

8-2018

Comparative Microbial Community Dynamics in a Karst Aquifer System and Proximal Surface Stream in Northwest Arkansas

Josue Rodriguez
University of Arkansas, Fayetteville

Follow this and additional works at: <https://scholarworks.uark.edu/etd>



Part of the [Geology Commons](#), [Hydrology Commons](#), [Microbiology Commons](#), and the [Sedimentology Commons](#)

Citation

Rodriguez, J. (2018). Comparative Microbial Community Dynamics in a Karst Aquifer System and Proximal Surface Stream in Northwest Arkansas. *Graduate Theses and Dissertations* Retrieved from <https://scholarworks.uark.edu/etd/2883>

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact uarepos@uark.edu.

Comparative Microbial Community Dynamics in a Karst Aquifer System and Proximal Surface
Stream in Northwest Arkansas

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science of Geology

by

Josue Rodriguez
Angelo State University
Bachelor of Science in Geoscience, 2015

August 2018
University of Arkansas

This thesis is approved for recommendation to the Graduate Council.

Matthew D. Covington, Ph.D.
Thesis Director

Kristen Gibson, Ph.D.
Committee Member

Ralph Davis, Ph.D.
Committee Member

Abstract

Northwest Arkansas has well-developed karst systems, with numerous sinking streams and springs. Karst conduits make it easy for contaminants to enter groundwater systems, degrading the water quality and destroying fragile karst ecosystems. With the increase of urbanization, potential threats in the form of fecal contamination may prove to be more of a problem. The purpose of this research is to compare the difference between microbial communities within two different settings, a karst aquifer and a surface stream. The microbial communities within Blowing Spring Cave and Little Sugar Creek were detected and identified in water and sediment samples. Samples were also analyzed using 16s rRNA metagenomic analysis to examine microbial diversity at the genus level. The results showed diversity was greatest between the water and sediment followed by the locations. Interaction of surface and groundwater allows for the introduction and transport of microbial communities, some of which are unique to urban sources.

Acknowledgments

Thank the Lord for I have finished.

I want to acknowledge Dr. James Ward for being a great teacher and mentor. Dr. John Brahana for telling me about the great opportunity I would have at the University of Arkansas and for being a great support on my research and life. Dr. Matthew Covington for giving me the chance to work under him and teaching me new tools to the future. Brandon Conlon for being the greatest friend and support in the office, in the field and in life. Kelli Schaedel for helping me understand microbiology. Sarah Chewing for helping me broaden my understanding of metagenomic sequencing. Max Cooper for helping with my python code. Holly Young, Alex Breeding and Hannah Gnoza for conducting field work alongside of me and helping in collecting my field samples. Sarah Williams and Joshua Blackstock, good friends who I enjoyed working alongside in the field and in the office. Ashlyn Haynes for being a great student, friend and support. Lastly to all the people I met along the way, thank you for being a part of this opportunity and journey.

Dedication

This thesis is dedicated to my professors who saw greater things in me than what I could ever see. To the ones who inspired me to work and strive to be the best that I could be. These were the ones who pushed me, disciplined me and spoke words of wisdom to me. Never giving up on me nor allowing me to give up on myself. Thank you.

Table of Contents

1. INTRODUCTION.....	1
1.1. PAST STUDIES	3
1.2. STUDY AREA.....	7
2. METHODS	8
2.1. SAMPLING SITE LOCATIONS	9
2. 2. DISCHARGE	10
2. 3. <i>E. COLI</i>	11
2. 4. DYE TRACE	11
2. 5. METAGENOMICS.....	12
2. 6. STATISTICAL ANALYSIS	15
3. RESULTS	16
3. 1. DISCHARGE	16
3. 2. HYDROLOGIC PARAMETERS.....	16
3. 3. <i>E. COLI</i>	17
3. 4. DYE TRACE	19
3. 5. METAGENOMICS.....	20
3. 5. 1. <i>General data</i>	20
3. 5. 2. <i>Raw Counts</i>	20
3. 5. 3. <i>Relative abundance</i>	21
3. 5. 4. <i>Alpha diversity</i>	22
3. 5. 5. <i>Beta diversity</i>	23
4. DISCUSSION	25
4. 1. WATER QUALITY PARAMETERS	25
4. 1. <i>E. COLI</i>	26
4. 1. 1. <i>Blowing Spring Cave</i>	26
4. 1. 2. <i>Little Sugar Creek</i>	27
4. 2. DYE TRACE	27
4. 3. METAGENOMICS.....	28
4. 3. 1. <i>Blowing Spring Cave</i>	30
4. 3. 2. <i>Little Sugar Creek</i>	31
5. CONCLUSION	32
6. REFERENCES.....	35
7. APPENDIX.....	43

1. Introduction

As the world's population increases, the need for clean freshwater also increases. However fresh water sources only make up 2.5% of the water in the world. Freshwater sources consist of surface water, groundwater, and glaciers, with each containing approximately 1.2%, 30.1%, and 68.7%, respectively (Perlman, 2016, Shiklomanov, 1998). Ultimately, life on Earth is maintained by approximately 1% of the freshwater worldwide, since most of the freshwater is locked up within the glaciers. Groundwater and surface water are the most available sources of freshwater, which have allowed for growth in agricultural, domestic and industrial usage (Döll et al., 2012). Agriculture itself accounts for 70% of the groundwater withdrawn worldwide. Groundwater also provides almost half of the drinking water (Smith et al., 2016). Overall, 982 km³/year are withdrawn from groundwater sources. Since 1960 groundwater withdrawal has increased, and if not managed properly, could exhaust groundwater reservoirs, reduce surface water levels, and dry wetlands in the process (Wada et al., 2010). Climate change and other anthropogenic impacts also influence our clean groundwater reserves. It is vital to protect and preserve these water sources from further exploitation and contamination. To create better preservation practices, it is important to identify and characterize the types of groundwater sources that are most impacted by depletion and contamination.

Karst is a landscape and specific hydrogeology that results from the dissolution of highly soluble rocks, such as limestone. Pipe like patterns are formed that can transport water faster than most groundwater media, such as porous media composed of detrital sediments. However, water storage is low within karst and contamination within the aquifers can move through these systems relatively rapidly, (Boyer and Pasquarell, 1999). Typically, with karst, little attenuation occurs, allowing contaminants to remain at elevated concentrations (Vesper et al., 2001). These

contaminated waters can be connected to public drinking water sources, including both surface water and groundwater, via sinkholes and springs which could pose health problems to the community. Since about 15% of Earth's ice-free land is covered in karst, it is important to consider the quantity and quality of freshwater within karst aquifers (Ford and Williams, 2007). A variety of types of contaminants exist, including dissolved nutrients (nitrate, phosphate, chlorides, etc.), dense non-aqueous phase liquids (DNAPLs), light non-aqueous phase liquids(LNAPLs), metals, and fecal matter, all of which behave differently and can come from both anthropogenic and natural sources (Fetter, 2018). Contaminant sources can be classified as point and non-point sources. Point sources consist of limited and fixed sources such as leaky septic tanks, sludge ponds, and spills. Nonpoint sources are those that issue from widely distributed origins, such as urban or agricultural runoff. Contaminants are often linked to urbanization (Kolpin et al., 2002; Naik et al., 2008; Misra, 2011; Kaushal and Belt, 2012; Knierem et al., 2015). Fecal contamination is a global concern, affecting water sources in both urban and rural areas regardless of the status of water source (e.g., improved or unprotected) (Bain et al., 2014). Bain et al. (2014) noted that rural water sources worldwide are more likely to be contaminated due to inadequate sanitation. Fecal contamination in wastewater can including nutrients such as nitrates, pharmaceuticals, metals, and a variety of microorganisms (e.g., viruses, bacteria, and protozoa) including pathogens (Kolpin et al., 2002). If exposed to pathogenic microorganisms, persons can experience an illness and even death. Bacterial pathogens commonly spread via contaminated water include enterotoxigenic *Escherichia coli*, *Vibrio cholerae*, and *Salmonella enterica* serovar Typhi; however, these are more commonly transmitted outside of the U.S. and other developed countries (Toze, 1999; Ashbolt, 2004). The pathogenic bacteria mentioned above can come from either human or animal sources of fecal

pollution as well as exist naturally in the environment (i.e. *V. cholerae*). For example, *Salmonella* Typhi—the cause of typhoid fever—is strictly associated with humans and higher-level primates. Additional pathogens transmitted by water include *Giardia* and Hepatitis A and E viruses (Ashbolt, 2004).

As the global freshwater supply is depleted, more water management will be needed. Therefore, it is important to characterize vulnerable freshwater sources. The characterization of karst systems will help address potential contamination sources within individual watersheds and in turn developing a better understanding of karst groundwater resources. The purpose of this study is to analyze the microbial dynamics within Blowing Spring Cave and Little Sugar Creek. Blowing Spring Cave is a karst system that discharges into Little Sugar Creek, a proximal surface water system. Monitoring microbial communities within both surface and subsurface systems will provide an understanding of how each system influences one another in terms of sources and transport. The objectives of the project are to: 1) characterize changes in microbial presence, abundance and diversity within the water and sediment column for both locations over a year by using 16s rRNA metagenomic analysis, and 2) constrain potential contaminant sources using the metagenomic data, and 3) further constrain the recharge area of the spring using dye tracing experiments.

1.1. Past Studies

Since microbes are wide spread through nature, it is difficult to pinpoint sources of contamination. Past studies have focused on identification and detection of bacterial sources in various systems from surface to subsurface and within rural to urban settings (Kendall, 1999; Ward et al., 2009; Mahler et al., 2000; Fields, 1999). Methods used for identifying sources range

from the analysis of chemical, physical, and biological parameters that correlate with fecal contamination.

Fecal contamination can be detected using fecal indicator bacteria (FIB), such as *E. coli*, a facultatively aerobic, Gram-negative, non-spore-forming, rod-shaped bacteria. In accordance with the Safe Drinking Water Act, the state of Arkansas Department of Environmental Quality (ADEQ) set standards for levels of *E. coli* that can be present for consumption of drinking water and for recreational contact (APCEC, 2016). The recreational water guidelines consist of the primary contact limit, the level at which full body contact is no longer safe, and secondary contact limit, the level at which contact in the form of wading, fishing, and boating is no longer allowed. The primary contact limit is 410 Most Probable Number (MPN)/100 mL and the secondary contact limit is 2050 MPN/100 mL. The ADEQ requires that MPN for no more than 25% of samples from no less than 8 samples per contact season may exceed the limit (ACA, 2011).

Alongside *E. coli*, other water quality parameters have been used to build correlations that enable identification of contamination within a given system. Past studies have analyzed the relationships between contamination and discharge (Knierem et al., 2015) or nutrients (Graening and Brown, 2003). Nitrogen-15 and oxygen-18 within nitrate have been used to track fecal sources (Burns et al., 2009). However, the division between human and animal waste becomes more difficult to interpret in a mixed setting of both rural and urban influences due to the overlapping isotopic signatures of manure and septic waste (Kendall, 1998).

Within karst systems, turbidity is often correlated with *E. coli* and other bacteria (Mahler et al., 2000). For instance, microorganisms can attach to sediment particles via adsorption and then be transported during resuspension of sediment due to storm water pulses (Mahler et al., 2000;

Jamieson et al., 2005; Ryan and Meiman, 1996). However, other research has shown high levels of microbes during periods with low turbidity levels (Bouchaou et al., 2002; Amraoui et al., 2003). Sediment is also thought to serve as a storage site for microorganisms including pathogenic microbes (Pronk et al., 2006; Rehman and Soupir, 2009). Given favorable conditions, pathogenic microbes have the potential to be present for several months (Davis et al., 2005).

Therefore, it is possible for sediment to be a source for fecal contamination in a karst system as storm water pulses resuspend these microbes into the water column potentially impacting water quality (Heinz et al., 2009).

Researchers have used tracer techniques to identify potential recharge points where contaminants could enter karst systems (Heinz et al., 2009). There are various forms of tracers; however, the most common tracers consist of fluorescent compounds that can be dissolved into injected waters and can be detected using in place fluorometers or activated charcoal packets (Davis et al., 1980; Fields, 1999). Charcoal packets are a cheap and effective method for collecting preliminary data including background fluorescence and qualitative results that verify connection between recharge and discharge points via sinkholes and springs, respectively. Other types of tracers consist of particulates which are used to represent large sized bacteria transport (Goppert et al., 2006); thermal pulses, which help characterize the hydraulic diameters of a conduit (Covington et al., 2011; Luhmann et al., 2015); and stable isotopes which can help identify sources and mixing of water sources.

Researchers have analyzed a variety of techniques to develop new ways to constrain contaminant sources to point sources. Further delineation has been attempted using AC/TC ratios, which compare concentrations of atypical coliforms (AC) and total coliforms (Reed et al., 2010; APHA, 1998). Typical total coliforms (TC) include fecal coliforms as a subset, with *E. coli* and

other enteric bacteria within the subset (Madigan and Martinko, 2005). AC/TC ratios, fecal age indicators, have been used to differentiate between agricultural, mixed, and urban sources (Brion et al., 2000). Lower AC/TC ratios indicate a young age of fecal input, therefore higher levels of contamination. Higher AC/TC ratios, indicate an older age of fecal input, meaning that the contamination is less concentrated (Black et al., 2007; Ward et al., 2009). However, AC/TC ratios are challenging to use in karst systems due to the favorable survival conditions for bacteria (Ward et al., 2009).

With large populations of microbes within water it is important to note what bacteria are present in a system. The culturing of samples limits the identification of bacteria to only culturable bacteria and masks any interactions of the various bacteria communities (Wooley et al., 2010). Metagenomic sequencing enables a better representation of microbial communities within the environment. Sequencing data, specifically 16s ribosomal RNA (rRNA) metagenomic analysis has become more widely used to understand microbial communities across the world (Klindworth et al., 2012; Gomes et al., 2007; Kozich et al., 2013, De Mandal et al., 2015). Next generation metagenomic sequencing has advanced greatly and is more cost effective than before. However, this form of sequencing only extends to the genus taxonomic level as it is designed to target only fragments of genomes instead of complete genomes (Wooley et al., 2010). Most of the genome sequence is required to identify individual species. With respect to microbial communities in water, DNA genome fragments come from the samples that have been filtered and extracted. These fragments are then cloned, sequenced and assembled through sequencers such as Illumina MiSeq. DNA fragments range from 20 base pairs (bp) to 700 bp. After assembly, the constructed sequence length is usually around 5,000 bp, only a fraction of the whole genome. Therefore, the technique is unable to produce a full genome for a specific species

since the species information is incomplete. (Wooley et al., 2010). However, genus taxonomic levels can be determined reasonably well, allowing the analysis of relationships between microbial communities, the locations in which they live, and how they change over (Ley et al., 2006; Von Mering et al., 2007; Qin et al., 2010). This increases sequencing quantity without decreasing quality.

1.2. Study Area

The study area contains both clastic and carbonate rocks of the Springfield plateau, which is a part of the Ozark plateau province. Rock units within the area are relatively flat-lying with gentle 2° dips towards the southeast. Carbonates dominate this region, which makes up one of the largest karst regions in the US (Weary and Doctor, 2014). However, surface karst development is poor due to the regolith created by the residual chert from the Boone Formation. This Formation is composed of a cherty limestone with interbedded chert and crinoidal limestone layers. Most conduits, springs, and sinkholes in the area develop in the Boone Formation. Springs in this Formation within the study area include Blowing Springs, Ford Spring, and Boy Scout Spring. These springs form at the base of the Boone in the St. Joe limestone member, which overlies the Chattanooga Shale. The Boone contains a set of orthogonal fractures, which are often enlarged by dissolution that can control the location and orientation of karst conduits. Chert layers within the Boone can serve as a flow surface where water will flow down dip

Blowing spring has a recharge area of approximately 2.9 km², as estimated using the normalized base flow method (Brahana, 1997; Knierim et al., 2013). The mapped passage in Blowing Springs Cave totals 2.4 km, and the spring ultimately discharges into Little Sugar Creek, a tributary of the Elk River. Blowing Spring is located within a park system managed by the

BVPOA. Mapped passages of the cave pass under residential houses and a forested area. Soils in the area are gravelly silt loam, roughly 3 m thick. These soils transmit water up to 15 cm/hr. (Natural Resources Conservation Service, 2013). Other springs in the area include Ford Spring, Trout Farm Spring, Boy Scout Spring, and Brassard Spring, all of which feed into Little Sugar Creek. Little Sugar Creek's drainage area is 508 km², and the creek flows northward (Figure 1). The creek flows through the city of Bella Vista, where golf courses and bike trails border the banks. A mixture of forest and pastures dominate the drainage area. Other cities in the basin area include Bentonville and Rogers.

The research area is in the town of Bella Vista in northwest Arkansas near the state line of Missouri. Bella Vista has developed into a retirement community and is now managed by the Bella Vista Property Owners Association (BVPOA) (Fite, 1993). The community was established in the 1920s. Residents of this community had been limited to the use of on-site septic systems while community buildings used sewage oxidation ponds. Bella Vista has since developed to accommodate for the increase of sewage by building and expanding a wastewater treatment plant. The Soil Survey Staff (2014) have stated soil areas around Blowing Springs Cave are "very limited" for the construction of septic tank absorption fields due to unfavorable conditions. In 2009, the City of Bella Vista placed an ordinance that requires homeowners to connect to a sewer system where it is available when property is sold (City of Bella Vista, 2009).

2. Methods

In this study we characterize the microbial community diversity in Blowing Spring Cave (BSC) and Little Sugar Creek (LSC) in both the sediment and water. We also attempted to identify sources of bacteria for BSC and LSC by delineating the stream channels and the recharge area for BSC. Water samples were collected for the analysis of *E. coli* to identify sites with

potentially elevated levels of fecal contamination. The same water samples were then processed for metagenomic analysis to quantify the taxonomic units at each location, showing the diversity of microbial communities in the sediment and water for BSC and LSC. A dye trace was conducted to delineate the recharge area of BSC to help constrain surface to subsurface interaction where contamination sources could enter a karst system. Methods used for the project are described below.

2.1. Sampling Site Locations

The three sample sites within Blowing Springs cave (BSC) were labeled from downstream to upstream as BS1, BS2, and BS3 (Figure 1). These three sites divide the cave into three sections: the front, middle, and back, respectively. Dividing the cave in this form helps to constrain whether contaminants are entering the cave through the main stream or infeeding streams. Both water and sediment samples were collected at each site. Water samples were collected using sterilized 500mL Nalgene bottles and sediment samples were collected using Whirl-Pak bags. The water samples were collected approximately 5 cm below the water surface and sealed with a cap. The sediment was collect 10 cm below the water with a scoop, and placed in Whirl-Pak bags. Most of the cave sediment ranges from fine grain sands to clays; however, the upstream location has some coarse-grained sand and gravel. Samples were collected in slow moving portions of the cave stream near the stream banks where sediment was deposited.

Similar methods were applied at sites within Little Sugar Creek. Sample sites within Little Sugar Creek (LSC) were labeled as LSC1, LSC2, LSC3 and LSC4, numbered from upstream to downstream. For LSC the sites span both rural and urban settings. Water from BSC intersects LSC between sites LSC2 and LSC3.

2. 2. Discharge

Water level was measured from a stilling well behind a V- notched weir at BSC using a HOBO, U20-001-04, pressure data logger that recorded water pressure. Barometric pressure changes were recorded in the air using a CR800 Campbell data logger. The pressure time series in the water was converted to water level by subtracting the values of air pressure.

Pressure time series data were used to calculate the head height of the water above the water pressure logger using the following equation:

$$h = \frac{p}{\rho g}$$

where h is the height of the water surface in meters, p is the pressure in kPa, ρ is the density of the water, 1000 kg/m^3 and g is the acceleration of gravity, 9.18 m/s^2 . Head height is then used to create a relationship with stage of the stream. With a continuous record for water height, the correlation between the logger and the 12 spot measurements for stage were then used to calculate the discharge of the stream, this also incorporated normalizing both the zero point of the weir and the head height of the water. The following weir formula was used to calculate discharge,

$$Q = 4.28C \left(\tan \frac{\theta}{2} \right) \left(\frac{h}{k} \right)^{5/2},$$

where Q is the discharge in m^3/s , and θ is the angle of the weir, 90° . C and k are constants of 0.5779 and 0.002903 (USBR, 2007). Discharge for BSC was also estimated using data from a USGS station on Little Sugar creek (USGS 07188838) and the established relationship by Knierim et al. (2015).

2. 3. *E. coli*

Water samples from each site were collected in 500 mL Nalgene bottles. Of those 500 mL, 100 mL were used for *E. coli* detection and enumeration using the Standard Method 9223B IDEXX Quanti-tray® 2000 system with Colilert™ reagent to determine the Most Probable Number (MPN). A negative control of 100 mL 0.1% peptone was analyzed for each batch of samples using the same method.

2. 4. Dye Trace

The approximate recharge basin has been estimated by Knierim et al (2013); however, we conducted a dye trace to verify and further constrain the recharge area. A sinkhole, located in the hypothesized recharge area, served as the injection point for the dye trace test. Five springs, within a two-mile radius, were also monitored for possible connections between the sinkhole and the springs.

Fluorescent dye, fluorescein, was utilized in the tracer experiment. Fluorescein was chosen for its non-absorptive properties as well as its low detection limit. The mass of dye used was 55.5g. This small quantity provided sufficient dye for detection while limiting the chances of visible water coloration. Before injection, 20 gallons were used to wet the sinkhole. After injection, dye was pushed through the system by adding approximately 500 gallons of water. Six sites were monitored until detection was verified. Charcoal packets were used to collect dye for analysis using a spectrofluorophotometer, Shimadzu RF-5301. Charcoal packets were eluted prior to fluorometric analysis. The elution process extracts dye from the charcoal packets using 70% isopropyl alcohol and potassium hydroxide. Approximately 30 mL of 70% isopropyl alcohol was poured into a 2-oz. plastic cup. Potassium hydroxide pellets were then added and mixed into the solution. The solution separates with the supernatant then being used to elute the dye. A field

fluorometer was placed in BSC for collection of time series data of dye concentration within the cave stream. Fluorescence was measured every 15 minutes from February 26, 2017 to April 4, 2017.

2. 5. Metagenomics

For total genomic DNA (gDNA) filtration and extraction, 200 mL of water for each sample was filtered through a 0.2 μm , 47 mm Supor-200 filter membrane to capture total bacterial cells.

Filter membranes were placed at -80°C in 500 μL of guanidine isothiocyanate buffer. The total genomic DNA (gDNA) was extracted from prepared filters using the Fast DNA Spin Kit for Soil (MP Biomedicals). For sediment the gDNA was extracted as described by Gomes et al. (2007), where 5 g of sediment was mixed for 30 minutes in a flask containing 5 g of 4-mm sterile glass beads, 45 mL of 0.1% Tween 80, and 0.1% sodium pyrophosphate extraction solution. 15 mL of the supernatant was centrifuged at 5000 rpm for 10 minutes. The pellet was resuspended in 100% ethanol to a final volume of 1.5 mL, frozen at -80°C followed by gDNA extraction. Total gDNA was then quantified using a NanoDrop UV spectrophotometer.

The 16S rRNA was then amplified to build a sequencing library targeting the V4 region of the 16S rRNA using the generic primers, including 16S Amplicon PCR Forward Primer, 16S Amplicon PCR Reverse Primer, and KAPA HiFi HotStart Ready Mix (2x) PCR Kit (KapaBiosystems). Primers are short strands of DNA that are required to catalyze the replication process of DNA. The appropriate primers were used to amplify the V4 region from the bacterial 16s rRNA gene (Kozich et al., 2013). The primers were prepared at a stock concentration of 100 μM , with a working concentration of 10 μM .

16S rRNA amplicons were then generated using a high-fidelity polymerase, which helps produce accurate replications of the desired gene. The purified amplicon's final concentration is then

determined using polymerase chain reaction (PCR), which clones the 16s rRNA region. The total reaction volume was 50 μL (Table 1). Gel electrophoresis was then used to verify that the size of the amplicon was approximately 500bp. 5 μL of the PCR DNA product was visualized by ethidium bromide staining on a 0.8% agarose gel (Promega cat # V3121, Madison, WI) electrophoresis (VWR MiniGel II) 90 V for 1 hour. The remaining 45 μL reaction volume of the PCR DNA products was stored at -20°C for sequencing. The DNA for each sample was readjusted to 10 ng/ μL before being submitted for sequencing as a pooled library.

Table 1: Reagents used in the total reaction for 16S rRNA amplification

REAGENTS	Volume (μL)
PCR- grade water	10
DNA (5 ng/ μl)	5
Amplicon PCR Forward Primer 1 μM	5
Amplicon PCR Reverse Primer 1 μM	5
2x KAPA HiFi HotStart ReadyMix	25
Total	50

A 20 nM of pooled library and 20 nM of PhiX control v3 (Illumina) were mixed with 0.2 N fresh NaOH and HT1 buffer (Illumina) to produce the final concentration at 12 pM each. The two were mixed to give a 95% 16S rRNA amplicon gene library and a 5% PhiX control v3. A total of 600 μL was loaded on a MiSeq1 v2 (500 cycle) reagent cartridge for sequencing. All the sequencing procedures were monitored through the Illumina BaseSpace[®] website. The

Illumina MiSeq sequencer allows for large numbers of high quality samples to be analyzed at once (Kozich et al., 2013). MiSeq assembles the cloned sequences and prepares them for taxonomic classification (Illumina, 2015). The files acquired from Illumina BaseSpace[®] were then processed through the program QIIME pipeline (version 2.0), a biodiversity analysis program, to calculate the alpha and beta diversity of the taxa in the samples. Taxa were assigned an operational taxonomic unit (OTU). These results can then be viewed through an html page. For alpha diversity, rarefaction was used. Rarefaction calculates the diversity metrics, which reflect the diversity within a sample based on the abundance of various taxa within a community (Kuczynski et al., 2012). The diversity metrics consist of Chao1, phylogenetic distance (PD_whole_tree), and observed OTUs. Chao1 estimates the taxa richness whereas PD_whole_tree is a phylogenetic metric to measure dissimilarities in the phylogenetic tree. The observed OTUs is the count of the different taxa found in the sample. Rarefaction curves plot OTUs as a function of number of individuals sampled (Wooley et al., 2010). Rarefaction can also be used to observe the different number of OTUs for a given number of sequences in a sample. This is helpful when comparing two or more different samples with different numbers of sequences. Relative abundance is used to show taxa for each sample. The α -diversity can be calculated using a Shannon's diversity index (H'_α),

$$H'_\alpha = -\sum_{i=1}^S p_i \ln p_i,$$

$$p_i = \frac{n_i}{N}$$

where p_i is the relative abundance of each OTU and N is the total number of individuals. n_i is the number of clones in each OTU. Lastly, S is the total number of OTUs. $H_{\alpha max} = \ln S$.

Beta diversity evaluates the difference between microbial communities. The dissimilarities between every pair of community samples are calculated using a square distance matrix. Principle coordinate analysis (PCoA) was used to indicate the phylogenetic distance between samples in 2-dimensional and 3-dimensional diagrams. Unique Fraction Metric (UniFrac) was used to consider both qualitative and quantitative phylogenetic beta diversity in order to assess differences in the overall bacterial community structure. UniFrac considers the distance of the branch length shared by communities in a phylogenetic tree of the sampled communities being compared (Lozupone et al., 2007). The unweighted UniFrac metric analyzes the community membership by the presence or absence of communities within the samples. The weighted UniFrac metric analyzes the community structure by weighting the branch length by the relative abundance of the communities within the samples. The environmental factors surrounding the samples are used for this comparison, since the metric is comparing communities by the given factors. Factors analyzed include sample type (water versus sediment), sample site, and sample location (stream versus cave).

2. 6. Statistical Analysis

All statistical analyses were performed using Anaconda Python3.6 for log-transformed values of *E. coli* concentrations. Nonparametric tests were used to compare the proportion of log *E. coli* in water samples under various physical water quality conditions as well as for each sampling location, type of sampling event (base flow vs. high flow), and season (Knierem et al., 2015). Level of significance for nonparametric test was set to $p < 0.05$.

3. Results

3. 1. Discharge

During the study period, BSC was had a mean discharge of 0.0211 CMS according to the local stream gage relationship. The mean discharge estimated from LSC (Q_{est}) was 0.0268 CMS with a 9.24% difference between the loggers. Both the estimated discharge and the observed discharge are used in this study. Estimated discharge is used to keep consistent values with the research that has been conducted in the past.

3. 2. Hydrologic parameters

Water quality parameters at each site were analyzed to characterize each location. Temperature and Specific Conductance were measured at both the surface and cave sites. Dissolved oxygen (DO), turbidity, and pH were also measured in LSC.

Table 2: Temperature and Specific Conductance data collected from both Little Sugar Creek and Blowing Spring Cave.

Table 2		LSC1	LSC2	LSC3	LSC4	BS1	BS2	BS3
Temperature (°C)	Mean	17.00	19.58	18.63	18.99	15.30	15.10	15.10
	Median	16.48	19.94	18.83	19.29	15.30	15.10	15.10
	Max	24.67	25.85	24.77	24.59	15.80	15.20	15.20
	Min	13.12	14.02	13.69	13.91	14.90	15.00	14.90
Specific Conductance ($\mu S/cm$)	Mean	360	360	370	360	210.11	202.86	223.45
	Median	360	360	370	360	215.00	202.75	231.00
	Max	380	400	400	390	242.00	243.00	298.00
	Min	330	260	290	310	162.10	164.40	169.60

Temperature dynamics at the cave and creek sites were distinct from one another (Table 2), with temperature at LSC varying between 13 to 26°C. Temperature at BSC was nearly constant at 15°C. As for Specific Conductance, BSC showed a distribution of measurements between 160 and 300 $\mu S/cm$. Specific Conductance at LSC varied between 260 and 400 $\mu S/cm$ (Table 2).

For the parameters specific to LSC, levels of DO were below 2.0 mg/L with higher values found at LSC1 and lower at the later sites Table 2. Values of temperature, and turbidity, were lower at LSC1 than the other Little Sugar Creek sites.

The median *E. coli* value for the sites at BSC was 40.15 MPN/mL with samples never exceeding the secondary contact limit, 2050 MPN/ 100 mL. Most sample concentrations were less than the primary contact limit, 410 MPN/ 100 mL. The values ranged from 0.9 to 920.8 MPN/100mL for March 2016 to December 2016. Though the ranges observed at the three cave sites were similar, BS2 had the highest concentrations and BS3 had the lowest (Figure 2).

Table 3: Dissolved oxygen and turbidity data collected at Little Sugar Creek.

Table 3		LSC1	LSC2	LSC3	LSC4
Dissolved Oxygen (ppm)	Mean	1.70	1.29	1.23	1.34
	Median	1.47	1.23	1.31	1.25
	Max	3.14	1.99	1.76	1.96
	Min	1.02	0.82	0.82	0.74
Turbidity (NTU)	Mean	10.68	27.90	23.60	17.99
	Median	11.50	26.70	23.05	18.10
	Max	17.10	41.90	44.30	33.10
	Min	0.39	11.20	3.26	8.10

3. 3. *E. coli*

E. coli data from each sample were plotted on the hydrograph for BSC to observe any qualitative changes throughout the sampling period (Figure 3). The data from the current study were added to the data set by Knierim et al. (2015) and a linear regression was conducted, which indicated that *E. coli* increased as discharge increased (Figure 4). The best fit line determined by the linear regression for the total data set was

$$\text{Log}_{10}(E. coli + 1) = 1.11\text{Log}_{10}(Q_{est}) + 3.45,$$

$$r^2 = 0.4641, p < 4.5E^{-8}$$

These slope and intercept values are slightly larger than determined by the previous study.

$$\text{Log}_{10}(E. coli + 1) = 1.10\text{Log}_{10}(Q_{est}) + 3.36,$$

$$r^2 = 0.5500, p < 0.0001$$

E. coli data were compared against other water quality parameters but was not found to have any other correlations that were statistically significant.

Two storm events were collected May 25-27, 2016 (Figure 5) and October 5-7, 2016 (Figure 6).

The storm event in May had a peak discharge of 0.057 CMS, while the October storm event had a peak discharge of 0.014 CMS within 24 hours. During the May storm event, water samples were collected one, two and three days after the peak discharge of the storm. Samples on the first two days were collected at the mouth of the cave due to high water levels restricting entrance into BSC. The samples on the third day were collected at all three sites designated within the cave. On the third day BS2 was recorded to have the highest *E. coli* concentration for this storm event, and BS3 had the lowest concentration during the event (Figure 5). The first two days show an increase in *E. coli* concentration as discharge recedes. For the storm event in October, discharge increases as several storm pulses occur throughout the storm sampling period, however, *E. coli* concentrations decrease the following two days (Figure 6).

For LSC, the median *E. coli* concentration was 120.4 MPN/100mL, with values ranging from 4.1 to 2419.6 MPN/100mL. These values exceed both contact limits at various times of the year. LSC2 was recorded to have the highest levels of *E. coli*, while LSC1 had the lowest values recorded (Figure 2).

Samples of *E. coli* were plotted on the hydrograph for LSC to observe any qualitative changes throughout the sampling period (Figure 7). The *E. coli* concentrations at LSC were not correlated with discharge or any other water quality parameters (Figure 8). At the LSC sites, samples for the May 5-7, 2016 storm event appear to have peaked prior to peak discharge, with LSC2 having the highest values and LSC1 having the lowest values (Figure 9). For October storm event, LSC2 again had the highest values and LSC1 having the lowest values (Figure 10).

3. 4. Dye Trace

Dye was injected at 11 a.m. on February 27, 2017 at the sinkhole location shown in Figure 11. Dye was detected at BSC on 7:15 a.m. on March 25, 2017. Both charcoal packets and the field fluorimeter detected the dye. The intensity measured by the fluorimeter was 0.10 AU without any background fluorescence prior the storm event. After the initial pulse, two more pulses of dye occurred on the 26th through the 28th (Figure 12). A charcoal packet placed on March 13, 2017 and collected on April 4, 2017 at Blowing Spring verified these results with positive traces of fluorescein with an intensity of 9.156 AU. The charcoal packets placed in Blowing Spring on April 4th, 2017 were washed away by a storm event. No other spring showed any sign of dyes (Figure 11).

Spatial data from Knierem et al (2013) show the most concentrated areas of on-site septic tanks within the hypothesized recharge area. The location of the injection point for the positive trace to Blowing Springs is within the area with a moderate concentration of potential septic influence (Figure 13).

3. 5. Metagenomics

3. 5. 1. General data

The metagenomics analysis was conducted on 181 samples resulting in 4,840,887 total raw counts. The raw counts account for each individual genomic sequence found in each of the samples. These values are used to calculate the relative abundance for all the individual samples. The mean raw count per sample was 28,270 with the maximum and minimum number of 58,269 and 3 counts, respectively. From these raw counts relative abundances were calculated to determine the dominant taxa in each sample (Figure 14). The dominant overall taxa were *Acinetobacter*, *Enterobacteriaceae*, *Psuedomonas*, and *Bacillaceae*. All taxa are identified at the genus level aside from *Enterobacteriaceae* and *Bacillaceae*, which are identified at the family level of taxa. *Enterobacteriaceae* is also the family for *E. coli*.

3. 5. 2. Raw Counts

Operational Taxonomic Units, OTUs, were filtered down to bacteria that showed greater than 850 counts. This was done to observe the largest microbial populations at each site in sediment and water. Sediment samples never surpassed more than 20,000 counts, and water samples ranged above 30,000 to 40,000 raw counts. However, due to different sampling and filtration methods, it is important to note that water and sediment samples cannot be compared by using the raw counts alone. Comparative analysis of sediment and water are discussed below.

According to the raw counts *Acinetobacter* and *Psuedomonas* were highly abundant in the water. *Enterobacteriaceae* and *Bacillaceae* were the most abundant in the sediment.

Enterobacteriaceae was also present at all sites. For the sediment, *Enterobacteriaceae* was the highest at BS3 and lowest at LSC3. For water samples, *Enterobacteriaceae* abundance was the highest at BS2 and lowest at LSC2. Other taxa found in the samples were *Flavobacterium*,

Comamonadaceae, *Oxalobacteraceae*, and two forms of *Clostridium*. These taxa were present in large numbers at some but not all sample sites (Figure 14).

Microbial community differences were observed at all seven sites. For sediment samples at Blowing Spring Cave, *Clostridiaceae Clostridium* was present at only two out of the three sites, BS2 and BS3. Other major taxa observed were the family *Enterobacteriaceae* and *Bacillaceae*. *Bacillaceae* was most prevalent in BS2 and least abundant in BS3. However, for *Enterobacteriaceae*, BS3 was higher in mean concentration compared to BS1 and BS2. Overall, BS2 had the highest mean concentration of microbes, with approximately 16,000 counts, while BS3 and BS1 only totaled at ~ 14,000 and ~12,000, respectively. As for the water samples at BSC, samples ranged from 32,000 to 35,000 counts with BS3 totaling at about 36,000, BS1 at 34,000 and BS2 at 3500. Taxa present at each site from most to least abundant consisted of *Acinetobacter*, *Psuedomonas*, *Enterobacteriaceae*, and *Oxalobacteriaceae*. *Flavobacterium* was only present at BS1 and BS2.

The sediment samples at Little Sugar Creek totaled around 20,000 counts for each site. Taxa for this location were the same as BSC with the addition of *Lachnospiraceae Clostridium* found in LSC2. LSC4 did not have *Clostridiaceae Clostridium* at a level above 850 counts. The lowest level of *Enterobacteriaceae* was found at LSC3. For *Bacillaceae*, the most abundant site was LSC4. The water samples contained levels between 30,000 and 40,000 counts. Taxa were consistent with those found in BSC excluding *Oxalobacteriaceae* but including *Comamonadaceae* at sites LSC2 – 4.

3. 5. 3. Relative abundance

Relative abundance represents the major populations in relation to one another at each site. Taxa that represented more than 5% of the population were plotted in a bar graph separated by site and

type (Figure 15). The sites that showed greater diversity in sediment were BS3, LSC1, and LSC3. LSC3 showed the greatest diversity in the water as well. Note that the percentages do not always total to 100% due to false positives that were produced by chimeras—genes formed by the amalgamation of gene fragments sequenced together to produce new genes (Figure 15).

3. 5. 4. Alpha diversity

The alpha diversity, which is the diversity at a given site (Kuczynski et al., 2011), was analyzed using rarefaction curves. Rarefaction also indicates how well the samples represent the environment by analyzing the curvature of the line. The flatter the line the more exhaustive the sampling was. Rarefaction curves were plotted to compare both water and sediment as well as BSC and LSC.

Sediment versus water was first analyzed by species richness using Chao1 (Figure 16). The sequences per sample, rarified to 5230.0, showed a mean chao1 value of 523 for the sediment and 1722 for the water. Therefore, the water showed a greater richness of species within itself compared to the sediment. According to the PD_whole_tree, the phylogenetic distance, showed an average value for the sediment and water of 30.7 and 80.3 respectively. Water shows a greater phylogenetic distance than the sediment (Figure 17). For the observed OTUs, the sediment had 309.942 different observed OTUs. Water contained 748.468 different observed OTUs, which is approximately 2.4 times more than the sediment samples. This also represents more diversity within the water compared to the sediment (Figure 18).

When analyzing the alpha diversity for the locations LSC and BSC, the mean Chao1 metric showed values for BSC to be 705.417 (Figure 19). The mean Chao1 metric value for LSC was 1488.399, roughly twice as rich as BSC. The PD_whole_tree values for BSC and LSC were 43.783 and 67.485, with BSC has smaller phylogenetic distances between taxa (Figure 20). BSC,

on average, had 366.960 different observed OTUs, while LSC had 670.818 different OTUs. Therefore, LSC showed more diversity within itself compared to BSC (Figure 21).

The locations were further broken down into the seven sampling sites with respect to the location. When BSC sites were specifically analyzed, according to all three metrics, the two sites BS1 and BS2 had richness levels of 641.316 and 645.937, respectively. However, BS3 was revealed to have the highest mean Chao1 value of 853.561 (Figure 22). Yet when analyzing the PD_whole_tree metric the mean values ranged 5.5 units of each other, with BS3 having the highest value at 46.645 (Figure 23). This means that the phylogenetic distances are not so different from one another. Ultimately, this is also reflected in the OTUs for each site with BS3 being the most diverse, 399.865. BS2 had the fewest OTUs at 346.815 and BS1 had 357.477(Figure 24).

The Little Sugar Creek sites when analyzed for richness showed that LSC1 was the least rich with 1138.553. LSC2 and LSC4 were 1516.140 and 1507.160, respectively. LSC3 presented the greatest species richness at 1814.011. The phylogenetic distances were in the same order from least to greatest, LSC1, LSC4, LSC2 and LSC3. The values ranged from 60.335 to 77.592. The observed OTUs displayed the same order as well with the fewest OTUs at LSC1 (575.438) and the most observed OTUs at LSC3 (811.016). Through this analysis, LSC3 is the most diverse site with the greatest richness and distance in the phylogenetic tree.

3. 5. 5. Beta diversity

The beta diversity, which is the diversity among habitats (Wooley et al., 2010), was characterized using Principal Coordinate Analysis with unweighted (uwPCoA) and weighted (wPCoA) UniFrac metrics. The uwPCoA focuses on community membership by the presence or absence of the taxa in the samples. The wPCoA focuses on community structure by weighting

the branch length through relative abundance of the lineage of the communities in the samples.

The samples' environmental factors were used in the PCoA plots to distinguish the differences of communities by these environmental factors. For this analysis 164 of the 181 samples were used.

The 17 unused samples were due to low values of raw counts produced.

In the uwPCoA analysis, the sample type (sediment versus water) contributed the greatest difference in taxa that were present or absent. Though these differences were not visually distant from each other, samples still clustered separately by sediment and water. The principal coordinates (PC1, PC2, and PC3), accounted for 25.89% of the dissimilarities. The location of the samples further divided the water and sediment samples into LSC and BSC, which represents the difference of communities present or absent within LSC and BSC. Within the water samples, BSC and LSC clustered well with some overlap. As for the sediment, BSC samples clustered tighter than the LSC samples. Overall, the uwPCoA analysis showed that the presence/absence of the taxa was primarily affected by the sample type (sediment versus water) and secondarily by the location (surface stream versus cave) (Figure 25).

The differences in the abundance of the common taxa within the samples were analyzed using wPCoA. The wPCoA results were summarized by the distribution of samples within the space of the first two principle coordinates which accounted for 61.52% of the sample dissimilarities.

Visualization of the samples in this space, divided between sediment and water, demonstrates that there are distinct differences in the abundances of microbial communities within the sediment and water (Figure 25). Though the major difference came from the type of sample collected, BSC and LSC water samples are also somewhat differentiated (Figure 26). The sediment samples for BSC displayed some clustering as did the location samples in the water, however there is some overlap with the LSC sampling, which had a broader range. Therefore, the

sediment and water are once again the primary reason for the dissimilarities of the samples. The location is the secondary contributor for the differences.

4. Discussion

4. 1. Water Quality Parameters

Water temperature measurements for each location showed how daily temperature variability and other sources of heat influenced each site. The cave and stream sites were quite different from one another with LSC showing a greater water temperature variability because of its influence by weather. On the other hand, the cave environment is relatively insulated from external weather, and therefore water temperatures in BSC were approximately constant. From the individual sites at LSC, LSC1's lower water temperatures could indicate mixing of different streams or even mixing with spring waters nearby, seeing how these values are close to the values in BSC.

For LSC, the Specific Conductance shows a large quantity of dissolved solids in the water. This results from extended interaction between the water and limestone bedrock. Within BSC the highest levels of specific conductance were at the back of the cave, indicating precipitation along the flow path or an input of water with a lower dissolved load along the flow path.

Sediment within the stream is mobilized at high discharge, increasing turbidity throughout Little Sugar Creek. Turbidity was found to be higher within the urban area of the study site downstream of the Lake Bella Vista dam. In this area, Trout Farm Spring discharges into Little Sugar Creek. Lake Bella Vista has been subject to overtopping on multiple occasions in the past few years. This overtopping has allowed for the transport of sediment to Little Sugar Creek and accumulation on the upstream portion of the dam (Jaynes and Hermely, 2015). After the last overtopping event in December 2015 the flood gates were left opened, bringing the water level

down and allowing for a constant flow of water to move through the system, in turn releasing more sediment from within Lake Bella Vista into Little Sugar Creek.

4. 1. *E. coli*

E. coli is used as an indicator of fecal contamination. *E. coli* were present at all sites under most conditions. Discharge had a strong correlation to the concentration of *E. coli* at all locations within Blowing Springs Cave. However, *E. coli* concentrations during high discharge exceeded the maximum level of detection, 2419.5 MPN/ 100 mL. Therefore, levels of *E. coli* in BSC for large storm events are lower limits. Future studies should note to dilute storm samples to avoid maxing out readings. In addition, both BSC and LSC surpassed the primary recreational contact limit, 410 MPN/ 100 mL, during the period. LSC also passed the secondary contact limit, 2050 MPN/ 100 mL. These cases indicate that people should not be in contact during or shortly after storm events. Fecal contamination influences for the two locations are different due to the high *E. coli* concentrations of LSC2, which is upstream of the point where BSC discharges into LSC. Though BSC does have some influence of fecal contamination on Little Sugar Creek, it is not the dominant contributor for the contamination seen. Further sampling of LSC2 and the surrounding area should be considered in future studies to identify or constrain fecal contamination influences.

4. 1. 1. Blowing Spring Cave

E. coli average concentrations at the sites within BSC were 40.15 MPN/ 100 mL indicating that there is a common influence at this location with little contribution from infeeders and additional sources that are in between the sites. *E. coli* levels increased as discharge increased, suggesting that contaminants are mobilized during high flow; however, during storm events, the highest *E. coli* concentrations occurred during discharge recession, indicating that contaminated water is

arriving late in the storm event. Low concentrations during baseflow suggest that baseflow water has had sufficient time within the subsurface to remove *E. coli* from the water via transport or by bacteria attaching to the sediment. Typically, levels for BSC site lie below the primary contact limit.

4. 1. 2. Little Sugar Creek

In LSC, *E. coli* concentrations do not correlate well with discharge; however, the highest levels were observed during storm events. The lack of correlation could be due to runoff water not containing high concentrations of contaminants. Therefore, the sources of fecal contamination are potentially in the stream itself or in the waters that are slowly discharging into the creek during baseflow conditions. The first site at LSC showed a decrease in *E. coli* concentration through the storm events while the rest of the sites showed increases in concentration. The decrease can be attributed to the dilution via runoff influences. Fecal contaminants are already present in the creek, upstream from the study area. Several potential contamination sources observed between LSC1 and LSC2 are McKisic and Spanker Creeks, Lake Bella Vista, and Big Spring, which flows through and provides water to a trout farm. The concentration of *E. coli* within each system is unknown. Urbanization around the creek may also contribute to the increase of the *E. coli* concentration in Little Sugar Creek, other potential sources could be from the wildlife surrounding these areas such as geese.

4. 2. Dye Trace

The qualitative dye trace verified a connection between a sinkhole located within the hypothesized recharge area and BSC. This also verifies a surface to subsurface connection within an area of moderately dense septic systems. None of the other springs monitored showed signs of dye. Future studies should look to expand traces outside of the hypothesized recharge area to

better constrain the area and characterize the groundwater hydrology of Bella Vista. Lastly, travel time for the dye was slower than expected, with dye being pushed through the system only after a series of storm events. The injected water may have taken some time to move through sediment filled pathways in the vadose zone before passing into a conduit carrying active flow. Alternatively, the sinkhole could be connected to an abandoned channel that only transports water during large storm events. If this is indicative of the hypothesized recharge area, the increase of *E. coli* concentrations could be caused by the connectivity of septic tanks in the vadose zone to the main conduit during large storm events.

4. 3. Metagenomics

Data for the metagenomic sequencing gave insight into the diversity of the microbial communities with respect to sample location, type, and site. Ultimately, the greatest differences between the bacterial communities were those between the sediment and water, where communities were distinctly different in terms of community membership (presence/absence) and structure (relative abundance). These key differences are likely due to the natural habitat of specific microorganisms, which are dependent on the favorable conditions of the environment and nutrients that are provided in these environments. The major bacterial communities identified in this study were the genera *Acinetobacter*, *Pseudomonas*, *Flavobacterium*, *Clostridium* and two unclassified genera from the families *Bacillaceae* and *Enterobacteriaceae*. The known genera are communities that can commonly be found in freshwater and sediment as well as the families of the two unclassified genera (Krieg and Holt, 1984; Vaz-Moreira et al., 2011). *Acinetobacter* and *Pseudomonas* have been classified as dominant taxa in urban sequence signatures (Fisher et al., 2015). These taxa align with the study site, which is found in the city of Bella Vista. Other studies have shown similar results with of presence of dominant taxa at

different taxonomic levels, ranging from phylum to genus (Toze, 1999; Vaz-Moreira et al., 2011; McLellan et al., 2015; Uyaguari-Diaz et al., 2016). The level of identification has become more detailed over time. This study alongside other current studies contribute to the increase in understanding of environmental systems through the use of metagenomic analysis.

The abundance of these taxa in the soil and water are different, which can be attributed to the nutrients available and the microbe's motility, allowing them to move and find nutrients.

Therefore, waterborne bacteria that have flagella are more likely to be found within the water column, while sediment-borne bacteria would likely be nonmotile or have gliding motility, where bacteria move over surfaces without the aid of a flagella. However, some bacteria can be found in both sediment and water, and bacteria that have been considered waterborne are found in the sediment at a higher abundance. This can be seen for unclassified genus in

Enterobacteriaceae, the family of *E. coli*. This study showed presence of *Enterobacteriaceae* in the sediment. This could indicate that this genus is becoming subject to sedimentation and resuspension, therefore sediment acts as a temporary storage for microbial communities, which could include fecal communities. In the present study, quantification of *E. coli* was only performed on water samples, but studies have shown that death rates of *E. coli* in the sediment are lower than those in water (Burton et al., 1987; Sherer et al., 1992; Jamieson et al., 2002; Craig et al., 2004; Garzio-Hadzick et al., 2010). This bacterial survival could then add back into a given system, such as Blowing Springs Cave where *E. coli* increases as discharge increases, because storm events can resuspend sediment and the bacteria within the sediment. Overall, the survival of *E. coli* in sediment would be enabled by favorable conditions. Conditions within the sediment are better-suited for survival of *E. coli* than within the water. For instance, the impacts of temperature, UV radiation, dissolved solids, and competition for nutrients are lessened within

the sediment when compared to water. Burton et al. (1987) found *E. coli* and *Salmonella* Newport to have longer survival rates within sediment with 25% clay compared to sand. Both LSC and BSC had higher compositions of clays than sands.

4. 3. 1. Blowing Spring Cave

The location of the samples was the second greatest contributor to the differences in microbial communities. The communities within the water were similar for BSC and LSC. Both sites had the same dominant taxa and differed only in the abundance of each taxa. Blowing Spring Cave is a cool and dark environment, where there is no solar energy input. This could potentially limit the amount of nutrients for microbes to feed on, however it also provides protection for the microbes. Nutrients are likely being introduced into the system from surface to subsurface interaction. This can be seen by the result of the dye trace for BSC. Blowing Spring Cave was less diverse than Little Sugar Creek but was more abundant in certain taxa, unclassified genera in *Enterobacteriaceae* and *Bacillaceae* in both sediment and water. The lack of diversity could be from the lack of nutrients and from a limited number of taxa entering the system from surface influences. With this lack of diversity, the dominant taxa that are entering the system are either surviving for a longer duration or are input at a high concentration.

Blowing Springs exhibited significantly higher concentrations of *Acinetobacter*, *Bacillaceae* and *Oxalobacteriaceae*. The additional presence of other bacteria could come from a unique source which infiltrates into Blowing Springs Cave. Lastly, *Enterobacteriaceae* average abundance was higher at the back of the cave in the sediment, but for the water, concentrations were reflective of those seen in the *E. coli* data, where BS2 had the highest average of *E. coli*. These similar concentration trend could indicate that most of the *Enterobacteriaceae* in Blowing Springs is *E. coli*.

4. 3. 2. Little Sugar Creek

Little Sugar Creek is influenced by seasonal changes and solar energy, allowing for the presence of nutrients on which larger microbial communities can feed. Little Sugar Creek showed a higher concentration of *Flavobacterium* compared to Blowing Springs Cave, which is due to the multiple types of nutrients that can be metabolized by the members of this genus (Bergey et al., 1923; Krieg and Holt, 1984; Madigan and Martinko, 2005). Little Sugar Creek would allow for various forms of nutrients to be present. This would explain the greater diversity observed in the metagenomic data. It is also likely that Little Sugar Creek has multiple influences where bacteria can be introduced to the creek via runoff. These different influences can potentially be attributed to urbanization and fecal contamination, which can be represented by the presence of *Lachnospiraceae Clostridium*. *Lachnospiraceae* alongside *Bacteroidaceae*, *Ruminococcaceae*, *Porphyromonadaceae*, *Rikenellaceae*, and *Prevotellaceae* have been shown to reflect human fecal contamination (McLellan et al., 2013; Fisher, Eren, et al., 2015; Newton et al., 2015). Newton et al. (2015) found that 97% of human fecal oligotypes were found in municipal sewage, with only 15% of the total sewage sequence reads originating from human fecal matter. Future studies should seek to exam the presence of these families at Little Sugar Creek and Blowing Springs Cave. Expected average for environmental sites would be no greater than 15% of the total abundance of the given sample at each site. LSC2 had high levels of *Lachnospiraceae Clostridium*, which could show there is human influence being introduced to this location (Fisher, Eren, et al., 2015; Newton et al., 2015).

The increasing presence of *Comamonadaceae* for the four sites moving downstream could be tied to this increase of urbanization through runoff. Overall the interaction between the microbial communities from the cave and creek can reflect those seen in the conceptual model of surface-

groundwater interaction (Figure 27). Precipitation can runoff directly into the stream or be infiltrated into karst systems, while carrying along bacteria found in soils and plants. For karst systems, these waters can also reconnect abandoned channels that are influenced by septic tanks, allowing enteric bacteria to also be added to the system. As storm waters recede, some bacteria can become subject to sedimentation, where the microbes can survive for a prolonged period. Over time, storm pulses introduce more bacteria into the cave systems, while also resuspending the bacteria found in the sediment of the cave stream. This storm water is then discharged out of the spring into the proximal stream where it can add to the present communities in both the water and sediment. Pathogenic bacteria, such as *E. coli* and *Pseudomonas aeruginosa*, could be included in this process. These pathogens compromise the quality of the water and can make it unsuitable for drinking or interaction with these waters through recreational activities.

5. Conclusion

The purpose of this study was to compare the microbial diversity within two different locations and analyze key differences using metagenomics and FIB, *E. coli*. The research revealed that the major distinction in biodiversity was due to the water and sediment. This was independent of site whether inside the cave or in the surface stream. Water samples did show a dissimilarity between the cave and surface stream locations. Sediment sample showed a tighter clustering for Blowing Springs Cave. These dissimilarities in the PCoA results are caused by the narrow range of values found in the cave compared to the stream, indicating that the cave is more of a closed system. Other key distinctions presented in this study were the concentrations of *E. coli* at the various sites within the cave and surface stream. These concentrations were different for the cave and stream. The changes in *E. coli* levels along the flow path of Little Sugar Creek were greater than in the cave, potentially a result of sources located within the flow path. *E. coli* concentrations in

Blowing Spring Cave were relatively uniform, indicating that the major source of contamination is upstream of the accessible portion of the karst conduit system. This is also supported by the *E. coli* relationship with discharge, which shows that *E. coli* levels increase as discharge increases. Using a dye tracer test, a sinkhole within the hypothesized recharge area was verified to connect to Blowing Spring. Future research should further constrain this recharge area. The dye trace confirms the connection between the surface and subsurface recharge, where potentially leaky septic systems could leach into Blowing Spring Cave and/or other cave systems in the area.

The presence of fecal contamination within Blowing Springs Cave was shown to be correlated to the increase of discharge within the cave, confirming the study by Knierim et al. (2015). The source of the fecal contamination is unknown, though it is hypothesized to originate from septic tanks within the recharge area. However, the additional increase of *E. coli* with discharge could also be from the contaminant being deposited in the sediment only to be resuspended during storm events, allowing sediment to serve as a storage location for bacteria and potentially other microbes.

One limitation of this work related to spot sampling of only two storm events. Since not every rain event produced the same result, it was difficult to see an exact response of contamination to changes in discharge. Furthermore, Blowing Spring Cave could not be entered during large storm events, making it impossible to examine changes along the flow path during storms. Future studies could benefit from placement of an auto sampler within this system for a shorter sampling period. Another limitation came from the analysis of *E. coli*, which maxed out for multiple samples in Little Sugar Creek. Future studies should dilute samples prior to analysis to avoid this and provide better insight into the relationship with discharge.

Overall this study has provided insight to the dynamics of microbial communities in this karst region, where a mixture of rural and urban influences can affect freshwater sources. This research shows the interconnectedness of the surface and groundwater systems by showing that the major microbiology is same but differ in abundance. The overlying differences are determined by the protection of sediment and the nutrients available in the individual stream and cave system. These variables also contribute to the understanding of sites where *E. coli* concentrations were the highest, by giving insight as to what the other types of microbes are there. However, our results do not specifically indicate the direct sources of contamination. It is a possibility that septic tanks are contributing to the presence of *E. coli* in Blowing Springs Cave due to the high transmissivity in karst. Since most of the southcentral and eastern United States is represented by a mixture of rural and urban settings, it can be assumed that these findings may be similar to the biodiversity of other cave and surface stream systems in this region.

6. References

- Aagaard, K., Riehle, K., Ma, J., Segata, N., Mistretta, T.A., Coarfa, C., Raza, S., Rosenbaum, S., Van den Veyver, I., Milosavljevic, A. and Gevers, D., 2012. A metagenomic approach to characterization of the vaginal microbiome signature in pregnancy. *PloS one*, 7(6), p.e36466.
- ACA-8-4-202, 2011, Arkansas regulations no. 2: regulations establishing water quality standards for surface waters of the State of Arkansas, p 124
- Amraoui, F., M. Razack, and L. Bouchaou, 2003, Turbidity dynamics in karstic systems. Example of Ribaa and Bittit springs in the Middle Atlas (Morocco): *Hydrological Sciences Journal*, v. 48, no. 6, p. 971–984.
- Antunes, L., P. Visca, and K. J. Towner, 2014, *Acinetobacter baumannii*: evolution of a global pathogen: *Pathogens and disease*, v. 71, no. 3, p. 292–301.
- APCEC, 2016, REGULATION NO. 2 REGULATION ESTABLISHING WATER QUALITY STANDARDS FOR SURFACE WATERS OF THE STATE OF ARKANSAS: Arkansas Pollution Control and Ecology Commission, p. 135.
- APHA, 1998, *Standard Methods for the Examination of Water and Wastewater*: American Public Health Association (APHA), Washington, DC
- Ashbolt, N. J., 2004, Microbial contamination of drinking water and disease outcomes in developing regions: *Toxicology*, v. 198, no. 1–3, p. 229–238.
- Bain, R., R. Cronk, R. Hossain, S. Bonjour, K. Onda, J. Wright, H. Yang, T. Slaymaker, P. Hunter, and A. Prüss-Ustün, 2014, Global assessment of exposure to faecal contamination through drinking water based on a systematic review: *Tropical Medicine & International Health*, v. 19, no. 8, p. 917–927.
- Baldani, J. I., L. Rouws, L. M. Cruz, F. L. Olivares, M. Schmid, and A. Hartmann, 2014, The family oxalobacteraceae, in *The prokaryotes*: Springer, p. 919–974.
- Baquero, F., J.-L. Martínez, and R. Cantón, 2008, Antibiotics and antibiotic resistance in water environments: *Current opinion in biotechnology*, v. 19, no. 3, p. 260–265.
- Bergey, D. H., F. C. Harrison, R. S. Breed, B. W. Hammer, and F. M. Huntoon, 1923, Genus II. *Flavobacterium* gen. nov: *Bergey's Manual of Determinative Bacteriology*, v. 97, p. 117.
- Bernardet, J.-F., and J. P. Bowman, 2015, *Flavobacterium*, in *Bergey's Manual of Systematics of Archaea and Bacteria*: American Cancer Society, p. 1–75, doi:10.1002/9781118960608.gbm00312.

- Black, L. E., G. M. Brion, and S. J. Freitas, 2007, Multivariate logistic regression for predicting total culturable virus presence at the intake of a potable-water treatment plant: Novel application of the atypical coliform/total coliform ratio: *Applied and environmental microbiology*, v. 73, no. 12, p. 3965–3974.
- Bouchaou, L., A. Mangin, and P. Chauve, 2002, Turbidity mechanism of water from a karstic spring: example of the Ain Asserdoune spring (Beni Mellal Atlas, Morocco): *Journal of hydrology*, v. 265, no. 1–4, p. 34–42.
- Boyer, D. G., and G. C. Pasquarell, 1999, Agricultural land use impacts on bacterial water quality in a karst groundwater aquifer: *JAWRA Journal of the American Water Resources Association*, v. 35, no. 2, p. 291–300.
- Brahana, J.V., 1997. Rationale and methodology for approximating spring-basin boundaries in the mantled karst terrane of the Springfield Plateau, northwestern Arkansas. *The engineering geology and hydrogeology of karst terranes*, pp.77-82.
- Brenner, D. J., 1984, *Enterobacteriaceae*: Bergey's manual of systematic bacteriology, v. 1, p. 353–361.
- Brion, G.M., Mao, H.H. and Lingireddy, S., 2000. New approach to use of total coliform test for watershed management. *Water science and technology*, 42(1-2), pp.65-69.
- Burns, D.A., Boyer, E.W., Elliott, E.M. and Kendall, C., 2009. Sources and transformations of nitrate from streams draining varying land uses: evidence from dual isotope analysis. *Journal of Environmental Quality*, 38(3), pp.1149-1159.
- Burton, G. A., D. Gunnison, and G. R. Lanza, 1987, Survival of pathogenic bacteria in various freshwater sediments.: *Applied and Environmental Microbiology*, v. 53, no. 4, p. 633–638.
- City of Bella Vista, Arkansas (2009) Ordinance no. 2009-03: city of Bella Vista, Arkansas. Connect Sewer 2009-03:2. <http://cityofbellavista.com/administration/ordinances.htm>. Accessed 7 April 2014
- Covington, M.D., Luhmann, A.J., Gabrovšek, F., Saar, M.O. and Wicks, C.M., 2011. Mechanisms of heat exchange between water and rock in karst conduits. *Water Resources Research*, 47(10).
- Craig, D. L., H. J. Fallowfield, and N. J. Cromar, 2004, Use of microcosms to determine persistence of *Escherichia coli* in recreational coastal water and sediment and validation with in situ measurements: *Journal of Applied Microbiology*, v. 96, no. 5, p. 922–930.
- Davis, J.C. and Sampson, R.J., 1986. *Statistics and data analysis in geology* (Vol. 646). New York et al.: Wiley.

- Davis, R.K., S. Hamilton, and J. Van Brahana. 2005. Escherichia coli survival in mantled karst springs and streams, Northwest Arkansas Ozarks, USA. 2005. Journal of the American Water Resources Association. 41(6):1279-1287.
- Davis, S.N., Thompson, G.M., Bentley, H.W. and Stiles, G., 1980. Ground-Water Tracers—A Short Review. Groundwater, 18(1), pp.14-23.
- De Mandal, S., Panda, A.K., Lalnunmawii, E., Bisht, S.S. and Kumar, N.S., 2015. Illumina-based analysis of bacterial community in Khuangcherapuk cave of Mizoram, Northeast India. Genomics data, 5, pp.13-14.
- Döll, P., H. Hoffmann-Dobrev, F. T. Portmann, S. Siebert, A. Eicker, M. Rodell, G. Strassberg, and B. R. Scanlon, 2012, Impact of water withdrawals from groundwater and surface water on continental water storage variations: Journal of Geodynamics, v. 59, no. 60, p. 143–156.
- Fetter, C.W., 2018. *Applied hydrogeology*. Waveland Press.
- Field, M.S., 1999. The QTRACER program for tracer-breakthrough curve analysis for karst and fractured-rock aquifers (Vol. 98). National Center for Environmental Assessment-- Washington Office, Office of Research and Development, US Environmental Protection Agency.
- Fisher, J. C., A. M. Eren, H. C. Green, O. C. Shanks, H. G. Morrison, J. H. Vineis, M. L. Sogin, and S. L. McLellan, 2015, Comparison of sewage and animal fecal microbiomes by using oligotyping reveals potential human fecal indicators in multiple taxonomic groups: Applied and environmental microbiology, v. 81, no. 20, p. 7023–7033.
- Fisher, J. C., R. J. Newton, D. K. Dila, and S. L. McLellan, 2015, Urban microbial ecology of a freshwater estuary of Lake Michigan: Elementa (Washington, DC), v. 3.
- Fite GC (1993) From vision to reality: a history of Bella Vista Village, 1915–1993. RoArk Printing, Rogers
- Ford, D.C. and P. Williams. 2007. Karst Hydrogeology and Geomorphology, John Wiley and Sons: West Sussex, England.
- Garzio-Hadzick, A., D. R. Shelton, R. L. Hill, Y. A. Pachepsky, A. K. Guber, and R. Rowland, 2010, Survival of manure-borne E. coli in streambed sediment: effects of temperature and sediment properties: water research, v. 44, no. 9, p. 2753–2762.
- Gomes, N.C.M., L.R. Borges, R. Paranhos, F.M. Pinto, E. Krögerrecklenfort, L.C.S. Mendonça-Hagler, and K. Smalla. 2007. Diversity of ndo genes in mangrove sediments exposed to different sources of polycyclic aromatic hydrocarbon pollution. Applied and Environmental Microbiology. 73(22):7392-7399.

- Göppert, N. and Goldscheider, N., 2008. Solute and Colloid Transport in Karst Conduits under Low-and High-Flow Conditions. *Ground water*, 46(1), pp.61-68.
- Graening, G.O. and A.V. Brown. 2003. Ecosystem dynamics and pollution effects in an Ozark cave stream. *Journal of the American Water Resources Association*. 39(6):1497-1507.
- Heinz, B., S. Birk, R. Liedl, T. Geyer, K.L. Straub, J. Andresen, K. Bester, and A. Kappler. 2009. Water quality deterioration at a karst spring (Gallusquelle, Germany) due to combined sewer overflow: evidence of bacterial and micro-pollutant contamination. *Environmental Geology*. 57:797-808.
- Hugh-Jones, M., and J. Blackburn, 2009, *The ecology of Bacillus anthracis: Molecular aspects of medicine*, v. 30, no. 6, p. 356–367.
- Illumina. 2015. 16S Metagenomics Sequencing Library Preparation Protocol. Part# 15044223 Rev.B. Available at: Kassim, I., and C. G. Ray, 2004, *Sherris Medical Microbiology*: McGraw Hill, v. 9, p. 8385–8529.
- Jamieson, R. C., R. J. Gordon, K. E. Sharples, G. W. Stratton, and A. Madani, 2002, Movement and persistence of fecal bacteria in agricultural soils and subsurface drainage water: A review.: *Canadian Biosystems Engineering*, v. 44, no. 1, p. 1–9.
- Jamieson, R., Joy, D.M., Lee, H., Kostaschuk, R. and Gordon, R., 2005. Transport and deposition of sediment-associated *Escherichia coli* in natural streams. *Water Research*, 39(12), pp.2665-2675.
- Jaynes, K., and A. Hermely, 2015, *City of Bentonville Improvements to Lake Bella Vista Dam, Draft Environmental Assessment FEMA-1975-DR-AR / PW1562*: Washington, DC, Federal Emergency Management Agency Department of Homeland Security, 61 p.
- Juni, E., 1984, *Acinetobacter*: *Bergey's Manual of Systematics of Archaea and Bacteria*.
- Kassim, I., and C. G. Ray, 2004, *Sherris Medical Microbiology*: McGraw Hill, v. 9, p. 8385–8529.
- Kaushal, S. S., and K. T. Belt, 2012, The urban watershed continuum: evolving spatial and temporal dimensions: *Urban Ecosystems*, v. 15, no. 2, p. 409–435.
- Kendall, C., 1998, Tracing nitrogen sources and cycling in catchments, in *Isotope tracers in catchment hydrology*: Elsevier, p. 519–576.
- Klindworth, A., E. Pruesse, T. Schweer, J. Pipiles, C. Quast, M. Horn, and F.O. Glöckner. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*. 41(1):e1

- Knierim, K.J., P.D. Hays, and D. Bowman. 2015. Quantifying the variability in *Escherichia coli* (*E. coli*) throughout storm events at a karst spring in northwestern Arkansas, United States. *Environmental Earth Sciences*. 74:4607-4623.
- Knierim, K.J., Pollock, E. and Hays, P.D., 2013. Using isotopes of dissolved inorganic carbon species and water to separate sources of recharge in a cave spring, northwestern Arkansas, USA Blowing Spring Cave. *Acta Carsologica*, 42(2-3).
- Koeleman, J. G., M. W. van der Bijl, J. Stoof, C. M. Vandenbroucke-Grauls, and P. H. Savelkoul, 2001, Antibiotic resistance is a major risk factor for epidemic behavior of *Acinetobacter baumannii*: *Infection Control & Hospital Epidemiology*, v. 22, no. 5, p. 284–288.
- Kolpin, D. W., E. T. Furlong, M. T. Meyer, E. M. Thurman, S. D. Zaugg, L. B. Barber, and H. T. Buxton, 2002, Pharmaceuticals, hormones, and other organic wastewater contaminants in US streams, 1999- 2000: A national reconnaissance: *Environmental science & technology*, v. 36, no. 6, p. 1202–1211.
- Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K. and Schloss, P.D., 2013. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and environmental microbiology*, 79(17), pp.5112-5120.
- Krieg, N. R., and J. G. Holt, 1984, *Bergey's manual of systematic bacteriology*, in *Bergey's manual of systematic bacteriology*: Williams & Wilkins.
- Ley, R. E., D. A. Peterson, and J. I. Gordon, 2006, Ecological and evolutionary forces shaping microbial diversity in the human intestine: *Cell*, v. 124, no. 4, p. 837–848.
- Luhmann, A.J., Covington, M.D., Alexander, S.C., Chai, S.Y., Schwartz, B.F., Groten, J.T. and Alexander, E.C., 2012. Comparing conservative and nonconservative tracers in karst and using them to estimate flow path geometry. *Journal of hydrology*, 448, pp.201-211.
- Maczulak, A., 2011, *Encyclopedia of Microbiology (Facts on File Science Library): Facts on file*.
- Madigan, M., and J. Martinko, 2005, *Brock biology of microorganisms*.(11thedn): SciELO Espana.
- Mahler, B.J., J-C. Personné, G.F. Lods, and C. Drogue. 2000. Transport of free and particulate-associated bacteria in karst. *Journal of Hydrology*. 238(3):179-193.
- Mandic-Mulec, I., P. Stefanic, and J. D. van Elsas, 2016, Ecology of *Bacillaceae*, in *The Bacterial Spore: from Molecules to Systems*: American Society of Microbiology, p. 59–85.

- McLellan, S.L., J. C Fisher, and R. J Newton, 2015, The microbiome of urban waters: *International Microbiology*, v. 18, no. 3, p. 141–149.
- McLellan, S. L., R. J. Newton, J. L. Vandewalle, O. C. Shanks, S. M. Huse, A. M. Eren, and M. L. Sogin, 2013, Sewage reflects the distribution of human faecal Lachnospiraceae: *Environmental microbiology*, v. 15, no. 8, p. 2213–2227.
- Misra, A. K., 2011, Impact of urbanization on the hydrology of Ganga Basin (India): *Water resources management*, v. 25, no. 2, p. 705–719.
- Munir, M., K. Wong, and I. Xagorarakis, 2011, Release of antibiotic resistant bacteria and genes in the effluent and biosolids of five wastewater utilities in Michigan: *Water research*, v. 45, no. 2, p. 681–693.
- Naik, P. K., J. A. Tambe, B. N. Dehury, and A. N. Tiwari, 2008, Impact of urbanization on the groundwater regime in a fast growing city in central India: *Environmental monitoring and assessment*, v. 146, no. 1–3, p. 339–373.
- Natural Resources Conservation Service. 2013 Custom soil report— Benton County. 19. <http://websoilsurvey.sc.egov.usda.gov/App/HomePage.htm>. Accessed 15 May 2012
- Newton, R. J., S. L. McLellan, D. K. Dila, J. H. Vineis, H. G. Morrison, A. M. Eren, and M. L. Sogin, 2015, Sewage reflects the microbiomes of human populations: *MBio*, v. 6, no. 2, p. e02574–14.
- Palleroni, N. J., 2015, *Pseudomonas: Bergey's Manual of Systematics of Archaea and Bacteria*.
- Pasquarell, G.C., and D.G. Boyer. 1995. Agricultural impacts on bacterial water quality in karst groundwater. *Journal of Environmental Quality*. 24:959-969.
- Perlman, H., 2016, Where is Earth's water? USGS Water-Science School: <https://water.usgs.gov/edu/earthwherewater.html> (accessed February 12, 2018).
- Peterson, E.W., Davis, R.K., Brahana, J.V. and Orndorff, H.A., 2002. Movement of nitrate through regolith covered karst terrane, northwest Arkansas. *Journal of Hydrology*, 256(1), pp.35-47.
- Pronk, M., N. Goldscheider, and J. Zopfi. 2006. Dynamics and interaction of organic carbon, turbidity and bacteria in a karst aquifer system. *Hydrogeology Journal*. 14:473-484.
- Qin, J., R. Li, J. Raes, M. Arumugam, K. S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, and T. Yamada, 2010, A human gut microbial gene catalogue established by metagenomic sequencing: *nature*, v. 464, no. 7285, p. 59.

- Reed, T.M., Fryar, A.E., Brion, G.M. and Ward, J.W., 2011. Differences in pathogen indicators between proximal urban and rural karst springs, Central Kentucky, USA. *Environmental Earth Sciences*, 64(1), pp.47-55.
- Rehmann, C.R., and M.L. Soupir. Importance of interactions between the water column and the sediment for the microbial concentrations in streams. *Water Research*. 43:4579-4589.
- Rice, L. B., 2008, Federal Funding for the Study of Antimicrobial Resistance in Nosocomial Pathogens: No ESKAPE: *The Journal of Infectious Diseases*, v. 197, no. 8, p. 1079–1081, doi:10.1086/533452.
- Ryan, M. and Meiman, J., 1996. An Examination of Short-Term Variations in Water Quality at a Karst Spring in Kentucky. *Ground water*, 34(1), pp.23-30.
- Schwartz, T., Kohnen, W., Jansen, B. and Obst, U., 2003. Detection of antibiotic-resistant bacteria and their resistance genes in wastewater, surface water, and drinking water biofilms. *FEMS microbiology ecology*, 43(3), pp.325-335.
- Setlow, P., 2006, Spores of *Bacillus subtilis*: their resistance to and killing by radiation, heat and chemicals: *Journal of applied microbiology*, v. 101, no. 3, p. 514–525.
- Sherer, B. M., J. R. Miner, J. A. Moore, and J. C. Buckhouse, 1992, Indicator bacterial survival in stream sediments: *Journal of Environmental Quality*, v. 21, no. 4, p. 591–595.
- Shiklomanov, I.A., 1998. World water resources. A new appraisal and assessment for the 21st century.
- Smith, M., K. Cross, M. Paden, and P. Laban, 2016, Spring–Managing groundwater sustainably: IUCN, Gland, Switzerland.
- Soil Survey Staff. 2014. Web Soil Survey. In: *Soil Data Explor. Sanit. Facil.*
<http://websoilsurvey.sc.egov.usda.gov/App/WebSoilSurvey.aspx>. Accessed 7 Apr 2014
- Toze, S., 1999, PCR and the detection of microbial pathogens in water and wastewater: *Water Research*, v. 33, no. 17, p. 3545–3556.
- US Environmental Protection Agency. 2009 Recreational Water Quality Criteria. USA
- Uyaguari-Diaz, M. I., M. Chan, B. L. Chaban, M. A. Croxen, J. F. Finke, J. E. Hill, M. A. Peabody, T. Van Rossum, C. A. Suttle, and F. S. Brinkman, 2016, A comprehensive method for amplicon-based and metagenomic characterization of viruses, bacteria, and eukaryotes in freshwater samples: *Microbiome*, v. 4, no. 1, p. 20.
- Van Ness, G. B., 1971, Ecology of anthrax: *Science*, v. 172, no. 3990, p. 1303–1307.

- Vaz-Moreira, I., C. Egas, O. C. Nunes, and C. M. Manaia, 2011, Culture-dependent and culture-independent diversity surveys target different bacteria: a case study in a freshwater sample: *Antonie Van Leeuwenhoek*, v. 100, no. 2, p. 245–257.
- Vazquez-Baeza Y, Pirrung M, Gonzalez A, Knight R. *Gigascience*. 2013. EMPeror: a tool for visualizing high-throughput microbial community data *Nov 26;2(1):16*.
- Vesper, D. J., C. M. Loop, and W. B. White, 2001, Contaminant transport in karst aquifers: *Theoretical and Applied Karstology*, v. 13, no. 14, p. 101–111.
- Visca, P., H. Seifert, and K. J. Towner, 2011, *Acinetobacter* infection – an emerging threat to human health: *IUBMB Life*, v. 63, no. 12, p. 1048–1054, doi:10.1002/iub.534.
- Wada, Y., L. P. van Beek, C. M. van Kempen, J. W. Reckman, S. Vasak, and M. F. Bierkens, 2010, Global depletion of groundwater resources: *Geophysical research letters*, v. 37, no. 20.
- Ward, J.W., Reed, T.M., Fryar, A.E. and Brion, G.M., 2009. Using the AC/TC ratio to evaluate fecal inputs in a karst groundwater basin. *Environmental & Engineering Geoscience*, 15(2), pp.57-65.
- Weary, D. J., and D. H. Doctor, 2014, *Karst in the United States: A digital map compilation and database*: US Department of the Interior, US Geological Survey.
- Willems, A., J. De Ley, M. Gillis, and K. Kersters, 1991, *Comamonadaceae*, a new family encompassing the acidovorans rRNA complex, including *Variovorax paradoxus* gen. nov., comb. nov., for *Alcaligenes paradoxus* (Davis 1969): *International Journal of Systematic and Evolutionary Microbiology*, v. 41, no. 3, p. 445–450.
- Williams, J. W., 2012, *Williams gynecology*: McGraw-Hill Education LLC.
- Wooley, J.C., Godzik, A. and Friedberg, I., 2010. A primer on metagenomics. *PLoS computational biology*, 6(2), p.e1000667.
- Amraoui, F., M. Razack, and L. Bouchaou, 2003, Turbidity dynamics in karstic systems. Example of Ribaa and Bittit springs in the Middle Atlas (Morocco): *Hydrological Sciences Journal*, v. 48, no. 6, p. 971–984.
- Yeom, J., J.-H. Shin, J.-Y. Yang, J. Kim, and G.-S. Hwang, 2013, ¹H NMR-Based Metabolite Profiling of Planktonic and Biofilm Cells in *Acinetobacter baumannii* 1656-2: *PLOS ONE*, v. 8, no. 3, p. e57730, doi:10.1371/journal.pone.0057730.

7. Appendix

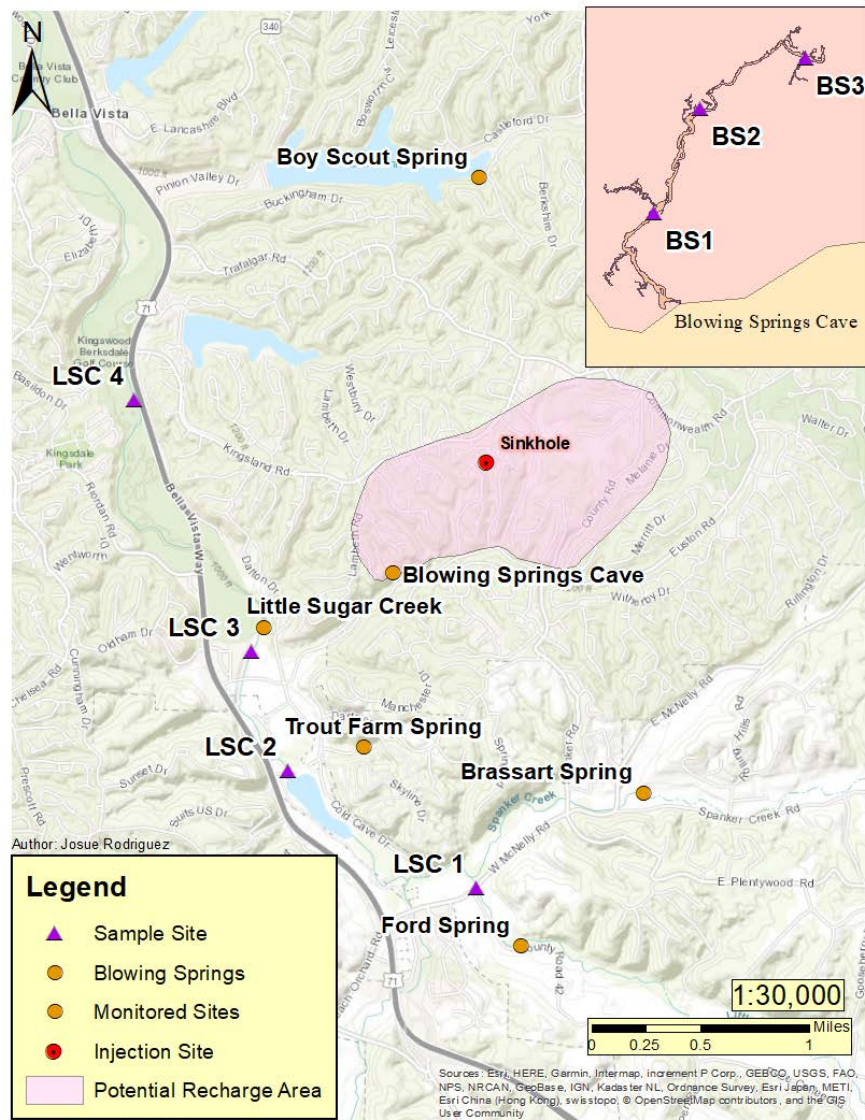


Figure 1: Research area with all sample sites.

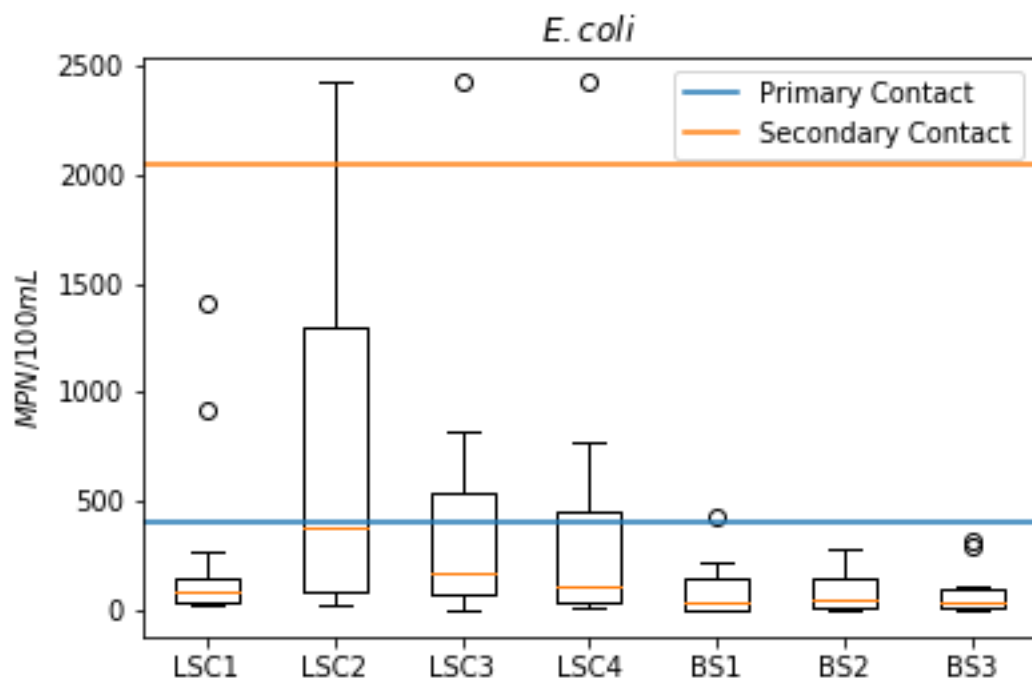


Figure 2: Box plot of the concentration of *E. coli* at each sample site. The blue line represents the level at which primary contact limits are implemented. The orange line is the level at which secondary contact limits are implemented. LSC 1-4 are samples from Little Sugar Creek and BS1-3 are samples from Blowing Spring Cave.

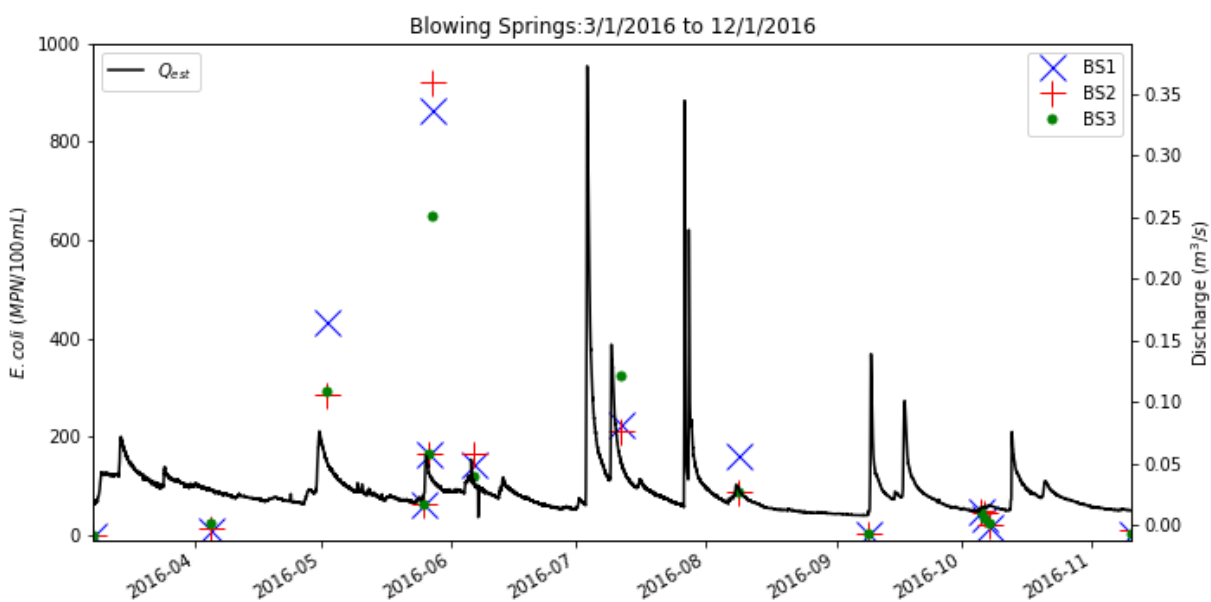


Figure 3: Hydrograph with spot measurements of *E. coli* samples collected from individual sites within Blowing Spring Cave for the entire sampling period. BS1 (blue X) is the most

downstream sample, BS2 (Red cross) is the middle sample in the stream and BS3 (green circle) is from the most upstream part of the stream.

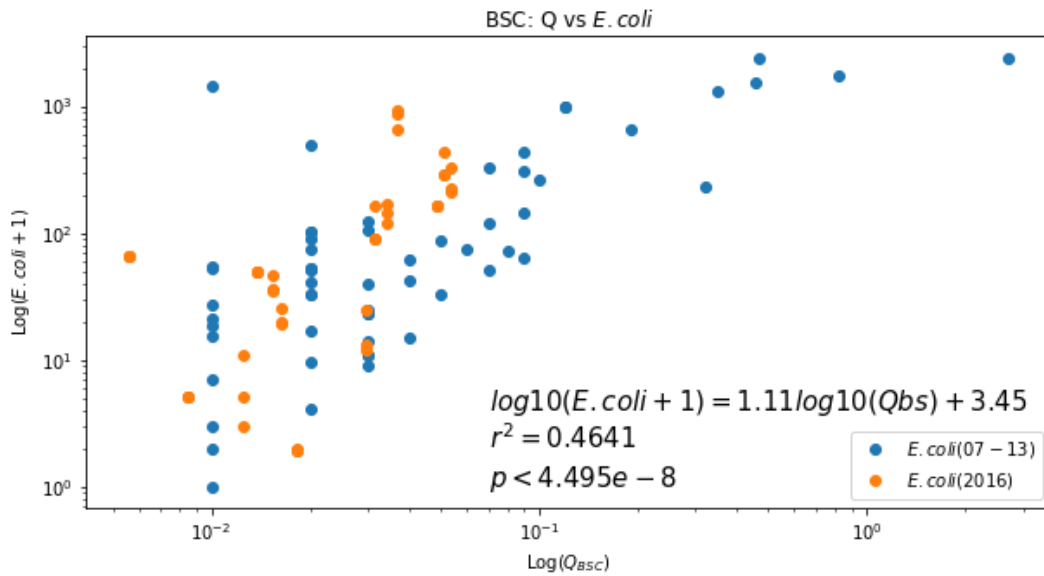


Figure 4: Relationship between discharge and *E. coli*. These data consist of all samples collected at Blowing Spring Cave, with blue representing samples from 2007-2013, and orange representing samples from 2016. The results showed a positive correlation using Spearman correlation that was statistically significant.

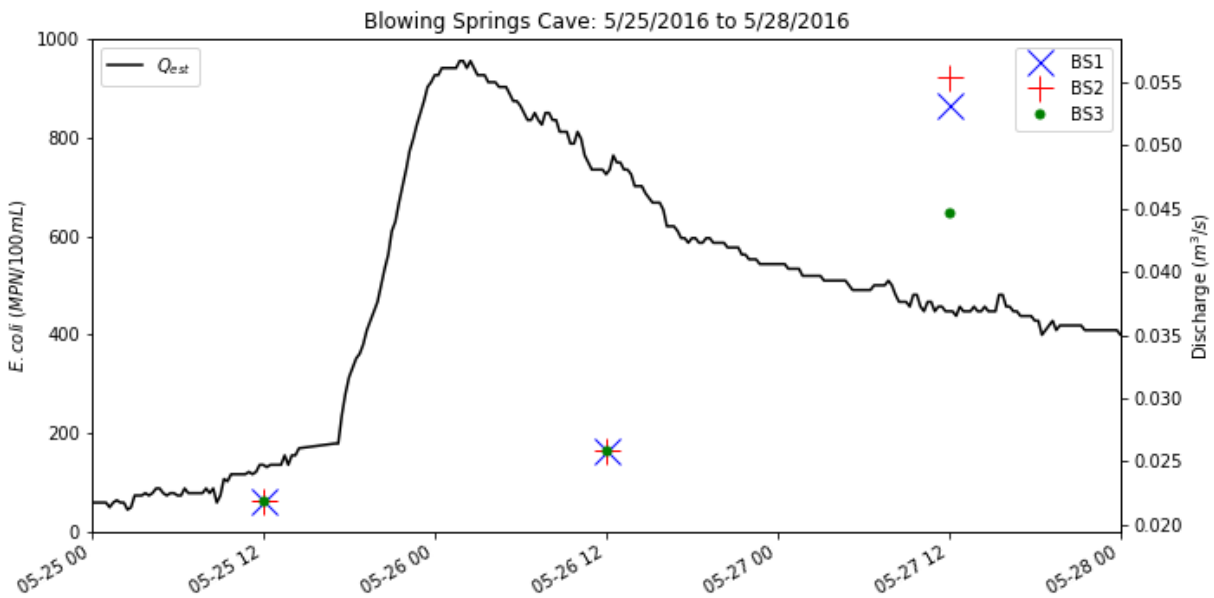


Figure 5: Hydrograph with spot measurements of *E. coli* samples collected from individual sites within Blowing Spring Cave for a storm event. BS1 (blue X) is the most downstream sample,

BS2 (red cross) is the middle sample in the stream and BS3 (green circle) is from the most upstream part of the stream. Peak discharge was 0.057 CMS.

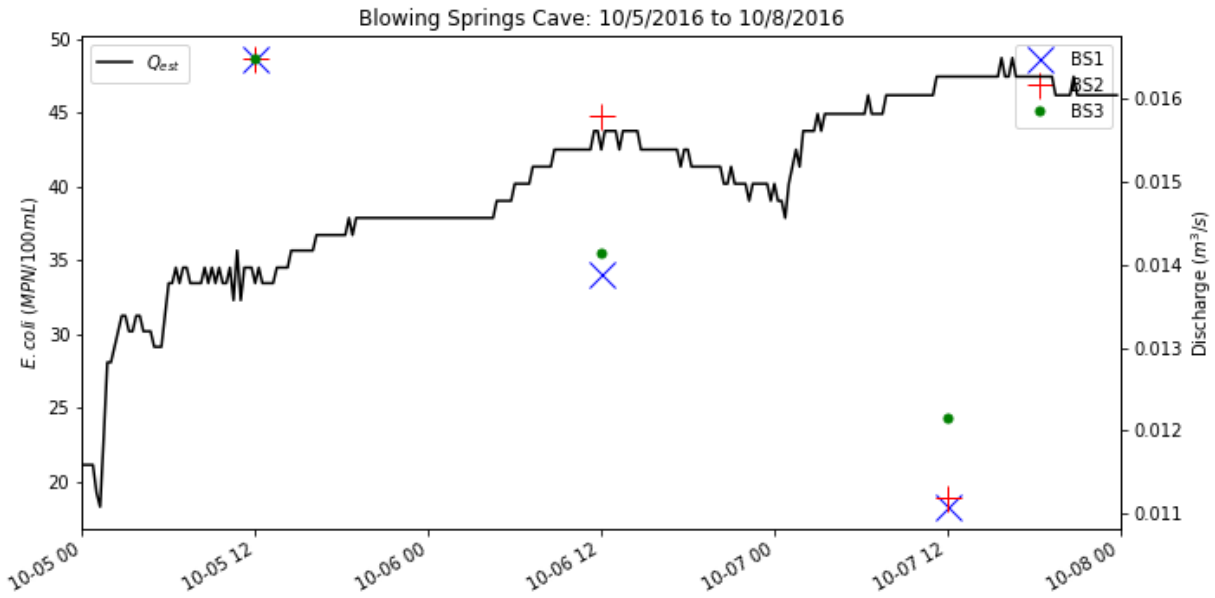


Figure 6: Hydrograph with spot measurements of *E. coli* samples collected from individual sites within Blowing Spring Cave for a storm event. BS1 (blue X) is the most downstream sample, BS2 (red cross) is the middle sample in the stream and BS3 (green circle) is from the most upstream part of the stream. Concentration at each site decreases over time.

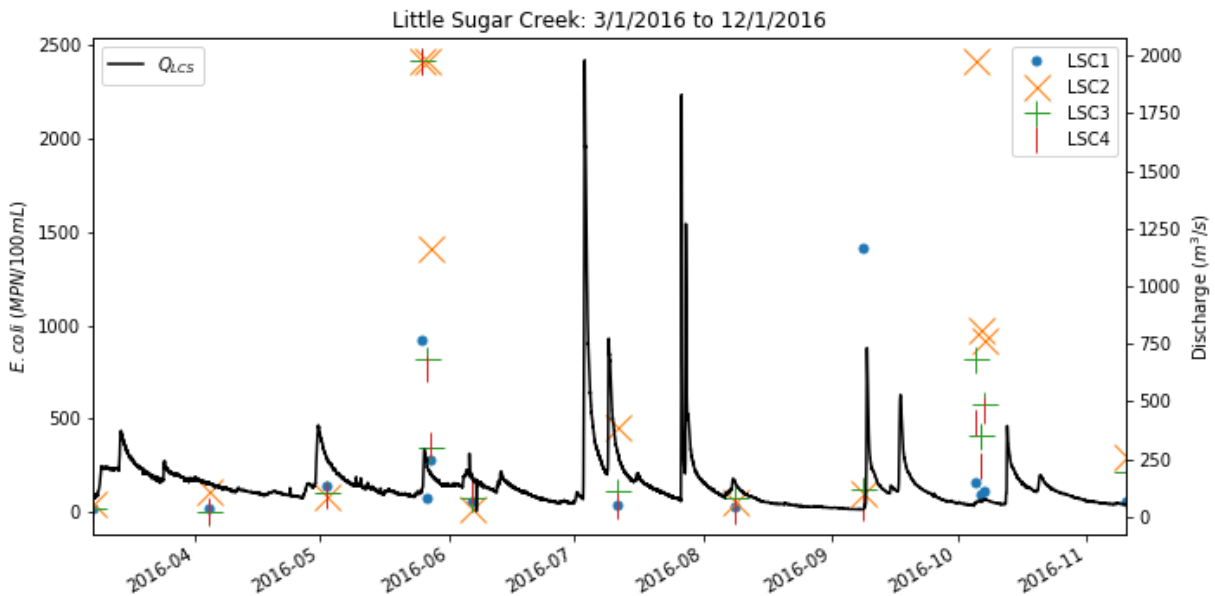


Figure 7: Hydrograph with spot measurements of *E. coli* samples collected from individual sites within Little Sugar Creek for the entire sampling period. Sample sites are in order from upstream (LSC1) to downstream (LSC4).

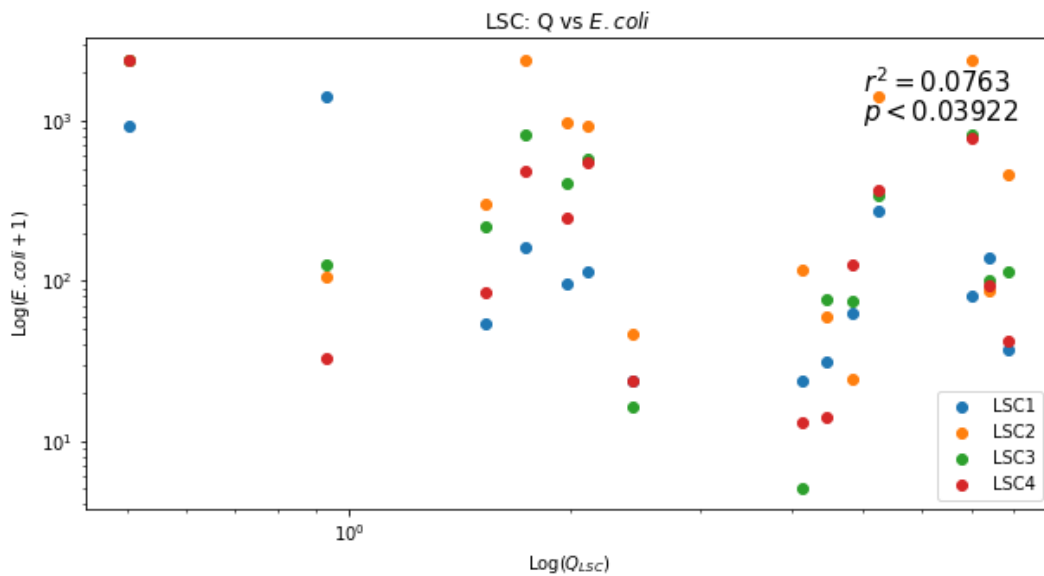


Figure 8: Relationship between discharge and *E. coli*. These data consist of the samples collected at Little Sugar Creek, where blue represents the samples from LSC1, orange represents LSC2, green represents LSC3 and red represents LSC4. The results showed little to no correlation between discharge and *E. coli* level.

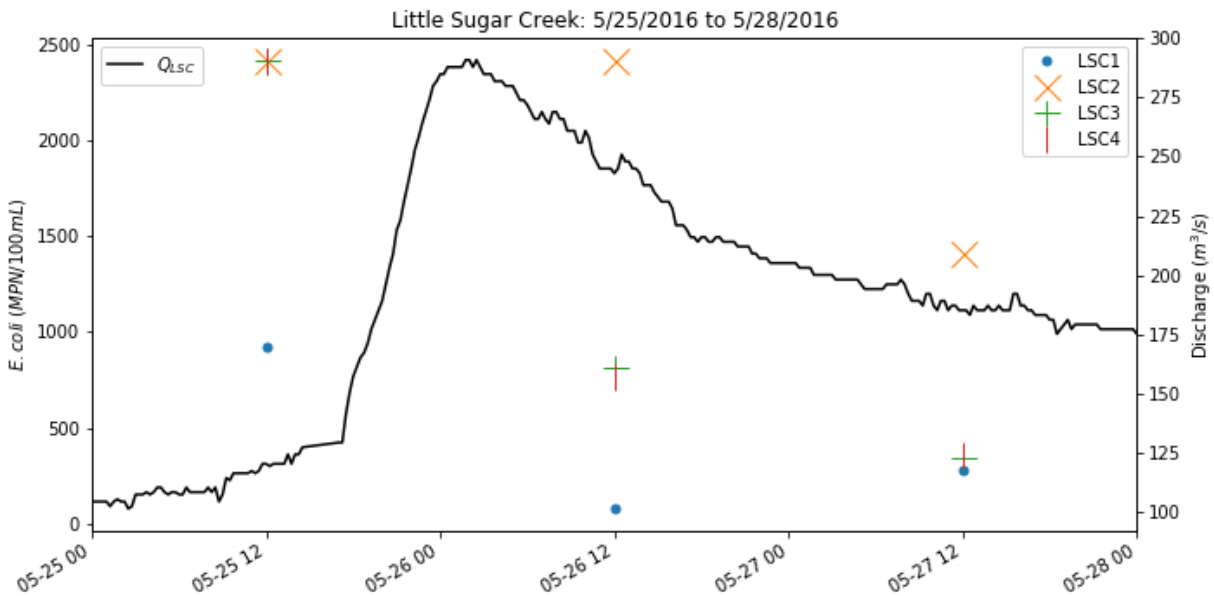


Figure 9: Hydrograph with spot measurements of *E. coli* in water samples collected from individual sites within Little Sugar Creek for a storm event. Sites are in order from upstream (LSC1) to downstream (LSC4).

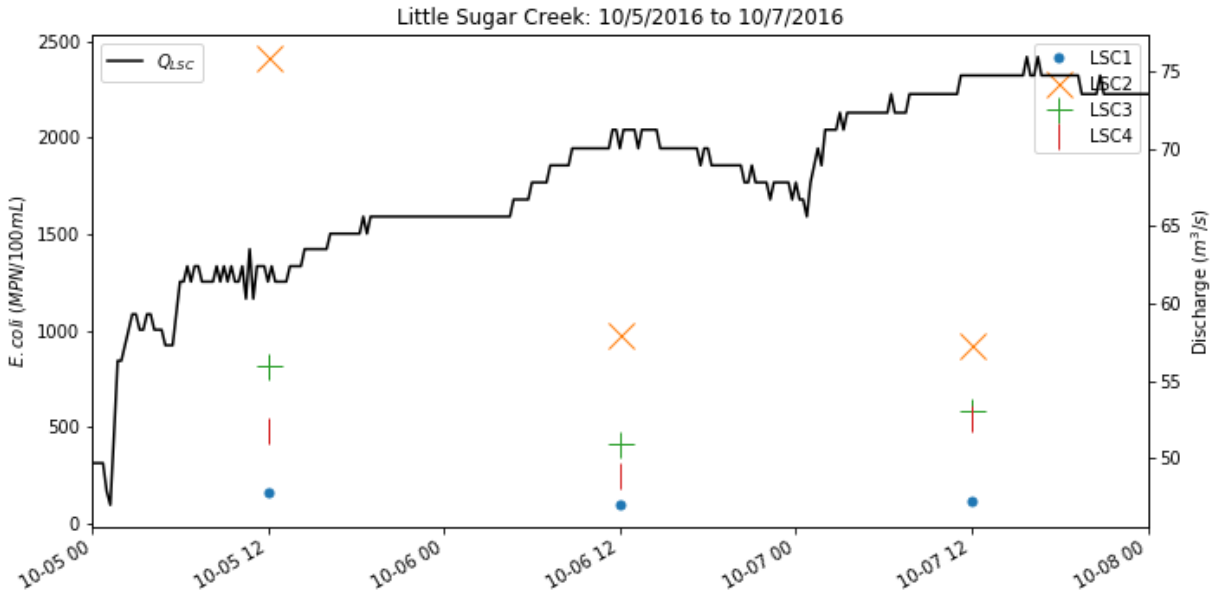


Figure 10: Hydrograph with spot measurements of *E. coli* samples collected from individual sites within Little Sugar Creek for a storm event. Sites are in order from upstream (LSC1) to downstream (LSC4).

Dye Trace

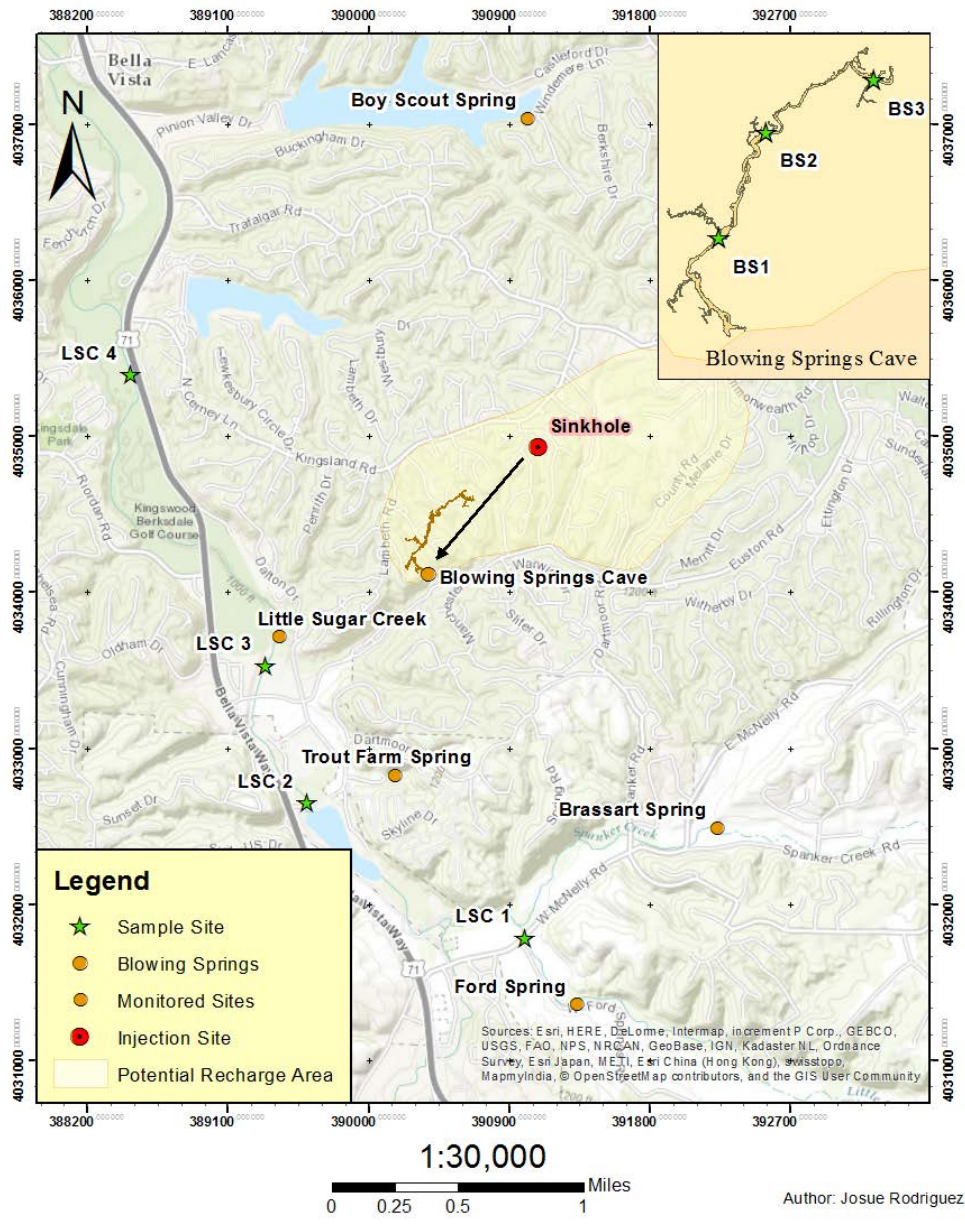


Figure 11: Map of the study area detailing locations where samples were collected (green stars) as well as springs that were monitored for the detection of dye (orange circles). The injection site was a sinkhole denoted as a red circle. This site lies within the hypothesized recharge area displayed as a tan boundary. Dye flowed from the sinkhole to Blowing Springs Cave.

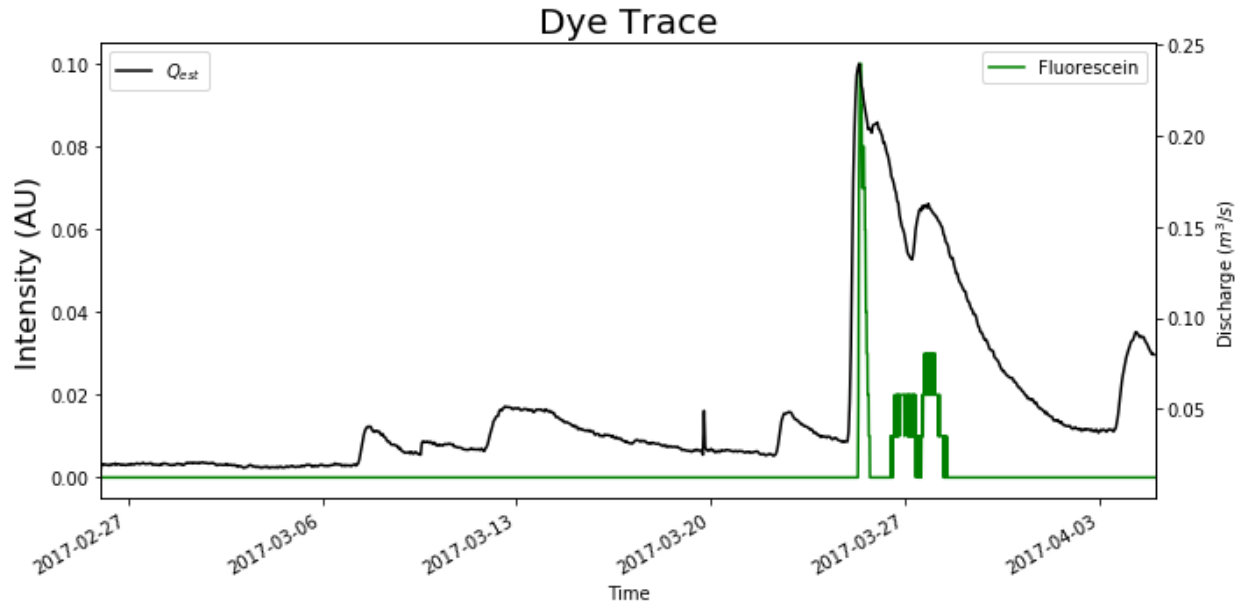


Figure 12: Hydrograph of the water (black line) in Blowing Spring Cave alongside the dye intensity (green line).

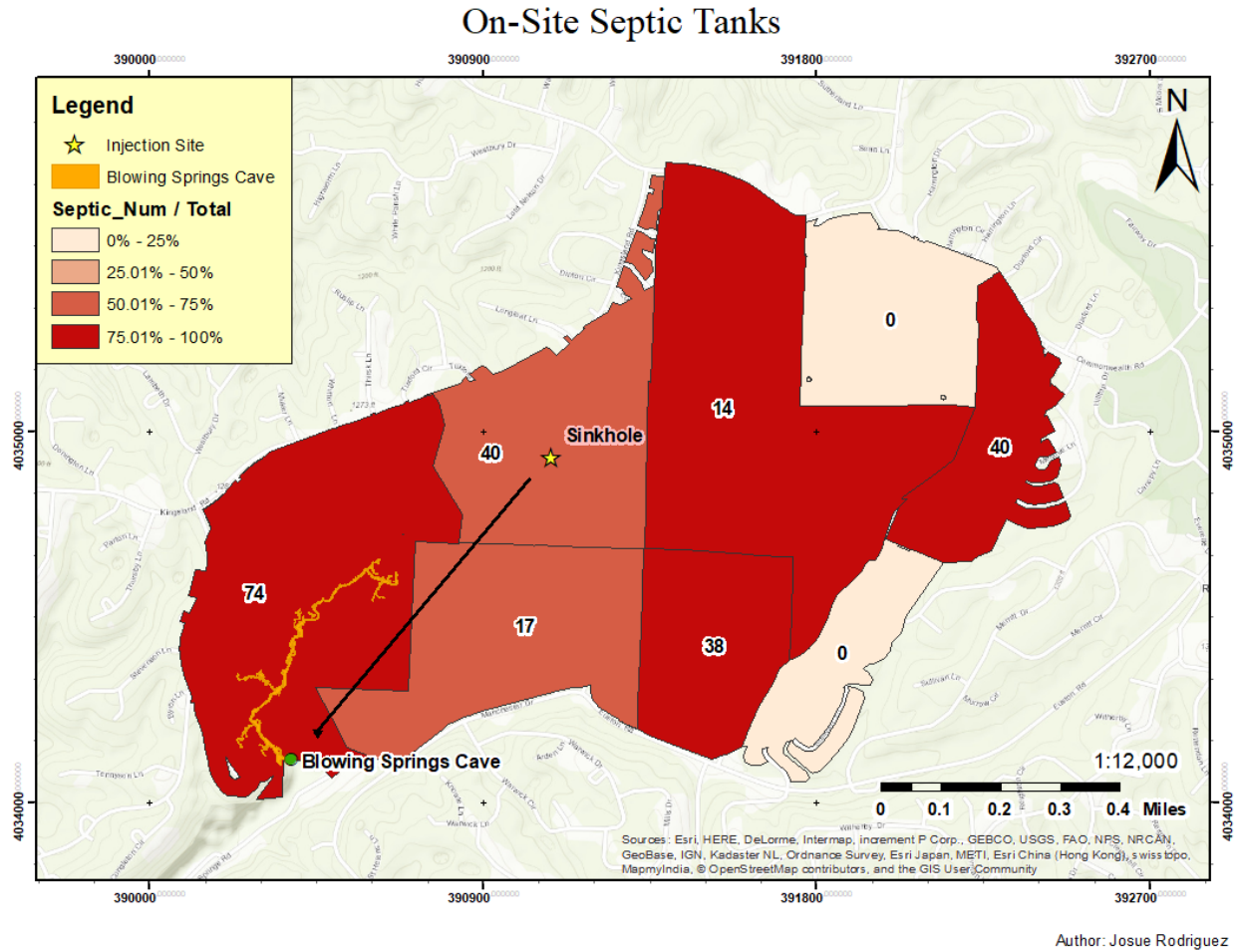


Figure 13: Map representing the concentration of septic tanks per area. The sinkhole (yellow star) is in a moderate section of septic tank coverage, while the known cave passage and spring (orange and green dot) lie in an area of high density septic tank coverage.

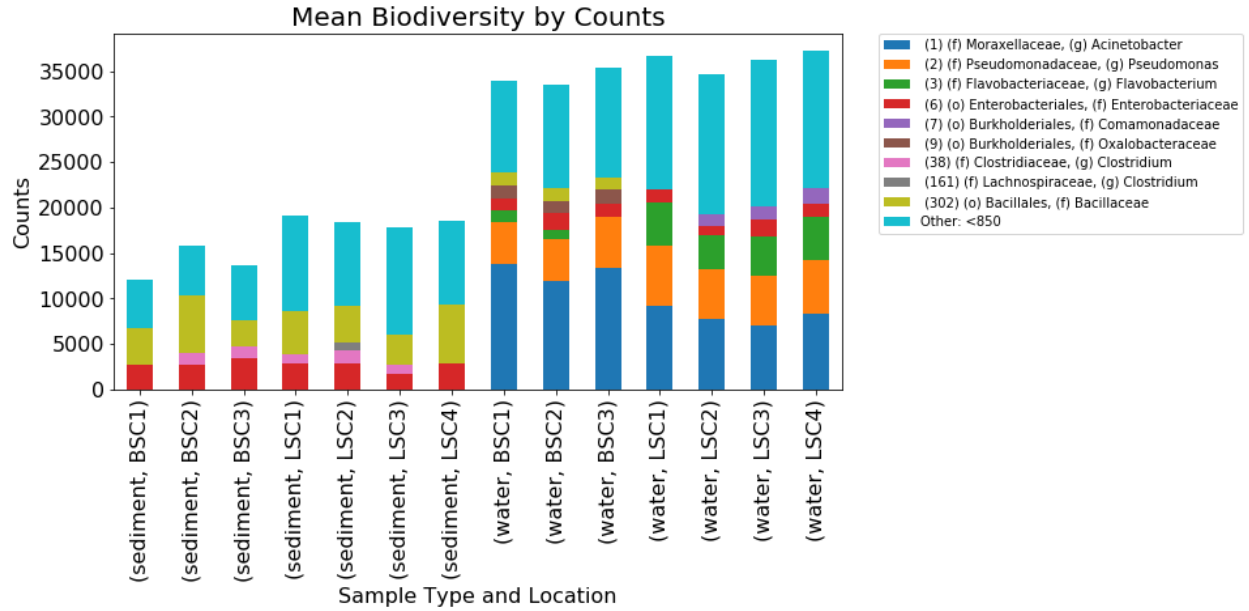


Figure 14: Stacked bar graph representing the mean concentration of most abundant taxa per sample site and type. These samples consisted of samples greater than 850 counts in a sample. The rest of the samples are grouped together in Other (light blue).

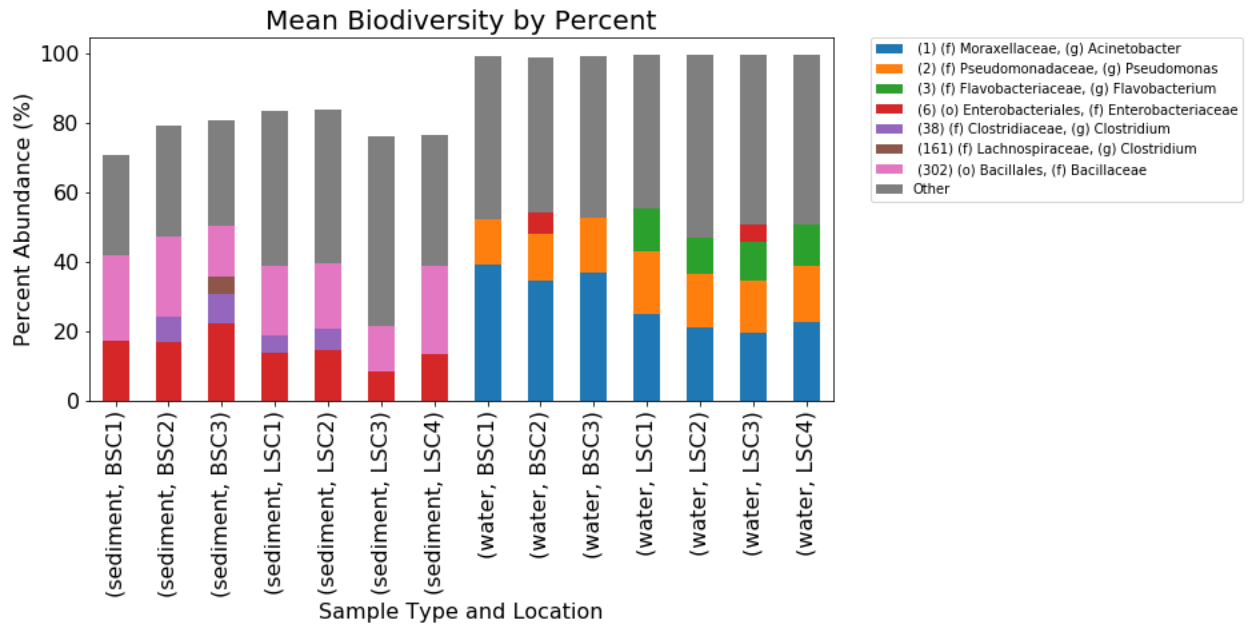


Figure 15: Stacked bar graph representing the mean relative abundance of taxa per sample site and type. These samples consisted of samples greater than 5% counts in a sample. The rest of the samples are grouped together in Other (grey).

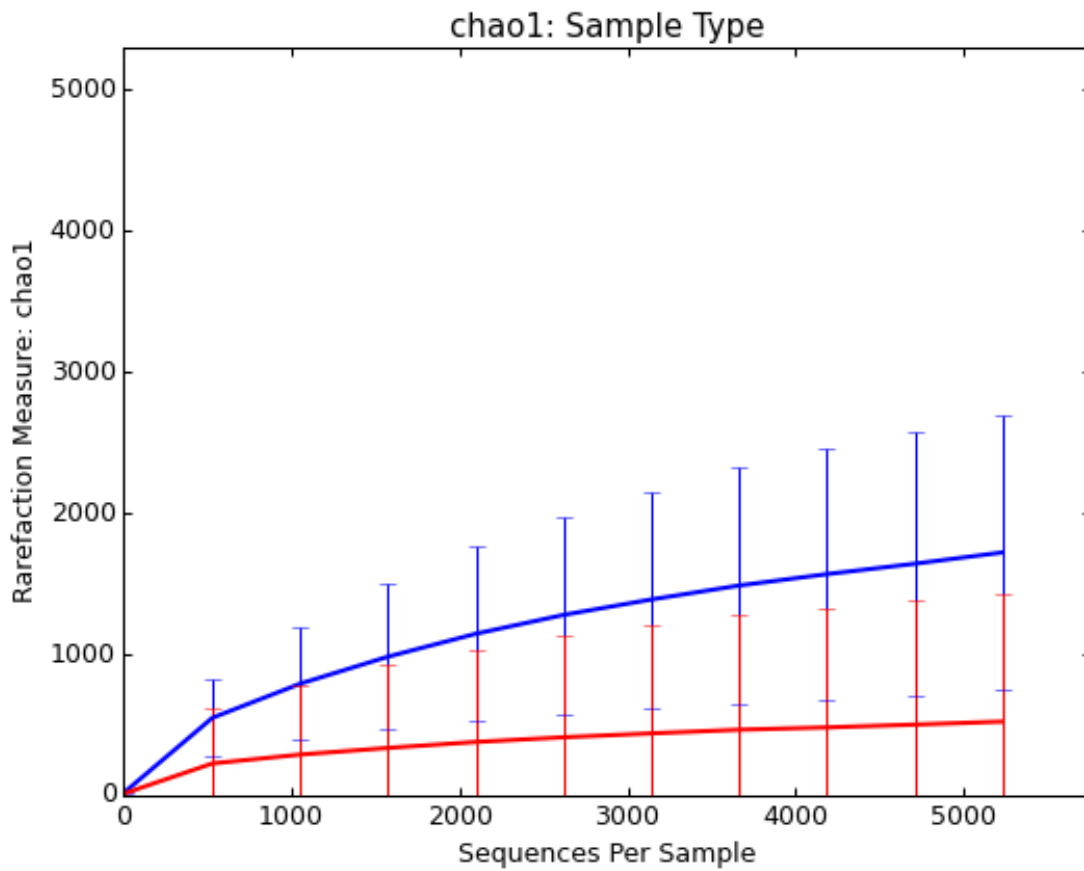


Figure 16: Alpha Diversity plot using the rarefaction measure: chao1 which measures the species richness in the sample types. Water is represented by blue and sediment is represented by red.

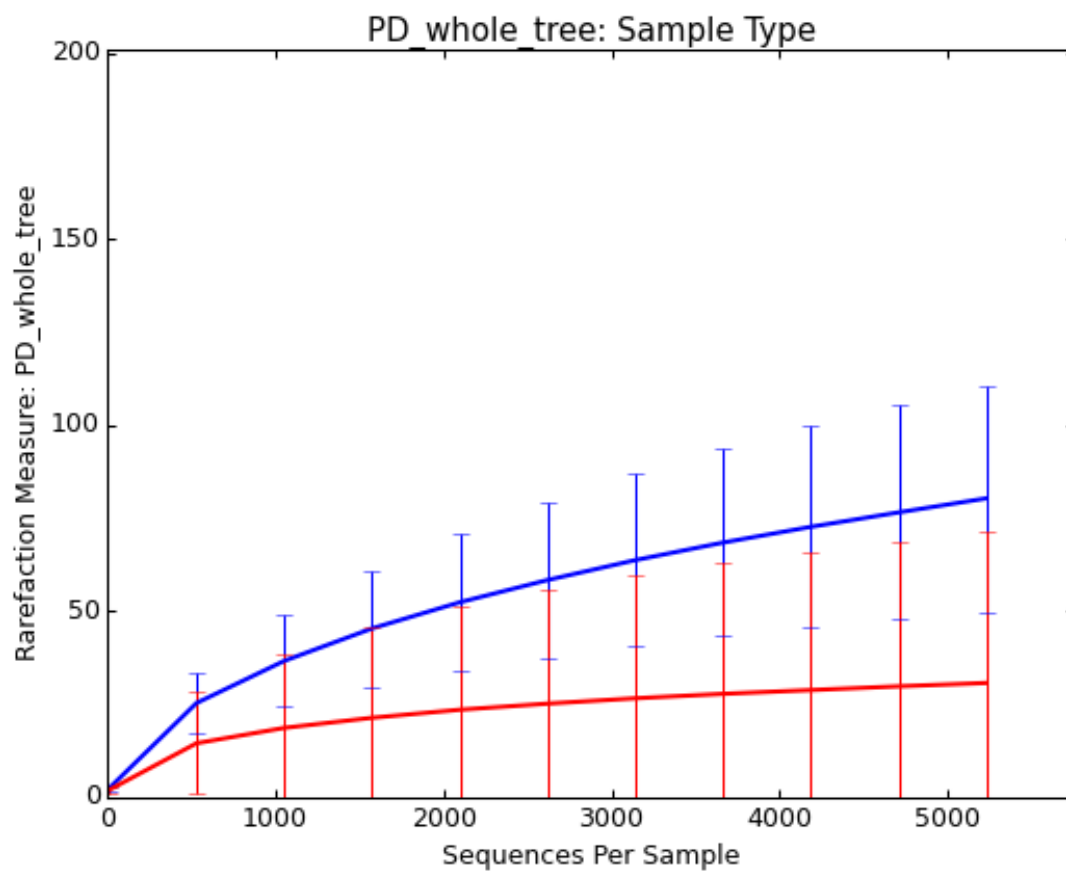


Figure 17: Alpha Diversity plot using the rarefaction measure: *PD_whole_tree*, which measures the phylogenetic distance between taxa in the sample types. Water is represented by blue and sediment is represented by red.

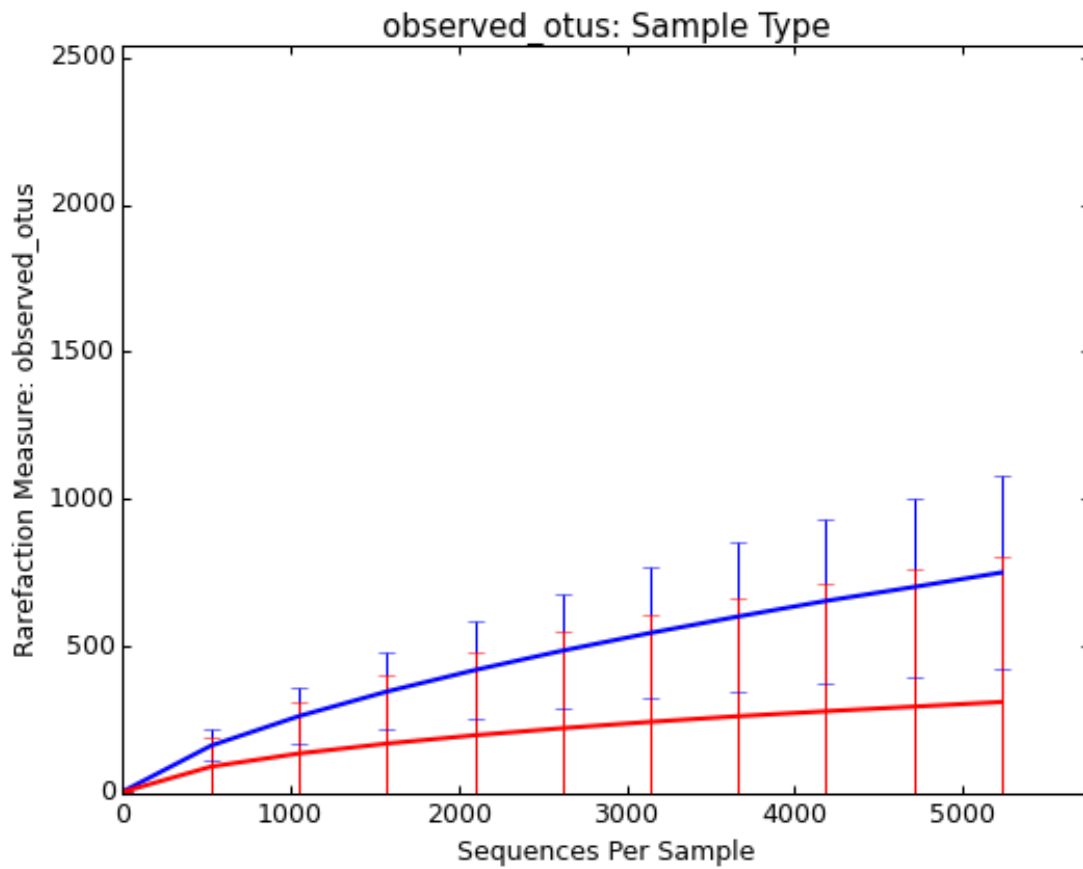


Figure 18: Alpha Diversity plot using the rarefaction measure: OTUs, which represents the operational taxonomic units in the sample types. Water is represented by blue and sediment is represented by red.

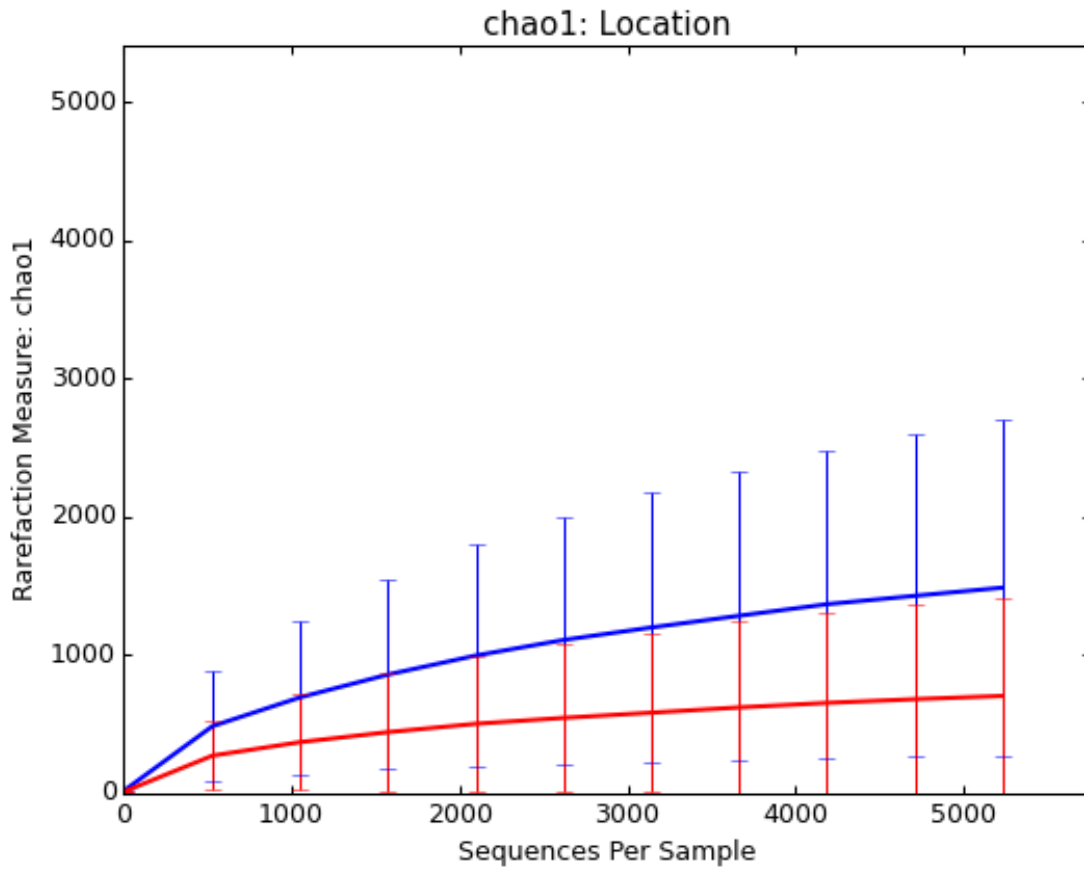


Figure 19: Alpha Diversity plot using the rarefaction measure: chao1 which measures the species richness in the sample location. Little Sugar Creek is represented by blue and Blowing Spring Cave is represented by red.

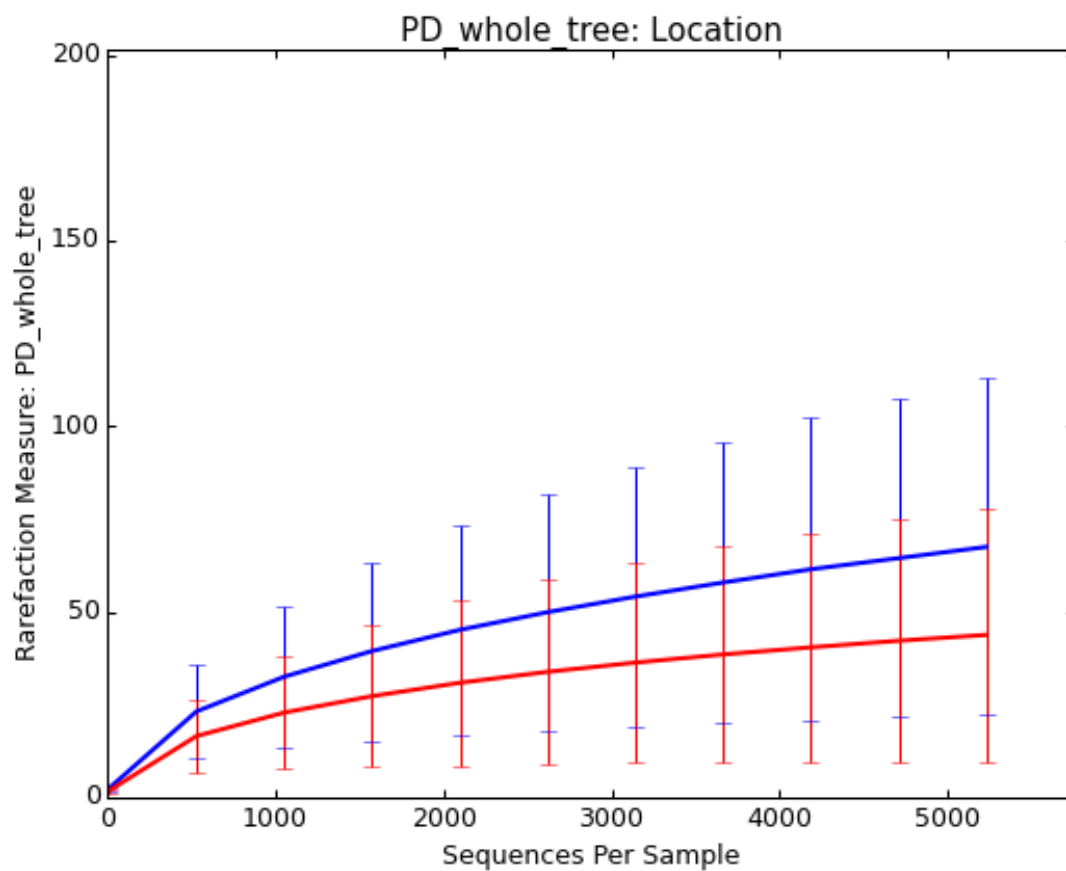


Figure 20: Alpha Diversity plot using the rarefaction measure: PD_whole_tree, which measures the phylogenetic distance between taxa in the sample location. Little Sugar Creek is represented by blue and Blowing Spring Cave is represented by red.

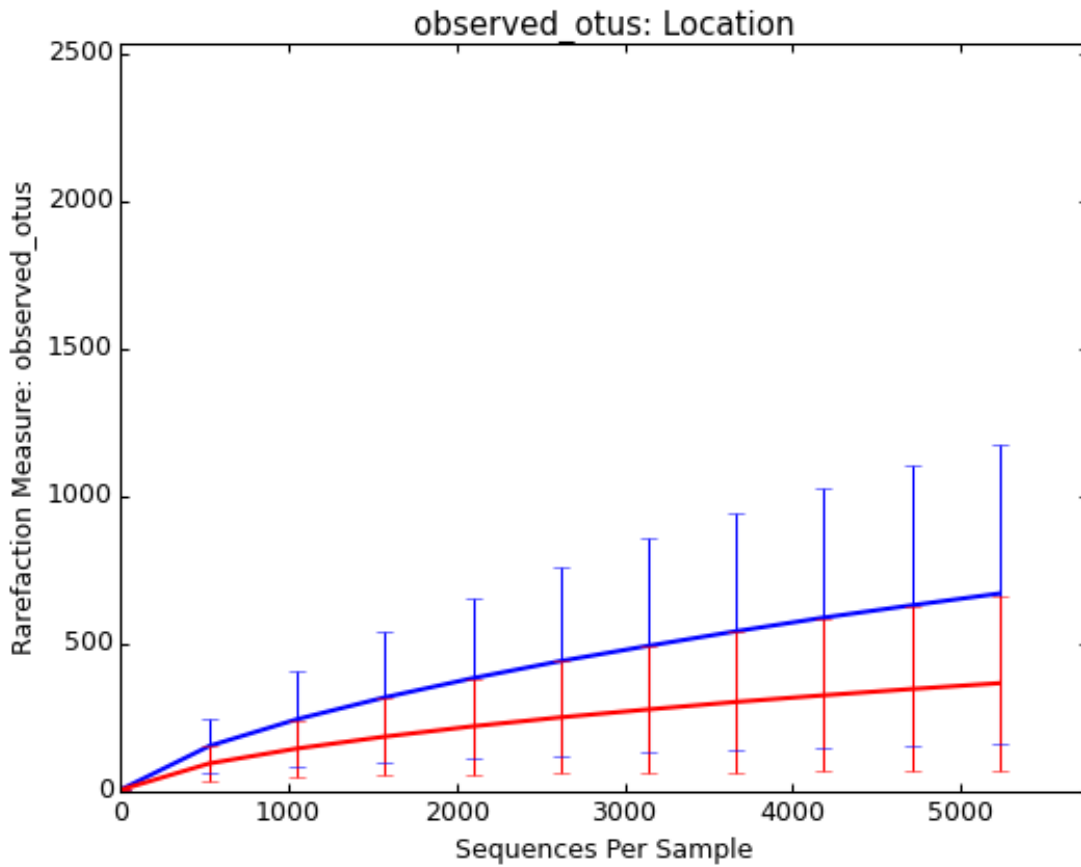


Figure 21: Alpha Diversity plot using the rarefaction measure: OTUs, which represents the operational taxonomic units in the sample location. Little Sugar Creek is represented by blue and Blowing Spring Cave is represented by red.

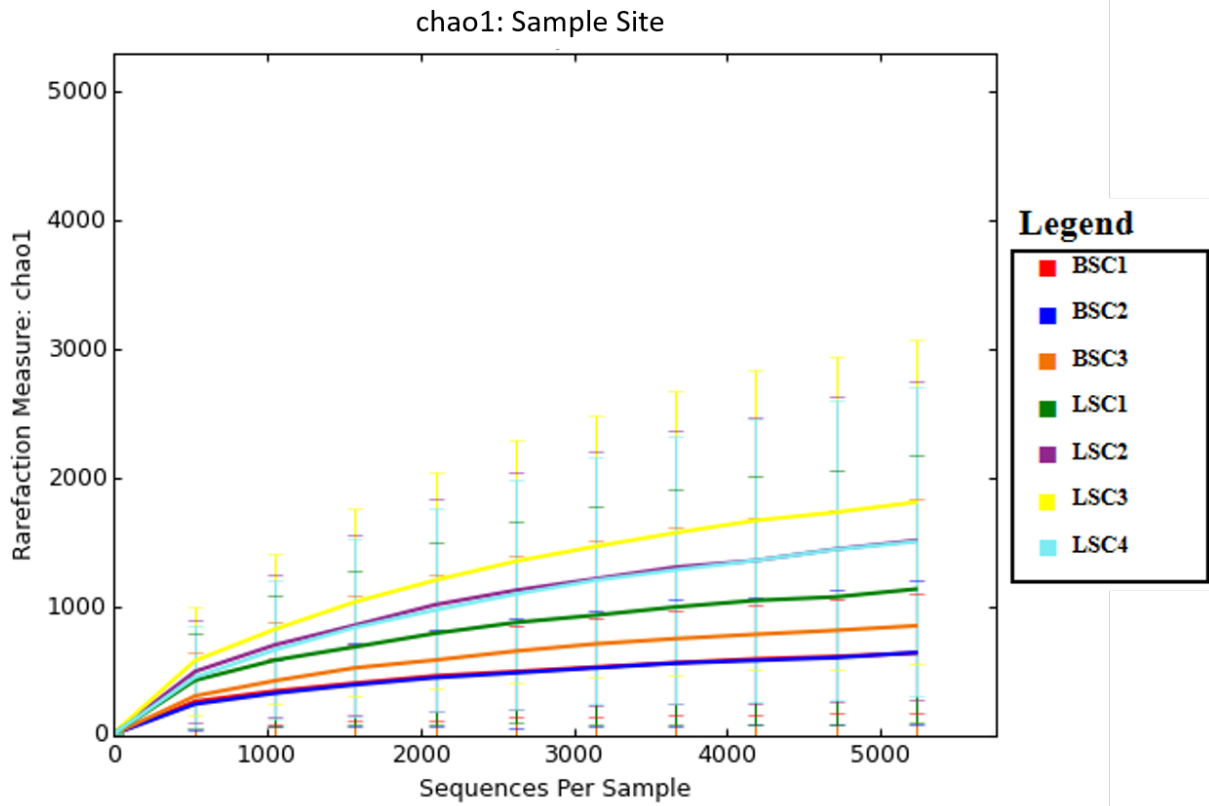


Figure 22: Alpha Diversity plot using the rarefaction measure: chao1 which measures the species richness in the sample sites.

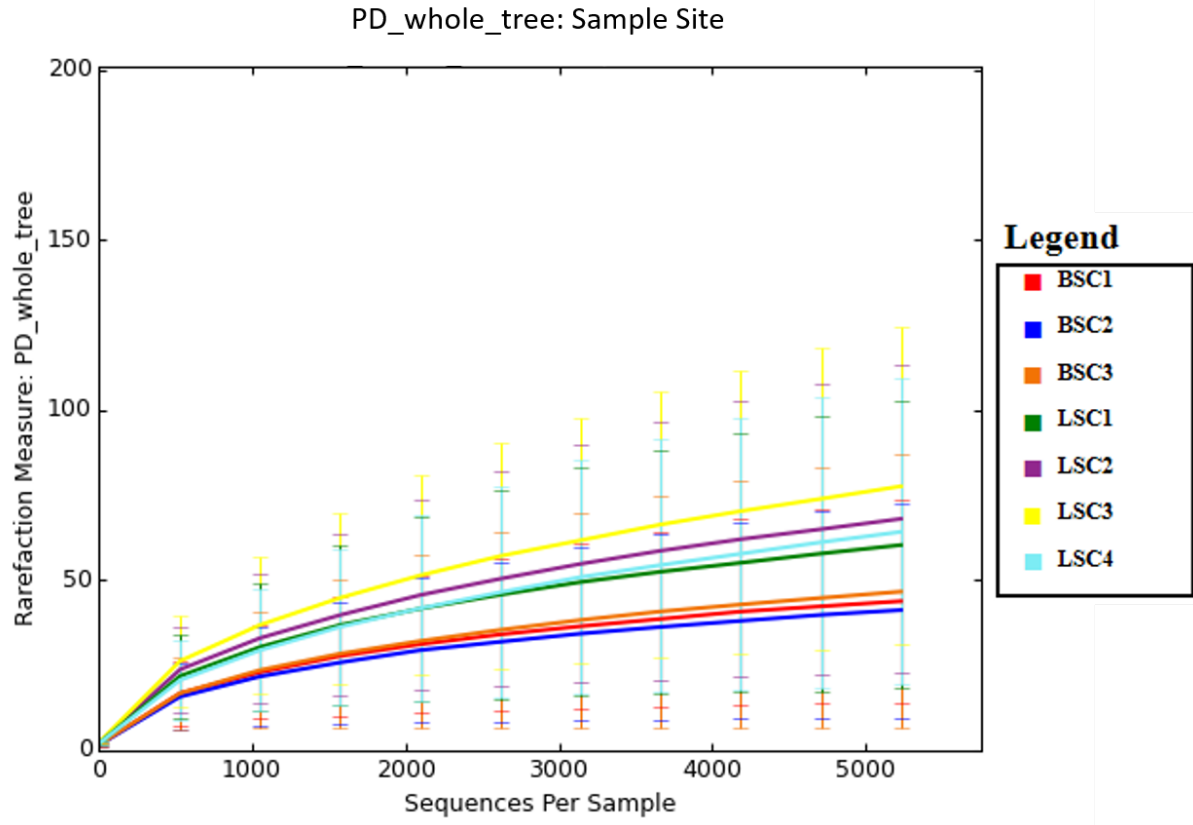


Figure 23: Alpha Diversity plot using the rarefaction measure: PD_{whole_tree}, which measures the phylogenetic distance between taxa in the sample location.

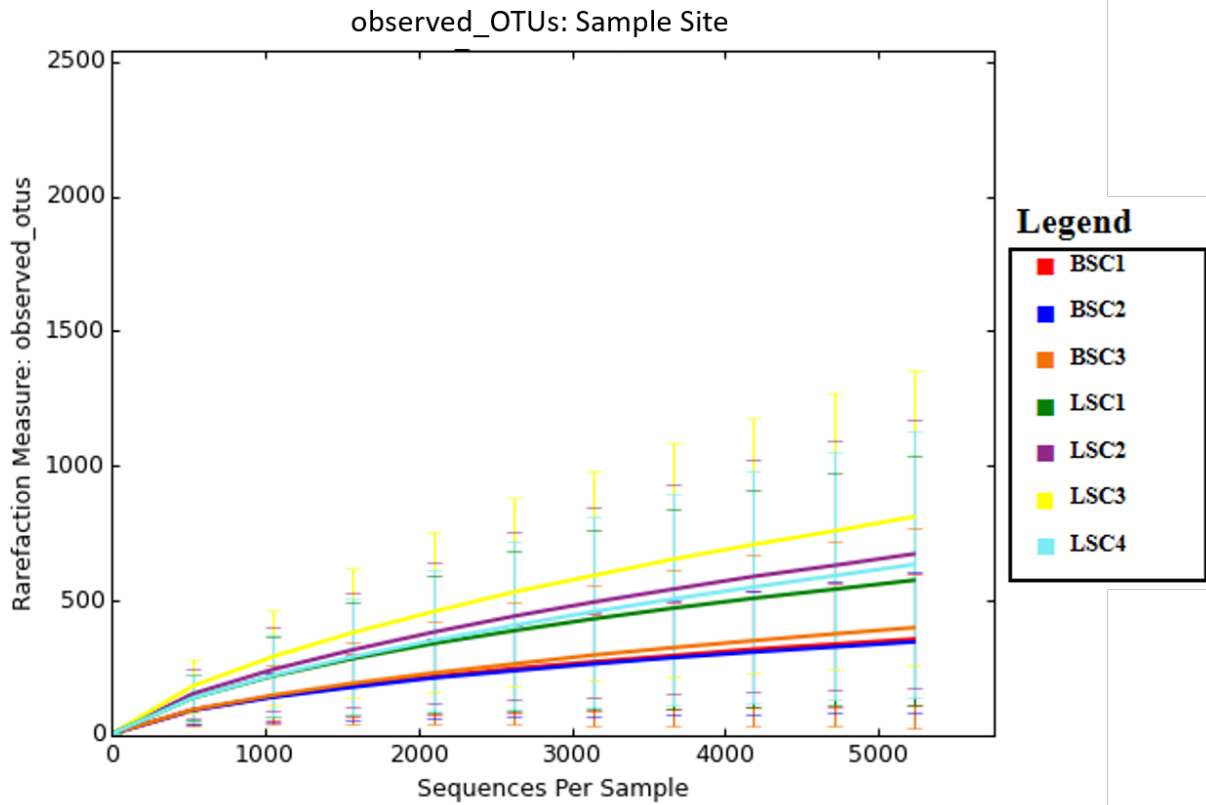


Figure 24: Alpha Diversity plot using the rarefaction measure: OTUs, which represents the operational taxonomic units in the sample location.

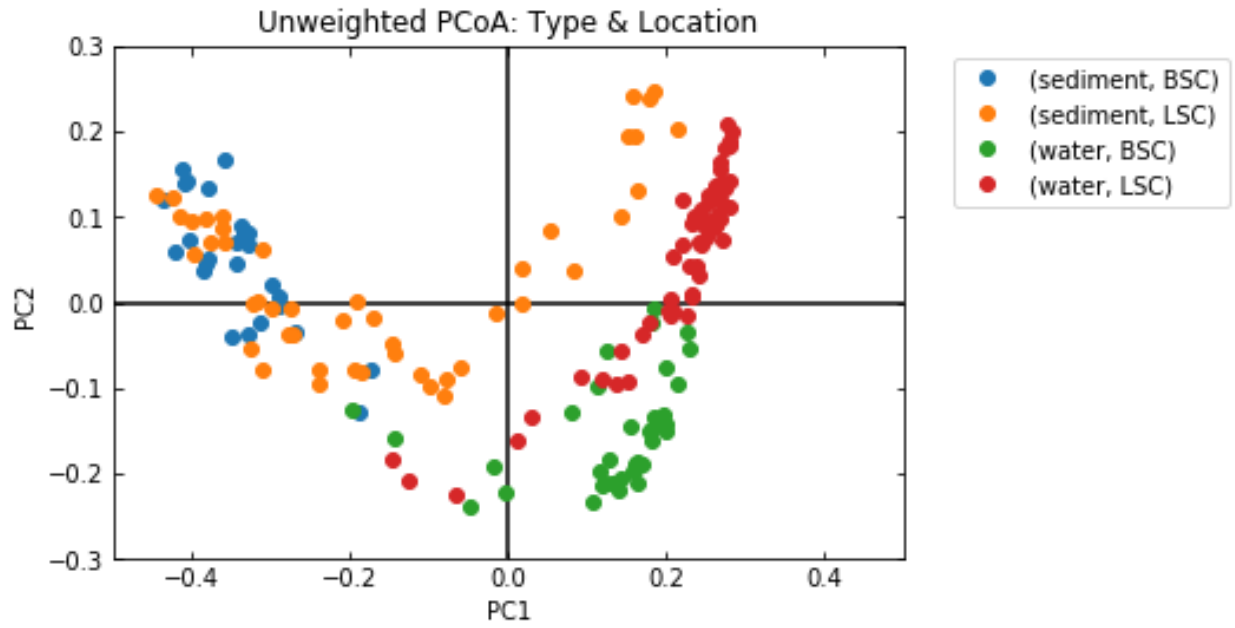


Figure 25: Beta diversity plot of the unweighted UNIFRAC principal coordinate analysis, which represents the difference between samples by presence/absence of taxa. The samples are colored by sample type and location.

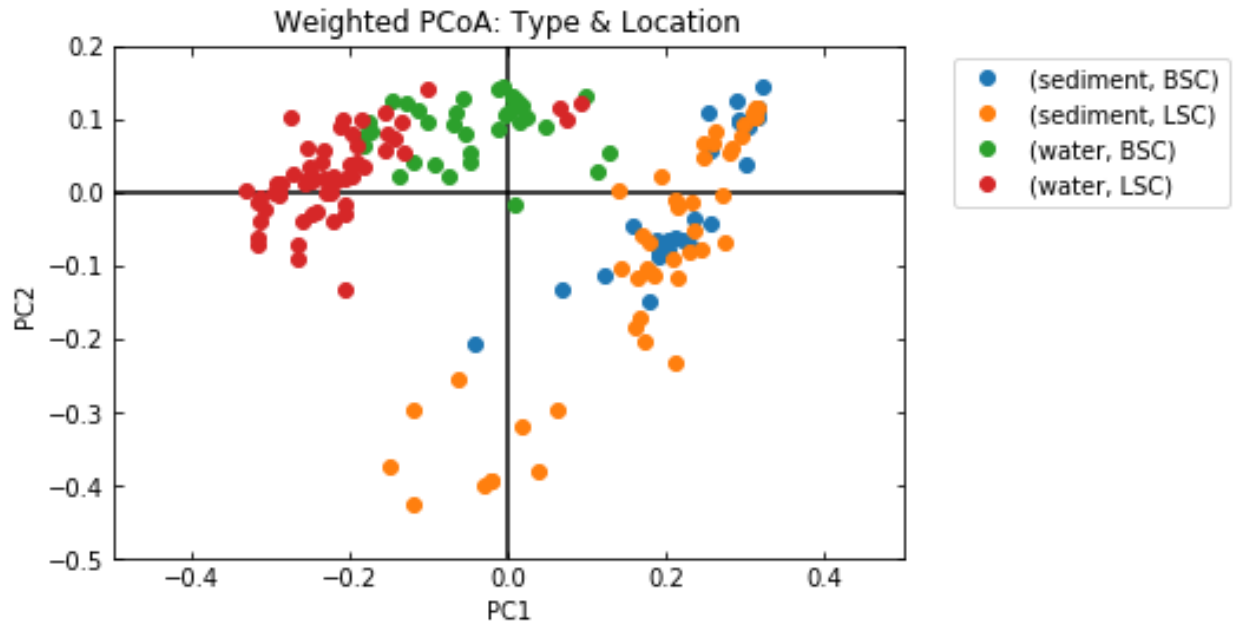


Figure 26: Beta diversity plot of the weighted UNIFRAC principal coordinate analysis, which represents the difference between samples by the abundance of shared taxa. The samples are colored by sample type and location.

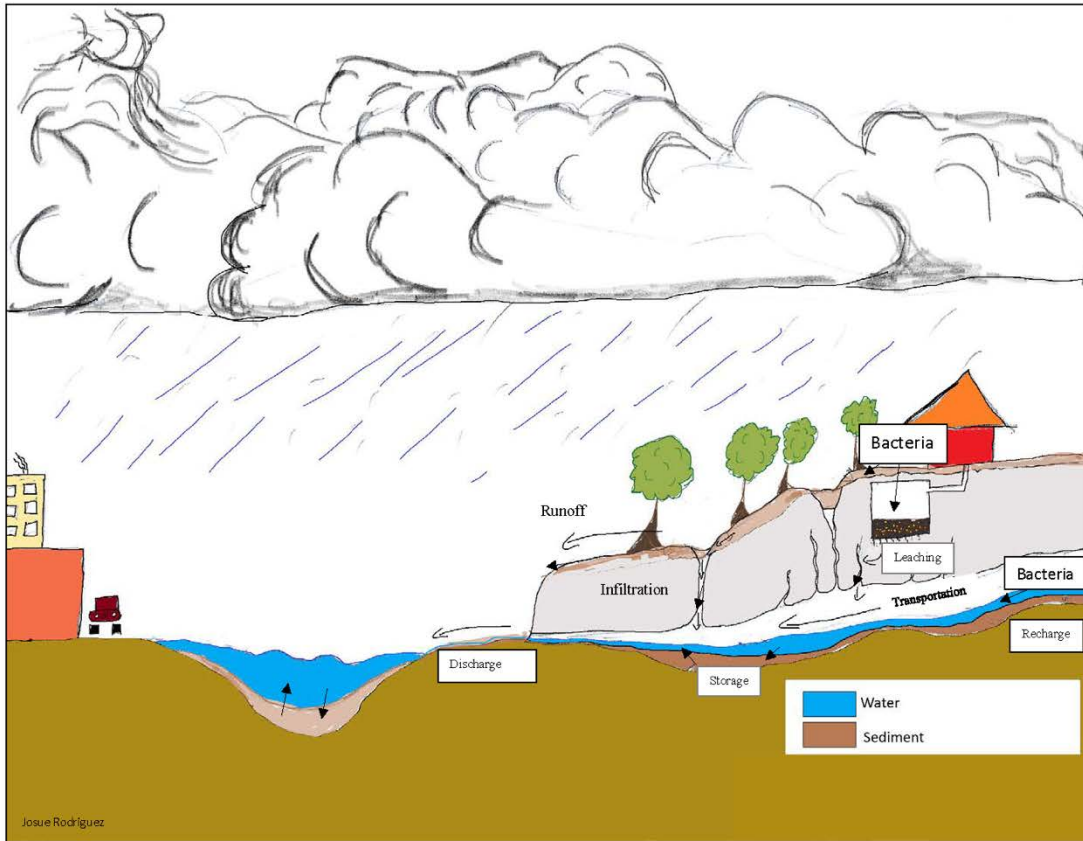


Figure 27: Conceptual model of the microbial interaction between surface water and groundwater via karst. Other interactions include the microbial interaction between sediment and water. Sources of the bacteria include anthropogenic (septic tanks) and natural influences (soil). Water is the primary means of transport of the bacteria. Therefore, movement of the bacteria is governed by the water cycle.