Fish Biodiversity Assessment of Abrams Creek, Tennessee: Illumina Sequencing of COI Amplicons vs. Electrofishing

by
WILLEM GLENN BALLARD

(Under the Direction of Travis Glenn)

ABSTRACT

Biodiversity assessments are routinely used to investigate and safeguard native ecosystems and natural resources. New assessment approaches based on DNA shed into the environment (environmental DNA, eDNA) are powerful tools, but most are constrained to surveying one species at a time via quantitative PCR. This study reports results from eDNA sampled via PCR and Illumina DNA sequencing from a model Appalachian stream in the Great Smoky Mountain National Park as a means of measuring fish biodiversity as compared to the currently employed sampling method of electrofishing. The eDNA sequencing method is a less invasive, less harmful, and potentially more informative survey technique that reports the same type of information collected by electrofishing. Sampling of two separate sites on Abrams Creek with 12 species of interest at each site revealed significant correlation between the results from electrofishing and eDNA sequencing (Spearman's Rho = 0.85, $p < 0.001$; Rho = 0.90, $p < 0.0001$ for the two sites).

INDEX WORDS: biodiversity survey, environmental DNA, Illumina, national park, Next Generation Sequencing, PCR

Fish Biodiversity Assessment of Abrams Creek, Tennessee: Illumina Sequencing of COI Amplicons vs. Electrofishing

by

WILLEM GLENN BALLARD
Bachelor of Science, Georgia College and State University, 2014

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA
2016

Fish Biodiversity Assessment of Abrams Creek, Tennessee: Illumina Sequencing of COI Amplicons vs. Electrofishing

by

WILLEM GLENN BALLARD

Major Professor:	Travis Glenn
Committee:	Marsha Black
	Greg Moyer

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2016

Acknowledgements

I would like to thank Dr. Glenn for his help and sponsorship throughout this project, Dr. Moyer for his guidance and partnership on this study, and Dr. Black for joining my committee and providing continual encouragement. I want to thank the Glenn Lab members for all the help and encouragement with experimental setup and analysis, EHS department faculty and graduate students for constant support and encouragement, Dr. Anna McKee for lab and experimental setup input and help, U.S. Fish and Wildlife Service for the use of equipment and lab spaces at Warm Springs Hatchery, Great Smoky Mountain National Park for allowing me to conduct research inside the park, Matt Kulp and the summer fisheries team for their partnership and help in the field sampling and logistics, Dr. Jeffrey Werner for help with the MacQiime analysis program, UGA Statistical Consulting Center for help with determining appropriate interpretations for the data, Ashantye Williams at the US Fish and Wildlife center in Warm Springs for all her assistance with borrowing equipment, Warnell School of Forestry aquatic biotech environmental lab staff for sampling of their fish tanks at Whitehall, Dr. Richard Chandler for help with analysis challenges, and Dr. Stephen Spear for advice from previous studies and brainstorming.

TABLE OF CONTENTS

List of Tables

List of Figures

1) Introduction/Literature Review

Biodiversity surveys are a vital part of the biological sciences and provide valuable and necessary insight into many fields such as conservation and ecology studies (Cadotte et al. 2011; Wiens et al. 2010). Current biodiversity survey methods involve collecting and counting specimens from multiple small sampling areas within the ecosystem being studied to determine presence and abundance (Flotemersch et al. 2014; Hense et al. 2010). Determining the species composition of freshwater ecosystems has traditionally been a time and labor-intensive practice that may produce variable results depending on numerous factors such as equipment, trained personnel, and field conditions. Additionally, these surveys are usually limited to only a few focal taxa (e.g., amphibians, fish or specific macro-invertebrate groups) rather than focusing on a more complete picture of the biodiversity of the sample area. Creating an assay that can quickly, accurately, and cheaply monitor biodiversity in a more broad way (e.g., all animals) would be beneficial in studying not only species of interest, but also the ecology and interactions in the ecosystem(Lodge et al. 2012; Meusnier et al. 2008).

Biodiversity sampling in freshwater streams is an important monitoring activity that helps to conserve the ecosystem and the relationships among all members while also keeping track of natural resources important to human civilization(Cadotte et al. 2011). These surveys help to monitor and protect the natural ecosystems around us to help sustain a healthy environment that helps

support global public health (Alves and Rosa 2007; Hough 2014; McMichael and Beaglehole 2000). Losing native species is not only a possibility of ecosystem disruption, but also the loss of valuable natural resources with developed uses and those with uses yet to be discovered, especially in the fields of medicine and ecology (Alves and Rosa 2007). The continued and improved monitoring of biodiversity allows for the continued use and development of known natural resources and the discovery of new taxa and their potential uses.

Among its multiple effective uses, biodiversity sampling in freshwater systems helps to track the native organisms and identify possible invasive species (Gozlan et al. 2010; Vrtilek and Reichard 2012). The ease and accessibility of travel in the present day has greatly increased global homogenization of different biotic regions as species are more commonly introduced into habitats where they are not native (Lodge et al. 1998). Invasive species that enter biomes where they are not native can be extremely disruptive and cause the native species to be out competed, leading to the disruption of the entire ecosystem. This is especially true in closed ecosystems such as freshwater streams and river systems where some organisms, given the time and space, can cause serious disruption to all the aquatic organisms in that freshwater system (Vrtilek and Reichard 2012). This highlights the importance of monitoring and tracking native and non-native organisms in freshwater streams and the need for consistent and accurate surveying protocols that are not limited detecting single target-species of interest.

Biodiversity sampling helps keep data accessible on the natural resources that are available and make sure that they are not over harvested or abused to the

point of ecosystem degradation. Therefore, developing a freshwater biodiversity survey that includes the vast majority of species present would allow for a much broader application of collected data to preserve and protect natural resources, investigate inter-species relationships, and explore new avenues for the ecosystems direct and indirect impact on public and global health.

Current Freshwater Biodiversity Survey Methods

Traditionally, many of the organisms surveyed, outside of conservational practices, are important to monitor because of their roles as bio-indicators of ecological health. Organisms including ostracods, fish, and plankton have all been used in water quality and contamination studies to indicate health of a freshwater body based on the behavior and presence/absence of these organisms. For instance, brown trout (*Salmo trutta*) were found to be sensitive to elevated levels of cyclophosphamide, colchicine, and cadmium and were found to have micronuclei in response to elevated levels of these toxins (Rodriguez-Cea et al. 2003). In numerous other studies, aquatic macro-invertebrates were collected at various sites along a Chilean stream and significant relationships were established between the macro-invertebrate diversity and water quality as well as under multiple sets of conditions in streams in Belgrad Forest in Istanbul ([Anonymous] 1996; Cordova et al. 2009; Yurtseven et al. 2016). However, biodiversity needs to be monitored and maintained due to the complexities of ecological relationships and the drastic impact that taking one species out of an ecosystem can facilitate (Dudgeon et al. 2006; Hooper et al. 2012). Current standards set by organizations such as the Environmental Protection Agency, Geological Survey, and National Park Service help to formulate current

guidelines for the management and monitoring of freshwater resources as well as formulate fish and invert biotic indices. These methods mandate the use of electrofishing, seining, and kick sampling as the most accurate and informative surveying measures in the study of freshwater macro-organism biodiversity (Service 2006; USGS 2002).

Fish are commonly surveyed because they are the predominant vertebrates in aquatic environments and are a valuable natural resource in economic, cultural, aesthetic, scientific, and educational fields (Dudgeon et al. 2006). The current gold standard method for fish surveys is electrofishing, a method that involves electrifying the water to momentarily stun organisms, acquire data from them, and then place them back after they have recovered (Hense et al. 2010; Kulp and Moore 2000; Pusey et al. 1998). On a national scale, these surveys are mandated by guidelines set down for conserving and monitoring natural resources by the National Park Service and other federal agencies, making them an essential tool for surveying fish (Service 2006). Even though this method is more accurate than observational studies, (Bozek and Rahel 1991) there are numerous drawbacks that hinder the efficiency, accuracy, and environmental friendliness of electrofishing. Electrofishing poses a potential threat to injure or kill organisms affected by the electricity based on their age, body size, and other physical factors (Ainslie et al. 1998; Densmore and Panek 2013). Voltages that are too high run the risk of killing or injuring different aquatic species that are meant to be stunned temporarily for the purpose of inventories (Densmore and Panek 2013; Habera et al. 1996; Nielsen 1998). Additionally, organisms trapped under rocks or debris that are not collected

while the voltage is applied may drown before they recover. In addition to the danger to the organisms themselves, disturbing much of the streambed habitat by surveyors as they move along also disrupts the microhabitat for species that interact directly with the bottom. Electrofishing can also endanger the survey team members via accidental shocking. Thus, it is important to develop sampling methods that are not only more efficient but also safer for the organisms being sampled as well as the team of researchers conducting the survey.

Environmental DNA, DNA Barcoding, and Next Generation Sequencing

Environmental DNA (eDNA) can be defined as any DNA that an organism sheds into its environment whether via sloughed skin cells, defecation, urination, open wounds, or other modes. eDNA is a relatively new approach in the sampling field that has shown great promise in its efficiency, accuracy, and speed since its initial conceptual introduction in the late 1980's (Ogram et al. 1987; Taberlet et al. 2012a). The overall approach makes use of eDNA collected from environmental samples such as soil, water, or air, amplifying specific regions of interest using the polymerase chain reaction (PCR), and either sequencing the regions of interest or verifying the sequence via taxon-specific probes. The mitochondrial cytochrome oxidase region (COI), which is found in most animals and has little sequence variation within species, but significant variation among species, is frequently used (Leray et al. 2013; Savolainen et al. 2005).

The general approach of using DNA sequences to identify the species of unknown samples is known as DNA barcoding (Hebert et al. 2003). For initial

characterization, barcoding regions are normally sequenced from several individuals of each species, using individuals that have been identified by taxonomic experts, to create a reference DNA sequence database to subsequently identify and categorize individuals of unknown identity (Benson et al. 2005). Traditionally unknown organisms have been individually subjected to PCR and capillary DNA sequencing (Cohen et al. 1990). Capillary sequencing requires that a sample is from a single individual, therefore it cannot be used on samples that are mixtures of individuals, such as those collected for eDNA analysis.

The most popular current use of eDNA as a method for finding and classifying organisms uses the method of real time quantitative polymerase chain reaction (qPCR). In eDNA surveys of aquatic systems, water is non-invasively collected, filtered and DNA extracted from the filters (Pilliod et al. 2013). Specific primers are then used to amplify a target region of DNA (typically mitochondrial DNA of animals) and fluorescent probes to verify the presence of desired amplicons (Gibson et al. 1996; Heid et al. 1996). Since its development, qPCR has been widely used and implemented in eDNA studies due to its accuracy, sensitivity, and specificity in detecting specific sequences of DNA in environmental samples (Pierson et al. In Press; Pilliod et al. 2014; Wilcox et al. 2013). There are, however, limitations of qPCR, including the need for taxon-specific primers and/or probes, as well as cost. On average, samples cost between $13 and $40 dollars ($15.36 and $47.25 as adjusted for inflation in 2016) (US) each for qPCR reaction when all the costs for equipment, probes, and outsourced work are totaled (Kriger et al. 2006). Additionally, the probes used for qPCR need to be highly specific to the target

organism to ensure accuracy and positive identification of the organism(s) that are studied. qPCR assays often make use of DNA barcoding loci, taking advantage of the Barcode of Life reference database. Broad range probes to assess biodiversity of large diverse groups of organisms are not usually used, thus researchers investigating multiple organisms of interest require individually designed assays with specific probes for each organism of interest, substantially increasing costs. While being a valuable tool for researchers looking at a small number of species, this method remains an expensive option for researchers attempting to look at biodiversity of multiple different organisms. It is here that next generation sequencing (NGS) can be applied for a multitude of applications at much lower cost per sample and with a greater return of data for analysis.

Massively parallel NGS offers a way to collected DNA barcoding sequences from eDNA in a powerful and cost-effective way. During the NGS process, mixtures of molecules are either sequenced as individual molecules or individual molecules are clonally amplified prior to sequencing. Thus, the DNA sequence reads from all NGS platforms derive from individual molecules, facilitating sequencing of complex mixtures. The sheer amount of data generated by next generation sequencing is revolutionizing how researchers look at genes and genomes as well as how to apply the information collected from the sequence data. Great strides have already been made in the use of NGS as a tool for individual organism and biodiversity surveys through the use of barcoding as well as eDNA found in aquatic and terrestrial systems (Hajibabaei et al. 2007; Hajibabaei et al. 2011; Medinger et al. 2010; Shokralla et al. 2012). However, its uses are not limited to biodiversity studies or

eDNA but also to fields such as medicine, genomics, variant discovery, and looking at non-coding RNA/DNA (Mardis 2008; Metzker 2010; Morozova and Marra 2008).

Developing NGS eDNA methods to accurately and less invasively conduct biodiversity assessments of streams would benefit not only the cataloging and conservation of the native organisms present in the stream, but also to monitor non-native species. This method can revolutionize how researchers identify endangered species, track invasive species, and keep records on population or species fluctuations over time that could point to ecological shifts (Ardura et al. 2015; Wilcox et al. 2013). It would also decrease human intervention in many of these habitats and lower the current injury/mortality rates that accompany electrofishing.

Although large scale use of eDNA has been suggested and many species scale studies have been performed, attempting to do a full-scale biodiversity assessment using eDNA and NGS has only been discussed along with the numerous challenges involved that would need to be overcome to make an eDNA biodiversity assessment a legitimate tool (Ficetola et al. 2008; Goldberg et al. 2011; Jerde et al. 2011; Lodge et al. 2012; Taberlet et al. 2012b). Developing large-scale approaches should start with the most well documented organisms in monitored and regularly sampled streams to determine the accuracy of the new methods versus traditional methods. Any additional data collected on less well-documented organisms, such as hard to find invertebrates or other small organisms, could then rely upon this foundation. The potential for eDNA and NGS to change the face of freshwater biodiversity surveying with a greater return of data, cheaper cost, and a less invasive sampling protocol makes it a technique well worth further exploring and perfecting.

Study Goals

The overarching goal of this study is to determine the effectiveness and potential of eDNA surveys for estimating biodiversity and relative abundance of fish in a model Appalachian stream. With the use of broad-spectrum primers to minimize the number of PCR reactions and utilizing the regions commonly used in DNA barcoding, this study will condense the workload for the surveyor and attempt to provide results that are comparable to traditional surveys. By comparing the results of this study to documented traditional surveys performed at the same sample site, we hope to present a new assay that is proof of concept for the use of eDNA as a tool in the management of fish in Great Smoky Mountain National Park (GSMNP) streams. Additionally, these broad-spectrum primers have the capability to amplify DNA from other aquatic organisms that are not assessed as regularly or accurately. Thus, this eDNA method has the potential to help provide insight into the species presence and abundance for organisms including small aquatic invertebrates, amphibians, and crayfish. Using this method would be more time efficient, less labor intensive, and less expensive for the U.S. National Park Service (NPS) while allowing them to gather a richer dataset about the streams within national parks.

For this study, we hypothesize that species detected by eDNA will not differ significantly from species detected by traditional electroshocking method used currently. We will test detection probability for each of the species of interest. Additionally, we hypothesized that data collected from eDNA surveying will yield comparable data to population estimates generated from electroshocking.

2) Methods:

Mock Communities:

Two sets of mock communities were established for other purposes, but sampled opportunistically to help determine primer efficiency and the efficiency with which the species known to be present are observed in the eDNA data. The first set of mock communities were constructed at the United States Fish and Wildlife Service fish hatchery, Warm Springs, GA. Three 75 liter tanks were filled and stocked with six bluegill (*Lepomis macrochirus*), four stonerollers (*Campostoma pauciradii*), two goldfish (*Carassius auratus*), six salamanders , two crayfish, and anywhere from five to fifteen clams (*Corbicula fluminea*). The three tanks were identical except for the number of Asiatic clams placed in each tank, which varied by 5 individuals between the tanks. Tank one had 5, tank two had 10, and tank three had 15 clams to help assess the Leray primer efficiency (Leray et al. 2013) with varied numbers of low eDNA emitting species. These tanks were left for three days before four, one-liter water samples were collected from each tank for DNA filtration, amplification, and sequencing.

The second set of mock communities were fish tanks set up at facilities maintained by the University of Georgia Warnell School of Forestry at Whitehall Forest where the compositions of the fish tanks were known. These tanks had the described species for an indefinite period of time before samples were collected. Five, one liter samples were taken with the first coming from a tank with bluegill

(*Lepomis macrochirus*) and largemouth bass (*Micropterus salmoides*), the second from a tank of catfish (*Ictalurus punctatus*), the third from a tank of goldfish (*Carrassius auratus*), and the fourth and fifth samples from the mixed runoff water collected from all three tanks. These one-liter samples were filtered through a combination of a "coarse" filter with a porosity of 1.5um (Whatman 934-AH Glass Microfiber Filter) and a "fine" filter beneath with a porosity of 0.45um (Millipore HAWG-A0 Mixed Cellulose Ester Gridded Filter). Samples from tanks one and three, and one of the mixed runoff samples were filtered through a combination of both the "coarse" and "fine" filters. The samples collected from tank two and the other mixed runoff sample were filtered only through the "fine" filter to assess any significant impact of either filter arrangement on eDNA collection. Following filtration, all the samples were extracted and amplified using the same protocols utilized in the Warm Springs community (see DNA extraction and PCR optimization) except that we also included another fish primer set (Ward et al. 2005) in addition to the (Leray et al. 2013) primers. The new primer set was brought onto the project after the Warm Springs mock community experiment was completed to capture more diversity from future eDNA samples. Post sample collection, DNA extraction, amplification, sequencing, and analysis techniques and workflows were homogenized across the UGA mock community as well as all the field samples.

Field Sample collection:

Abrams Creek has long been a highly studied stream inside the boundaries of the national park due to its species richness and endangered species and samples were collected from two documented study sites inside the Great Smoky Mountain

National Park (King 1942; Lennon and Parker 1959). Due to the numerous endangered fish species found in its waters, along with the abundance of general biodiversity that can be found, Abrams Creek remains a highly studied and documented site (Gibbs et al. 2014; Shute et al. 2005). Within the boundaries of the park, Abrams Creek runs from the Southwest corner of the park in Tennessee, through Cades Cove, and up until the state line with North Carolina near the center of the park. The two sites selected for study are both located in Cades Cove with the first being downstream from the fork with Mill Creek on the Western edge and the second being located on the Eastern side of the cove.

Two methods were used to compare the DNA capture efficiency. 1) on-site filtering and 2) lab filtering. The method of on-site filtering involved using a peristaltic pump run by car batteries to draw water directly out of the stream using clear, half-inch tubes, through a capsule containing two filters, and out another tube into a bucket. In the filtering capsule, the filters used were a "coarse" filter with a porosity of 1.5μm (Whatman 934-AH Glass Microfiber Filter) and the one beneath a "fine" filter with a porosity of 0.45μm (Millipore HAWG-A0 Mixed Cellulose Ester Gridded Filter). These filters were stacked with the idea of removing overly abundant microorganisms (i.e. algae) from the "fine" filters to avoid non-targeted DNA overshadowing any target organismal DNA present. Suction was provided by a pair of peristaltic pumps connected to separate car batteries on the edge of the stream with the intake tubes for each pump strung out at two different points across the width of the stream. These intake points were attached to a rope strung across the stream and anchored by carabineers attached to a floating PVC pipe with a mesh

screen to avoid the intake of large debris found in the stream. For each sample, approximately one liter of water was drawn across the pair of filters and measured by the output into a bucket or an empty Nalgene container. Once enough water had been filtered, the filters were removed using sterilized tweezers from the filter capsule and stored separately in 1.5mL tubes in 95% ethanol and frozen until DNA extraction.

The second method of sample collection (practiced only on one of the two study sites) was performed by taking one liter Nalgene bottles and filling them with water, directly sampled upstream from where the sampler stood in the stream. Once the bottle was filled, a small amount was poured back out of the bottle and 33mL of ethanol and 1mL of sodium acetate were added to act as preservatives for the DNA until the samples could be brought back to the lab. In the lab these water samples were then vacuum filtered through a MicroFunnel Filter Unit with GN-6 Membrane Gridded Filter that has a porosity of 0.45μm. This method was used only on the first sampling trip and failed to produce any DNA product after amplification and so was not employed for the duration of field sampling. All of the filters from all of the collection techniques were stored in enough 95% ethanol to cover the filter in 1.5mL tubes and kept frozen until DNA extraction.

DNA Extraction:

The DNA was extracted from the filters using QIAgen DNeasy Blood and Tissue DNA extraction kits along with QIAshredder tubes, following an adjusted protocol for filter DNA extraction. The first step involved taking one half of the filter out of the ethanol it was stored in and ripping it into fourths to let dry overnight.

The following day, lysis buffer and proteinase K were added to digest the organic material on the filter and break open individual cells to make the DNA accessible for extraction. The filters were left overnight in the lysis solution to prepare for the DNA extraction the next day following the manufacturer's directions ((Qiagen 2015). To extract the DNA from the lysed solution the following day, the filter and all of the lysis solution were pipetted into QIAshredder spin columns and were spun at 8,000rpms for 5 minutes to draw the liquid off of the filter and into a collection tube. Once the liquid is removed, the filter was discarded and 200uL of another lysis buffer was added to the solution and incubated at 70 degrees Celsius for 10 minutes. Next, 200 uL of 100% ethanol was added to act as a preservative for the now exposed DNA in the solution. The solution was then filtered through a QIAmp spin column where the DNA binds to the filter in the column while all the contaminants pass through and are discarded. The filter column was then moved into a new collection tube where it was washed with another buffer and then moved into another collection tube where it was washed again with a secondary buffer to remove remaining contaminants and enzyme inhibitors. The DNA was then eluted in 100uL of EB buffer, measured for concentration using a Life Technologies Invitrogen Qubit 2.0 Fluorometer and diluted or concentrated as needed to obtain a working concentration and stored at -20C.

Primer Selection:

Primers for this study were selected to be as broad as possible and amplify a vast range of organisms found in freshwater aquatic habitats. All the selected primers amplify the COI region of the mitochondrial DNA (mtDNA). The first primer

set was designed by Leray and colleagues (Leray et al. 2013) and spans a 313 bp length that amplifies DNA from organisms spanning much of metazoan diversity. The second set was designed by Zeale and colleagues (Zeale et al. 2011) and amplifies a target segment of 157 bp that span 11 arthropod orders which include many invertebrates with an aquatic life stage. The third set was designed by Ward and colleagues (Ward et al. 2005) and amplifies a target segment of 655 bp region from 207 different fish species and was used by Hubert and colleagues (Hubert et al. 2008) to amplify the target DNA of 194 species of freshwater fish. By using all three of these primer sets, we could then compare and contrast results obtained.

PCR Optimization:

PCRs were carried out using Kapa Biosystems reagents and Bio Rad T100 Thermal Cyclers. The PCR mix of reagents was from the Kapa Biosystems 2G Robust HS PCR Kit and was optimized to a quarter of the recommended reaction volume for a total final volume of 12μL per reaction. In one such reaction there is 2.5μL Kapa 2G Buffer A, 0.25μL 10mM dNTP mix, 0.1μL Kapa 2G Robust Hotstart, 4.15μL molecular grade water, 1μL of the forward primer, 1μL of the reverse primer, and 3μL of the sample DNA. This reaction could then be run individually for different primer sets using DNA from the same sample to amplify a broad spectrum of DNA collected on individual filters.

Thermal cycler conditions followed the KAPA protocol of 95˚C for 3 minutes, then 25 or 35 cycles of 95°C for 30 seconds, 60°C for 30 seconds, and 72°C for 30 seconds, followed by a final incubation of 72°C for 1 minute. We used 35 cycles for the Leray and Zeale primers and 25 cycles for the Fish primers. The PCR products

were then loaded onto a 1.5% agarose gel where 4μL of finished PCR product and 2μL of loading dye were run at 100 volts and 40milliAmps for 40 minutes next to a Bionexus Hi-Lo DNA marker. Successfully amplified products were viewed under UV light where they produce a band at the spot indicated by the marker ladder to be the length of the target amplicon for that primer.

Samples with a product of the expected size, were cleaned by mixing the PCR product with an equal volume of diluted Sera-Mag Speedbeads (Rohland & Reich, 2012). These beads bind to the amplicons and draw them to a magnet, while the remaining DNA fragments formed from primer-dimers and contaminants are washed away. These beads are only active in a PEG solution; when that liquid is drawn off while the beads are on the magnet, DNA above a size determined by the percentage of PEG in solution is trapped on the surface of the beads until after an ethanol wash to remove any remaining contaminants. Finally, TLE is added in the original volume of the PCR product solution and the beads release the amplicon DNA back into the solution (while the other constituents of the PCR were washed away). This cleaning step can be made easier by pooling PCR products from the same sample regardless of their primer set because after the PCR, the target DNA has already been amplified and so cannot be contaminated by other amplicons made with different primer sets. This allows for the cleaning of multiple PCR products at once and simplifies the second round of PCR's for all of the samples.

The second round of PCRs for all the samples is not for the purpose of amplification but rather for tagging the amplicons from each individual sample with a unique sequence of nucleotides (i.e., adding indexes) and adding the remaining

adapter sequences for Illumina libraries (Glenn et al. In Prep). This system allows samples destined for sequencing to be individually tagged with up to four unique tags, which allows for an immense number of samples to be sequenced at the same time (Glenn et al. In Prep). This second round of PCR adds the iTru adaptors that are compatible with Illumina sequencing instruments. For this round of PCR, a Kapa Hifi Hotstart PCR Kit was used as well as the same Bio Rad T100 Thermal Cycler model from the first round PCR. These reactions were run at a total volume of 25µL made up of 5µL Kapa Hifi Buffer, 0.75µL 10mM dNTPs, 0.5µL Hifi Hotstart, 10.75µL molecular grade water, 1.5µL of the forward primer, 1.5µL of the reverse primer, and 5µL of the cleaned first round PCR product. This mix is made individually for each different sample and each sample will get a unique combination of the iTru5 and iTru7 second round primers that contain their indexing tags known as the i5 and i7 respectively. This ensures all the samples are uniquely tagged and can be demultiplexed by their outer tags when the sequencing run is complete. Thermal cycler settings were 95°C for 3 minutes followed by 10-12 cycles of 95°C for 30 seconds, 60°C for 30 seconds, 72°C for 30 seconds, followed by 72°C for 5 minutes, and an infinity hold at 10°C. Samples were again run on a 1.5% agarose gel with a DNA Marker Ladder to determine if the reaction worked as expected. Successful samples were then pooled together and cleaned using the diluted Sera-Mag Speedbeads as before. Finally, the concentration of each sample was measured using the Life Technologies Invitrogen Qubit 2.0 Fluorometer prior to pooling for sequencing.

Sequencing:

Samples were sequenced at the Georgia Genomics Facility using an Illumina MiSeq sequencer using v3 600 cycle kits for paired end 300 reads. For the Warm Springs mock community samples, 50,000 reads per sample was targeted whereas for the UGA wet lab mock community samples and the Abrams Creek field samples, that number was increased to 100,000 reads per sample. Samples collected from the field were assigned 100,000 reads per sample for each successful primer set. Due to a sequencing bias that favors shorter DNA fragments over longer ones, the samples pooled together on the same sequencing run were mixed in proportions favoring the larger fragments to achieve uniform coverage of all the DNA. Sequences generated from the MiSeq were received via Illumina BaseSpace.

Analysis:

The sequences were imported into Geneious version R8.1.6 (Kearse et al. 2012), a software package developed for organizing, annotating, and searching raw sequences, including next generation sequencing platforms. Using Geneious, the forward and reverse reads were paired, the sequences are trimmed to remove the primer sequences from the amplicons, and fragments with the proper amplicon length were extracted from each set of sequences inside a single sample. After the sequence preparation in Geneious, Qiime software (Caporaso et al. 2010) was used to obtain the number of species represented in each sample and the percentage of DNA in that sample that those species comprise. From Geneious, the sequences for each sample were exported in FASTA files that were combined into a mapping file in Excel where Qiime combines all the sequences into a single file that is categorized

by the mapping file. Qiime uses this file along with a reference file of known species-specific sequences and a taxonomic file that holds all the taxonomic organization data for the reference species to map the sequences from the individual samples to their species of origin. The similarity percentage for the matching of unknown sequences to the known reference database was manipulated from the default setting of 90% to a higher specification set at 94%, to produce more confidence in the matches. Once the matching is complete, Qiime can produce numerical outputs that determine how many sequences in each sample match to a certain sequence by taxonomic level of organization from kingdom to species. For this study, the results were examined by viewing the matches in the species level of taxonomic organization.

These numerical results were then graphically displayed by creating line plots in Microsoft Excel that display the detection probabilities of each monitored species using both the traditional electrofishing technique and the eDNA technique. These detection probabilities were calculated separately for each sampling technique and then compared as different points and lines on the same graph. For the eDNA samples, detection probability was calculated by dividing the number of samples in which the species of interest was detected by 20 or more reads by the total number of samples taken from that site and amplified with primers able to amplify that species. This technique yields values between zero and one for each species, with more detectable species having values closer to one and less detectable species having values closer to zero. These values were calculated at 94% similarity matching strength in Qiime and are displayed to highlight differences in

the results based on the sample site and primer. For invertebrates, any significant detection of 20 or more reads of a genus was determined to be a significant detection and the number of unique genera detected by eDNA between Abrams Creek sites 3 and 4 was compared to the total number of observed or recorded genera of invertebrates present. In the same manner, species detection probability was calculated for electrofishing by dividing the number of passes a species was detected on by the total number of three passes in the used electrofishing survey. This also yields a value between zero and one following the same principles as the detection probability for each species from the eDNA data. These data can then be plotted together with detection probabilities for each species from the historical data kept in the GSMNP records for a direct comparison between the current method in use the past few decades, and the newer eDNA method.

Additionally, the number of reads for each species of fish from the samples from each site were translated into percentages by taking the total number of reads from a species and dividing it by the total number of reads assigned to any of the 12 species. Once done for all the samples collected at that site, the percentages from each sample were compared to the percentages calculated from the population estimates calculated from the shocking data. These percentages were also derived from taking the total number of fish estimated to be present for a single species and dividing it by the total number of estimated fish. All of these percentages for the individual eDNA samples as well as the shocking estimates were compiled into a bar chart in Microsoft Excel to provide a simple visualization of the species distribution according to each method. We then averaged all of the eDNA sample percentage

results and converted the averages for each species to decimals along with the shocking population percentages. Each species was then ranked from the lowest to the highest abundance estimate with a value from 1 to 12 for each of the two methods. These two columns of rankings were then compared via Spearman's Rho test to determine the strength of the correlation between the population estimates of the two methods with a corresponding P-value to determine statistical significance.

3) Results

Mock Communities

Results from the mock community in Warm Springs (Figure 1) showed that almost all of the species present in each community were detected by eDNA, although the vast majority of the reads from each of the samples in each community were unassigned (i.e., they did not match sequences in the reference list used). There were no reads detected for the salamanders in any of the three tank communities and crayfish were detected in only one of the three tanks and only with a relatively low detection probability.

Results from the UGA wet lab mock communities indicated that each of the fish species was detected in each of the samples at a detection probability of at least 0.5 or higher (Figure 2). Once again, the majority of reads from each of the samples were unassigned. Interestingly, every species was detected in every sample, which does not reflect the separation of the species in the tanks with the exception of the two samples taken from the mixed runoff water. However, water among the fish tanks at the facility sampled is shared which explains the presence of DNA from each species in samples from each tank.

Primer Efficiency

The efficiency of the primers selected for this study was tested by utilizing the mock communities, which had known species compositions. The eDNA collected

from these communities was compared by contrasting different tanks with the same species, different tanks with different species, different primers amplifying the same DNA, and different filters to extract the eDNA from the water. The results for the mock communities are displayed in Figures 3 and 4 for the Warm Springs and UGA wet lab mock communities respectively. The Warm Springs primer efficiency deals only with the (Leray et al. 2013) primer and highlights the high number of unassigned reads for each sample in Figure 3 (an average of 76% of reads went unassigned across all three tanks). The average percentage of reads belonging to any individual species in the mock community was capped at 17% of the reads belonging to the stonerollers with all of the other species' percentages falling below 6%.

The UGA mock community results also identified a high number of unassigned reads for each of the samples but was significantly different when examining the two different primers. Figure 4 highlights the differences between the two sets of primers used and how drastically they differ at amplifying different species, as well as their effectiveness in amplifying those species. The Leray primer had a much larger percentage of unassigned reads (95%) when compared to the Fish primer (45%). Additionally, the eDNA captured on each of the two types of filters is compared and emphasizes the differences in DNA capture for each of the individual species as well as unassigned reads. From the five samples taken, after combining the reads from both primers and filters, the highest percentage of reads allotted to a single species is 23%, while that number jumps to 41% allotted to a single species when examining only the reads from the Fish primer set.

When the results for both mock communities are compared in Figures 1-4, it can be seen that the addition of a second primer set allows for more amplification and that using a combination of "coarse" and "fine" filters or just the "fine" filter does not outstandingly impact eDNA capture. The addition of a supplementary primer set in the UGA wet lab mock community allowed for more diverse species detection and at much more significant levels than using only the one primer set in the Warm Springs mock community.

Abrams Creek Field Samples

Species detectability was calculated for 15 species of fish found in Abrams Creek at sites designated 3 and 4 by the Great Smoky Mountain National Park (Figure 5). Of the 15 known species in Abrams Creek sites three and four, only one species detected by shocking was not detected in the eDNA and of the seven known species not detected at all by shocking, none were detected with eDNA either (Tables 1+2). Each site has 12 species of fish that are monitored and known to occur on that stretch of the creek, with three species unique to only one of the two sites and the majority of the fish being represented at both study sites. Detection probabilities were calculated from the traditional shocking data by noting the presence or absence of each species and the number of passes on which it was detected in the standard three-pass shocking run used by the park fisheries branch. Using this three pass protocol, species that are detected in all three passes earn a detection probability of 1, species detected in 2 of 3 passes earn a 0.66, and species detected in 1 of 3 passes earn a 0.33. In the same way, species detected from the numerous eDNA samples at each site were assigned a detection probability based

on the number of samples that they were detected in divided by the number of samples collected wherein they could be detected. Species' common and scientific names that are referenced to by abbreviations in the data are noted in Table 3.

The results illustrate that the eDNA detection probabilities tend to follow the approximate detection probability offered by the shocking data from this year. Results from Abrams Creek Site 3 (ABC3) are displayed in Figure 6 and the results for Abrams Creek Site 4 (ABC4) are illustrated in Figure 7. While there were substantial differences between the different primers used to amplify the DNA for certain species, at least one of the primers amplified every species within a comparable range to the shocking detection probabilities, with the exception of two species in ABC3. When compared to the historical data, eDNA detection probabilities from at least one primer set were also following the general detection trend for each species as calculated from the historical data.

Additionally, data was collected for the invertebrate community present in Abrams Creek with the broad taxa coverage of the Leray and Zeale primer sets. Thirty-five unique genera of invertebrates out of 184 observed genera in Abrams Creek were detected between ABC3 and ABC4 with 29 genera observed at ABC3 and 24 observed at ABC4 (Figure 8).

Abundance estimates

Abundance estimates for species detected at both sites three and four on Abrams Creek were calculated by examining the percent of assigned reads assigned to each of the species of interest at each site. These percentages were then compared to the species population estimates calculated using the data from the

shocking survey, which were also converted to percentages (Figures 9+10). Both sites three and four show noticeable variation between the population estimates calculated from shocking and the percent of reads assigned to individual species in the eDNA data. Differences among eDNA samples from the same site were also noted but less pronounced than the differences between the eDNA and shocking abundance estimates. For the percentages from eDNA samples from ABC3, the eDNA samples have rainbow trout (Oncorhynchus mykiss), blacknose dace (Rhynichthys atratulus), and white suckers (Catostomus commersonii) as the three species having been assigned the most reads. The estimates calculated from the shocking data, once also converted to percentages, has the blacknose dace (Rhynichthys atratulus) as the most populous species followed by the rainbow trout (Oncorhynchus mykiss), and then the creek chub (Semotilus atromaculatus). In Figure 9 it can be seen that from the eDNA samples from ABC4, the rosyside dace (Clinostomus funduloides), stoneroller (Campostoma sp.), and blacknose dace (Rhynichthys atratulus) are the three species with the most assigned reads. The estimates calculated from the shocking data have the blacknose dace (Rhynichthys atratulus) as the most prevalent species followed by the stoneroller (Campostoma sp.) and the rosyside dace (Clinostomus funduloides).

To fully understand the relationship between the shocking population estimates and the eDNA read percentages, the values for each site were converted to decimals and ranked from 1-12 for the 12 species in each site. Using a Spearman's Rho test, the relationship between the species rankings' from the shocking and eDNA methods was summarized by the calculation of a correlation coefficient and a

corresponding P-value. For ABC3, the correlation coefficient was calculated to be 0.852 with a P-value of 0.0004 (Table 4) while for ABC4, the correlation coefficient was calculated to be 0.901 with a corresponding P-value of <0.0001 (Table 5).

4) Discussion

Primer efficiency

The primers in this study were selected to amplify as many organisms as possible that have been studied, collected, or surveyed in Abrams Creek in the past few decades. This requires the targeting of a species-specific region of DNA that allows for very few primer sets to bind and amplify DNA from a multitude of different species. There has been a recent large-scale effort to categorize species by specific regions of DNA unique to each species and one such effort known as the Barcode of Life project (Savolainen et al. 2005) has been making use of the Cytochrome Oxidase 1 (COI) region of the mtDNA for species barcoding. Species barcoding works very much the same way as barcoding items at a grocery store works, meaning each barcode, or in this case the COI region of the DNA, corresponds to only one species and determining that species is as easy as comparing the DNA to the reference database. This same region of the DNA was then selected as the target region for this study and three sets of primers chosen that amplified short segments of CO1 DNA. A large part of selecting this region of the DNA for this project is that much of the genome data currently available on many organisms is limited and the CO1 region is quickly becoming known as a reliable genetic marker to differentiate organisms on the species level (Che et al. 2012; Derycke et al. 2010; Hebert et al. 2003). The benefits of just a few primer sets that can amplify a broad range of organisms reach to the logistical aspects of research as well as the practical

applications. While keeping the costs down by only purchasing a limited number of primers, researchers will still have the ability to study a diversity of organisms by using the same primers and thereby, the same workflows, for each organism.

Target species amplification

While these primers do standardize the workflow for all the taxa studied in this project and amplify a very broad range of organisms, the target organisms comprise a very small piece of the total amplifying capability of the primers. The mock community experiments at Warm Springs and the UGA wet labs demonstrated the primers ability to amplify numerous species from different taxa but also highlighted how many background organisms were amplified as well. For samples taken from both sites on Abrams Creek, the majority of reads went unassigned to the reference database provided to the Qiime software and belonged to a broad span of organisms ranging from fungi to unicellular eukaryotes. In other words, the primers selected for this study are too universal to only be counted on to amplify the ~330 species provided in the reference database of known organisms in Abrams Creek. While the ability of the primers to amplify many other organisms is impressive and seems useful, it can be detrimental to data collection on targeted species due to all the background noise of DNA that mostly belongs to non-targeted organisms living in the stream. This may be attributed to the primer-template mismatch that is often associated with universal primers targeting species-specific regions of the DNA such as the mitochondrial CO1. In many cases, primers have a higher tendency to bind the DNA of some species than they do other species and this can be very detrimental to an accurate representation of the population by DNA

collection and amplification (Pinol et al. 2015). For the purposes of this study, the primers did well to amplify many of the target species but the broad sweep of other organisms amplified took reads away from the species of interest and minimized the usable data.

On average, approximately 3.3% of the reads from the samples collected from ABC3 were assigned to the ~330 species in the reference database. From the samples collected from ABC4, the average percentage of reads assigned to the reference database totaled approximately 26.9%. This large disparity between the reads assigned at each site may be due to any number of reasons ranging from site differences between ABC3 and ABC4 at the times of sampling, to the month of sample collection, to the amount of water at each site. Additionally, the samples from ABC4 were sequenced initially by combining the products of the three individual primer PCR's with ~33,333 reads per primer set while samples from ABC3 were allotted 100,000 reads per primer set per sample. In this case then, the high number of reads per primer per sample from ABC3 may have contributed to the higher number of unassigned reads considering the extent of the primers' universal capabilities. Many of these unassigned reads when BLAST searched against the entire Genbank online database (Benson et al. 2005) did not return any conclusive results that were strongly matched to any specific taxa or species group. Many of those unassigned sequences failed to match to any known sequence on Genbank with a matching percentage higher than 70%, suggesting a lack of genome data for many lesser-known, smaller organisms. To maximize the data available for analysis, eliminating the amplification of unwanted sequences would drastically

increase the number of reads assigned to the database. One possible way of achieving this would be to run a very similar survey project with blocking primers to eliminate DNA from untargeted taxa and ensure maximal amplification of the species of interest (Vestheim and Jarman 2008). Using this method, primers are designed to block a certain species or taxa based on an extension of the primer that disallows annealing to an unwanted DNA strand. If universal primers could be engineered to a region of the DNA that is very taxa specific, it is also possible that they may become "universal" blocking primers to stop an entire range of undesired taxa from being amplified.

Combination of Primers

While finding one primer set that can successfully amplify any species of interest is highly unlikely, the combination of the three primer sets on this project proved highly effective. With one very broad range primer set (Leray et al. 2013), one set devoted specifically to fish (Ward et al. 2005), and one devoted to 13 major orders of invertebrates with aquatic life stages (Zeale et al. 2011), the entire scope of known macro-organisms in Abrams Creek was covered. The impact of multiple primers is readily seen in the results of the fish detections at each site where one primer may have performed poorly while another primer was excellent for one species and the trend is then reversed for the next species. In other studies, combinations of primers have proven effective for examining DNA from multiple species and taxa (Hajibabaei et al. 2012; Robeson et al. 2009) and can greatly increase the chances of target DNA amplification where examining diversity is the goal of the study. Additionally, the use of multiple primers decreases the amount of

false negatives from DNA unable to be amplified by a single primer set by providing double or triple coverage of the same group of taxa. In this study the benefits of multiple primer pairs becomes apparent when examining species such as the creek chub (CKC, *Semotilus atromaculatus*) which at both ABC3 and ABC4 failed to amplify in any samples using the Leray primer (Leray et al. 2013), but successfully amplified in all of the samples using the Fish primer (Ward et al. 2005)(Tables 1+2). Furthermore, species like the rosyside dace (RSD, *Clinostomus funduloides*) amplified much more frequently and abundantly using the Leray primer than it was able to amplify when using the Fish primer. These detections can only be seen as a result of using multiple primer sets to increase coverage of the targeted taxa and ensure the highest rate of detection by ensuring double coverage of target organisms. The Leray and Fish primers both amplify the fish species of interest while the Leray and Zeale primer (Zeale et al. 2011) both have the capability to amplify the large numbers of invertebrates known to be present in Abrams Creek. Of the 35 unique genera of invertebrates identified at ABC3 and ABC4, 15 of them were detected solely by the Leray primer and not at all by the Zeale primer. Most of the data collected in this study was only an outcome of the combined amplicons from three different primer sets that span not only a broad range of organisms, but many of the same organisms.

Head to Head Comparison

Despite the high numbers of unassigned reads, the eDNA samples from both ABC3 and ABC4 were shown to be strongly correlated to the results gathered from the shocking data at each of the sites. Speaking in terms of presence/absence only,

the eDNA samples detected all but one of the species detected by the shocking surveys and did not detect a significant number of sequences from any of the species not detected by shocking (Tables 1+2). In terms of abundance and how each method ranked the species in terms of prevalence at the site, the data from both sites still showed a very positive correlation with the shocking data (Tables 4+5). As the current gold standard remains the method of shocking, these results do validate the further pursuit of eDNA as an invaluable tool in the process of creating more accurate and less environmentally invasive survey techniques. Although many of the estimates for each species individual populations differ by what seems a large degree from the shocking data, it is important to note that the shocking population numbers are also just estimates. These estimates are calculated by looking at the numbers of each fish species collected from a site with a measured area and then applying those estimates to the rest of the stream site. While shocking does well to prove the presence or absence of species by physical identification, the population data calculated from those presences/absences are nothing more than our best guess based on what was surveyed. If the only definite data we can collect from shocking surveys is the presence or absence of species and anything past that is the best estimate, then the eDNA surveying method performed comparably to the definite data.

Environmental DNA also delivers an invaluable conservation service by providing a means of early detection for new species in an ecosystem, whether invasive or endangered. While shocking will also eventually detect invasive or unusual species in a certain ecosystem, eDNA allows for the possibility that the

species can be detected and dealt with before it is fully established in the watershed. In a study done by (Wilson et al. 2014), eDNA sampling was combined with electrofishing techniques to monitor the possible encroachment of Asian carps into the Ontario waters of Lake Erie and Lake St. Clair. While all their samples came back negative for Asian carp, using both eDNA and electrofishing, their study highlights the benefits of the use of eDNA combined with current study techniques to investigate the widespread issue of invasive species by doubling chances of detection. Additionally, eDNA allows for species detection where locating the physical organisms may be impossible. In another study done by (Ardura et al. 2015), eDNA was collected from the ballast water used in ocean going cargo ships to locate and confirm the presence of European mudsnails in the ballast tanks. These species are then carried worldwide allowing for new avenues of speciation and big problems for areas of the globe where these snails are not native. This study validates the use of eDNA as a screening tool for biodiversity in shipments that cross the globe and carry the risk of introducing non-native species into new environments which can devastate the native ecosystem.

While each method has merit, the combination of eDNA and shocking techniques for the use of biodiversity surveying may provide more reliable data, much like the principal of combining primers to ensure multiplied coverage of each species' DNA. Looking at the positive correlation calculated between both methods using the Spearman's Rho test, it is clear that their data outputs are very similar and can be used together to solidify any findings. Even when both methods do not get

similar results it would be worthwhile to explore and determine what the cause of the discrepancy is and possibly reveal new information.

eDNA Bias

While the correlations between the two methods were strongly positive, it must also be noted that the results from the eDNA samples have the potential to be directly and indirectly manipulated at certain points in the workflow. During the initial eDNA sampling, the possibility of acquiring more of a specific organism's DNA because of its location in the sampling site, its physiology, or even its most recent bowel movement are all factors that may end up indirectly influencing the results. Similarly, the use of two filters may also have influenced the quantity of DNA that was extracted from the filters as only DNA from the bottom, "fine" 0.45 micron filter was collected as there appeared to be an inhibitor that prevented successful PCR reactions with any DNA collected off of the top, "coarse" 1.5 micron filter. These coarser filters had significantly higher amounts of DNA after extraction as measured by the Qubit fluorometer than the finer filters underneath and so many may have contained sequences from species whose DNA did not pass through to the fine filter below the coarser filter. The most apparent direct method of manipulation is the primer(s) selected to amplify the collected DNA. As seen in the results from this study, one primer set however "universal" it may be, still has the possibility to miss certain species that may end up skewing the final results by missing an organism that was actually present. It also leads to the indirect source of bias that comes from the PCR reaction itself and the affinity of primers to bind DNA from certain species over other species (Pinol et al. 2015). In a study published by (Bellemain et al.

2010), the internal transcribed spacer (ITS) region of DNA was used to study fungal diversity and significant PCR bias was described between primer sets and their affinity among different species. The study concluded that the use of combinations of primers would help to cut back on the PCR bias of individual primer sets.

Additionally, a much more easily manipulated source of bias is the Qiime software used in the eDNA sequence analysis (Caporaso et al. 2010). Using the command line format, Qiime allows the researcher to set a matching percentage for grouping similar sequences as well as matching sequences to the reference database. By manipulating these matching percentages, the researcher can essentially alter the outcome of the analysis. For this study, operational taxonomic units (OTU's) were grouped using the default 97% matching criteria and the matching to the reference database was set at 94% similarity due to no fish species of interest containing COI sequences that matched within a 94% similarity range. Once the data was processed through Qiime, the results were scrutinized to help eliminate false data. To be classified as a positively assigned read, a sample had to have 20+ sequences registered for a single species for it to count as a detection. This was done with the hope of eliminating false positives that may have arisen from cross contamination with sources of DNA not originally collected from the study sites (i.e., lab contamination or misassignment of samples from the Illumina).

Shocking Bias

However, similar to eDNA, shocking methods also have potential sources for error and bias at different stages in the survey workflow. The experience of a crew member can play a large role in the accuracy of the survey as the more time and

practice one has had using an electro shocker, the better the performance tends to be. The percent of the fish actually removed from the stream during a multi-pass electroshocking run can vary substantially but also never really exceeds ~75% of the fish estimated to be in the stream (Penczak 2013; Peterson et al. 2004). Additionally, the most critical part of the survey is identifying the fish, which can be very easily misidentified by an inexperienced crewmember and even by a seasoned ichthyologist. In a study performed by (Tillett et al. 2012), scientific observers were asked to identify five species of morphologically similar shark species in a northern Australian fishery. The results indicated an overall misidentification rate of 19.8% averaged among the misidentification rates collected for each of the five species and highlighted the need to incorporate species identification errors into survey results. Among the fish species surveyed in Abrams Creek, there are numerous species which are morphologically similar and have a high potential to be misidentified. One genus in particular, the stonerollers (*Campostoma sp.*), has very morphologically similar species that could be easily misidentified by even a seasoned park crewmember. In the data from this study, the visual identifications made by the shocking team identified the stonerollers captured during the survey as central stonerollers (*Campostoma anomalum*) while the eDNA data assigned the collected DNA more closely to the bluefin stoneroller (*Campostoma pauciradii*). With the potential for visual misidentifications, genomic identification holds promise to more accurately identify present species and distinguish between morphologically similar organisms.

Current Limitations of eDNA

One of the current most crippling limitations to the implementation of eDNA is the amount of available reference data on the genomes of different organisms. As research progresses and more researchers sequence organism genomes they can compile them in large online libraries such as Genbank (Benson et al. 2005) where other researchers can access them. However, the data currently available pales in comparison to the unknown genome information for many organisms, making a reference database something that can be difficult to build. If samples of the species of interest are available, the researcher can sequence DNA extracted from the tissue and use that as a reference but then limits the references to however many species the researcher can collect tissue samples from. Having a publically available genome library is essential to compiling a more complete reference database that can encompass species seen or known to be present in the study site that are not available for tissue sampling, or organisms that are unknowingly present. Additionally, there are many concerns over the validity of eDNA due to things such as the range of eDNA flow and that it can not be used for all aquatic species (Jane et al. 2014; Rees et al. 2015). While these concerns are issues that do need to be further studied and addressed, they are not limitations that disqualify the use of eDNA in the realm of aquatic biodiversity surveys. The current bias involved in the sampling, amplification, and analysis of eDNA samples may also limit the convincingness of the resulting data to the scientific community. However, these issues also highlight the benefits of combining both methods when examining

biodiversity, in order to maximize the accuracy of the results and research any controversies between identifications.

5) Future work

To further validate this eDNA method as an accurate and reliable surveying tool, there are measures that can be taken to improve the performance and data output. First, while the COI region is reliable for species identification, using another commonly studied marker in the genome, such as the 12s region, in combination with the COI region would provide reassurance of the results. Utilizing a region of the DNA that may also have variation among individuals within species would allow the possibility of determining the number of individuals that the collected DNA represents. Combining multiple primers with multiple regions of DNA would greatly reinforce the results gathered from eDNA by validating species presence by primer and region of the DNA. It would also help to construct more accurate population estimates without having to collect the specimens themselves. Adding more primer sets to ensure double or even triple coverage of species across different regions of DNA would lend more credibility to this method of surveying.

However, another issue that would need to be addressed is primer-template matches and mismatches among different species while using the same broad range primers (Pinol et al. 2015). To build the most accurate reference database possible, any species of interest from Abrams Creek should be sampled and sequenced to construct a known database based on organisms present in the ecosystem and species populations. Then these DNA samples, taken directly from the populations of interest, can be screened with multiple primer sets to ensure that each species'

DNA can be amplified efficiently. Fish species of interest should have fin clips taken and sequenced to build a database and other macro-invertebrates can be collected and have DNA extracted as well to provide the reference database with the most exact species information possible. Species for which primer-template mismatch is problematic can be tested among multiple primer sets and may be represented by one of numerous primers used.

In addition to having multiple primer sets over multiple regions of the DNA, engineering these primers to block unwanted taxa would be beneficial in assigning the majority of reads from each sample to the reference database. If the primers are only able to amplify the DNA of the target taxa, then the vast majority of reads sequenced from a successful PCR will pertain to the study. While this presents unique challenges for primer-template match for universal primers to allow the binding of numerous different taxa while still preventing binding of other taxa, it would be a significant step forward for the development of this method. Deeper sequencing would be an added benefit of the removal of background DNA as samples could be allotted more reads if the results would e from the targeted DNA.

Further studies could also be performed to analyze the flow, distribution and range of eDNA in the specific waterway or stream of interest to examine its effects on survey results. Understanding how DNA moves through the study site under similar or different conditions would be a vital part of establishing the credibility of eDNA as a legitimate surveying tool. Similar to studies done by (Jane et al. 2014) to measure the capture efficiency of DNA and how that efficiency varies at range,

further studies done along the same lines in likely study sites could prove invaluable.

6) Conclusion

In comparing the currently used method of electroshocking with eDNA collection and analysis to survey freshwater organisms in Abrams Creek in the GSMNP, the results support the ability of both methods to produce positively correlated results. As the current gold standard, shocking remains the most acknowledged and supported method of conducting freshwater fish biodiversity surveys but has numerous drawbacks that should be acknowledged. Environmental DNA offers a union between accurate survey data and a minimally invasive surveying protocol that combined, can offer a wealth of data with a fraction of the work. With the interchangeable options offered by multiple primer sets being used in numerous combinations, eDNA with paired next generation sequencing can reach far beyond the information gathered from shocking surveys. With the primers selected for this study, eDNA was able to examine not only fish species present, but numerous other taxa of freshwater vertebrates and macro-invertebrates with a far less invasive protocol. There is still much work to be done to resolve the challenges faced by a universal eDNA method but this study has shown its effectiveness in doing as much and more than current methods.

Referencing the Spearman's Rho test results, it is apparent that the two methods are closely positively correlated and while not a perfect comparison, it points toward the realistic adoption of widespread eDNA surveying in the near future. The combination of shocking and eDNA could offer a surveying protocol that

is overall less invasive and more informative by alternating frequent eDNA collection with infrequent shocking. This would allow both methods to reinforce each other by relying more on eDNA and following up with visual checks by shocking every few years to validate the finding from the collected DNA. The use of eDNA can also expand the understanding of the entire community being studied by examining the information gathered from the DNA of more overlooked organisms instead of solely examining a single taxa.

While there is still work to be done to improve the accuracy and reliability of eDNA as a tool for biodiversity studies, there is much promise that it can provide data of similar or greater utility to data collected from electroshocking. Streamlining and optimizing the use of eDNA will provide a more accurate survey at a lower cost and with a far less environmentally invasive protocol than currently employed methods. It will also allow for easier sampling in hard to reach places and isolated locations by a smaller team and greatly expand the work accomplished when there are limited personnel and resources. The applications of eDNA are also not limited to biodiversity surveying but has the potential to branch into fields of study that cannot be accessed by current surveying methods. Further study of eDNA and its uses in combination with NGS will only expand our ability to understand and interact with the organisms and environments around us.

Literature Cited

[Anonymous]. 1996. Water quality assessments - a guide to use of biota, sediments and water in environmental monitoring - second edition University Press in Cambridge, Great Britain:UNESCO/WHO/UNEP.

Ainslie BJ, Post JR, Paul AJ. 1998. Effects of pulsed and continuous dc electrofishing on juvenile rainbow trout. North American Journal of Fisheries Management 18:905-918.

Alves RR, Rosa IM. 2007. Biodiversity, traditional medicine and public health: Where do they meet? Journal of ethnobiology and ethnomedicine 3:14.

Ardura A, Zaiko A, Martinez JL, Samuiloviene A, Borrell Y, Garcia-Vazquez E. 2015. Environmental DNA evidence of transfer of north sea molluscs across tropical waters through ballast water. Journal of Molluscan Studies 81:495-501.

Bellemain E, Carlsen T, Brochmann C, Coissac E, Taberlet P, Kauserud H. 2010. Its as an environmental DNA barcode for fungi: An in silico approach reveals potential pcr biases. Bmc Microbiology 10:9.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2005. Genbank. Nucleic Acids Research 33:D34-D38.

Bozek MA, Rahel FJ. 1991. Comparison of streamside visual counts to electrofishing estimates of colorado river USA cutthroat trout fry and adults. North American Journal of Fisheries Management 11:38-42.

Cadotte MW, Carscadden K, Mirotchnick N. 2011. Beyond species: Functional diversity and the maintenance of ecological processes and services. J Appl Ecol 48:1079-1087.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. 2010. Qiime allows analysis of high-throughput community sequencing data. Nature Methods 7:335-336.

Che J, Chen HM, Yang JX, Jin JQ, Jiang K, Yuan ZY, et al. 2012. Universal coi primers for DNA barcoding amphibians. Molecular ecology resources 12:247-258.

Cohen AS, Najarian D, Karger BL. 1990. Rapid DNA sequencing using gel-filled capillary and laser-induced fluorescence detection. Abstracts of Papers American Chemical Society 199:58.

Cordova S, Gaete H, Aranguiz F, Figueroa R. 2009. Water quality assessment in the limache stream (central chile), using bioindicators and bioassays. Lat Am J Aquat Res 37:199-209.

Densmore CL, Panek FM. 2013. Effects of depletion sampling by standard three-pass pulsed dc electrofishing on blood chemistry parameters of fishes from appalachian streams. North American Journal of Fisheries Management 33:298-306.

Derycke S, Vanaverbeke J, Rigaux A, Backeljau T, Moens T. 2010. Exploring the use of cytochrome oxidase c subunit 1 (coi) for DNA barcoding of free-living marine nematodes. PloS one 5:9.

Dudgeon D, Arthington AH, Gessner MO, Kawabata ZI, Knowler DJ, Leveque C, et al. 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges. Biol Rev 81:163-182.

Ficetola GF, Miaud C, Pompanon F, Taberlet P. 2008. Species detection using environmental DNA from water samples. Biol Lett 4:423-425.

Flotemersch JE, North S, Blocksom KA. 2014. Evaluation of an alternate method for sampling benthic macroinvertebrates in low-gradient streams sampled as part of the national rivers and streams assessment. Environmental monitoring and assessment 186:949-959.

Gibbs WK, Miller JE, Throneberry JK, Cook SB, Kulp MA. 2014. Summer habitat use and partitioning by two reintroduced rare madtom species. Journal of Freshwater Ecology 29:243-258.

Gibson UEM, Heid CA, Williams PM. 1996. A novel method for real time quantitative rt pcr. Genome Res 6:995-1001.

Glenn TC, Pierson TW, Kieran TJ, Hoffberg S, Rushmore J, Wong K, et al. In Prep. Adapterama ii: Universal amplicon sequencing on illumina platforms (taggimatrix). Molecular ecology resources.

Goldberg CS, Pilliod DS, Arkle RS, Waits LP. 2011. Molecular detection of vertebrates in stream water: A demonstration using rocky mountain tailed frogs and idaho giant salamanders. PloS one 6:e22746.

Gozlan RE, Britton JR, Cowx I, Copp GH. 2010. Current knowledge on non-native freshwater fish introductions. J Fish Biol 76:751-786.

Habera JW, Strange RJ, Carter BD, Moore SE. 1996. Short-term mortality and injury of rainbow trout caused by three-pass ac electrofishing in a southern appalachian stream. North American Journal of Fisheries Management 16:192-200.

Hajibabaei M, Singer GAC, Clare EL, Hebert PDN. 2007. Design and applicability of DNA arrays and DNA barcodes in biodiversity monitoring. Bmc Biology 5:7.

Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ. 2011. Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. PloS one 6:7.

Hajibabaei M, Spall JL, Shokralla S, van Konynenburg S. 2012. Assessing biodiversity of a freshwater benthic macroinvertebrate community through non-destructive environmental barcoding of DNA from preservative ethanol. Bmc Ecology 12:10.

Hebert PDN, Ratnasingham S, deWaard JR. 2003. Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. Proceedings of the Royal Society B-Biological Sciences 270:S96-S99.

Heid CA, Stevens J, Livak KJ, Williams PM. 1996. Real time quantitative pcr. Genome Res 6:986-994.

Hense Z, Martin RW, Petty JT. 2010. Electrofishing capture efficiencies for common stream fish species to support watershed-scale studies in the central appalachians. North American Journal of Fisheries Management 30:1041-1050.

Hooper DU, Adair EC, Cardinale BJ, Byrnes JEK, Hungate BA, Matulich KL, et al. 2012. A global synthesis reveals biodiversity loss as a major driver of ecosystem change. Nature 486:105-U129.

Hough RL. 2014. Biodiversity and human health: Evidence for causality? Biodivers Conserv 23:267-288.

Hubert N, Hanner R, Holm E, Mandrak NE, Taylor E, Burridge M, et al. 2008. Identifying canadian freshwater fishes through DNA barcodes. PloS one 3:e2490.

Jane SF, Wilcox TM, McKelvey KS, Young MK, Schwartz MK, Lowe WH, et al. 2014. Distance, flow and pcr inhibition: Edna dynamics in two headwater streams. Molecular ecology resources.

Jerde CL, Mahon AR, Chadderton WL, Lodge DM. 2011. "Sight-unseen" detection of rare aquatic species using environmental DNA. Conserv Lett 4:150-157.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. 2012. Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-1649.

King W. 1942. Trout management studies at great smoky mountains national park. Jour Wildlife Management 6:147-161.

Kriger KM, Hero JM, Ashton KJ. 2006. Cost efficiency in the detection of chytridiomycosis using pcr assay. Dis Aquat Org 71:149-154.

Kulp MA, Moore SE. 2000. Multiple electrofishing removals for eliminating rainbow trout in a small southern appalachian stream. North American Journal of Fisheries Management 20:259-266.

Lennon RE, Parker PS. 1959. The reclamation of indian and abrams creeks great smoky mountains national park. U S Fish and Wildlife Serv Spec Sci Rept Fish 306:1-22.

Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, Ranwez V, et al. 2013. A new versatile primer set targeting a short fragment of the mitochondrial coi region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. Frontiers in zoology 10:34.

Lodge DM, Stein RA, Brown KM, Covich AP, Bronmark C, Garvey JE, et al. 1998. Predicting impact of freshwater exotic species on native biodiversity: Challenges in spatial scaling. Aust J Ecol 23:53-67.

Lodge DM, Turner CR, Jerde CL, Barnes MA, Chadderton L, Egan SP, et al. 2012. Conservation in a cup of water: Estimating biodiversity and population abundance from environmental DNA. Molecular ecology 21:2555-2558.

Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. Trends Genet 24:133-141.

McMichael AJ, Beaglehole R. 2000. The changing global context of public health. Lancet 356:495-499.

Medinger R, Nolte V, Pandey RV, Jost S, Ottenwalder B, Schlotterer C, et al. 2010. Diversity in a hidden world: Potential and limitation of next-generation sequencing for surveys of molecular diversity of eukaryotic microorganisms. Molecular ecology 19:32-40.

Metzker ML. 2010. Applications of next-generation sequencing sequencing technologies - the next generation. Nat Rev Genet 11:31-46.

Meusnier I, Singer GAC, Landry JF, Hickey DA, Hebert PDN, Hajibabaei M. 2008. A universal DNA mini-barcode for biodiversity analysis. Bmc Genomics 9:4.

Morozova O, Marra MA. 2008. Applications of next-generation sequencing technologies in functional genomics. Genomics 92:255-264.

Nielsen JL. 1998. Scientific sampling effects: Electrofishing california's endangered fish populations. Fisheries 23:6-12.

Ogram A, Sayler GS, Barkay T. 1987. The extraction and purification of microbial DNA from sediments. J Microbiol Methods 7:57-66.

Penczak T. 2013. Drawbacks of successive electrofishing in a lowland stream. Journal of Applied Ichthyology 29:1036-1039.

Peterson JT, Thurow RF, Guzevich JW. 2004. An evaluation of multipass electrofishing for estimating the abundance of stream-dwelling salmonids. Trans Am Fish Soc 133:462-475.

Pierson TW, McKee A, Spear SF, Maerz JC, Camp CD, Glenn TC. In Press. Detection of an enigmatic plethodontid salamander using environmental DNA. Copeia.

Pilliod DS, Goldberg CS, Arkle RS, Waits LP. 2013. Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. Can J Fish Aquat Sci 70:1123-1130.

Pilliod DS, Goldberg CS, Arkle RS, Waits LP. 2014. Factors influencing detection of edna from a stream-dwelling amphibian. Molecular ecology resources 14:109-116.

Pinol J, Mir G, Gomez-Polo P, Agusti N. 2015. Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. Molecular ecology resources 15:819-830.

Pusey BJ, Kennard MJ, Arthur JM, Arthington AH. 1998. Quantitative sampling of stream fish assemblages: Single- vs multiple-pass electrofishing. Aust J Ecol 23:365-374.

Qiagen. 2015. Dneasy blood & tissue kit. Available: https://www.qiagen.com/us/shop/sample-technologies/dna/dna-preparation/dneasy-blood-and-tissue-kit - productdetails.

Rees HC, Gough KC, Middleditch DJ, Patmore JRM, Maddison BC. 2015. Applications and limitations of measuring environmental DNA as indicators of the presence of aquatic animals. J Appl Ecol 52:827-831.

Robeson MS, II, Costello EK, Freeman KR, Whiting J, Adams B, Martin AP, et al. 2009. Environmental DNA sequencing primers for eutardigrades and bdelloid rotifers. BMC Ecology 9:25.

Rodriguez-Cea A, Ayllon F, Garcia-Vazquez E. 2003. Micronucleus test in freshwater fish species: An evaluation of its sensitivity for application in field surveys. Ecotox Environ Safe 56:442-448.

Savolainen V, Cowan RS, Vogler AP, Roderick GK, Lane R. 2005. Towards writing the encyclopaedia of life: An introduction to DNA barcoding. Philos Trans R Soc B-Biol Sci 360:1805-1811.

Service USNP. 2006. Management policies: The guide to managing the national park system, chapter 4.4: Biological resource management. Fed Reg Chapter 4.4.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M. 2012. Next-generation sequencing technologies for environmental DNA research. Molecular ecology 21:1794-1805.

Shute JR, Rakes PL, Shute PW. 2005. Reintroduction of four imperiled fishes in abrams creek, tennessee. Southeastern Naturalist 4:93-110.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH. 2012a. Environmental DNA. Molecular ecology 21:1789-1793.

Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E. 2012b. Towards next-generation biodiversity assessment using DNA metabarcoding. Molecular ecology 21:2045-2050.

Tillett BJ, Field IC, Bradshaw CJA, Johnson G, Buckworth RC, Meekan MG, et al. 2012. Accuracy of species identification by fisheries observers in a north australian shark fishery. Fisheries Research 127:109-115.

Vestheim H, Jarman SN. 2008. Blocking primers to enhance pcr amplification of rare sequences in mixed samples - a case study on prey DNA in antarctic krill stomachs. Frontiers in zoology 5:11.

Vrtilek M, Reichard M. 2012. An indirect effect of biological invasions: The effect of zebra mussel fouling on parasitisation of unionid mussels by bitterling fish. Hydrobiologia 696:205-214.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD. 2005. DNA barcoding australia's fish species. Philosophical transactions of the Royal Society of London Series B, Biological sciences 360:1847-1857.

Wiens JJ, Ackerly DD, Allen AP, Anacker BL, Buckley LB, Cornell HV, et al. 2010. Niche conservatism as an emerging principle in ecology and conservation biology. Ecol Lett 13:1310-1324.

Wilcox TM, McKelvey KS, Young MK, Jane SF, Lowe WH, Whiteley AR, et al. 2013. Robust detection of rare species using environmental DNA: The importance of primer specificity. PloS one 8:e59520.

Wilson C, Wright E, Bronnenhuber J, MacDonald F, Belore M, Locke B. 2014. Tracking ghosts: Combined electrofishing and environmental DNA surveillance efforts for asian carps in ontario waters of lake erie. Management of Biological Invasions 5:225-231.

Yurtseven I, Serengil Y, Pamukcu P. 2016. Seasonal changes in stream water quality and its effects on macroinvertebrate assemblages in a forested watershed. Applied Ecology and Environmental Research 14:175-188.

Zeale MR, Butlin RK, Barker GL, Lees DC, Jones G. 2011. Taxon-specific pcr for DNA barcoding arthropod prey in bat faeces. Molecular ecology resources 11:236-244.

Table 1: Detection probabilities for each known and monitored species at Abrams Creek site 3 for the 2015 shocking data, historical data, and eDNA data measured at 94% similarity and separated by primer.

	Leray eDNA	**Fish eDNA**	**Shocking 2015**	**Historical Data**
SAS	0	0	0	0.071428
TNS	0	0	0	0.21428
NHS	0	0	0	0.5
RIC	0.73	0.2	1	0.78571
WPS	0.06	0	1	0.85714
CKC	0	1	1	1
STR	0.93	0.13	0.67	1
RSD	0.53	0	0.67	1
TSD	0.06	0.2	0.67	1
BND	1	0.86	1	1
WHS	0.93	0.33	1	1
RBT	1	0.93	1	1

Table 2: Detection probabilities for each known and monitored species at Abrams Creek site 4 for the 2015 shocking data, historical data, and eDNA data measured at 94% similarity and separated by primer.

	Leray eDNA	Fish eDNA	Shocking 2015	Historical Data
TNS	0	0	0	0.07142
WPS	0	0	0	0.07142
TND	0	0	0	0.142857
NHS	0	0	0	0.42857
FTD	0	0	0.67	0.92857
TSD	0.4	1	1	0.92857
RBT	0.75	1	1	0.92857
STR	1	1	1	1
CKC	0	1	1	1
LND	1	0.75	1	1
RSD	1	0.25	1	1
BND	1	1	1	1

Table 3: Common and scientific names for each species abbreviated in figures and tables.

Abbreviation	Species
BND	Blacknose Dace (Rhinichthys atratulus)
STR	Stoneroller (Campostoma sp.)
CKC	Creek Chub (Semotilus atromaculatus)
FTD	Fantail Darter (Etheostoma flabellare)
NHS	Northern Hogsucker (Hypentelium nigricans)
LND	Longnose Dace (Rhinichthys cataractae)
RBT	Rainbow Trout (Oncorhynchus mykiss)
RIC	River Chub (Nocomis micropogon)
RSD	Rosyside Dace (Clinostomus funduloides)
SAS	Saffron Shiner (Notropis rubricroceus)
TND	Tennessee Dace (Chrosomus tennesseensis)
TSD	Tennessee Snubnose Darter (Etheostoma simoterum)
TNS	Tennessee Shiner (Notropis leuciodus)
WPS	Warpaint Shiner (Luxilus coccogenis)
WHS	White Sucker (Catostomus commersonii)

Table 4: Percentage values converted to decimals for shocking and eDNA results from ABC3 and then ranked in order from lowest to highest with the resulting Spearman's Rho correlation coefficient and p-value

	Shocking	eDNA	RankShocking	RankeDNA
SAS	0.0000	0.0000	2	1.5
TNS	0.0000	0.0000	2	1.5
NHS	0.0000	0.0003	2	3
RIC	0.0219	0.0691	8	8
WPS	0.0110	0.0023	6.5	4
CKC	0.2326	0.0318	10	6
STR	0.0110	0.1022	6.5	9
RSD	0.0037	0.0387	4	7
TSD	0.0055	0.0080	5	5
BND	0.4250	0.1480	12	11
WHS	0.0348	0.1292	9	10
RBT	0.2545	0.4685	11	12

Spearman's Rho:
0.851611
P-value (2-tailed):
0.00044

Table 5: Percentage values converted to decimals for shocking and eDNA results from ABC4 and then ranked in order from lowest to highest with the resulting Spearman's Rho correlation coefficient and p-value

	Shocking	eDNA	RankShocking	RankeDNA
TNS	0.0000	0.0000	2.5	3
WPS	0.0000	0.0000	2.5	3
TND	0.0000	0.0000	2.5	3
NHS	0.0000	0.0000	2.5	3
FTD	0.0117	0.0000	6	3
TSD	0.0234	0.0540	7	9
RBT	0.0025	0.0120	5	6
STR	0.2982	0.3051	11	11
CKC	0.0927	0.0389	8	8
LND	0.1170	0.0169	9	7
RSD	0.1412	0.4082	10	12
BND	0.3133	0.1644	12	10

Spearman's Rho:
0.900522
P-value (2-tailed):
0.00006

Figure 1: The detection probability for each species calculated by dividing the number of samples that returned reads for that species by the total number of samples from that tank.

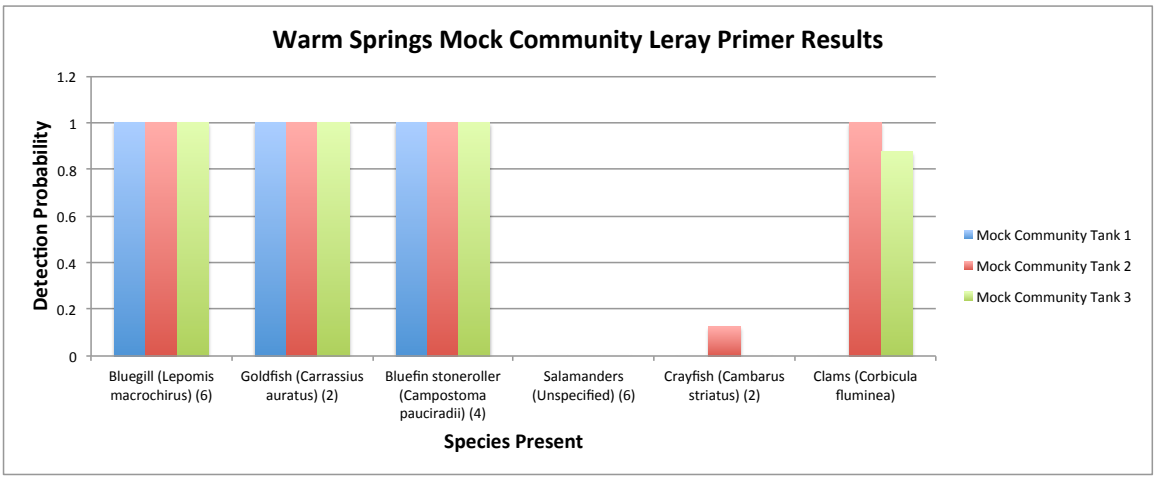


Figure 2: The detection probability for each species calculated by dividing the number of samples that returned reads for that species by the total number of samples from that tank.

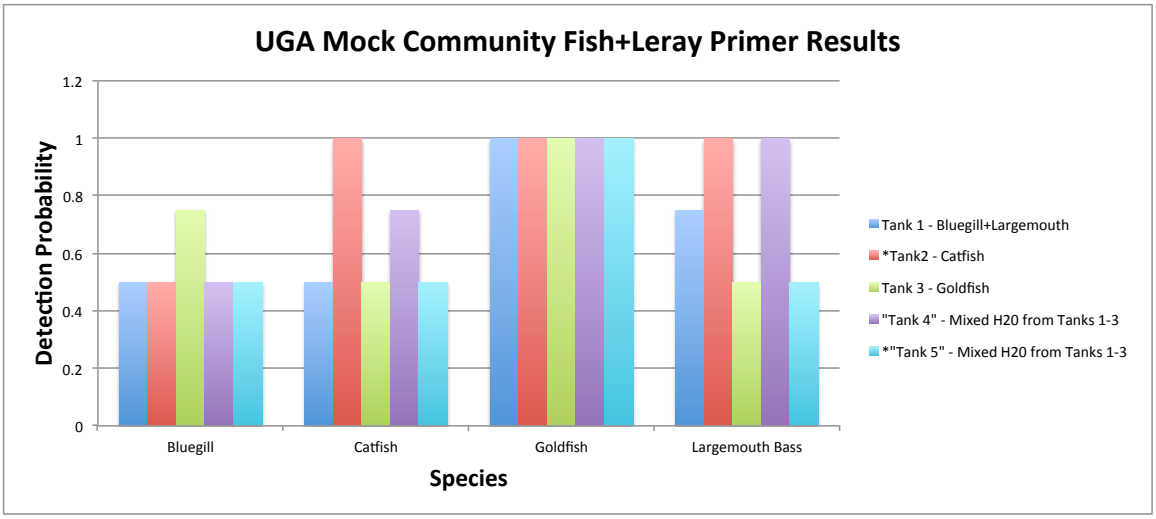
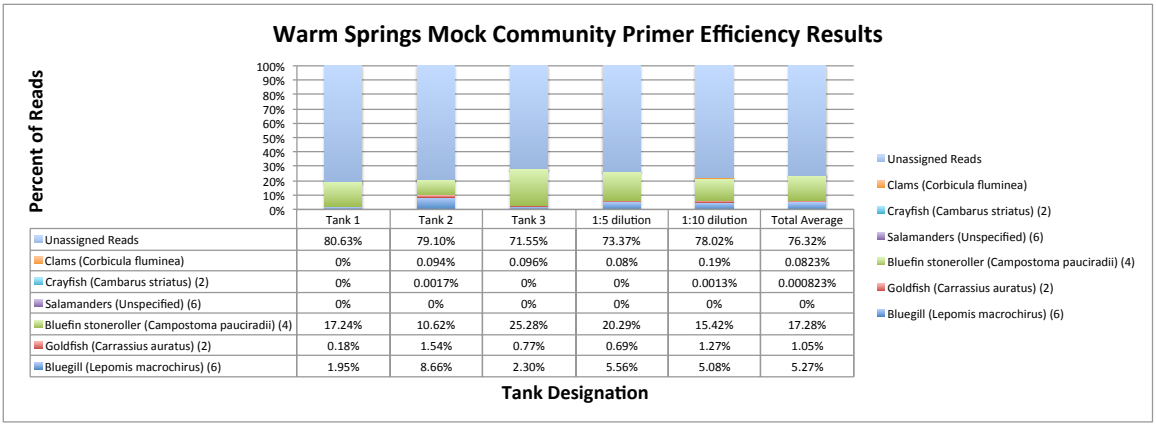
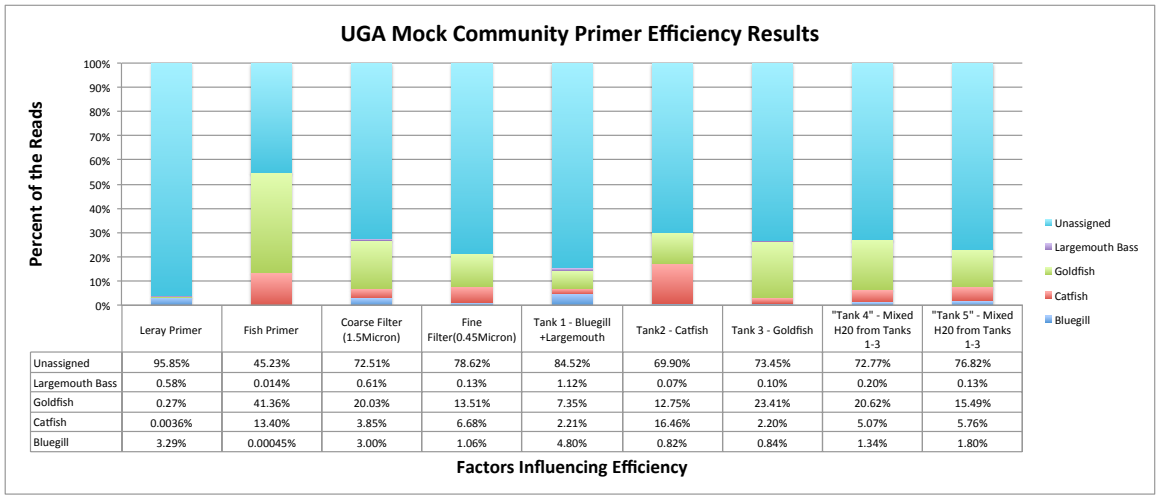


Figure 3: The reads returned for the combined samples from each tank separated by the percentage representing each species. The total average column represents that reads from all three tanks



	Tank 1	Tank 2	Tank 3	1:5 dilution	1:10 dilution	Total Average
Unassigned Reads	80.63%	79.10%	71.55%	73.37%	78.02%	76.32%
Clams (Corbicula fluminea)	0%	0.094%	0.096%	0.08%	0.19%	0.0823%
Crayfish (Cambarus striatus) (2)	0%	0.0017%	0%	0%	0.0013%	0.000823%
Salamanders (Unspecified) (6)	0%	0%	0%	0%	0%	0%
Bluefin stoneroller (Campostoma pauciradii) (4)	17.24%	10.62%	25.28%	20.29%	15.42%	17.28%
Goldfish (Carrassius auratus) (2)	0.18%	1.54%	0.77%	0.69%	1.27%	1.05%
Bluegill (Lepomis macrochirus) (6)	1.95%	8.66%	2.30%	5.56%	5.08%	5.27%

Figure 4: The reads from all the tank samples as well as reads separated by different primers and filters separated by the percentage of each species represented in each column. The four columns for the two types of primers and two types of filters include all the samples from all five tanks, while the columns for each tank include only the samples from that specific tank.



	Leray Primer	Fish Primer	Coarse Filter (1.5Micron)	Fine Filter(0.45Micron)	Tank 1 - Bluegill +Largemouth	Tank2 - Catfish	Tank 3 - Goldfish	"Tank 4" - Mixed H20 from Tanks 1-3	"Tank 5" - Mixed H20 from Tanks 1-3
Unassigned	95.85%	45.23%	72.51%	78.62%	84.52%	69.90%	73.45%	72.77%	76.82%
Largemouth Bass	0.58%	0.014%	0.61%	0.13%	1.12%	0.07%	0.10%	0.20%	0.13%
Goldfish	0.27%	41.36%	20.03%	13.51%	7.35%	12.75%	23.41%	20.62%	15.49%
Catfish	0.0036%	13.40%	3.85%	6.68%	2.21%	16.46%	2.20%	5.07%	5.76%
Bluegill	3.29%	0.00045%	3.00%	1.06%	4.80%	0.82%	0.84%	1.34%	1.80%

Figure 5: Map of Abrams Creek Sample Sites ABC3 and ABC4

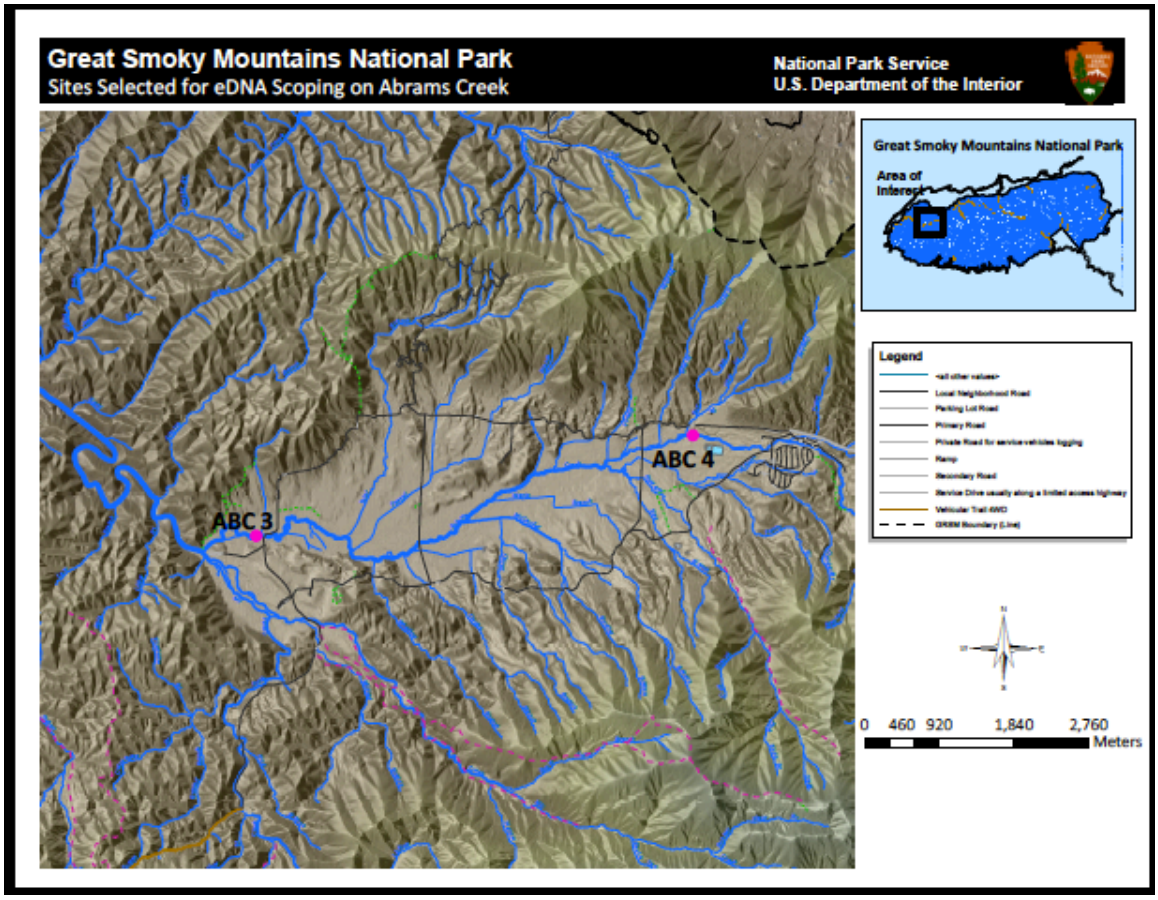


Figure 6: The detection probability for species at Abrams Creek Site 3 for historical data, 2015 shocking data, and eDNA data at 94% match similarity to the reference database

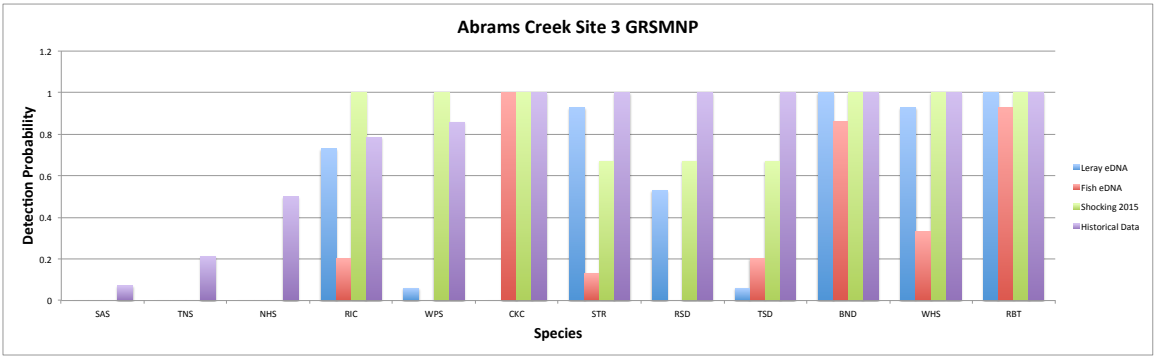


Figure 7: The detection probability for species at Abrams Creek Site 4 for historical data, 2015 shocking data, and eDNA data at 94% match similarity to the reference database

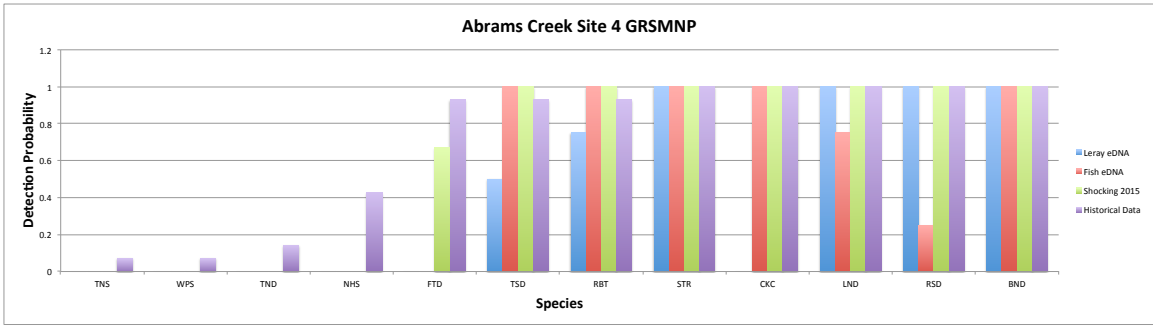


Figure 8: The number of invertebrate genera detected between both study sites out of the total number of observed invertebrate genera in Abrams Creek

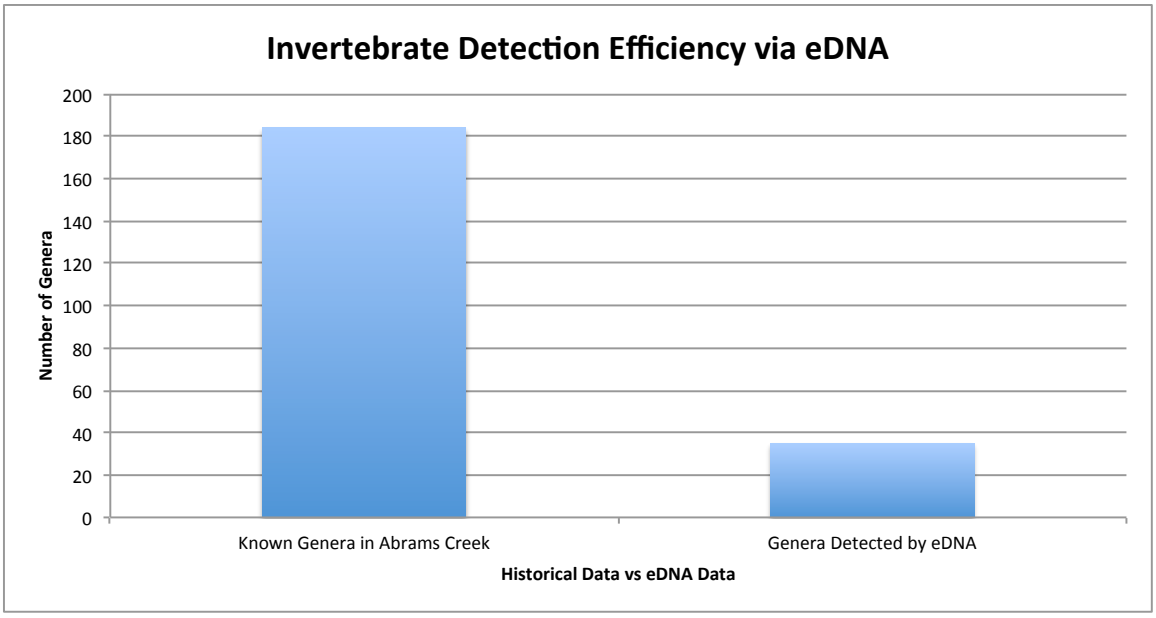


Figure 9: From the assigned reads, the percentage of reads assigned to each species for samples 1-15 from Abrams Creek Site 3 and the population estimates from the shocking data collected from Abrams Creek Site 3

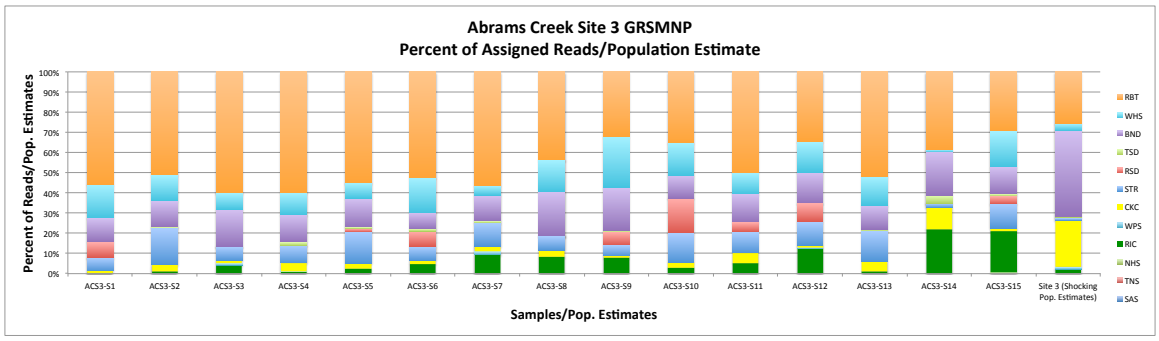
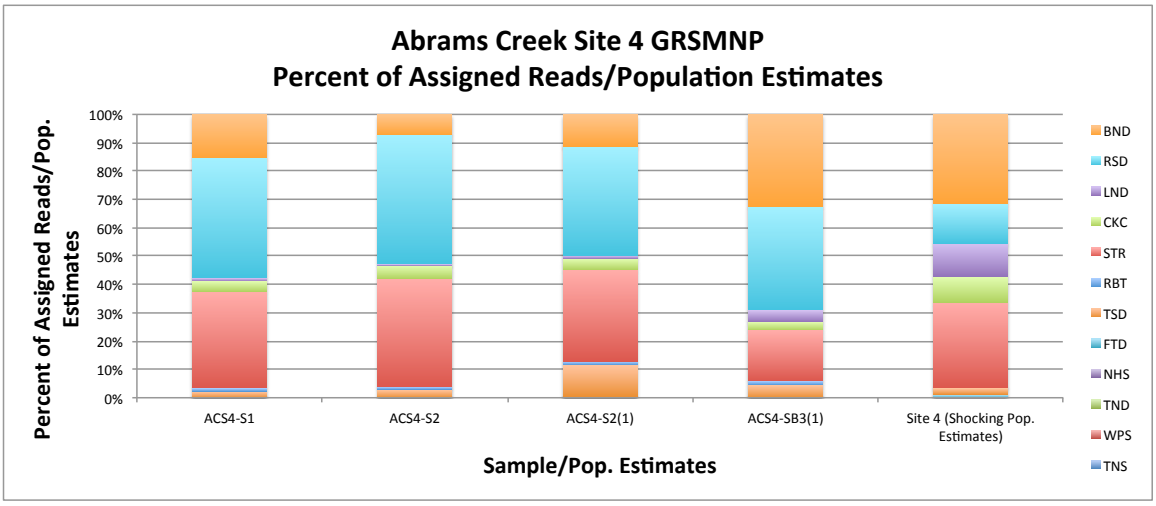


Figure 10: From the assigned reads, the percentage of reads assigned to each species for samples 1-4 from Abrams Creek Site 4 and the population estimates from the shocking data collected from Abrams Creek Site 4



Appendix A
Supplementary Figures

Figure 1A: Total number of reads for each sample from Abrams Creek Site 3 categorized in two percentages as having been assigned to the reference database or being unassigned.

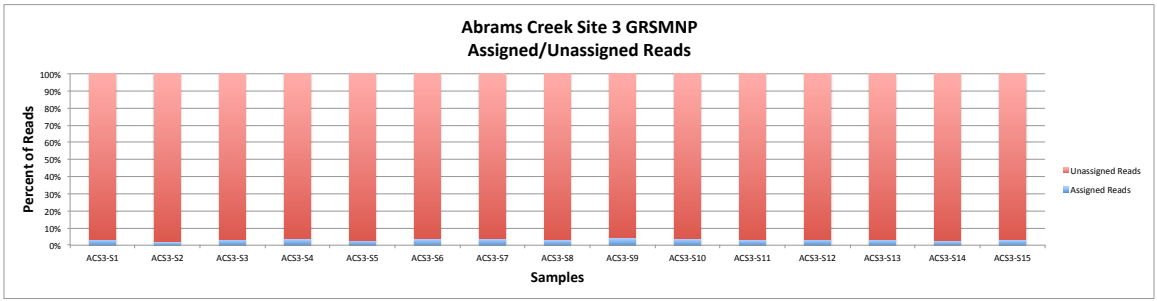


Figure 2A: Total number of reads for each sample from Abrams Creek Site 4 categorized in two percentages as having been assigned to the reference database or being unassigned.