al. 2003)

Chapter 4 Results

Samples were collected on a monthly basis from Fonner Run, the control stream, and Bates Fork, the stream potentially impacted by Marcellus Shale drilling activity. The stream water temperature, pH, dissolved oxygen, conductivity, and chloride concentration were tested with an YSI Multimeter in the field in tandem with the sample collection. In the lab, the water samples were tested for chloride, nitrate, sulfate, and bromide on an ion chromatography analyzer. In addition, DNA was extracted from the water samples and then amplified using bacterial-specific primers; with the PCR products, ARISA, RISA, and 16S sequencing was performed. The summary of those results is displayed in Table 4.

Sample	Location	Date	pH	Temp (°C)	DO (mg/L)	Spec. Cond. (uS/cm)	Chloride (mg/L)	Chloride (ppm)	Nitrate (ppm	Sulfate (ppm)	Bromide (ppm)	RISA data	ARISA data	16S data
FR_7/10	Fonner Run	7/28/2010	7.7	20.57	7.64	257.00	na	3.10	0.11	23.90	ND	+	+	+
BF_7/10	Bates Fork	7/28/2010	7.9	22.92	8.83	312	120	na	na	na	na	-	-	+
FR_8/10	Fonner Run	8/3/2010	7.3	21.36	4.38	266.00	10.00	na	na	na	na	-	-	-
BF_8/10	Bates Fork	8/3/2010	7.2	21.90	5.74	302.00	10.00	na	na	na	na	-	-	-
FR_9/10	Fonner Run	9/29/2010	7.4	15.13	5.69	214.00	12.50	1.30	0.88	7.24	ND	+	+	+
BF_9/10	Bates Fork	9/29/2010	7.5	15.01	8.66	254.00	12.50	na	na	na	na	+	+	+
FR_10/10	Fonner Run	10/20/2010	7.5	11.02	10.87	266.00	10.00	2.69	0.18	23.85	ND	-	-	-
BF_10/10	Bates Fork	10/20/2010	7.3	11.02	13.7	337.00	35.00	15.52	0.18	35.93	ND	-	-	-
TM1_10/10	10 Mile Run	10/14/2010	na	13.10	na	873	na	165.14	1.85	228.56	0.87	+	+	-
TM2_10/10	10 Mile/Mon	10/14/2010	na	na	na	319	na	13.55	2.02	148.35	ND	+	+	-
TM3_10/10	Mon Free Tech	10/14/2010	7.2	17.84	11.19	380	na	15.26	2.33	183.38	0.19	+	+	-
FR_5/11	Fonner Run	5/5/2011	7.3	9.35	13.15	141.50	2.25	1.32	0.56	26.90	ND	+	+	+
BF_5/11	Bates Fork	5/5/2011	7.6	10.65	11.98	172.55	9.81	10.56	0.19	50.94	ND	+	+	+
FR_6/11	Fonner Run	6/14/2011	na	na	na	na	11.65	ND	ND	20.99	ND	+	+	+
BF_6/11	Bates Fork	6/14/2011	na	na	na	na	2.19	ND	ND	16.495	ND	+	+	+
DC_5/11	Dunkard at Bobtown	6/22/2011	na	na	na	na	42.13	0.265	2.84	622	ND	+	-	-
WC_6/11	Whiteley Creek	6/22/2011	na	na	na	na	44.61	ND	ND	ND	ND	+	+	-
FR_7/11	Fonner Run	7/7/2011	7.2	19.95	6.21	233.00	2.32	1.65	0.16	21.79	ND	+	+	+
BF_7/11	Bates Fork	7/7/2011	7.5	21.60	6.44	303.00	18.47	14.23	0.32	17.48	ND	+	+	+
FR_8/11	Fonner Run	8/5/2011	7.1	na	na	272.50	3.47	4.41	0.22	30.93	ND	+	+	+
BF_8/11	Bates Fork	8/5/2011	7.5	na	na	391.50	45.87	68.88	0.45	40.55	0.17	+	+	+
BL_8/11	Blacklick Creek	8/5/2011	na	na	na	na	na	38290.00	ND	ND	378.20	+	-	-
FR_9/11	Fonner Run	9/9/2011	7.6	17.40	8.28	209.50	2.44	1.24	0.96	12.57	ND	-	-	-
BF_9/11	Bates Fork	9/9/2011	7.5	17.30	8.58	270.00	12.58	8.21	1.16	15.35	0.07	-	-	-
BS_9/11	Bleeder Shaft	9/8/2011	na	na	na	2440.00	na	1870.00	ND	ND	ND	-	-	-
FR_10/11	Fonner Run	10/28/2011	6.8	9.35	12.10	380.50	1.68	2.25	1.29	18.40	ND	-	-	-
BF_10/11	Bates Fork	10/28/2011	7.2	9.25	12.5	452.5	5.625	3.53	1	10.75	ND	-	-	-
FR_11/11	Fonner Run	11/1/2011	7.9	9.10	13.64	404.00	na	0.56	0.15	6.14	ND	-	-	-
BF_11/11	Bates Fork	11/1/2011	8.1	8.13	15.84	476.00	na	3.09	0.21	9.80	ND	-	-	-
FR_12/11	Fonner Run	12/16/2011	7.9	4.3	13.95	387	32.1	0.88	0.58	11.7	ND	-	-	-
BF_12/11	Bates Fork	12/16/2011	7.6	4	13.6	481	37.5	3.5	0.45	7.48	ND	-	-	-
FR_1/12	Fonner Run	1/16/2012	7.2	2.9	14.4	89.2	0.05	1.3	1.41	18.46	ND	-	-	-
BF_1/12	Bates Fork	1/16/2012	7.2	2.8	14.3	113	0.09	6.27	2.235	21.2	ND	-	-	-
FR_2/12	Fonner Run	2/16/2012	7.6	3.7	13.23	227	1.63	1.845	0.67	25.375	ND	-	-	-
BF_2/12	Bates Fork	2/16/2012	7.5	3.5	13.39	268	6.46	9.725	1.355	25.99	ND	-	-	-
FR_3/12	Fonner Run	3/20/2012	7.6	9.9	11.33	226	4.05	na	na	na	na	-	-	-
BF_3/12	Bates Fork	3/20/2012	7.4	10.5	11.36	251	7.57	na	na	na	na	-	-	-
FR_4/12	Fonner Run	4/26/2012	7.9	12.3	10.52	271	14.245	na	na	na	na	-	-	-
BF_4/12	Bates Fork	4/26/2012	7.7	12.9	13.08	317	32.255	na	na	na	na	-	-	-
FR_5/12	Fonner Run	5/29/2012	7.5	20.6	6.945	295	3.447	na	na	na	na	-	-	-
BF_5/12	Bates Fork	5/29/2012	7.5	23.2	6.39	318.5	11.055	na	na	na	na	-	-	-

Table 4. Complete results for all of the samples collected in this study ("ND" means not detected, "na" means data not available).

4.1 Water Chemistry

To gain a monthly snap-shot of the stream conditions, chemical characteristics were evaluated in the field, using an YSI Multimeter, and in the lab, using ion chromatography. Both Fonner Run, the control stream, and Bates Fork, the stream with drilling in its watershed, were monitored on a monthly basis for pH, temperature, conductivity, dissolved oxygen, sulfate, chloride, nitrate, and bromide (Figures 7 and 8). The pH of both streams remained relatively constant throughout the duration of the study. As expected, the temperature of the streams fluctuated with the seasons; the dissolved oxygen concentration of the water seemed to have a classic inverse relationship with temperature. The specific conductance, which is a measure of how well water can hold an electric current, oscillated and seemed to peak in the fall of 2011.

The chloride levels for the stream samples, as determined by the ion chromatography analyzer, were generally consistent over time, with Bates Fork having consistently higher concentrations than Fonner Run, on average of 15 ppm and 2 ppm, respectively. A very strong peak in Bates Fork of nearly 70 ppm was observed in August 2011, coinciding with newly detectible levels of bromide (0.17ppm); the black stars in the chloride graph in Figure 7 indicate detectible levels of bromide in Bates Fork. The month following the chloride peak, September 2011, the chloride levels were back down to where they had been, but there was still trace levels of bromide detectible in the stream water (0.07ppm). Bromide was not detected in Fonner at any point in time or at any other time in Bates Fork.

Nitrate levels in both Bates and Fonner were fairly similar and fluctuated with no observable seasonal pattern (Figure 8). Sulfate levels were much higher in Bates than in

Fonner for the end of 2010 and the beginning of 2011, but became more similar over time (Figure 7); again, no seasonal patterns were observed.

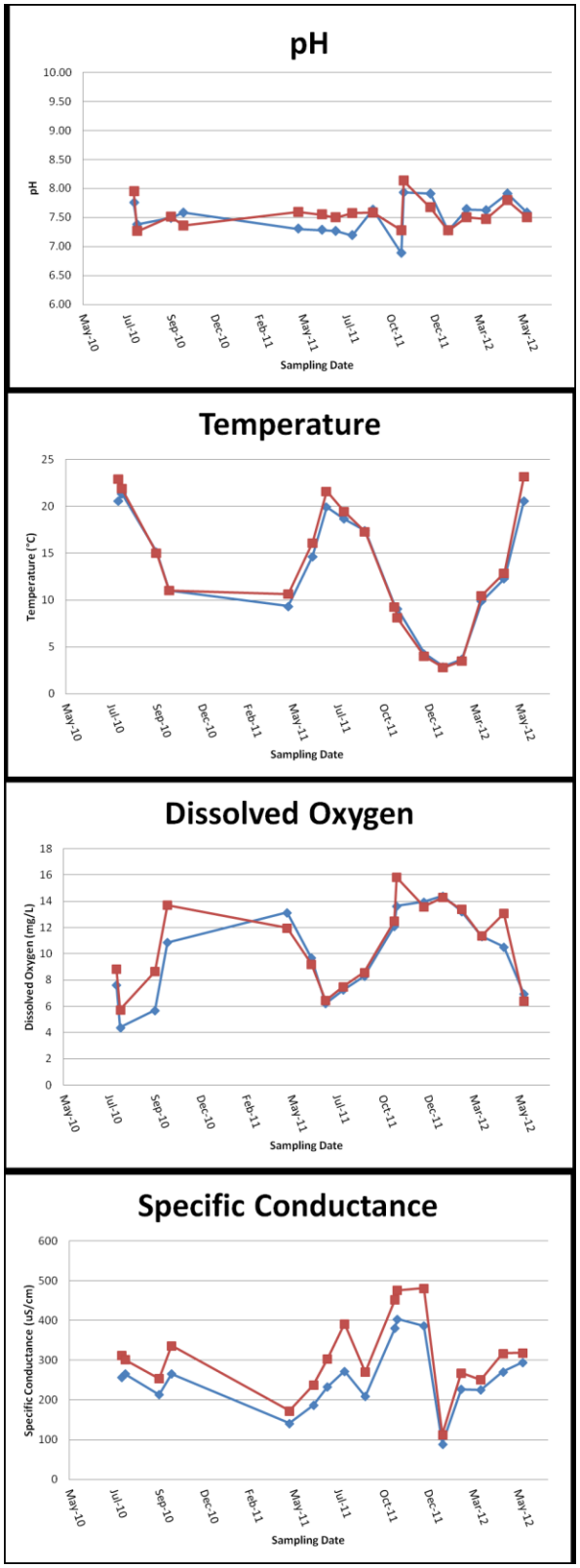


Figure 7. Chemical data obtained using the YSI plotted over time; the red line represents Bates Fork and the blue like represents Fonner Run.

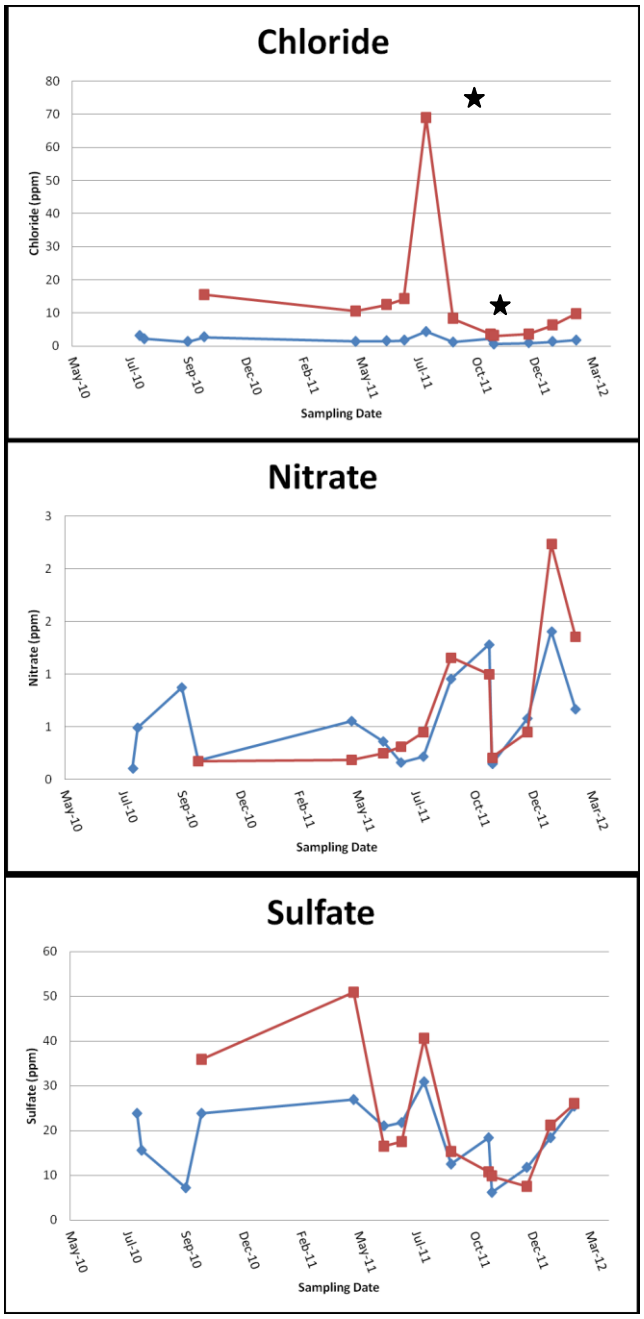


Figure 8. Chemical data obtained using ion chromatography plotted over time; the red line represents Bates Fork and the blue like represents Fonner Run. Black stars indicate the two months when Bromide levels were detectible in Bates Fork (August 2011: 0.17ppm, September 2011: 0.07ppm).

4.2 Bacterial Community Composition as Determined by Molecular Fragment Analyses (ARISA/RISA)

The bacterial community composition of the stream water samples were determined using molecular fragment analyses, which involved DNA extraction, ARISA and RISA PCR amplification, and performing the methods as described above. Both the ARISA and RISA analyses were performed using the same primer set: ITSF/ITSReub. The resulting peaks and bands from the analyses are considered to be bacterial operational taxonomic units (OTUs).

4.2.1 ARISA

The spacer region between the 16S and 23S rRNA was amplified from the DNA extracted from the collected stream samples and analyzed using ARISA. The DNA fragments that were generated were evaluated using a genetic analyzer, resulting in an electropheragram that represents the fragment lengths of the bacterial species of each sample, as depicted in figure 9.

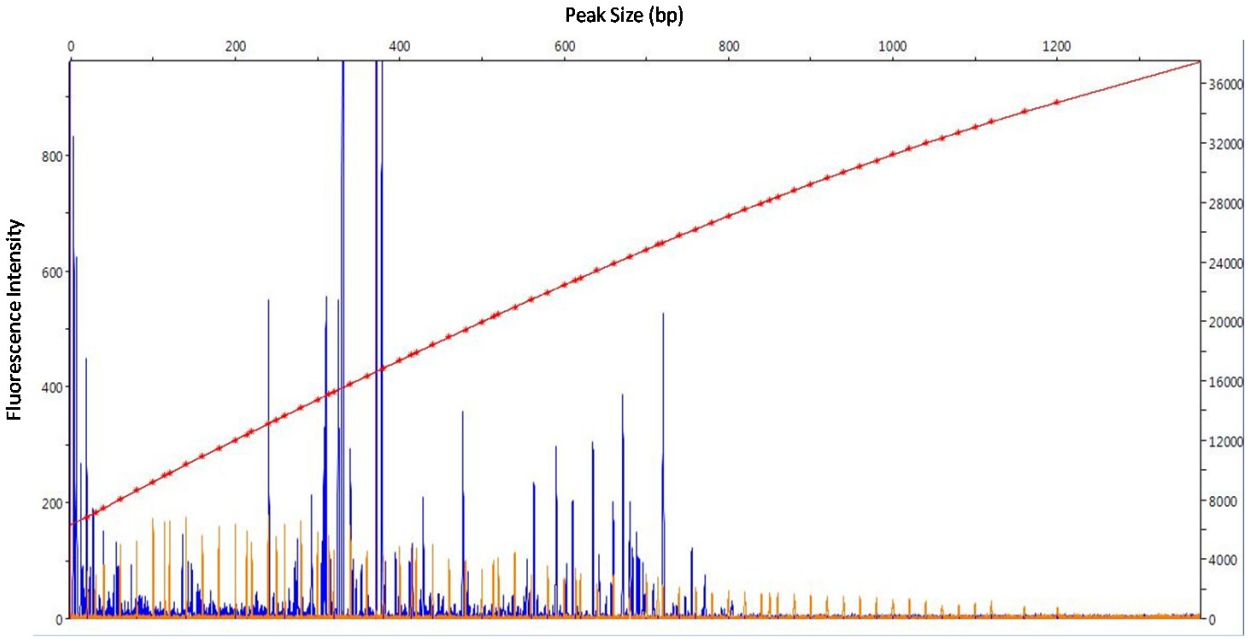


Figure 9. A typical bacterial community ARISA electropheragram obtained from Bates Fork; the x-axis represents the peak size in basepairs, the y-axis represents the relative fluorescence intensity; the orange peaks represent the LIZ1200 size standard, the blue peaks represent the DNA fragments labeled with 6-FAM.

ARISA electropheragrams for the stream samples showed ranges in product size from 209 bps to 836 bps, with several product sizes being more abundant than others (Figure 10). The primers used to amplify the ITS region also amplify approximately 120 bps from the 16S and 23S rRNA genes, therefore the intergenic spacer length is determined by subtracting 120 from the product range (Cardinale 2004). As a result, the actual fragment size range is approximately 89 bps to 716 bps in length. There was high reproducibility of the ARISA profiles among the replicates of the same sample. To reduce background noise and minor variability among the same samples, peaks with an intensity of less than 50 fluorescence units relative to total peak height were not included in the analyses.

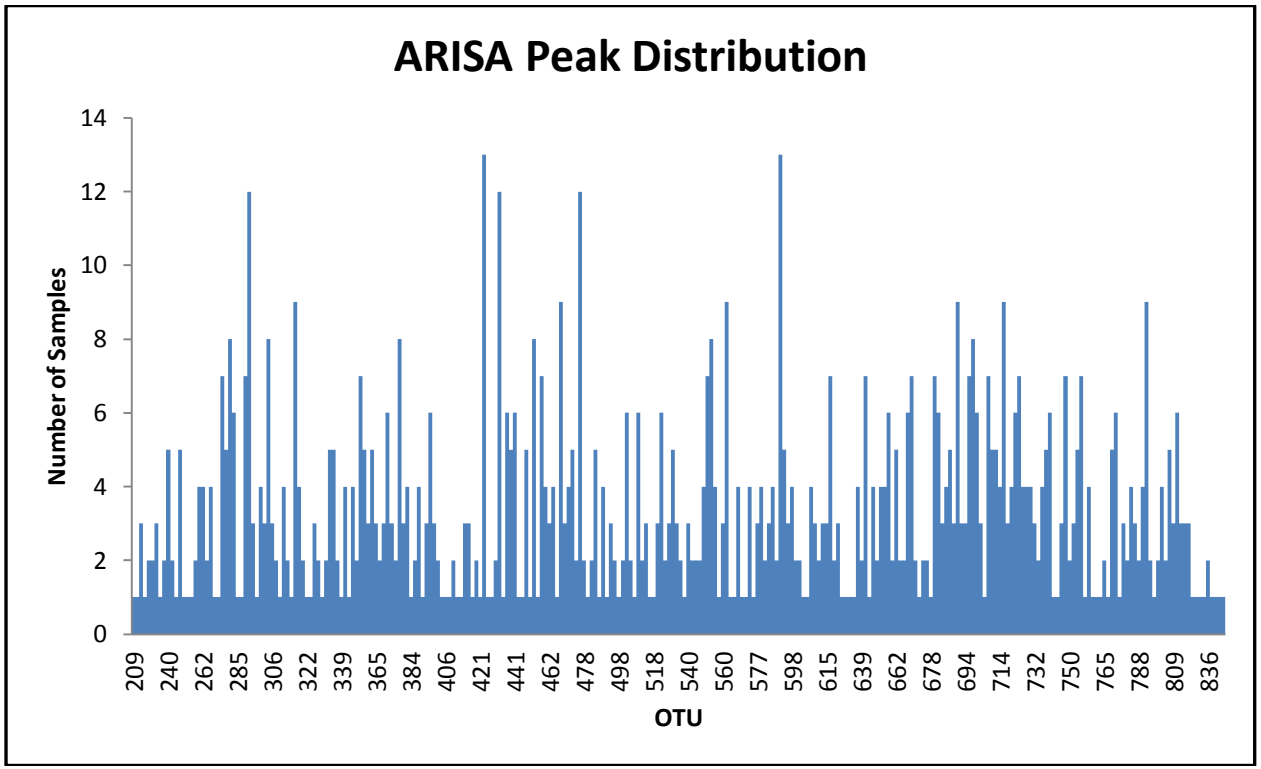


Figure 10. ARISA peak distribution among all of the samples.

The total number of OTUs of each sample is represented in Table 5. Overall, Fonner and Bates had similar total and average numbers of OTUs, in addition to the calculated diversity indices for each stream. The Shannon Diversity Index is out of one, with zero being infinitely diverse; the Simpson Diversity Index is out of one, with one being infinitely diverse. Both Fonner Run and Bates Fork were found to have highly diverse bacterial populations.

Table 5. Number of OTUs for each sample with calculated averages and diversity indices.

Bates Sample	# OTUs	Fonner Sample	# OTUs
BF_9/10	54	FR_9/10	55
BF_5/11	123	FR_5/11	60
BF_6/11	115	FR_6/11	128
BF_7/11	44	FR_7/11	9
BF_8/11	33	FR_8/11	54
Total OTUs	205	**Total OTUs**	211
Avg. OTUs	73.8	**Avg. OTUs**	52.8
Shannon	0.136	**Shannon**	0.132
Simpson	0.993	**Simpson**	0.994

Figure 11 shows the peak area, peak height, and binary data (peak presence/absence) for each sample, as determined by ARISA, represented by heat maps created in Permut Matrix. Some bands are clearly common among several of the samples, presumably meaning that the organism represented by that OTU is present at all of those locations. Dendograms were generated using McQuitty's Method and show how closely related each sample is by using the peak area (A), peak height (B), and binary data (C) (Figure 11). For each heat map, Fonner Run (8/11) is the least related to any of the other samples. Ten Mile (10/10), Bates Fork (6/11), and Fonner Run (9/10) are closely related, according to the dendogram generated using the data from the peak area.

Similarly, the peak height and binary dendograms also show Ten Mile (10/10) and Bates Fork (6/11) being closely related. The Whiteley Creek sample was determined to be not considerably different from most of the Bates, Fonner, and Ten Mile Run samples.

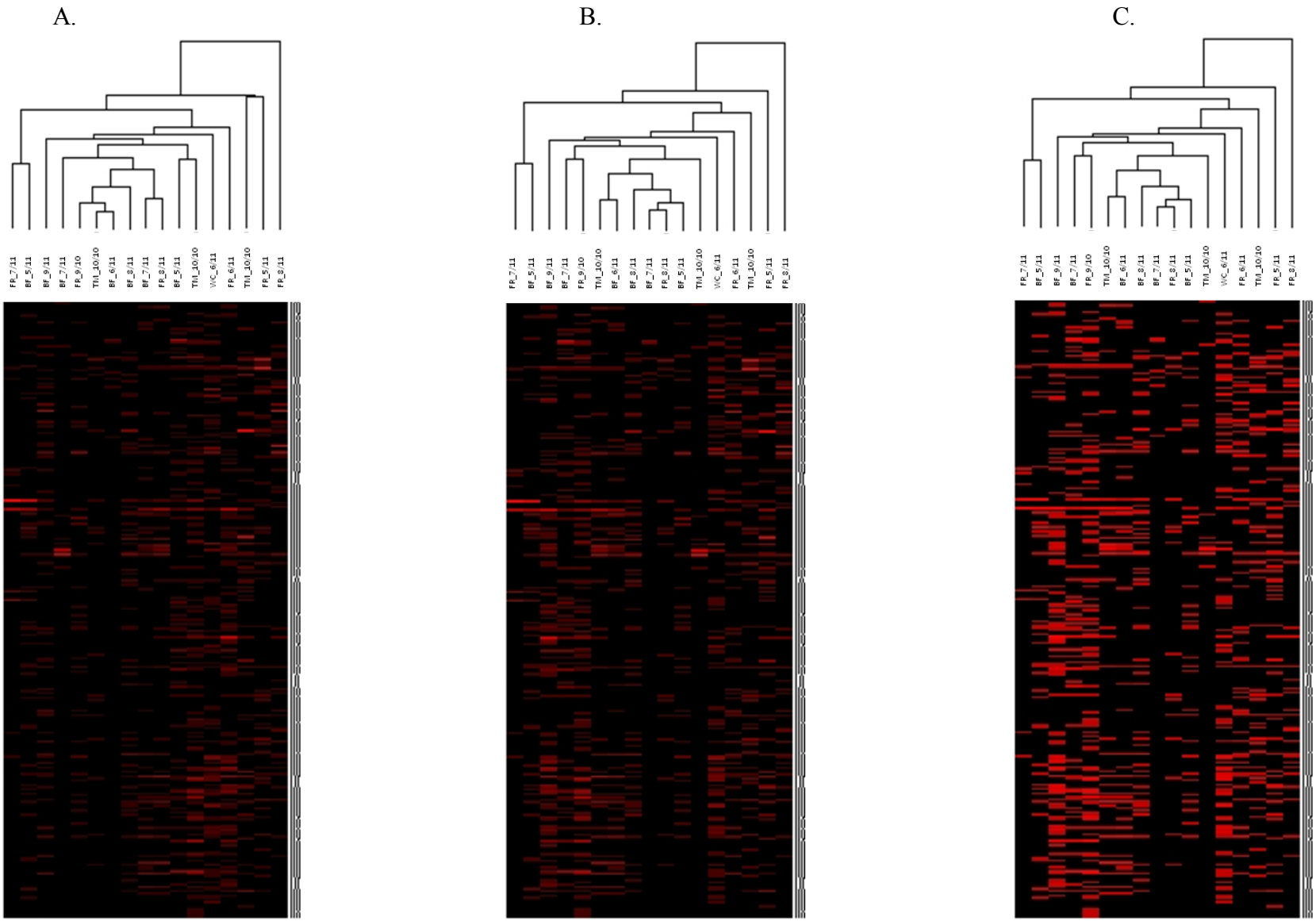


Figure 11. ARISA ITSF/ITSReub heatmaps for (A)Peak Area, (B) Peak Height, and (C) Binary (peak presence/absence).

Heatmaps use many varying banding patterns to assess associations between samples, and can become very complex when the number of samples, and thus the number of OTUs, increases. Principal component analysis (PCA) can assist with those associations, allowing for a much easier way to visualize the data. A PCA plot of the ITS peak area data was generated to assess the similarities between the samples.

The PCA plot generated different associations between than samples than did the dendograms based on the heatmaps (Figure 12). In the PCA plot, both of the Bates Fork (7/11) samples grouped together, with the Bates Fork samples from 6/11 and 9/10 nearby. In addition, Bates Fork (8/11) and both Fonner Run (8/11) samples clustered with the Fonner Run sample from 7/11. The peak area PCA plot shows clear clustering among all three Ten Mile Run samples, as well as both Bates Fork (7/11) samples.

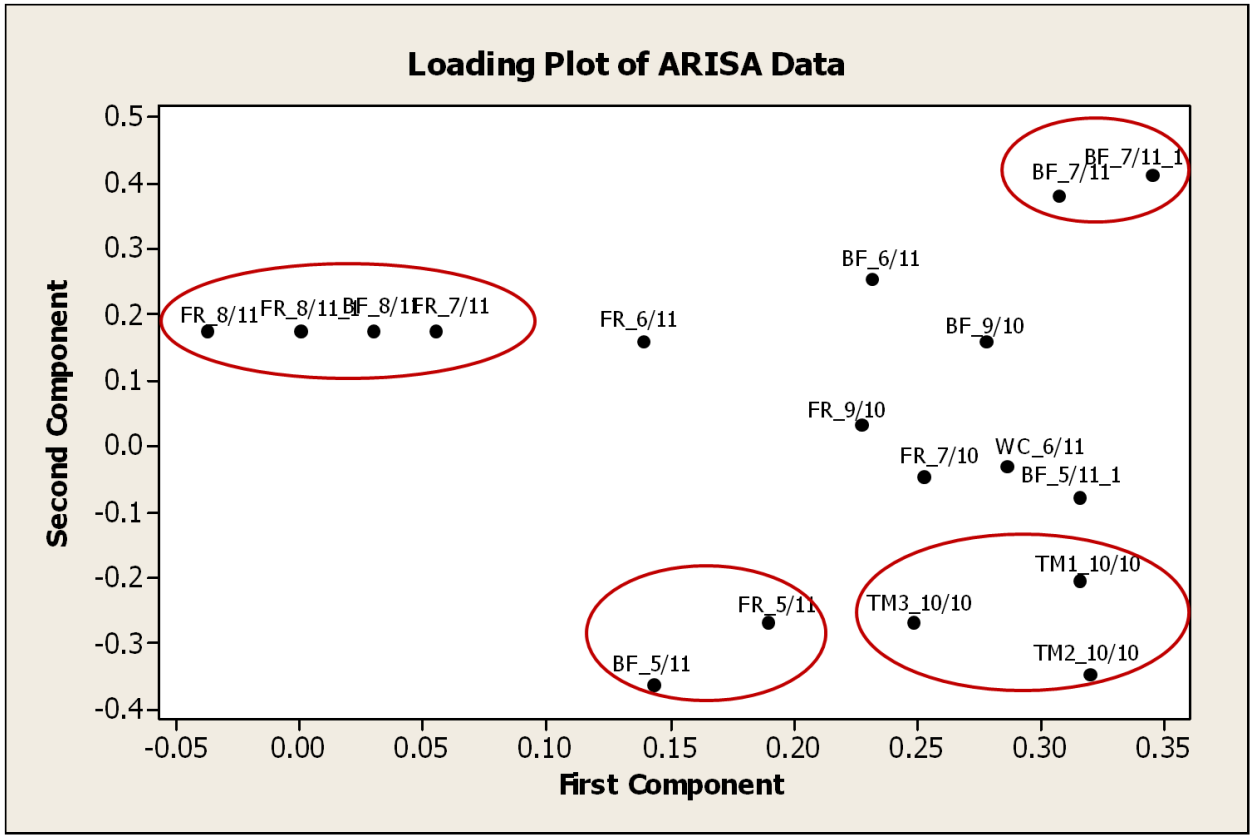


Figure 12. PCA plot from the ITSF/ITSReub ARISA data of the stream water samples using the peak area data; the clustering data points are circled in red.

4.2.2 RISA

The spacer region between the 16S and 23S rRNA was amplified from the DNA extracted from the collected stream samples using RISA. The DNA fragments that were generated were evaluated by running the PCR product out on a gel; the RISA gel image was used for banding pattern analysis in Quantity-One software (Bio-Rad, Hercules, CA, USA) (Figure 13). Bands with lengths greater than 1500 bps were disregarded. Some lanes clearly have more bands than others, and some lanes have bands that are brighter and stronger than others. These bands presumably represent the number and density of OTUs in each sample. Hierarchical cluster diagrams based on the variations in banding size and intensity were generated using the banding patterns of the RISA gel. The dendograms were created using the unweighted pair group method (UPGMA) and weighted pair group method (WPGMA) using the dice coefficient of similarity (Figure 14).

In both the UPGMA and WPGMA dendograms, Bates Fork (8/11) and Blacklick creek (8/11), clustered in a distinct branch. In addition, both dendograms show a branch that includes two Ten Mile Run sites, both Bates (7/11) samples, Whiteley Creek (6/11), and the Bleeder Shaft coal mine effluent (9/11). Another branch in both dendograms shows Bates Fork (5/11), two Fonner Run (8/11) samples, and Dunkard Creek (6/11) clustering.

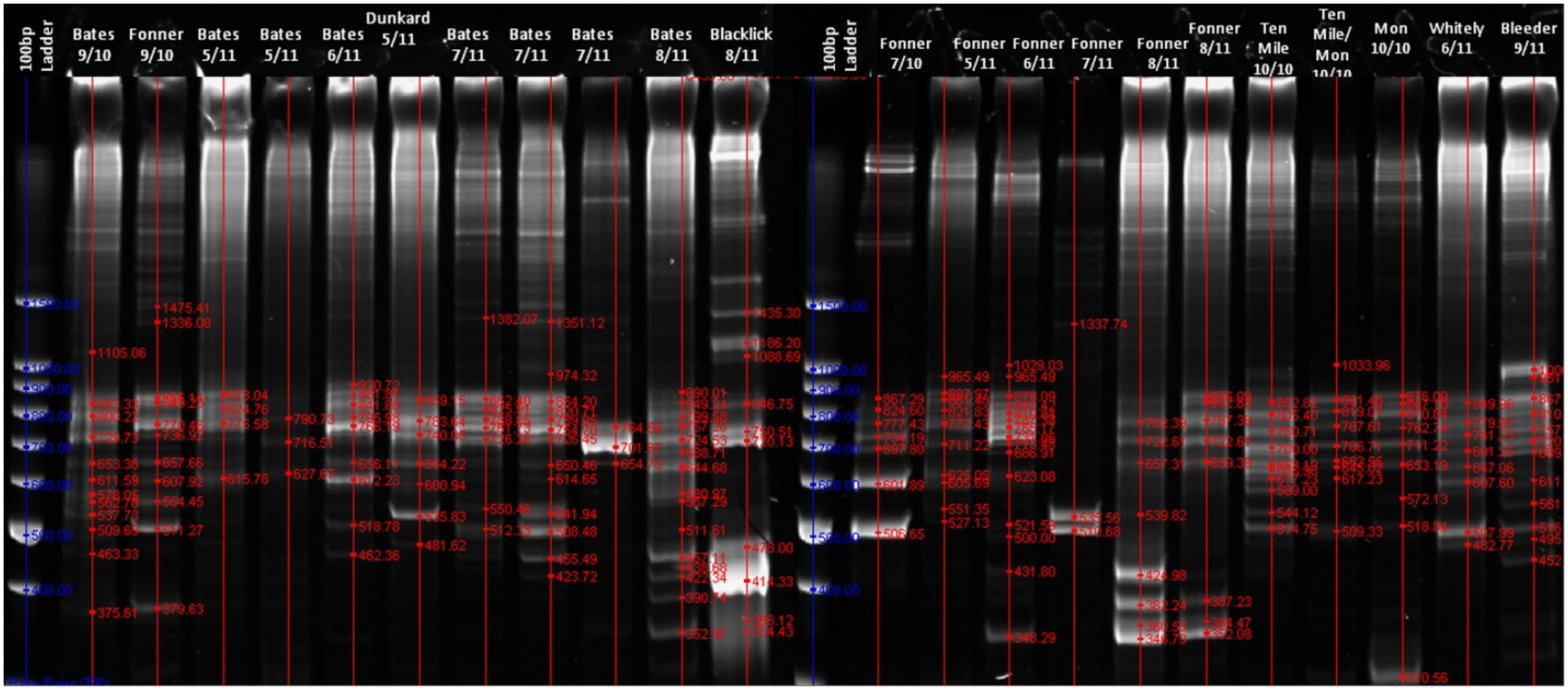
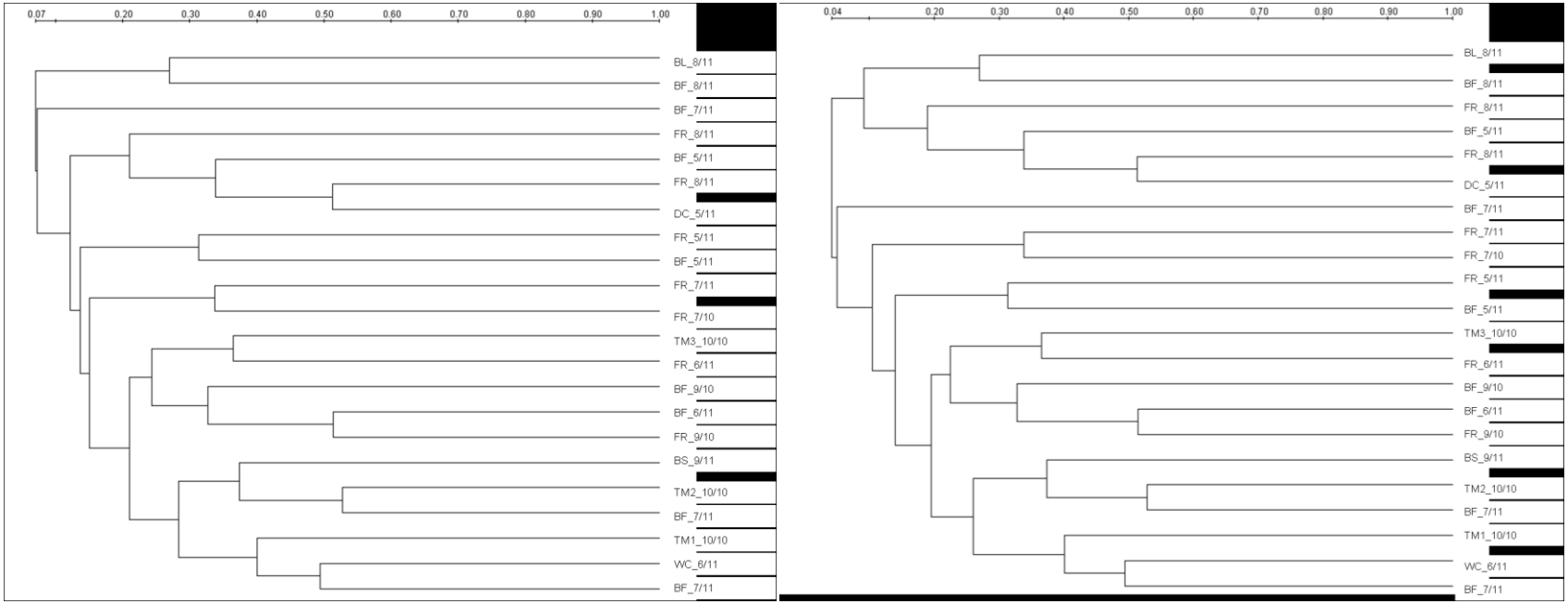


Figure 13. RISA gel displaying banding pattern analysis performed in Quantity One.



UPGMA, dice coefficient

WPGMA, dice coefficient

Figure 14. Mid-point rooted hierarchical clustering of the banding analysis using UPGMA and WPGMA clustering methods to compare RISA banding patterns.

4.3 Molecular 16s rRNA Sequencing

Following the RISA and ARISA studies, 16s rRNA sequencing was necessary to narrow down the taxonomic level of the bacteria that were present in the streams. Most samples that were amplified using the 16S primers resulted in clones that were unable to be successfully sequenced, so the total species were pooled for each stream sampling location. The clones that were successfully sequenced for each sampling location were confidently identified to the genus level; the closest species match for each sequence was also recorded, and all data is represented in Table 6.

Fonner Run produced the most clones that were successfully amplified and thus has the most number of organisms represented in Table 6. Genera found in Fonner Run include *Escherichia*, *Shigella*, *Bacillus*, *Arthrobacter*, *Massilia*, *Duganella*, and *Curvibacter*. The closest species match for the clones from the Fonner Run samples have also been isolated from sources such as goat feces, human feces, several types of soil, and well water (Hariharan et al 2007, Pichel et al 2012, Pellerier and Sygush 2012, Byrne-Bailey et al 2010, Zhang et al. 2006, Li et al. 2004, Ding and Yakota 2004). All of the identified bacterial species are environmental in nature and correspond with the fact that there are farms with livestock directly upstream from the sampling site on Fonner Run.

Bates Fork did not have as many clones that were successfully amplified as Fonner Run. The three genera that were found in Bates Fork were *Halothermothrix*, *Arthrobacter*, and *Micrococcus*. The closest species match for the clones from Bates Fork samples have also been isolated from sources such as salted lake sediment, manured soil, and plant roots (Cayol et al. 1994, Byrne-Bailey et al. 2010, Chen et al. 2009). The identified bacterial species found in Bates Fork suggest that, like Fonner, there is farm

land in the watershed, but some higher concentrations of salt have allowed for the presence of a *Halothermothrix*.

Table 6. Source information for the genus of the sequenced clones.

Class	*Genus*	*Closest Species Match*	*BLAST Score*	*Sample*	*Source*	*Reference*
Gammaproteobacteria	*Halomonas*	*Axialensis*	99%	BL	Solar Saltern	Lee et al. (2005)
	Escherichia	*fergusonii*	92%	FR	Goat Feces	Hariharan et al. (2007)
	Shigella	*flexneri*	95%	FR	Diarrhea Pathogen	Pichel et al. (2012)
Fermicutes	*Bacillus*	*megaterium*	91%	FR	Soil Samples	Pelletier and Sygush (1990)
	Halothermothrix	*orenii*	90%	BF	Salted Lake Sediment	Cayol et al. (1994)
Actinobacteria	*Arthrobacter*	*arilaitensis*	91%, 87%	BF, FR	Manured Soil	Byrne-Bailey et al. (2010)
	Micrococcus	*endophyticus*	90%	BF	Plant Roots	Chen et al. (2009)
Betaproteobacteria	*Massilia*	*plicata*	91%	FR	Soil	Zhang et al. (2006)
	Duganella	*violaceinigra*	92%	FR	Forest Soil	Li et al. (2004)
	Curvibacter	*lanceolatus*	92%	FR	Well Water	Ding and Yakota (2004)

A neighbor-joining, bootstrapped, midpoint-rooted phylogenetic tree was generated in Mega 4 software using the ClustalX alignment of both the 8F and 534R sequence analyses; the trees are shown in Figure 15. The reference species were the closest species matches listed in table 6, and analyzed alongside the stream sample sequences. In the phylogenetic tree, nine of the ten reference species clustered together on a branch and were not obviously closely related to any of the sample sequences. The one reference specie that clustered with several of the stream samples was *Escherichia fergusoni*; it grouped with Bates Fork samples from July and August 2011. The individual branches were broken up according to both location and date sampled.

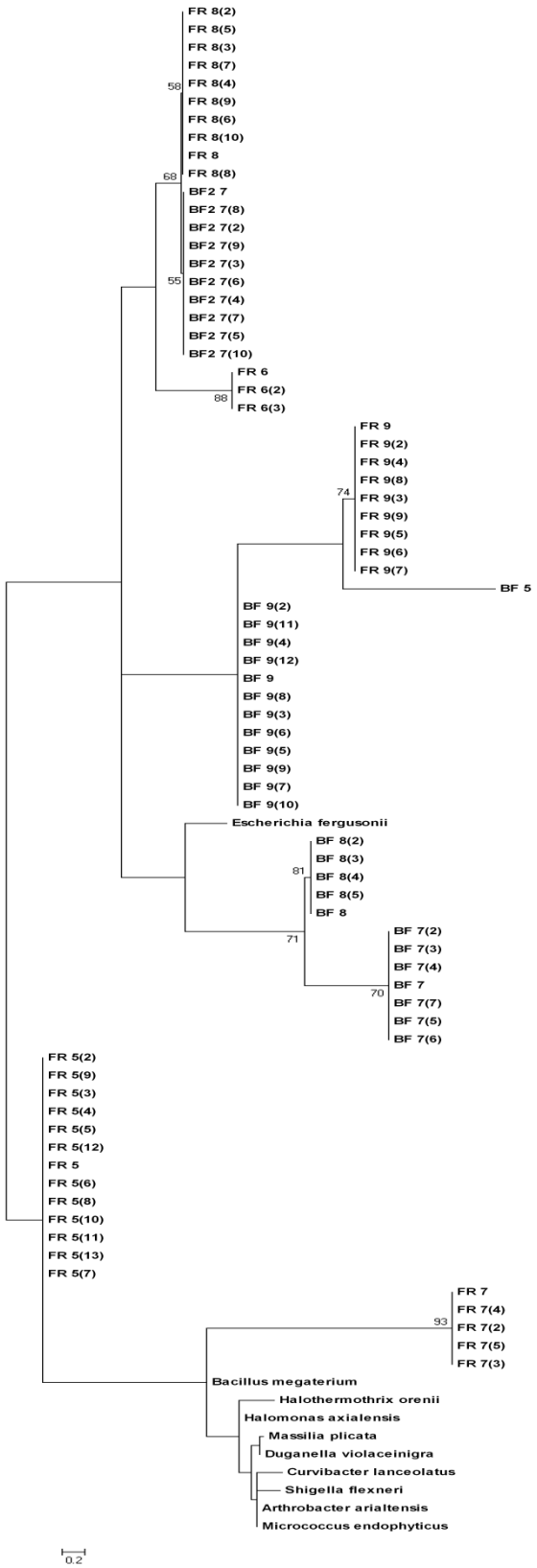


Figure 15. Neighbor-joining, mid-point rooted, bootstrap consensus phylogenetic tree generated using the aligned long sequence reads; includes Fonner, Bates, and several reference groups.

Chapter 5 Discussion

5.1 Water Analysis

The baseline study performed on the two representative streams included the assessment of both chemical and biological parameters. Valuable chemical data was collected on a monthly basis and allowed for a snap-shot of the stream conditions; the plotting of chemical data over time can provide a baseline and reveal the stream's seasonal norms. Abnormal peaks or dips that occur in the chemical data of a stream that is relatively constant otherwise can indicate an effect caused by an anthropogenic source within the watershed.

Both of the study streams, Fonner Run and Bates Fork, were tested each month for pH, temperature, conductivity, dissolved oxygen, sulfate, chloride, nitrate, and bromide. Fonner and Bates both displayed relatively constant pH levels throughout the study, fluctuating between 6.8 and 7.9, which is within the optimal pH range as determined by the EPA (2010). Many of the other streams in the area, particularly in the South Fork of Ten Mile Run, are impacted by net alkaline mine drainage and have a pH of about 8.3; acid mine drainage is also a concern in the region (Kimmel and Argent 2009). The temperature, dissolved oxygen, nitrate, and sulfate seemed to peak and dip at the same times for both streams, most likely reflecting the weather and agricultural patterns for that region. As expected, the dissolved oxygen decreased with increased temperatures.

The specific conductance, or conductivity, is defined as the ability of water to conduct electricity, which increases with increasing concentrations of total dissolved

solids (TDS) (Kimmel and Argent 2009). The TDS of a stream is influenced by its watershed land use, with increases in TDS often coming from anthropogenic sources such as industry and resource extraction, including natural gas drilling (Kimmel and Argent 2009). The conductivity of Fonner and Bates fluctuated at the same time points, but was always an average of 50 $\mu S/cm^3$ higher in Bates Fork. The consistently higher TDS levels in Bates compared to Fonner could be attributed to one or several explanations: 1) Bates Fork has a watershed of nearly double that of Fonner Run, with more surface area to influence it, 2) Bates Fork has more roads in its watershed, contributing more salt from the winter as deicing runoff, and 3) Bates Fork has shale gas drilling activity in its watershed, with contamination coming from spills of frack and flowback fluid.

Higher average concentrations of chloride levels were also seen in Bates Fork; although the overall chloride concentrations for Bates are not high enough to cause concern, as 250 ppm is the chronic toxicity level to freshwater life (US EPA 2010), the fact that the adjacent stream, Fonner Run, has concentrations ten times lower indicates that there could be an anthropogenic source in the Bates Run watershed that is contributing to the notably higher chloride concentrations. In addition to the elevated average chloride concentrations observed in Bates, a chloride peak of 70 ppm was observed in August 2011 and was nearly five times higher than the mean chloride concentration of 15 ppm. This peak in chloride was also seen with detectible levels of bromide. Although bromide occurs naturally in the environment in the ocean and underground rock formations, freshwater lotic systems like the study streams would not typically have detectable levels. According to the EPA (1993), bromide is found in

inorganic halide pesticides, such as methyl bromide and sodium bromide, and it is known to be a component in the brine mixture that drilling operations use. Sodium bromide is a microbicide and prevents bacteria and other microbes from growing and causing unintentional emulsification of the hydrocarbons (Albemarle 2011). Bromide was detected in Bates Fork in the months of August and September 2011, with concentrations of 0.17 ppm and 0.07 ppm, respectively. Volz et al. (2011) state that there is a general agreement that level of bromide in fresh water should be kept below 100 ppb, or 0.1 ppm. The levels in Bates are above this threshold, signifying that authorities should become concerned and begin an investigation to find the source. There are two explanations for the peak of both chloride and bromide in Bates in August 2011: 1) the pesticide methyl bromide that is put on crops ran off the field into the stream, 2) there was a release of frac fluid or flowback fluid containing sodium bromide from one of the wells in the watershed (Flury and Papritz 1993). This chemical data supports the first hypothesis that Bates Fork will have higher chloride and bromide levels than Fonner run, but the reason for the high levels is not yet verified.

5.2 Bacterial Community Richness and Diversity

The chemical conditions observed in Bates and Fonner could, over time, be reflected in the populations of organisms that live in those streams. The fauna of streams, particularly fish, benthic macroinvertebrates, and bacteria, make excellent bioindicators of water quality and stream health; while chemical characteristics only provide a snap-shot in time of the water quality, bioindicators show the conditions over time, incorporating chemical, biological, and physical properties (OH EPA 2010). In this study, bacteria were used as biomonitors and the population composition of the study

streams was approximated using the 16S-23S intergenic spacer region in conjunction with gel electrophoresis and fragment analysis.

Automated Ribosomal Intergenic Spacer Analysis (ARISA)

After the DNA was extracted from the monthly stream samples, the bacterial-specific ITS region was amplified and analyzed using a technique called automated ribosomal intergenic spacer analysis (ARISA). The resulting DNA fragments were analyzed and displayed in electropheragrams for each sample, showing peaks that represent the fragment lengths. Product sizes for the stream samples ranged from 209 bps to 836 bps, with a corresponding fragment size range of approximately 89 bps to 716 bps. The ranges of fragment and peak size, as well as the number of peaks for each sample, correspond to the ranges and number of peaks found by Cardinale et al. (2004) for natural and polluted soil.

Fonner Run was found to have slightly higher total and average numbers of OTUs than Bates Fork, but the difference for both is very small. The Shannon and Simpson diversity indices were calculated for Fonner and Bates using the fragment lengths determined by the ARISA data; both sampling sites were found to have very diverse bacterial populations. This data does not support the second hypothesis of this study, that Bates would have a less diverse bacterial community than Fonner due to the drilling activity in its watershed. Several studies suggest that harsh conditions, such as highly acidic or toxic waters, result in less biomass and less diverse bacterial communities, with specialists ultimately dominating the ecosystem (Lear et al. 2009, Rasmussen et al. 2011). In addition, natural process occurring in the stream that are augmented by bacteria, such as leaf decomposition, can affected and the rates of activity can decrease, creating a

damaging ripple effect in the ecosystem (Rasmussen et al. 2011). In contrast, Hewson and Furhman (2004) performed a study that looked at bacterial diversity along an estuarine gradient and found that there was not a strong correlation between diversity and salinity. The chemical conditions of Fonner and Bates are relatively similar, and are therefore not expected to have much difference in diversity. If the conditions at the sites change over time, especially if the Marcellus Shale drilling activity increases, then a decrease in diversity of the stream could be expected.

The fragment lengths for each sample were compared using both heatmaps and a PCA plot, with trees being generated for each data analysis method. The trees generated using the data from the heatmaps show the sample taken from Fonner Run in August 2011 being the least related to all of the other samples. The reasoning for this branch placement on the trees is unclear from the data, although it would be expected that adjacent streams would have similar species of bacteria at the same time of year, thus similar community profiles, as bacterial populations are known to vary seasonally in freshwater systems (Leff and Lemke 1998). This was the case with the Bates and Fonner samples from May 2011 and August 2011, as well as the Ten Mile Run samples, which included Ten Mile Run, the Ten Mile/Monongahela confluence, and the Free Tech site on the Monongahela River; the samples were taken from adjacent sites on the same day and resulted in strong clustering in the PCA plot using the peak area data.

Ribosomal Intergenic Spacer Analysis (RISA)

RISA was performed in conjunction with ARISA by using the same primers. The resulting DNA fragments were run out on a 10% TBE gel and analyzed in Quantity-One software. The gel image showed bands for each sample, representing the fragment

lengths. Dendogram trees were generated based on the banding patterns. In both of the dendograms that were created, the Bates Fork sample collected in August 2011 and the Blacklick Creek sample from the same month clustered in a distinct branch. The Blacklick sampling site is located at the outflow of the Josephine POTW that has an accepted brine treatment facility for Marcellus drilling wastewater; the chloride concentration was measured at 38,290 ppm and the bromide was 378 ppm (Volz et al. 2011). August 2011 was also the month that Bates showed a chloride concentration peak and detectible bromide levels. It is possible that the same bacterial population that thrives in high salinity and high bromide conditions was making itself evident in Bates Fork.

As shown in figure 13, the number and size of the bands shown on the RISA gel, which ranged from 310bps to 1475bps, were highly variable in all of the samples; both seasonal and stream location variation can been seen in the banding patterns. This was confirmed by the unpredictable clustering band analysis profiles. The high richness and low resolution of the community profiles did not allow for tight clustering of replicates of the samples using the RISA method.

Community Member Identification via 16S Sequencing

To study the microbial community structure in more detail, PCR amplification of the 16S rRNA genes was performed, followed by the subsequent sequencing of the amplicon. The taxonomic level of the bacterial clones that were successfully sequenced could be reliably identified to the genus level; the closest species match for each sequence was also identified. Fonner Run produced the most successfully sequenced clones, most likely to do the high nutrient level resulting from the upstream farmland.

The genera that were found in Fonner Run were *Escherichia*, *Shigella*, *Bacillus*, *Arthrobacter*, *Massilia*, *Duganella*, and *Curvibacter*. The closest species matches for each of the sequenced clones derived from the Fonner Run samples have also been found in goat feces, human feces, manured soil, forest soil, and well water (Hariharan et al 2007, Pichel et al 2012, Pellerier and Sygush 2012, Byrne-Bailey et al 2010, Zhang et al. 2006, Li et al. 2004, Ding and Yakota 2004). The identified bacterial species found to live in Fonner Run are constant to those that might live downstream from forests and farms with livestock. The farm directly upstream from the Fonner Run sampling site may have influenced which species were most prevalent, particularly with livestock in such close proximity to the stream; the animal excrement may have substantially increased the organic substrate load, resulting in a higher abundance of organisms (Merkley 2004).

Bates Fork, conversely, did not have as many clones that were successfully amplified. There were three genera that were found in Bates Fork: *Halothermothrix*, *Arthrobacter*, and *Micrococcus*. The closest species matches for each of the sequenced clones derived from the Bates Fork samples have also been found in salted lake sediment, manured soil, and on plant roots (Cayol et al. 1994, Byrne-Bailey et al. 2010, Chen et al. 2009). The identified bacterial species that were found to live in Bates Fork suggest that, like Fonner, there is both forested and agricultural land use. The genera *Halothermothrix* is salt-loving and could therefore be expected to be found in highly saline environments. Bates Fork does not have the high salt concentrations that the *Halothermothrix* would be expected to be found in, but the bacteria may have come from

the brine used in the Marcellus Shale drilling. One possibility is that the *Halothermothrix* was generated in a frac-water spill upstream, and then washed downstream and diluted.

Dunkard Creek is tributary of the Monongahela River and flows along the border of West Virginia and Pennsylvania that experienced a fish-kill in September 2009 caused by a bloom of salt-loving *Prymnesium parvum*, or golden algae (Renner 2009). Dunkard Creek regularly experiences high levels of TDS due to discharges from the active and abandoned coal mines in the area, but is considered to be a good quality fishing stream (Renner 2009). Immediately before the algal bloom, observed levels of chloride hit 300 ppm, TDS hit 9500 ppm, and the conductivity was greater than 50,000 µS/cm (Renner 2009). Biologists suspect that the soaring salt levels initiated the bloom of golden algae, and they believe that the source of the chemical contamination could be from hydraulic fracturing or coal-bed methane wastewaters (Renner 2009). Unlike coal mine discharges, hydraulic fracturing wastewater is not regulated by the Clean Water Act, so the fate of the liquids resulting from horizontal drilling activity cannot be tracked. The *Prymnesium parvum* is commonly found in East Texas and coastal waters, and some suspect that it may have been transported to southwestern Pennsylvania in a water tanker truck or other drilling equipment that came from the Barnett Shale fields (Renner 2009). Halophilic algae like *Prymnesium parvum* are now justifiable indicator species for biomonitioring in streams in this region, their presence signifying that extremely high chloride and TDS are present, possibly indicating that a release of fracking fluid occurred.

5.3 Potential Bias in Microbial Assessment Methods

Bias in the analysis of gene sequences is introduced though differential amplification caused by variation in the efficiency of the primer binding, interference by

sequences flanking primer regions, and differences in kinetics of the PCR reaction (Baker et al. 2003). These sampling issues can make rare sequences difficult to detect, resulting in PCR bias toward organisms that are more prevalent or have preferentially amplified sequences (Blackwood et al. 2005). The consequence is that many 16S rRNA clone libraries and ARISA fragment presence are not entirely representative of the microbial community being analyzed, especially quantitatively (Baker et al. 2003). In addition, co-extracted contaminants from environmental samples can inhibit PCR amplification altogether, resulting in a bias of which samples can be analyzed (Witzingerode et al. 1997).

Fisher and Triplett (1999) admit that ARISA may underestimate the diversity since unrelated microorganisms may have identical ITS lengths and are thus represented in the ARISA profile by a single peak; genera with identical ITS lengths were generally unrelated, often belonging to different phyla. Conversely, operons within the same genome may differ in spacer length, leading to overestimation of richness (Kovacs et al. 2010). For example, *Escherichia coli* has 15 strains and is represented by a total of nine different ARISA fragment lengths (Kovacs et al. 2010). Environments often contain multiple strains of the same species, resulting in an overestimation of species richness (Kovacs et al. 2010). Although ARISA has several limitations, it remains to be a reasonable method for exploring microbial diversity and creating complex, easy-to-analyze molecular fingerprints (Cardinale et al. 2004, Kovacs et al. 2010). The ITSF/ITSReub ARISA primers used in this study were suggested by Cardinale et al. (2004) to be the best at reducing the effects of PCR biases and providing the most reliable view of bacterial communities.

The 16S rRNA clone library method of microbial community quantification and assessment is also another area in which biases can be observed. The amplification of environmental 16S clones was found to be strongly dependant on the primers used and the number of replication s cycles in the PCR (Witzingerode et al. 1997). In addition, the concentration of DNA from environmental samples can affect the 16S rRNA amplification efficiency; very low DNA concentrations, around 10 picograms, can cause the efficiency of the PCR amplification to fluctuate, resulting in a variation among clone libraries from the same sample (Witzingerode et al. 1997).

5.4. Recommended Future Studies

Future studies of the affect of Marcellus Shale drilling activity on stream bacterial populations should include more matched-pairs of streams to allow for stronger statistical data. More samples should also be taken from the streams, and more often to provide a more seamless view of the chemical and biological health of the aquatic systems over time. Diversity information of other organisms should be used in conjunction with the bacterial diversity data to provide a broader perspective of the ecosystem and how the populations influence each other and interact, particularly when faced with an impact. There are currently two companion studies underway in the Bates Fork and Fonner Run paired stream system, investigating the potential impacts of Marcellus shale drilling activity on fish and salamanders (Porter 2012, Pascuzzi 2012).

The samples should be analyzed in triplicate for ARISA, RISA, and sequencing. Ideally, the microbial diversity of the bacterial species would be calculated using a complete clone library or metagenomics with high throughput sequencing, for example

Roche 454FLX, Illumina Genome Analyzer. More statistical analyses should be used in addition to the diversity indices applied in this study.

Fonner Run and Bates Fork should continue to be monitored for both chemical data and bacterial diversity. The bromide and chloride levels in Bates Fork indicate a source within the watershed; constant observation could allow for an early alert to the authorities if the levels become threatening. Additionally, a site is currently being cleared on a property adjacent to Fonner Run, directly upstream from the sampling site, which may be the location of a new unconventional well pad; this provides an opportunity to keep an eye on the stream health, now that there is an established two-year baseline. This study provided a baseline for Fonner and Bates, and the work should continue for many years in order to truly observe the potential effects of horizontal drilling and hydraulic fracturing on lotic systems.

Chapter 6 Conclusions

Both chemical and molecular and analyses of the study streams have provided a valuable baseline that will hopefully be used in the future to determine the impact of the Marcellus Shale drilling activity in the region. The preliminary chemical data obtained from 2010 to 2012 indicate that there is a difference between the streams, although it has not made itself evident in the stream health as represented by the bacterial diversity. The high chloride and detectible bromide levels are not yet a cause for concern, but should continue to be monitored.

The ARISA, RISA, and 16S sequencing analyses have proved to be important tools for bacterial community fingerprinting, especially from mixed microbial communities in environmental water samples. Each molecular technique has its own limitations in regard to the biases for the investigation of microbial richness and diversity, but the information that each can provide outweighs the weaknesses. Recent advances in molecular techniques, such as high throughput sequencing, should be considered for future studies in order to obtain more accurate clone sequences for each sample and in a higher number; this will allow for a more representative assessment of the community structure in each stream. Fragment analyses like ARISA and RISA should also continue to be used as efficient, low-cost, highly-precise analyses of bacterial richness and diversity.

Chapter 7 References

Albemarle. 2011. WELLBROM Clear Brine Fluids. Albemarle Corporation. Accessed on June 18, 2012 at www.albemarle.com

Araya R, Tani K, Takagi T, Yamaguchi N, Nasu M. 2003. Bacterial activity and community composition in stream water and biofilm from an urban river determined by fluorescent in situ hybridization and DGGE analysis. *FEMS Microbiology Ecology.* Volume 43: 111-119.

Arthur JD, Bohm B, Layne M. 2008. Hydraulic fracturing considerations for natural gas wells of the Marcellus Shale. *The Ground Water Protection Council*: *2008 Annual Forum.*

Baker GC, Smith JJ, Cowan DA. 2003. Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods.* Volume 55: 541-555.

Behlke MA, Huang L, Bogh L, Rose S, Devor EJ. 2005. Fluorescence quenching by proximal G-bases. *Integrated DNA Technologies*. 1-3.

Benlloch S, Acinas SG, Anton J, Lopez-Lopez, Luz SP, Rozriguez-Valera F. 2000. Archaeal biodiversity in crystallizer ponds from a solar saltern: culture verses PCR. *Microbial Ecology*. 41: 12-19.

Blackwood CB, Oaks A, Buyer JS. 2005. Phylum- and class-specific PCR primers for general microbial community analysis. *Applied and Environmental Microbiology.* 71 (10): 6193-6198.

Brown MV, Schwalbach MS, Hewson I, Fuhrman JA. 2005. Coupling 16S-ITS rDNA clone libraries and automated ribosomal intergenic spacer analysis to show marine microbial diversity: development and application to a time series. *Environmental Microbiology.* 7 (9): 1466-1479.

Byrne-Bailey KG, Gaze WH, Zhang L, Kay P, Boxall A, Hawkey PM, Wellington EMH. 2010. Integron prevalence and diversity in manured soil. *Applied and Environmental Microbiology.* 77 (2): 684-687.

Caraux G, Pinloche S. 2004. Permut Matrix: a graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics*. 21 (7): 1280-1281.

Cayol JL, Ollivier B, Patel BKC, Prensier G, Guezennec J, Garcia JL. 1994. Isolation and characterization of *Halothermothrix orenii* gen. nov., sp. nov., a halophilic, thermophilic, fermentative, strictly anaerobic bacterium. *International Journal of Systematic Bacteriology*. 534-540.

Chen HH, Zhao GZ, Park DJ, Zhang YQ, Xu LH, Lee JC, Kim CJ, Li WJ. 2009. *Micrococcus endophyticus* sp. nov., isolated from surface-sterilized *Aquilaria sinensis* roots. *International Journal of Systematic and Evolutionary Microbiology.* 59: 1070-1075.

Clark R. 2011. *Water Quality Terminology*. Delaware River Basin Commission. Accessed on November 13, 2011 at http://www.state.nj.us/drbc/snapshot_terms.htm

Cole JR, Chai B, Farris RJ, Wang Q, Kulam-Syed-Mohideen AS, McGarrell DM, Bandela AM, Cardenas E, Garrity GM, Tiedje JM. 2007. The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nucleic Acids Research.* 35: D169-172.

Culman, S.W., Bukowski, R., Gauch, H.G., Cadillo-Quiroz, H., Buckley, D.H. 2009. T-REX: Software for the Processing and Analysis of T-RFLP data. *BMC Bioinformatics.* 10:171

Ding L, Yokota A. 2004. Proposals of *Curvibacter gracilis* gen. nov., sp. nov. and *Herbaspirillum putei* sp. nov. for bacterial strains isolated from well water and reclassification of *[Pseudomonas] huttiensis, [Pseudomonas] lanceolata, [Aquaspirillum] delicatum* and *[Aquaspirillum] autotrophicum* as *Herbaspirillum huttiense* comb. nov., *Curvibacter lanceolatus* comb. nov., *Curvibacter delicates* comb. nov., and *Herbaspirillum* comb. nov. *International Journal of Systematic and Evolutionary Microbiology*. 54: 2223-2230.

Eldridge WH, Arscott DB, Jackson JK. 2010. Stroud Water Research Center Expert Report on the Proposed Rulemaking by the Pennsylvania Environmental Quality Board for Ambient Water Quality Criterion; Chloride. *Stroud Water Research Center*. Report #2010004.

Fisher MM, Triplett EW. 1999. Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied and Environmental Microbiology*. 65 (10): 4630-4636.

Flury M, Papritz A. 1993. Bromide in the natural environment: occurrence and toxicity. *Journal of Environmental Quality.* 22 (4): 747-758.

Galluzzi L, Bertozzini E, Penna A, Perini F, Pigalarga A, Graneli E, Magnani M. 2007. Detection and quatification of *Prymnesium parvum* (Haptophyceae) by real-time PCR. *Letters in Applied Microbiology*. 46: 261-266.

Garcia-Martinez J, Acinas SG, Antion AI, Rodriguez-Valera F. 1999. Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *Journal of Microbiological Methods*. 36: 55-64.

General Electric (GE). 2012. Natural gas: the bridge to renewable energy. Accessed on June 24, 2012 at http://www.txchnologist.com/volume/natural-gas

Goodfellow WL, Ausley LW, Burton DT, Denton DL, Dorn PB, Grothe DR, Herber MA, Norberg-King TJ, Rodgers JH. 2000. Major ion toxicity effluents: a review with permitting recommendations. *Environmental Toxicology and Chemistry*. 19 (1): 175-182.

Ground Water Protection Council and ALL Consulting. 2009. Modern shale gas development in the United States: a primer. *United States Department of Energy, Office of Fossil Energy.*

Hambright KD. 2010. *Prymnesium parvum* growth studies using the Dunkard Creek isolate (WANA Strain). West Virginia Department of Environmental Protection Division of Water and Waste Management, Charleston, West Virginia.

Hariharan H, Lopez A, Conboy G, Coles M, Muirhead T. 2007. Isolation of *Escherichia fergusonii* from the feces and internal organs of a goat with diarrhea. *The Canadian Veterinary Journal*. 48: 630-631.

Harrison SS. 1983. Evaluating system for ground-water contamination hazards due to gas-well drilling on the glaciated Appalachian Plateau. *Ground Water*. 21 (6): 689-700.

Hewson I, Fuhrman JA. 2004. Richness and diversity of bacterioplankton species along an estuarine gradient in Moreton Bay, Australia. *Applied and Environmental Microbiology.* 3425-3433.

Jackson PE. 2006. Ion chromotography in environmental analysis. *Encyclopedia of Analytical Chemistry.* John Wiley & Sons Ltd.

Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Reseach.* 36: W5-9.

Kargbo DM, Wilhelm RG, Campbell DJ. 2010. Natural gas plays in the Marcellus Shale: challenges and potential opportunities. *Environmental Science and Technology*. 44 (15): 5679-5684.

Kimmel WG, Argent DG. 2009. Stream fish community responses to a gradient of specific conductance. *Water, Air, and Soil Pollution.* Online publication at http://workforce.calu.edu/argent/WhoAmI/kimmelandargentconduct.pdf

Kovacs A, Yacoby K, Gophna U. 2010. A systematic assessment of automated ribosomal intergenic spacer analysis (ARISA) as a tool for estimating bacterial richness. *Research in Microbiology*. 161: 192-197.

Lear G, Niyogi D, Harding J, Dong Y, Lewis G. 2009. Biofilm bacterial community structure in streams affected by acid mine drainage. *Applied and Environmental Microbiology*. 75 (11): 3455-3460.

Lear G, Boothroyd IKG, Turner SJ, Roberts K, Lewis GD. 2009. A comparison of bacteria and benthic invertebrates as indicators of ecological health in streams. *Freshwater Biology*. 54: 1532-1543.

Lee JC, Che OJ, Lim JM, Lee SM, Lee JM, Song SM, Park DJ, Li WJ, Kim CJ. 2005. *Halomonas taeanensis* sp. nov., a novel moderately halophilic bacterium isolated from a solar saltern in Korea. *International Journal of Systematic and Evolutionary Microbiology.* 55 (5): 2027-2032.

Leff LG , Lemke MJ. 1998. Ecology of aquatic microbial populations: lessons from applied microbiology. *Journal of the North American Benthological Society.* 17 (2): 261-271.

Lemke MJ, Brown BJ, Leff LG. 1997. The response of three bacterial populations to pollution in a stream. *Microbial Ecology*. 34: 224-231.

Leuko S, Goh F, Allen MA, Burns BP, Walter MR, Neilan BA. 2007. Analysis of intergenic spacer region length polymorphisms to investigate the halophilic archaeal diversity of stromatolites and microbial mats. *Extremophiles.* 11: 203-210.

Li, WJ, Zhang YQ, Park DJ, Li CT, Xu LH, Kim CJ, Jiang CL. 2004. *Duganella violaceinigra* sp. nov., a mesophilic bacterium isolated from forest soil. *International Journal of Systematic Evolutionary Microbiology.* 54:1811-1814.

Lopez-Garcia P, Lopez-Lopez A, Moreira D, Rodriguez-Valera F. 2001. Diversity of free-living prokaryotes from a deep-sea site at the Antarctic Polar Front. *FEMS Microbiology Ecology*. 36: 193-202.

Malik S, Beer M, Megheraj M, Naidu R. 2007. The use of molecular techniques to characterize the microbial communities in contaminated soil and water. *Environment International*. 34: 265-276.

Merkley M, Rader RB, McArthur JV, Eggett D. 2004. Bacteria as bioindicators in wetlands: bioassessment in the Bonneville Basin of Utah, USA. *Wetlands*. 24 (3): 600-607.

National Energy Technology Laboratory (NETL). 2012. *Produced Water Management Information System (PWMIS)*. Accessed on March 21, 2012 at http://www.netl.doe.gov/technologies/pwmis/intropw/index.html

New York State Department of Environmental Conservation, Division of Mineral Resources. 2009. *DRAFT Supplemental Generic Environmental Impact Statement On The Oil, Gas and Solution Mining Regulatory Program*, available on line at: http://www.dec.ny.gov/energy/58440.html

Ohio Environmental Protection Agency (OHEPA). 2010. Biological criteria for the protection of aquatic life: volume 1: the role of biological data in water quality assessment. *The Ohio EPA Division of Surface Water.* 122 South Front Street, Columbus, Ohio 43215.

Pelletier A, Sygusch J. 1990. Purification and characterization of three chitoanase activities from *Bacillus megaterium* P1. *Applied and Environmental Microbiology.* 844-848

Pichel M, Brengi SP, Cooper KLF, Ribot EM, Al-Busaidy S, Araya P, Fernandez J, Vaz TI, Kam KM, Morcos M, Nielson EM, Nadon C, Pimentel G, Perez-Gutierrez E, Gerner-Smidt P, Binsztein N. 2012. Standardization and international multicenter validation of a PulseNet Pulsed-Field gel electrophoresis protocol for subtyping *Shigella flexneri* isolates. *Foodborne Pathogens and Disease*. 9 (5): 418-424.

Qu X, Mazeas L, Vavilin VA, Epissard J, Lemunier M, Mouchel JM, He PJ, Bouchez T. 2008. Combined monitoring of changes in δ^{13} CH_4 and archaeal community structure during mesophilic methanization of methanization of municipal solid waste. *FEMS Microbial Ecology*. 68: 236-245.

Rasmussen JJ, Wiberg-Larson P, Baattrup- Pedersen A, Monberg RJ, Kronvang B. 2011. Impacts of pesticides and natural stressors on leaf litter decomposition in agricultural streams. *Science of the Total Environment.* 416: 148-155.

Raytheon Employees Wildlife Habitat Committee (REWHC). 2000. Biodiversity Measures. Portsmouth, RI. Accessed on June 14, 2012 at http://rewhc.org/biomeasures.shtml

Renner R. 2009. Salt-loving algae wipe out fish in Appalachian stream. *Environmental Science and Technology*. 9046-9047.

River Network and PA DEP Citizen Volunteer Monitoring Program. 2001. *Designing Your Monitoring Program: A Technical Handbook for Community-Based Monitoring In Pennsylvania*.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 24: 1596-1599.

United States Energy Information Administration (US EIA). 2012. *What is shale gas and why is it so important?* Accessed on March 22, 2012 at http://www.eia.doe.gov/energy_in_brief/about_shale_gas.cfm

US Environmental Protection Agency (US EPA). 2010. CADDIS Volume 2: Sources, Stressors, and Responses. Accessed on June 15, 2012 at http://www.epa.gov/caddis/ssr_urb_wsq2.html

US Environmental Protection Agency (US EPA). 1993. R.E.D. Facts: Inorganic Halides. *Office of Prevention, Pesticides, and Toxic Substances*.

Volz CD, Ferrar K, Michanoqicz D, Christen C, Kearney S, Kelso M, Malone S. 2011. Contaminant characterization of effluent from Pennsylvania Brine Treatment Inc., Josephine Facility being released into Blacklick Creek, Indiana County, Pennsylvania: Implications for disposal of oil and gas flowback fluids from brine treatment plants. *EPA Workshop for the Hydraulic Fracturing Study.* March 28, 2011.

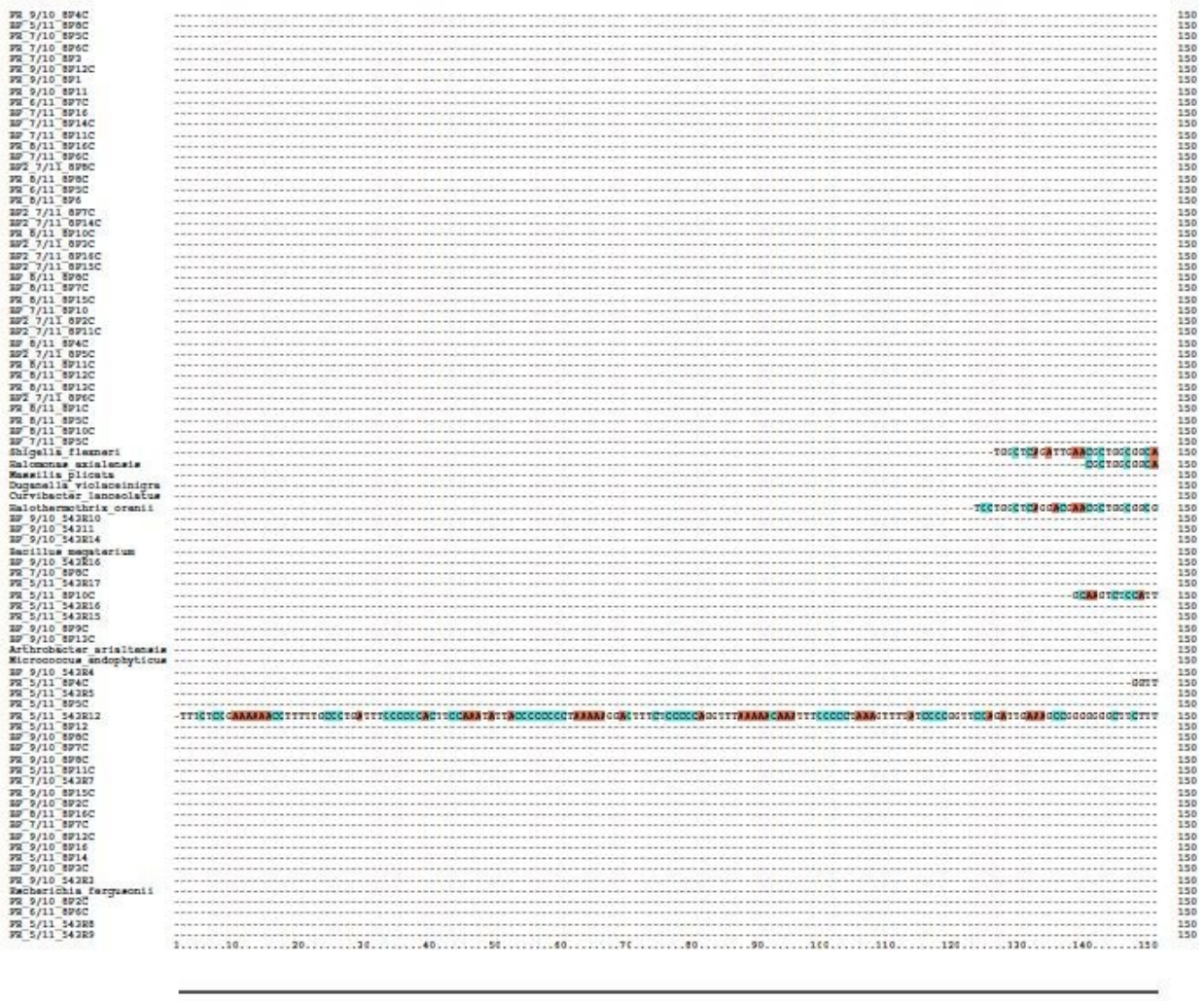
Weber-Scannell PK, Duffy LK. 2007. Effects of total dissolved solids on aquatic organisms: a review of literature and recommendation for salmonid species. *American Journal of Environmental Sciences*. 3 (1): 1-6.

Witzingerode FV, Gobel UB, Stackenbrandt E. 1997. Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiology Reviews.* 21: 213-229.

Yannarell AC, Triplett EW. 2004. Geographic and environmental sources of variation in lake bacterial community composition. *Applied and Environmental Microbiology*. 71 (1): 227-239.

Zhang YQ, Li WJ, Zhang KY, Tain XP, Jiang Y, Xu LH, Jiang CL, Lai R. 2006. *Massilia dura* sp. nov., *Massilia albidiflaca* sp. nov., *Massilia plicata* sp. nov., and *Massilia lutea* sp. nov., isolated from soils in China. *International Journal of Systematic and Evolutionary Microbiology.* 56: 459-463.

Appendix 1: Aligned Sequences




PR_9/10_8P4C
BF_5/11_8P6C
PR_7/10_8P5C
PR_7/10_8P6C
PR_7/10_8P3
PR_9/10_8P12C
PR_9/10_8P1
PR_9/10_8P11
PR_6/11_8P7C
BF_7/11_8P16
BF_7/11_8P14C
BF_7/11_8P13C
BF_7/11_8P16C
BF_7/11_8P6C
BF2_7/11_8P8C
PR_8/11_8P8C
PR_6/11_8P5C
PR_8/11_8P6
BF2_7/11_8P7C
BF2_7/11_8P14C
PR_8/11_8P10C
BF2_7/11_8P2C
BF2_7/11_8P16C
BF2_7/11_8P15C
BF_8/11_8P8C
BF_8/11_8P7C
PR_8/11_8P15C
BF_7/11_8P10
BF2_7/11_8P2C
BF2_7/11_8P11C
BF_8/11_8P4C
BF2_7/11_8P5C
PR_8/11_8P11C
PR_8/11_8P12C
PR_8/11_8P13C
BF2_7/11_8P6C
PR_8/11_8P1C
PR_8/11_8P5C
BF_8/11_8P10C
BF_7/11_8P5C
Shigella_flexneri
Halomonas_axialensis
Massilia_plicata
Duganella_violaceinigra
Curvibacter_lanceolatus
Halothermothrix_orenii
BF_9/10_543R10
BF_9/10_5431I
BF_9/10_543R14
Xanthlum_saxamterium
BF_9/10_543R16
PR_7/10_8P8C
PR_5/11_543R17
PR_5/11_8P10C
PR_5/11_543R16
PR_5/11_543R15
BF_9/10_8P9C
BF_9/10_8P13C
Arthrobacter_arilaitensis
Micrococcus_endophyticus
BF_9/10_543R4
PR_5/11_8P4C
PR_5/11_543R5
PR_5/11_8P5C
PR_5/11_543R12
PR_5/11_8P12
BF_9/10_8P8C
BF_9/10_8P7C
PR_9/10_8P8C
PR_5/11_8P11C
PR_7/10_543R7
PR_9/10_8P15C
BF_9/10_8P2C
BF_5/11_8P16C
BF_7/11_8P7C
BF_9/10_8P12C
PR_9/10_8P16
PR_5/11_8P14
BF_9/10_8P3C
PR_9/10_543R3
Escherichia_fergusonii
PR_9/10_8P2C
PR_6/11_8P6C
PR_5/11_543R8
PR_5/11_543R9
160
170
180
190
200
210
220
230
240
250
260
270
280
290
300

FR_9/10_SP4C
SP_5/11_SP9C
FR_7/10_SP9C
FR_7/10_SP6C
FR_7/10_SP3
FR_9/10_SP12C
FR_9/10_SP1
FR_9/10_SP11
FR_6/11_SP7C
SP_7/11_SP16
SP_7/11_SP14C
SP_7/11_SP12C
FR_8/11_SP16C
SP_7/11_SP6C
SP2_7/11_SP8C
FR_6/11_SP9C
FR_6/11_SP5C
FR_8/11_SP6
SP2_7/11_SP7C
SP2_7/11_SP14C
FR_8/11_SP10C
SP2_7/11_SP3C
SP2_7/11_SP16C
SP2_7/11_SP15C
SP_8/11_SP9C
SP_8/11_SP7C
FR_8/11_SP15C
SP_7/11_SP10
SP2_7/11_SP2C
SP2_7/11_SP11C
SP_8/11_SP4C
SP2_7/11_SP5C
FR_8/11_SP11C
FR_8/11_SP12C
FR_8/11_SP13C
SP2_7/11_SP6C
FR_8/11_SP1C
FR_8/11_SP3C
SP_8/11_SP10C
SP_7/11_SP3C
Shigella_flexneri
Halomonas_axialensis
Haasella_plicata
Duganella_violaceinigra
Cufvibacter_lanceolatus
Halothermothrix_orenii
SP_9/10_542R10
SP_9/10_542I1
SP_9/10_542R14
Bacillus_megaterium
SP_9/10_542R16
FR_7/10_SP8C
FR_5/11_542R17
FR_5/11_SP10C
FR_5/11_542R16
FR_5/11_542R15
SP_9/10_SP9C
SP_9/10_SP13C
Arthrobacter_arilaitensis
Micrococcus_endophyticus
SP_9/10_542R4
FR_5/11_SP4C
FR_5/11_542R5
FR_5/11_SP5C
FR_5/11_542R12
FR_5/11_SP12
SP_9/10_SP8C
SP_9/10_SP7C
FR_9/10_SP8C
FR_5/11_SP11C
FR_7/10_542R7
FR_9/10_SP15C
SP_9/10_SP2C
SP_8/11_SP16C
FR_9/10_SP7C
SP_9/10_SP12C
FR_9/10_SP16
FR_5/11_SP14
SP_9/10_SP3C
FR_9/10_542R3
Escherichia_fergusonii
FR_9/10_SP2C
FR_6/11_SP6C
FR_5/11_542R8
FR_5/11_542R9
460
470
480
490
500
510
520
530
540
550
560
570
580
590
600