January 2019

The use of environmental DNA for the detection of Palaemonias ganteri (Hay, 1901), a federally endangered cave species

Andrew Joseph Stump
Eastern Kentucky University

THE USE OF ENVIRONMENTAL DNA FOR THE DETECTION OF PALAEMONIAS GANTERI (HAY, 1901), A FEDERALLY ENDANGERED CAVE SPECIES

BY

ANDREW J STUMP

THESIS APPROVED:

Chair, Advisory Committee

Member, Advisory Committee

Member, Advisory Committee

Dean, Graduate School

STATEMENT OF PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Master of Science degree at Eastern Kentucky University, I agree that the Library shall make it available to borrowers under rules of the Library. Brief quotations from this document are allowable without special permission, provided that accurate acknowledgements of the source are made. Permission for extensive quotation from or reproduction of this document may be granted by my major professor. In his absence, by the Head of Interlibrary Services when, in the opinion of either, the proposed use of the material is for scholarly purposes. Any copying or use of the material in this document for financial gain shall not be allowed without my written permission.

Signature:

X 

Date: 7/1/2019

THE USE OF ENVIRONMENTAL DNA FOR THE DETECTION OF PALAEMONIAS GANTERI (HAY, 1901), A FEDERALLY ENDANGERED CAVE SPECIES

BY

ANDREW J STUMP

Submitted to the Faculty of the Graduate School of
Eastern Kentucky University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

2019

DEDICATION

I would like to dedicate this thesis to my family. My wife, Jill, and my two children, Andrew and Lillian, have been exceedingly reassuring and patient with me through this process. In addition, I would like to thank Dr. David M. Hayes for his guidance and help and my parents for their support and encouragement throughout my life. I am truly blessed.

ACKNOWLEDGEMENTS

This thesis was made possible through the work and assistance of several helpful volunteers, agency personnel, and institutions. Firstly, I would like to acknowledge the US Fish and Wildlife Service for providing the funding that made this project possible. In addition, I would like to thank my advisor Dr. David M. Hayes and my committee members (Dr. Stephen Richter and Dr. Sherry Harrel) for their patience and guidance. Also, collections of samples and reference DNA would not have been possible without the help of Richard Olsen (NPS), Rickard Toomey (NPS), and Carrie Allison (USFWS) and the access granted by Mammoth Cave National Park and Western Kentucky University's Green River Nature Preserve. The University of Kentucky's Advanced Genetic Techniques Center (AGTC) provided the assistance with high-throughput sequencing preparation and several others helped with field collections and laboratory work including: Brajaan Hayes, Heidi Braunrider, Jill Stump, Nicolas Revetta, Spencer Mattingly, and others. Thank you all for your help.

ABSTRACT

Kentucky's limestone caves and karst water systems are an ecologically important part of its natural heritage and home to many unique subterranean species. In addition to being geologically interesting, it is also home to one, of only three, extant North American atyids: the federally endangered Kentucky Cave Shrimp (*Palaemonias ganteri*, Hay 1901). However, access for monitoring and management objectives involving this species and other cave inhabitants can be difficult, highly contingent upon environmental conditions, and time consuming with low yields using traditional techniques. Advancements in metabarcoding and Next Generation Sequencing (NGS) technologies provide tools that may allow researchers and managers to address some of the hurdles posed by the difficult environment where *P. ganteri* live.

This study was able to identify the presence of *P. ganteri* DNA at seven locations in addition to inferring six distinct variations between homologous COI shrimp sequences and confirming the presence of one known shrimp variant. Metabarcoding identified some close variations to reference sequences for cave obligates, but more importantly, highlighted the need for better references when conducting cave faunal surveys with eDNA techniques. These findings not only have implications for better surveillance of *P. ganteri* and other cave inhabitants, but also open up the possibility for improving management goals by incorporating population-level genetic information that can be considered for each groundwater basin.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

I. Introduction

Environmental DNA and Metabarcoding

Recent global diversity estimates make it clear that new technologies will be needed to assist biologists in describing, investigating, and preserving the world's biodiversity (Hebert *et al.* 2003). The development and expansion of molecular data has advanced our ability to discern both higher taxonomic categories as well as distinguish between closely related species (Hebert *et al.* 2003; Blaxter 2004; Avise 2009; Souza *et al.* 2016). When integrated with other areas of study (i.e. morphometrics), molecular systematics can provide a more complete understanding of taxa groups and how they may be related. This provides an opportunity to broaden our resolution when considering populations or communities in an environment and may aid in shaping or guiding management decisions.

While fairly precise, molecular procedures can be intensive and require the use of high-quality tissue extractions from several gene regions to aid in the systematic description of an organism. This can be limiting when surveying species that are protected, difficult to locate with traditional techniques, or may be sensitive to tissue biopsies. In addition, collection of tissues can be time consuming and may only allow researchers and managers to focus attentions toward one organism at a time. Therefore, techniques that allow for the incorporation of data from multiple groups or communities across a landscape may allow for the development of more comprehensive management goals.

Environmental DNA (eDNA) has recently become a popular addition to projects concerning illusive organisms, invasive species, conservation management, and ecosystem biodiversity. It has proven to be a highly sensitive tool and provides researchers the benefit of collecting information about a community as well as a single organism under consideration while using less invasive techniques. This practice was originally exploited in a 1987 publication, which concerned methods for extracting microbial DNA from soil samples (Taberlet *et al.* 2012). However, it has now been broadly adapted for the study of plants and metazoan diversity and is becoming a powerful tool that can supplement many traditional research programs (Fonseca *et al.* 2010; Hajibabaei *et al.* 2011; Baird & Hajibabaei 2012b).

Environmental DNA can be categorized broadly and references environmental samples containing extra-cellular DNA, living cells, shed cells, and small organisms contained within a sample unit (Nielsen *et al.* 2007; Taberlet *et al.* 2012). Traces of DNA can be isolated from these samples (i.e. soil, water, or feces) without the need to directly handle the specific organism(s) intended for study (Baird & Hajibabaei 2012a; Taberlet *et al.* 2012; Lodge *et al.* 2012; Vörös *et al.* 2017). The deoxyribonucleic acids extracted from an environmental sample, can contain multitudes of different molecular sequences, useful when surveying and identifying specific taxa or groups amongst a landscape. However, DNA found in the environment can often be degraded because it exists either outside of or within a dying cell. Because of this, short, abundant gene regions which are also taxonomically informative are necessary targets when searching for specific organisms amongst such data-rich samples.

Genetic barcodes are short segments of DNA (often around 500 base pairs (bp) in length) that can be used to describe a unique organism when compared alongside the same gene region in other taxa groups. These select genes are both conserved and contain variability that improves taxonomic organization at the sub-terminal and terminal nodes of phylogenies (Blaxter 2004). Metabarcoding studies leverage the power of barcode libraries and metagenomics pipelines to expand taxa descriptions to entire communities identified from DNA found in environmental samples. Typically, the total length of these gene regions are relatively short and several barcodes have been identified as ideal markers to aid in taxonomic resolution when considering degraded samples (often seen with eDNA), museum specimens, and biodiversity surveys within complex [environmental] samples containing DNA from multiple organisms (Blaxter 2004; Hajibabaei *et al.* 2006; Layman & Mayden 2012; Ji *et al.* 2013).

However, along with the utility of complex environmental samples comes a need for our ability to sequence, read, and analyze large amounts of data in a cost-effective way. Traditional DNA sequencing was introduced by Sanger *et al.* 1977 and was capable of producing 1kilobase (kb) of sequence data from a single specimen (cited in Shokralla *et al.* 2012). The automation and technical variation of this process using dye-termination technologies decreased read time and expanded chain-termination capabilities (~96kb maximum) (Shokralla *et al.* 2012). However, this procedure was still only capable of handling 96 individual reads and simply did not have the capacity to adequately process a complex environmental sample, which can

contain millions of DNA fragments. However, since the integration of sequencing technologies during the Human Genome Project in the mid-1980s, the use of chemistry, high-resolution optics, hardware expansion, and software engineering has pushed sequencing pipelines towards several high-throughput systems (Mardis 2008).

Next-generation sequencing (NGS) platforms are capable of massively parallel DNA sequencing and, in some cases, streamline results using internal algorithms to assess data quality. These capabilities make NGS ideal for reading fragmented libraries from a specific genome (i.e. genome sequencing), a pool of reverse transcription RNA molecules (i.e. RNAseq and transcriptome sequencing), or a pool of PCR amplified molecules (i.e. amplicon sequencing) such as those generated using environmental samples (Shokralla *et al.* 2012). With the expansion of gene libraries (such as the cytochrome c oxidase subunit I), the advancement of NGS technologies, and decreases in costs, metabarcoding studies have combined the versatility of barcode regions with the power of high-throughput sequencers in order to utilize eDNA for documenting and identifying taxa within a community using a variety of indirect sampling methods.

To date, studies have utilized the combination of eDNA and NGS when investigating microscopic eukaryotes, meiofauna, and macro-organisms (Nielsen *et al.* 2007; Bik *et al.* 2012; Taberlet *et al.* 2012). This expansion includes the analysis of complex DNA extracts from stomach contents, feces, sediments, water, and other mediums with the goal to identify individual organisms or taxonomic groups simultaneously; this has allowed researchers to use environmental samples for biodiversity measures, presence-absence, and the detection of rare species, or early

invasions (Olson *et al.* 2012; Bik *et al.* 2012; Yoccoz 2012; Lodge *et al.* 2012; Bohmann *et al.* 2014; Mächler *et al.* 2014). Moreover, DNA from open water samples have been particularly useful and are thought to be good indicators of current fauna existing within an aquatic system due to the rapid environmental break down of nucleic acids in aqueous environments which lead to decreases in detectability with time (Alvarez *et al.* 1996; Ficetola *et al.* 2008; Dejean *et al.* 2011; Jerde *et al.* 2011; Thomsen *et al.* 2011; Lodge *et al.* 2012; Takahara *et al.* 2013). Thus, eDNA techniques and metabarcoding pipelines can be particularly powerful tools for management and research objectives involving aquatic resources, especially if the species or communities in question are difficult to access, in low abundance, or require more remote observation.

Karst and Cave Ecosystems

The term "karst" became prevalent in the late 19th century when geomorphology emerged as a scientific discipline (White *et al.* 1995). Karst lands are the product of chemical and physical erosion on soluble rocks, namely limestone and dolomite. Acids found in water come into contact with soluble rock strata to form unique dissolution landscapes with complex and distinctive geological formations (White *et al.* 1995). In the absence of light, ecological systems found in these passages are highly dependent on external, allogenic recharge to carry in organic materials for primary, trophic level uptake (White *et al.* 1995). This lack in primary production, along with other high stressors, place very strong selective forces on potential subterranean colonists and cave inhabitants (Howarth 1993).

Most subterranean organisms in the temperate regions are thought to originate from wide-spread, surface dwelling ancestors that frequently exploited and colonized underground passages (Holsinger 1988). This proposition has been strengthened by multiple bodies of work since Roach and colleagues studied the karst drainage basin flowing through St. Catherine's Cave, France. He and colleagues found that approximately 25% of observed crustaceans filtered from spring water at the lower end of the karst system were species regularly found in surface waters (White *et al.* 1995). Extirpation of parent populations after species integration in a subterranean community suggests that cave systems may act as refugia for isolated relicts partially suited for hypogean environments (Holsinger 1988; Howarth 1993).

Observations of species richness in cave fauna have been observed to follow patterns similar to other discontinuous habitats such as islands and mountain tops (Barr Jr & Holsinger 1985). Thus, obligate organisms inhabiting aquatic underground spaces are often endemic and raise concerns for management and conservation efforts (Asmyhr *et al.* 2014). Due to their dependence on the subterranean environment and their inability to readily migrate, most management programs identify the need to protect the water and land use around karst water basins (Elliott 2000). However, to properly manage and conserve obligate cave species, it is necessary to know, at a minimum, their presence and distribution within a basin or surrounding basins. Unfortunately, many cave systems are difficult to study due to issues such as poor understanding of subterranean interconnectedness, a need for specialized skills to access and investigate cave passages, and an inability to explore

sections that are too small for humans to pass through. These issues have led to a general lack of knowledge when discussing cave diversity and species distributions (Asmyhr *et al.* 2014).

The Mammoth Cave and Flint-Ridge systems in Kentucky are unique examples of karst formations present in a limestone belt (Figure 1, Appendix B) that extends from southern Indiana to Tennessee (White *et al.* 1970). While Mammoth Cave National Park encompasses parts of both of these, there are a number of sizeable systems and subterranean streams that surround the area that may, or may not be connected to the larger system. With a number of environmental issues associated with groundwater contamination along several basins surrounding the park (May *et al.* 2005), it is important to have good documentation of vulnerable, obligate, subterranean fauna within the karst water systems. Currently there are several obligate karst species living within the Mammoth Cave system, but only one federally endangered atyid shrimp: the Kentucky Cave Shrimp.

Atyidae: Palaemonias ganteri (Hay, 1901)

North America is inhabited by three, extant atyid species (Hobbs & Lodge 2010; von Rintelen *et al.* 2012): *Syncaris pacifica* (Holmes, 1895), *Palaemonias alabamae* (Smalley, 1961) and *Palaemonias ganteri* (Hay, 1901). The latter, *P. ganteri*, is listed as federally endangered and is only found in the limestone caves of the Mammoth Cave system and adjacent karst areas bordering the western coal fields of Kentucky (Culver *et al.* 2003; Hobbs *et al.* 2003; von Rintelen *et al.* 2012). *P. ganteri* is one of only two troglobitic North American atyid shrimp. The two species are likely “thermophilic

relicts" derived from a widely distributed common ancestor that underwent independent cave invasions in each geographic location (Barr Jr & Holsinger 1985; Hobbs & Lodge 2010). The other stygobiotic species, *P. alabamae*, is geographically separated by several hundred kilometers and is isolated to only three groundwater basins in northern Alabama (Hobbs & Lodge 2010; von Rintelen *et al.* 2012).

P. ganteri was described by Hay in 1901 using 12 specimens collected from the Roaring River passage in the Echo River basin (Lisowski 1983). Between 1967 and 1979 *P. ganteri* was not observed in any of its historically known locations, prompting further investigation into additional basins where the species might be found [1](Table 1, Appendix A). The shrimp was thought to be close to extinction due to pollution of local ground water from sewage, hydrocarbons, and oil brine runoff (Lisowski 1983; Brown 1991; National Park Service 2006). In addition, it was suggested that modifications to habitat and flood regimes from dam construction on the Green River adversely affected shrimp populations through lowered reproductive success and increased risk of predation (Lisowski 1983).

P. ganteri was first proposed as a threatened species in 1977, but the application was withdrawn to comply with amendments to the Endangered Species Act and then resubmitted in March, 1980. *P. ganteri* was officially listed in October of 1983 with Roaring River passage listed as the shrimp's critical habitat (U.S. Fish and Wildlife Service 1988). A twelve month review of the shrimp's biology and distribution was proposed two years prior to its listing by Holsinger and Leitheuser and was

[1] All figures and tables are presented in appendices at the end of this thesis.

awarded by the National Park Service (U.S. Fish and Wildlife Service 1988). The grant was extended for an additional five years and provided the base majority of information currently known about *P. ganteri*.

Study Objective

The primary intent of this study was to determine if eDNA, paired with NGS technologies, would be useful in detecting the presence of *P. ganteri* in several karst drainages throughout the Mammoth Cave area. In addition, it was my intent to determine if environmental DNA fragments could be used to infer genetic variation of shrimp DNA found at each site. Finally, I intended to identify additional species found within these cave communities using metagenomics methods. If successful, these pipelines may be a useful framework for remotely surveying cave shrimp populations, in additional to other troglobitic species. This could potentially provide a platform for prioritizing management locations for particular sub-basins and/or inform management strategies.

II. Methods

Location Description

Mammoth Cave lies approximately 160km south of Louisville, KY and 160 km north of Nashville, TN in between the Green and Barron rivers (Glennon & Groves 2002; May *et al.* 2005) (Figure 2, Appendix B). The park is located in the central portion of western Kentucky and spans Barren, Edmonson, and Hart counties (Palmer 1995). Mammoth Cave itself is comprised of over 629km of explored passages (US Department of the Interior 2012), making it the longest known cave in the world (May *et al.* 2005). The Mammoth Cave subterranean basins consists of 28 karst watersheds that drain throughout the park boundaries (National Park Service 2006). This extensive karst aquifer developed throughout three separate layers of Mississippian limestone and sits under an insoluble layer of sandstone and shale (Glennon & Groves 2002). The passages were created as the result of extensive weathering, provided by the Green River system, and a series of natural springs and stream drainages that flow through the karst system or submerge through sinkholes (Palmer 1995). The karst water networks act as tributaries to the Green River and subterranean flow is ultimately controlled by its location and behavior. Thus, water draining the majority of these groundwater basins can typically be accessed from the Green River directly, or at spring sites along its banks.

Eleven sites were sampled between September 2012 and September 2013. Nine of the eleven sampled sites resided within Mammoth Cave National Park (MCNP). Western Kentucky University's Green River Preserve (GRP), located northeast of the

park boundary, provided access to two additional historically recognized shrimp basins. These are currently the farthest known sites upriver in the watershed to house shrimp populations (Table 1, Appendix A; Figure 3, Appendix B).

Sterilization and Quality Control

Sterilization of field equipment using best practices suggested in the literature and other protocols was essential to keep extraneous DNA off of field equipment and reduce the likelihood of false detections (Kemp & Smith 2005; Blankenship *et al.* 2011; Jerde *et al.* 2011). All equipment was thoroughly cleaned with soap and warm water and scrubbed to remove dirt, dust, and other debris. All parts were then rinsed completely with distilled water to clear soap and surfactants from the surface of equipment. Pumps, hoses, filter heads, accompanying connectors, and forceps were sterilized using a 20-minute chlorine bath comprised of a 5-10% bleach solution. All equipment was rinsed thoroughly with distilled lab water to wash away residual bleach residue and then used to create one negative lab control per sampling site. Negative control filters were created by filtering two liters of distilled lab water through newly decontaminated equipment and stored for later assessment. Equipment was allowed to air dry in a clean environment and placed into sealed, plastic equipment bags. These individual equipment bags were placed into larger site bags that were further sealed for added protection and quick access in the field.

Field Sampling

Environmental DNA was collected from eleven sites by filtering karst water through a 0.7µm glass fiber filter contained in an Advantec In-Line Filter Holder using a

sterilized liquid transfer pump (Figure 4, Appendix B). Springs or surface accesses that drain the majority of nine karst basins were targeted due to their relative ease of access and a desire to draw samples from a lower (i.e. most downstream) point in each karst drainage basin. Two sites were accessed inside of caverns underground: the Roaring River Shrimp Pool and Owl Cave. At each site, two liters of water were pumped through a glass fiber filter before it was preserved for laboratory DNA extraction. To prevent further breakdown of environmental samples and encourage cell lysis, each filter was treated separately using heated cell lysis buffer from the MoBio PowerWater extraction kit and placed into individual, sterile whirl-pak bags. Twenty filters in all were used to sample each site, making the total volume of water extracted from each location 40 liters. All filter bags were packed in ice until samples could be placed into a laboratory freezer for further DNA extraction.

Laboratory Methods

DNA was extracted from filters using the MoBio PowerWater extraction kit (MO BIO Laboratories 2016). Extraction consisted of mechanically breaking up filters in bead tubes and running the resultant supernatant through a series of spin-columns to filter and remove PCR inhibitors from the extraction product for downstream amplification. Isolated DNA was then quantified using a Thermo Fisher Scientific NanoDrop spectrophotometer. Three reads were taken per sample to obtain an average DNA quantitation per sample. Samples were sorted from highest to lowest DNA concentration for each site. The samples containing the highest quantities of DNA were selected for PCR trials.

In order to amplify invertebrate metazoan DNA, in addition to shrimp DNA, the Folmer primers (LCO1490: 5'-GGTCAACAAATCATAAAGATATTGG-3", HCO2198: 5'-TAAACTTCAGGGTGACCAAAAAATCA-3') were selected to amplify the mitochondrial cytochrome C oxidase subunit 1 (COI) gene region (Folmer *et al.* 1994). Two tissue samples from *P. ganteri* pleopods, each from a separate shrimp, were collected and supplied by US Fish and Wildlife employees. These were used to obtain reference DNA after a standard CTAB extraction protocol (Murray & Thompson 1980; Allen *et al.* 2006; von Rintelen *et al.* 2012). The polymerase chain reaction (PCR) method was used to amplify extracted eDNA in a 20μl reaction using a Phusion Taq polymerase, 1.0μM concentrations of the forward and reverse COI Folmer primer solutions, 2μl of extracted eDNA sample, and 10μl the manufacturers recommended buffer solution. Each reaction was carried out under the following conditions: 98°C for 5 minutes as an initial denaturing step; 30 cycles of denaturing at 98°C for 1 second, annealing at 45°C for 5 seconds, and elongation at 72°C for 15 seconds; and a final extension step at 72°C for 60 seconds. Separately, *P. ganteri* reference DNA was amplified in the same manner and Sanger sequenced after amplification with COI primers. The sequence reads were provided as a bioinformatic reference in this project.

Three site samples with the highest DNA concentrations from each sampling location were selected and amplified using the Folmer primers and PCR. Negative controls were also run in reactions to ensure DNA did not amplify; an indication that sterilization methods were successful at removing any contaminant DNA between field sampling events. All PCR products were filtered using agarose gel electrophoresis

through a 1% agarose gel mixed with the intercalating agent, ethidium bromide. Samples were visualized using florescent light and product bands were noted for each sample. Amplified products from the same sites were pooled and cleaned up using ExoSAP-IT (Thermo Fisher Scientific 2017) and the manufacturer's recommended protocol. Cleaned PCR products were then amplified again using PCR to produce 30uL of sample for downstream library preparation.

Amplicons were sent to UK's genetics lab (Advanced Genetic Techniques Center: AGTC) for amplicon library preparation and Illumina MiSeq high-throughput sequencing. Samples were normalized using qPCR to dilute sample concentrations to equal ratios. The Nextera XT library preparation kit was then used to tagment amplicons with ligation adaptors using a transposition reaction. Specifically engineered transposases were used to fragment double-stranded DNA in order to covalently bind a complementary oligonucleotide to the 5' end of each DNA strand. Then a PCR reaction allowed the application of site-specific markers and Illumina p5 and p7 sequencing adaptors to be added to fragmented DNA amplicons. Amplicons were bound to an Illumina flow-cell and sequenced as paired-end reads on an Illumina MiSeq High-Throughput sequencer.

Bioinformatics Analyses: Quality Filtering and Formation of OTUs

Quality filtering and formation of operational taxonomic units (OTU) were preformed using the open-source, web-based public server, Galaxy version 19.01 (Afgan *et al.* 2018). Forward and reverse read qualities for individual sites were visualized using FastQC v.0.72 (Andrews). FASTQ read-pairs were trimmed to

eliminate low quality reads (PHRED = > 25) and short amplicons less than 125 base pairs were discarded using Trimmomatic (Bolger *et al.* 2014). Paired End reAd mergeR (PEAR) (Zhang *et al.* 2014) was used to assemble paired-end reads with a minimum overlap of 12 bases and a minimum assembled read length of 150 bp total. Assembled FASTQ files were then converted to FASTA using the FASTQ/A short reads pre-processing tool (Gordon 2010) from the FASTX-toolkit. Sequences were sorted by abundance and then examined for chimeras using VSearch (Rognes *et al.* 2015). Non-chimeric sequence files were then passed to VSearch clustering to generate operational taxonomic units (OTU). A known lab contaminant file was created and used to query out contaminant OTU groups that were likely picked up due to any contamination of lab equipment. After this last quality control step, VSearch was used to search the site lists for sequences that resembled *P. ganteri* DNA within a 97% pair-wise match. Sites containing OTUs that matched known *P. ganteri* sequences were marked as positive detections.

Bioinformatic Analyses: Identification of Community Cave Fauna

Genus and species level invertebrate taxa were identified as potential targets for community exploration using a list generated and maintained by the Karst Waters Institute (Hobbs *et al.* 2003). Additionally, some sequences of interest were added to include species known to inhabit or frequent cave systems in the area (i.e. *Amblyopsis spelaea*, *Cambarus tenebrosus*, *Cottus sp.*, etc…). A multiple sequence alignment was created to serve as a database containing homologous COI references from aquatic taxa found in Barren, Edmonson, or Hart counties in Kentucky. All sequences were

downloaded from GenBank and the Barcode of Life Database (BoLD) online libraries. VSearch was used to filter through OTUs generated from eDNA extracted at each site to find approximate pairwise matches with a minimum of 80% sequence similarity to those contained in the reference database. Matches were only considered if sequence similarity was above 80%, alignment length was greater than 100 base pairs, and there were little to no gaps in alignment between a matched OTU and a group of reference sequences.

Bioinformatic Analyses: Identification and Comparison of Allelic Groups

Individual, quality-filtered COI forward and reverse reads from sites with positive shrimp detections were converted to FASTA format directly using the FASTX-toolkit (Gordon 2010). In order to re-orient reads in the 5' to 3' direction, reverse reads were converted to their reverse-complements using the FASTX-toolkit (Gordon 2010). VSearch version 1.0.16 (Rognes *et al.* 2015) was used to search individual forward and reverse (reverse-complemented) read files for amplicons resembling shrimp COI reference DNA. All sequences with a pair-wise match of 98% or greater were aligned with known PG1 and PG5 Sanger sequenced shrimp haplotypes using MAFFT 7.0 (Katoh *et al.* 2002). The alignment was manually sorted so that sequencing pairs from each site were grouped together. AliView version 1.18 (Larsson 2014) was used to visualize and merge read pairs that met the following criteria: 1. Overlapped each other by > 12 base pairs; 2. Had overlapping sections that agreed 100% between forward and reverse reads and; 3. When combined, were greater than 150 base pairs in total length. After sequences were merged, unique alleles were filtered out and

examined. Variants that only appeared a single time (singletons) in the data were removed because they could not reliably be distinguished from potential sequencing errors. Variations that appeared multiple times either within sites or across sites were saved. DECIPHER (Wright 2015, 2016) in Program R (R Core Team 2018) was used to form a consensus sequence representing unique cave shrimp variants. These were then translated in DECIPHER using NCBI's invertebrate mitochondrial code SGC4 (Elzanowski & Ostell 2019). Reference sequences were used to determine the reading frame for the COI sequence fragments and any sequence translations from Illumina data which created stop codons that conflicted with references were removed and the alignment was saved. The final list of unique alleles were renamed to represent their unique variants, identified by site, and mapped in QGIS version 3.4 Madeira (QGIS Development Team 2018).

The Molecular Evolutionary Genetics Analysis (MEGA version 7) software package (Kumar *et al.* 2016) was used to generate an evolutionary distance matrix between haplotypes using a maximum composite likelihood model (Tamura *et al.* 2004) with pairwise deletion to deal with gaps and missing data in the alignment. These evolutionary distances between the haplotypes were used to infer an optimal neighbor-joining tree (Saitou & Nei 1987) with branch lengths equal to the evolutionary distances computed by the number of base substitutions per site. The Templeton, Crandall, and Sing (TCS) method was used to infer a population level genealogy using the TCS program designed by Clement *et al.* 2000 and a haplotype network was generated to represent estimated population level relationships among

shrimp variants (Clement *et al.* 2000). Available COI sequences, representative of Atyids from the proposed and closely related Paratya and Typhlatya clades (von Rintelen *et al.* 2012) were used to construct a neighbor-joining consensus tree (Felsenstein 1985) from 1000 iterations in an attempt to place *P. ganteri* haplotypes in context with other closely related shrimp genera.

III. Results

Sterilization and Quality Control

Best practices for equipment sterilization and the addition of negative controls provided preliminary quality assessments at the primary PCR step. After initial PCR amplification, negative controls from nine of the eleven sites showed no bands. This demonstrated that equipment used at each of these nine sites was unlikely to have contained persistent DNA, which could lead to false positive detections. Negative controls created prior to sampling for two of the sites produced bands in the primer target regions after initial PCR replication. Those samples could possibly have contained DNA from a previous sampling event and were excluded from further analyses.

Laboratory Results

DNA concentrations from sample extractions were generally low and ranged from 3 ng/mL – 10 ng/mL. Ganter Bluehole was the exception with several samples producing concentrations greater than 25 ng/uL (Figure 5, Appendix B). Despite relatively low concentrations of sample DNA, the thermal cycling reaction with the COI Folmer primers yielded product in 71% of sample replicates and amplification across all nine sampling sites. Bands ranged from faint to strong and were pooled together by sampling location after PCR cleanup. Each pooled sample produced strong bands across all sites with a product length of approximately 740 base pairs and DNA concentrations ranging from 4.5 ng/mL to 20 ng/mL.

Bioinformatics Results

Forward and reverse read files contained ~ 1.5 million sequences on average. Data quality failed several tests in the FastQC module indicating a need for data filtering and quality control measures; this is a common analysis step with Illumina MiSeq outputs and was expected for these sequencing libraries. Pre-filtered reads contained sequences between 35bp and 300bp and the sequence length distributions within each sample were extremely variable. Percent duplication was high (~ 86% on average) in most samples, indicating sample diversities were relatively low in comparison with the number of total sequenced reads (Table 2, Appendix A). Quality scores (PHRED) for sequencing reads were generally high, but declined as sequencing reads increased in length (Figure 6, Appendix B). Forward reads greater than 185bp were typically much higher in quality when compared to the reads from the reverse strands. After quality filtering with Trimmomatic, short sequence reads were removed, truncating read lengths between 125bp and 300bp. Only sequences that contained no ambiguous base calls and maintained an median score of PHRED = > 25 were advanced in the analysis pipeline (Figure 7, Appendix B).

PEAR filtered out a number of additional reads either by discarding sequence pairs which did not overlap by a minimum of 12 bases or had too many mis-matched pair-wise bases in overlapping regions (Table 3, Appendix A). After reads were assembled, an average of 70% of the original sequences still remained for downstream analysis. VSearch found no chimeras in assembled sequence files. The number of identified operational taxonomic units varied widely, but in general, were relatively

proportional to the initial number of sequencing reads (typically between 1%-2% of initial sequencing reads). The exception was Roaring River Shrimp Pool, which produced 29,769 unique OTU clusters, approximately 4% of the total number of original sequenced reads. This gave Roaring River the highest number of operational taxa units of all the sites, while Pike Spring produced the lowest number of OTU clusters. Known contaminant OTUs made up an extremely small proportion of samples, ranging from two to twenty-nine variants in all sites except Sud's Cave and the River Styx samples, which contained 87 and 582 contaminant OTUs respectively. VSearch detected *P. ganteri* matches from seven of the nine sites with pair-wise matches all above 99.4% similarity (Table 4, Appendix A). These sites were mapped to show the distribution of shrimp positive samples across the karst water drainages around Mammoth Cave National Park and the WKU Green River Nature Preserve (Figure 8, Appendix B). In addition, precipitation and flow information from the Brownsville gauge on the Green River were plotted along with sampling dates to hypothesize if conditions may have had an effect on detections (Figure 9, Appendix B).

Using the aquatic community reference database compiled using the Karst Water Institute taxa list, VSearch matched 83 Illumina amplicons with a 100 bp pair-wise match or better to compiled reference sequences. Inferred detections included sequences found across all nine sampling sites with sequence similarity ranging from 80 – 100 percent. All sites previously identified as shrimp positive sites contained sequences with similarities of 97.7% or greater. Several sequences resembling the reference for *Cottus bairdii* were identified from the Echo River and McCoy Bluehole

basins with over 91% sequence similarity and a pairwise alignment of over 400 base pairs. These are likely detections of *Cottus carolinae,* for which no published COI reference sequences currently exist. However, several sightings, captured in field notes, corroborate the presence of Banded Sculpin in one of the cave locations and at several of the spring sites where samples were collected. In addition, similar sequences (> 85% similarity) for both *Amblyopsis spelaea* and *Typhlichthys subterraneus* were found in Echo River and Turnhole Spring basins. DNA matches for *Cambarus tenebrosus* were found at 6 sampling locations, but all were relatively poor matches in relation to reference sequences (< 85% sequence similarity). This was similar for *Orconectes pellucidus*, which was found at three site locations; yet, all pair-wise matches for *O. pellucidus* were less than 83% similar to available references. A sequence resembling Crangonyx and several others with low matches to *Stygobromus* species were found at six locations indicating possible DNA matches to cave dwelling amphipod species. The closest matches to cave obligate amphipods were seen from DNA found in Echo River (87.6% match with 105 pair-wise alignment to *Stygobromus hayi* and 86.9% match with 107 bp alignment to *Stygobromus allegheniensis*), and Running Branch (85.7% match with 112 bp alignment to *Stygobromus ozarkensis*). Other sites containing lower matches for cave obligate amphipod genera were Ganter Bluehole, Pike Spring, Sud's Cave, and McCoy Bluehole.

As explained above, individual forward and reverse read files at shrimp-positive sites were queried for shrimp sequences independently in an effort to obtain a higher resolution for differences in sequenced variants. Matching sequences often

overlapped by 100 base pairs or more. Several sequence reads did not overlap, and were excluded from further analysis. After sequences were merged and aligned with the CTAB extracted reference sequences, 27 variations containing one or more single nucleotide polymorphisms (SNP) were found to represent additional differences between similar variants. Nine singletons were removed from the data and all additional sequences were able to be combined into distinct variant groups (Table 5, Appendix A). Three of those variants, Haplotype A, Haplotype C, and Haplotype G, were found in more than one location (Table 6, Appendix A; Figure 8, Appendix B). Haplotypes D, E, F, and H were all found at a single location among five of the karst water basins sampled. Haplotype B was not exactingly identified at any location, but its existence was known because it was one of the reference sequences isolated from the CTAB extracted tissue samples. Both Ganter Blue Hole and Running Branch had positive detections that could easily be recognized as shrimp DNA; however, the sequenced regions fell between relatively conserved sections of DNA, making the identification of several variants equally likely.

The TCS model showed a tight grouping of genealogies between the majority of the inferred genetic divisions (Figure 10a, Appendix B). Haplotype D was slightly removed, contingent on a difference of two base pairs between it and the next closest alleles: haplotypes A and E. Haplotype G was quite removed from the other allelic groups and differed from Haplotype C (its closest neighbor) by five SNPs. Equally-parsimonious connections from Haplotype H with variants E and A are represented in the network and are likely the result of too little variation and missing data towards

the 3' end of Haplotype H. This is a similar case for Haplotype D, which is equally removed from variants E and A by two substitutions, but it is unclear, due to gaps in the data, which group might be closer.

Analyses in MEGA showed some similar relationships between haplotype variants branched among a neighbor-joining tree when compared to the genetic network (Figure 10b, Appendix B). Genetic distances were relatively close for most variants excluding Haplotype G (Table 7, Appendix A). Haplotypes E and H clustered together with a potential common ancestor, also shared with Haplotype D. Haplotype A appeared to be the next closest evolutionary descendent off the node. This grouping basically makes up the lower half of the TCS gene network. Haplotypes B and F appear to be as removed from Haplotype C as Haplotype A. Haplotype G is still considered the farthest removed of all variants in the tree. The broader phylogenetic analysis, using a neighbor-joining tree considering the Kentucky Cave Shrimp within the context of other atyids, placed all *Palaemonias ganteri* branches as derivations off the same node (Figure 11, Appendix B). This cluster, relative to the other atyid groups, was well supported when considering the bootstrap values. In addition, all *P. ganteri* haplotypes were monophyletic with their next closest sister group being two Australian species within the genus *Stygocaris*. Most other nodes were highly supported toward the terminal and sub-terminal levels; however, nodes above those were relatively poorly supported and indicate a need for additional nuclear data to aid in clarifying higher level relationships. Proposed Paratya and Typhlatya groups were paraphyletic and subterranean freshwater species were interspersed with

subterranean anchialine and the surface dwelling freshwater atyids (Figure 12, Appendix B).

IV. Discussion

Theoretically, organismal DNA released into the environment through materials such as skin cells, mucus secretions, and feces, should infer species presence (Ficetola *et al.* 2008; Darling & Mahon 2011). Unfortunately, this is not always true, as the detection of an organism's DNA does not necessarily confirm its presence in the environment. However, with cave and karst water systems being fairly isolated and cave obligate species being far less ubiquitous than surface species, environmental DNA samples provide good opportunities for targeted detection. Nevertheless, many physical and chemical conditions can compromise DNA integrity. Chemical mutagens can alter the structure of DNA, heat can separate and fragment molecules, and radiation can inactivate DNA making it difficult to amplify during downstream processing (Nielsen *et al.* 2007). Additionally, DNA is subjected to breakdown by saprophytes, extracellular DNases, and bacterial communities in aqueous environments (Nielsen *et al.* 2007).

Subterranean karst water systems are likely good candidates for eDNA studies primarily because of their relatively stable environment. Water flowing through underground karst systems is protected from ultraviolet radiation, drastic changes in temperature, and buffered against chemical acidity by limestone. With most stygobitic organisms being relatively recluse, eDNA has the potential to augment monitoring and management programs by addressing the data insufficiencies of traditional sampling techniques. This study was successful in identifying shrimp DNA from seven of the ten historical sub-basins known to house shrimp populations using only filtered water

samples. This shows that water exiting a subterranean system can be utilized to collect, isolate, and amplify relatively large (650 – 700 bp) fragments of DNA, useful in detecting the Kentucky Cave Shrimp. An additional benefit of this remote monitoring was the ability to use spring locations for sampling. Surface springs were comparably easy to access in contrast to subterranean sites and provided DNA capture from a large majority of each basin at a singular point where water exited the system to join with the Green River.

Community analysis using VSearch and the compiled references, downloaded from Genbank and BoLD, were fairly unsuccessful. The identification of species that had close pairwise matches to reference sequences were few and far between. The only matches that were relatively close to reference sequences were those identified as *P. ganteri*, *Cottus bairdii* (likely *C. carolinae*), and *Amblyopsis spelaea*. This was disappointing, but not particularly surprising considering that only two reference sequences were available for download out of the 27 aquatic obligate, karst-water species identified by Hobbs *et al.* (2003) from Barren, Hart, and Edmonson counties. In order to supply VSearch with additional reference sequences, taxa within the same genus or of some of interest (i.e. the southern cavefish) were included in the alignment. While I do not know the variability of COI at the genus level for any of these cave obligates, I would assume that they did not accurately reflect the genetic variation for species present in the Mammoth Cave region. With several species in the analysis having close approximations (> 90% pair-wise matches), community analysis may have some promise for future applications. However, baseline references are

currently needed to provide accurate genetic databases, useful in identifying species diversity within these specific subterranean systems.

The only sub-basin where shrimp DNA was not detected in Illumina data was Echo River. This was interesting due to the fact that the Roaring River Passage was originally where the species was first described, in addition to being one of the areas where its initial decline was first noticed. This basin was sampled twice, once from the Roaring River Spring opening, near Green River Ferry Road, and within the basin, at the Roaring River Shrimp pools. Access to the surface spring was fairly easy and flow out of the spring didn't appear to be under river influence at the time samples were collected. Samples from the shrimp pools were under more lentic conditions and no shrimp were sighted during the time of sampling. While eDNA sampling during high flow has been discouraged due to lower yields seen in other studies, I would argue that some flow from cave systems might actually be necessary for proper species detection. Many organisms inhabiting these epigean environments are primarily scavengers or saprotrophs and the cellular sloughing, secretions, excretions, and extracellular DNA discarded by organisms may be highly sought after as a nutrient source in such a resource limited environment. This could make the detection of DNA in still water or low flow difficult in the absence of the target organism. Water movement and flooding of subterranean environments might facilitate the transport of cells and DNA to spring locations allowing it to be picked up before its consumption by bacteria and other organisms.

Rainfall during the sampling periods did not provide clear answers of whether flow and water movement out of karst springs affected eDNA detection. However, on most occasions, steady flow was visible exiting spring sites and some rainfall had preceded field activities. One consideration for future eDNA sampling when concerning *P. ganteri* detections is that samples taken at groundwater openings in late spring or late fall may produce best results. This recommendation is purely based on a better understanding of the shrimp's biology rather than observation. Both Cooper & Cooper (2011) and Hobbs Jr *et al.* (1977) describe *P. ganteri* and *P. alabamae* as residing in still pools throughout the year, which are then flooded annually, being recharged only by seasonal rains. Without this recharge, there may be less opportunity to remotely gather enough DNA from primary shrimp habitat without being right at the source pool where the organism is located. Thus, seasonal rains may push DNA, via flowing water, to spring mouths before it is taken up by other organisms in the environment.

Flow may also be an important contributor to more than just our ability to detect *P. ganteri* in these systems, it is likely important in population distributions. While my knowledge of sub-basin interconnectedness is limited, I do know that several systems connect during high water levels. However, considering the typical, remote nature and relative isolation of each karst water sub-basin, I had expected a more clustered geographic separation between allelic variants. Surprisingly, no clear patterns existed between haplotype distribution and the current sub-basin drainages. In addition, several haplotypes appeared highly disjunct from each other. There could

be several explanations for this observation: 1) COI variation existed before the formation of the current sub-basins and groups inferred from Illumina data represent the distribution of alleles that have been recently separated; 2) There are a series of unknown passages and interconnectedness that allow for interspersion of reproducing females between basins and sub-populations; or 3) during some situations (possibly high flow scenarios) female shrimp are carried out of upriver sub-basins and deposited in locations further downstream where they migrate into new cave systems and reproduce. The first circumstance is probably unlikely as COI is known to have a relatively fast mutation rate (Avise 2009) and shrimp have been cave inhabitants in the region for a very long time. The second instance is more believable, but would require an extensive network of passages to exist and potentially cross under the Green River system itself. The third hypothesis seems the most likely and could account for similar haplotypes appearing highly separated from each other. This is interesting as it may indicate that efforts placed into exploring karst systems farther down river from Turnhole basin may reveal additional shrimp populations.

Some additional points are also of interest when considering the distribution of inferred haplotypes discovered using these eDNA results. Firstly, Haplotype A was the most frequent allele, appearing with positive detections in four of the seven basins and distributed throughout the study range. Secondly, the Turnhole basin produced the highest number of allelic variants. This system is not only the farthest downstream sub-basin considered here, but is also recognized as the largest karst-water basin sampled in this project. In addition, McCoy Bluehole was the second largest basin

under consideration in this study and contained the second highest number of inferred haplotypes. Higher numbers of haplotypes within Turnhole and McCoy Bluehole sub-basins may indicate that there is a relationship between sub-basin size and allelic diversity. Also, Turnhole basin lies adjacent to several other basins, which also contained DNA from haplotypes A and C. Under certain conditions, the Turnhole system could act as a repository for source populations that may spread to adjacent basins if connections are present, or are opened under certain flow conditions, or temporal changes. Lastly, several of the alleles were unique and only found in one location. Haplotypes D, E, and H were only observed in data taken from McCoy Bluehole, Pike, and Turnhole sub-basins respectively. More work is necessary to investigate whether these basins may hold unique or less frequent variants of shrimp, as it may have real implications in terms of groundwater or land management strategies and regulations on future cave access or use.

While eDNA has not often been used to study genetic diversity at the population level, NGS technologies have progressed to the point where population genetics can be inferred from environmental samples (Adams *et al.* 2019). In fact, Thomsen *et al.* (2016) demonstrated that after correcting for sequencing error rates, Illumina data identifying several haplotypes within whale sharks could be used independently for population genetic inferences (such as haplotype diversity and even frequency) without prior knowledge of the study population. Their findings were verified and found to be consistent with reference sequences available from traditional tissue samples. In addition, a recent study collecting DNA from sea water in

the wake of killer whales was able to identify the pod's regional ecotype using a 700bp region of mtDNA (Baker *et al.* 2018). This discovery matched pod vocalizations picked up on hydrophones and direct field observations from sampling encounters. Finally, Parsons *et al.* (2018) used eDNA to overcome challenging sampling limitations for harbor porpoises in order to determine population-level estimates of genetic diversities for pods located in coastal waters of southeast Alaska. Findings revealed previously unknown haplotype diversity and lead to the suggestion that management strategies be applied separately across two distinct harbor porpoise populations. These pioneering studies are not an exhaustive list of how eDNA is enriching our ability to study unique species in challenging environments, but they provide a solid groundwork for expanding the use of environmental DNA when considering population level ecological research and management goals.

Shrimp haplotypes A – H showed relatively little divergence and many of the single nucleotide polymorphisms seen across the eight alleles appear to be carried along through some of the gene variants (Table 5, Appendix A). This resemblance is apparent in the haplotype network generated using these data (Figure 10a, Appendix B). Only Haplotype G appeared far removed from other sequences, while most variants formed a fairly close network with little difference between adjacent haplotypes. Originally, Haplotype G was considered for removal from results because it differed so widely from other groups and contained nine nucleotide substitutions over a 331 bp fragment. However, alignment with reference sequences and translation of DNA revealed only two amino acid changes among the nine base pair

substitutions; indicating that the polypeptide structure of Haplotype G was still relatively conserved.

Unfortunately, haplotype groups identified using Illumina MiSeq data are only inferred and are otherwise difficult to verify. However, several points might be made in validating the results contained here. When considering data quality, only sequences with a low probability of incorrect base calls (< 1 in 1,000) were utilized for this analysis. Also, haplotypes A, C, and G were all identified from multiple locations, which means that if replication mistakes were made during the PCR process, they would have had to occur at the same positions in separate PCR reactions. While this is highly unlikely, it is also important to note that I used a proof-reading taq polymerase (Phusion Taq) for initial amplification of extracted DNA samples. Theoretically this made it less likely that point mutations seen in the data might be attributed to replication errors during initial PCR thermal cycling. Moreover, two variants (Haplotype A and Haplotype B) were known from reference sequences obtained from CTAB-extracted samples. One of these variants, Haplotype A, was confirmed at several sites and may have been identified independently, without the known COI reference. Considering these points, I feel confident that the haplotype groups identified from sampled locations reflect part of the true genetic diversity present in *P. ganteri* populations.

While the use of Illumina MiSeq alleles provided enough resolution to begin investigating intraspecies relationships among shrimp haplotypes in these populations, the data is limited in its ability to relate *P. ganteri* among other atyids. This limitation is

due to the fact that COI is a mitochondrial gene with a moderately high mutation rate. As seen in most animal groups, it is an excellent marker for distinguishing between closely related taxa, but preforms poorly when used to relate more distant species. MEGA 7 analysis of the Paratya and Typhlatya groups demonstrate this clearly as atyid species appeared to group well regionally, but showed poor support for relationships between those groups globally (Figure 11, Appendix B). Nuclear data is necessary for clarifying relationships at earlier nodes in phylogenies, but is likely present in the environment in smaller quantities and would be more difficult to target using DNA found in the environment.

Additional detections of shrimp, as well as any other species of interest, may improve with the development or use of smaller mini-barcode primers. In fact, several primers are in existence today that amplify, in part, a 150 to 200 bp region of the COI gene, while still providing high species coverage with good taxonomic resolution (Meusnier *et al.* 2008). Unfortunately, as amplicon length begins to decrease, so does the resolution for population-level genetic analyses. Ultimately, there is a balance between simply being able to detect taxa in the environment and obtaining enough genetic information about those populations or communities to look at intraspecies variations. Future development and use of eDNA for detecting and studying *P. ganteri* populations will depend on management and conservation objectives.

The Folmer primers (Folmer *et al.* 1994) used in this study may not be considered ideal when simply seeking to detect taxa of interest. Because eDNA can be subjected to considerable damage or breakdown after being released from an

organism, the amplified products used here are considered quite long. In most cases, the need to amplify longer environmental sequences likely decreases the probability that target sequences can be obtained and amplified from environmental samples. In addition, Folmer's primers are known to amplify bacterial gene regions that may introduce problems when bacterial fragments outnumber targeted metazoan DNA. Accordingly, primer design and specificity can place considerable challenges on the ability of researchers and managers to detect target organisms as well as sequence environmental samples.

The use of the Nextera XT library preparation kit complicated the study of population-level diversity, but was necessary for two major reasons: 1) its ease of use and time efficiency when coupled with Illumina MiSeq technologies created a concise workflow for handling site specific multiplexing of pooled samples; and 2) it addressed read length limitations between the Folmer primer product lengths (650 – 700 bp) and current Illumina sequencing limitations. The Nextera XT kit allowed for site specific multiplexing of PCR pooled samples, while also decreasing total amplicon length so that products could be sequenced on the Illumina platform smoothly as paired-end reads. MiSeq systems can only generate 300 bp paired-end reads in a single sequencing run. The larger amplicons produced by the Folmer primers, product lengths needed to be shortened. However, the drawback was that random products of the Nextera XT transposases produced fragmented sequences at random lengths varying from 35bp to 300bp. Only sequences greater than 125bp were kept for analysis and some information was likely lost through initial filtering. Because of this,

the alignment location of shrimp amplicons from two locations (Ganter Bluehole and Running Branch) could be identified as shrimp DNA, but could not be used to assign haplotypes.

Despite this study's use of an indiscriminate primer pair with broad taxonomic specificity, shrimp DNA was still able to be identified from seven of the ten historically recognized shrimp basins using only water samples. This process circumvented potential hurdles caused by species rarity or access and required no previous DNA references for primer design. Additionally, the data provided enough information to infer six potentially unique shrimp alleles across amplified sections of the COI gene region. Also, sequencing data identified one additional haplotype in several basins that was a verified reference sequenced using tissue extractions from harvest shrimp pleopods. One additional known haplotype (Haplotype B) was not able to be distinguished at any site, but provides evidence that results displayed here are only the beginning of what could be a much larger population genetics initiative.

With associated NGS technologies advancing and becoming more cost effective, eDNA is rapidly approaching a point where it may become an integral step when surveying rare, vagile, endangered, threatened, or sensitive species. Species with data deficiencies or those that require extensive permitting or specific skillsets to access can cost both time and effort, which may be hurdles for progress toward understanding and conserving cave biodiversity. In addition, *P. ganteri's* endangered status contributes to a reluctance for acquiring direct tissue samples, making it difficult to obtain valuable population information, which could be useful when considering

management decisions. As seen here, DNA from water samples can be used to aid in surveying this difficult landscape for species that are often hard to find.

Future advancements, which may build on the information presented here, should be directed towards improving and expanding several points of interest when considering the use of eDNA for targeted monitoring of cave populations. When considering *P. ganteri*, the design of better gene target regions should be a priority. It would be interesting to attempt to include sections that contain the variation seen here, but only require a fragment that incorporates all or most of the sites where point mutations were observed using this sequencing data. Also, other gene regions or mini-barcoding sections should also be investigated to ensure that the best options for detection of cave fauna (i.e. *P. ganteri*) are being used. Moreover, as was apparent when investigating the use of this data to identify additional cave fauna, it is clear that better reference databases are badly needed for troglobitic species in the Mammoth Cave region. Without these references, metabarcoding pipelines will likely not provide the resolution necessary for monitoring and conservation decisions. It is my recommendation that some effort be placed into sampling and sequencing a broad range of cave fauna in the area so that appropriate reference databases can be compiled for additional monitoring activities using eDNA. Finally, with the possibility that allelic distributions may be influenced by downriver movement, efforts to survey systems farther west of the Turnhole sub-basin that are connected to the Green River should be investigated for additional shrimp populations.

References

Adams C, Knapp M, Gemmell NJ *et al.* (2019) Beyond Biodiversity: Can environmental DNA (eDNA) cut it as a population genetics tool? *Genes*, **10**, 192.

Afgan E, Baker D, Batut B *et al.* (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Research*, **46**, W537–W544.

Allen GC, Flores-Vergara MA, Krasynanski S, Kumar S, Thompson WF (2006) A modified protocol for rapid DNA isolation from plant tissues using cetyltrimethylammonium bromide. *Nature Protocols*, **1**, 2320–2325.

Alvarez AJ, Yumet GM, Santiago CL, Toranzos GA (1996) Stability of Manipulated Plasmid DNA in Aquatic Environments. *Environmental Toxicology and Water Quality: An International Journal*, **11**, 129–135.

Andrews S A quality control tool for high throughput sequence data.

Asmyhr MG, Linke S, Hose G, Nipperess DA (2014) Systematic conservation planning for groundwater ecosystems using phylogenetic diversity. *PLoS ONE*, **9**, e115132.

Avise JC (2009) Phylogeography: retrospect and prospect. *Journal of Biogeography*, **36**, 3–15.

Baird DJ, Hajibabaei M (2012a) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular Ecology*, **21**, 2039–2044.

Baird DJ, Hajibabaei M (2012b) Biomonitoring 2.0: a new paradigm in ecosystem assessment made possible by next-generation DNA sequencing. *Molecular ecology*, **21**, 2039–44.

Baker CS, Steel D, Nieukirk S, Klinck H (2018) Environmental DNA (eDNA) from the wake of the whales: droplet digital PCR for detection and species identification. *Frontiers in Marine Science*, **5**, 1–11.

Barr Jr TC, Holsinger JR (1985) Speciation in cave faunas. *Annual Review of Ecology and Systematics*, **16**, 313–337.

Bik HM, Porazinska DL, Creer S *et al.* (2012) Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, 1–11.

Blankenship S, Teply M, Dominguez L, Schumer G (2011) *Sampling and analysis to assess brook trout (Salvelinus fontinalis) population trends in High Lake (Oregon) using environmental DNA monitoring*. Malheur National Forest.

Blaxter ML (2004) The promise of a DNA taxonomy. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **359**, 669–79.

Bohmann K, Evans A, Gilbert MTP *et al.* (2014) Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, **29**, 358–367.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.

Brown RF (1991) *Hydrology of the cavernous limestones of the Mammoth Cave area Kentucky*. United States Government Printing Office, Washington D.C.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Cooper JE, Cooper MR (2011) Observations on the biology of the endangered stygobiotic shrimp Palaemonias alabamae, with notes on P. ganteri (Decapoda: Atyidae). *Subterranean Biology*, **8**, 9–20.

Culver DC, Christman MC, Elliott WR, Hobbs III HH, Reddell JR (2003) The North American obligate cave fauna: regional patterns. *Biodiversity and Conservation*, **12**, 441–468.

Darling JA, Mahon AR (2011) From molecules to management: adopting DNA-based methods for monitoring biological invasions in aquatic environments. *Environmental Research*, **111**, 978–88.

Dejean T, Valentini A, Duparc A *et al.* (2011) Persistence of environmental DNA in freshwater ecosystems. *PloS one*, **6**, e23398.

Elliott WR (2000) Conservation of the North American cave and karst biota. In: *Ecosystems of the World (Book 30) Subterranean Ecosystems* (eds Wilkens H, Culver DC, Humphreys WF), pp. 665–689. Elsevier, Amsterdam.

Elzanowski A, Ostell J (2019) The genetic codes.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, **39**, 783–791.

Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. *Biology letters*, **4**, 423–5.

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, **3**, 294–299.

Fonseca VG, Carvalho GR, Sung W *et al.* (2010) Second-generation environmental sequencing unmasks marine metazoan biodiversity. *Nature communications*, **1**, 98.

Glennon A, Groves C (2002) An examination of perennial stream drainage patterns within the Mammoth Cave watershed, Kentucky. *Journal of Cave and Karst Studies*, **64**, 82–91.

Gordon A (2010) FASTQ/A short-reads pre-processing tools.

Hajibabaei M, Shokralla S, Zhou X, Singer G a C, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. *PloS one*, **6**, e17497.

Hajibabaei M, Smith MA, Janzen DH *et al.* (2006) A minimalist barcode can identify a specimen whose DNA is degraded. *Molecular Ecology Notes*, **6**, 959–964.

Hebert PDN, Ratnasingham S, DeWaard JR (2003) Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings. Biological sciences / The Royal Society*, **270**, S96-9.

Hobbs HH, Culver DC, Elliott WR (2003) A list of cave-limited species in the United States and Canada. *Karst Waters Institute Electronic Documents and Databases*.

Hobbs Jr HH, Hobbs III HH, Daniel MA (1977) *A review of the troglobitic decapod crustaceans of the Americas*. Smithsonian Institution Press, Washington D.C.

Hobbs HH, Lodge DM (2010) Decapoda. In: *Ecology and Classification of North American Freshwater Invertebrates* (eds Thorp JH, Covich AP), pp. 901–967. Elsevier Inc.

Holsinger JR (1988) Troglobites : the evolution of cave-dwelling organisms. *American Scientist*, **76**, 146–153.
Howarth F (1993) High-stress subterranean habitats and evolutionary change in cave-inhabiting arthropods. *The American Naturalist*, **142**, S65–S77.
Jerde CL, Mahon AR, Chadderton WL, Lodge DM (2011) "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, **4**, 150–157.
Ji Y, Ashton L, Pedley SM *et al.* (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. *Ecology letters*, **16**, 1245–57.
Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*, **30**, 3059–3066.
Kemp BM, Smith DG (2005) Use of bleach to eliminate contaminating DNA from the surface of bones and teeth. *Forensic Science International*, **154**, 53–61.
Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, **33**, 1870–1874.
Larsson A (2014) AliView: A fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*, **30**, 3276–3278.
Layman SR, Mayden RL (2012) Morphological Diversity and Phylogenetics of the Darter Subgenus Doration (Percidae: Etheostoma), with Descriptions of Five New Species. *Bulletin: Alabama Museum of Natural History*, **30**, 83.
Lisowski EA (1983) Distribution, habitat , and behavior of the Kentucky Cave Shrimp Palaemonias ganteri, Hay. *Journal of Crustacean Biology*, **3**, 88–92.
Lodge DM, Turner CR, Jerde CL *et al.* (2012) Conservation in a cup of water: estimating biodiversity and population abundance from environmental DNA. *Molecular Ecology*, **21**, 2555–8.
Mächler E, Deiner K, Steinmann P, Altermatt F (2014) Utility of environmental DNA for monitoring rare and indicator macroinvertebrate species. *Freshwater Science*, **33**, 1174–1183.
Mardis ER (2008) Next-generation DNA sequencing methods. *Annual review of genomics and human genetics*, **9**, 387–402.
May MT, Kuehn KW, Groves CG, Meiman J (2005) Karst geomorphology and environmental concerns of the Mammoth Cave region, Kentucky. In: *American Institute of Professional Geologists: 42nd Annual Meeting "Geologic Information:Racing into the Digital Age."* (eds Spalding T, Rhodes L, Lumm D, et al.), p. 49. tpmgroup.
Meusnier I, Singer GAC, Landry J-F *et al.* (2008) A universal DNA mini-barcode for biodiversity analysis. *BMC genomics*, **9**, 214.
MO BIO Laboratories (2016) PowerWater ® DNA Isolation Kit.
Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Research*, **8**, 4321–4326.
National Park Service (2006) *Water Resources Management Plan - Mammoth Cave National Park*.

Nielsen KM, Johnsen PJ, Bensasson D, Daffonchio D (2007) Release and persistence of extracellular DNA in the environment. *Environmental Biosafety Research*, **6**, 37–53.

Olson ZH, Briggler JT, Williams RN (2012) An eDNA approach to detect eastern hellbenders (Cryptobranchus a. alleganiensis) using samples of water. *Wildlife Research*, **39**, 629–636.

Palmer AN (1995) *A geological guide to Mammoth Cave National Park* (RA Watson, Ed,). Zephyrus Press, Teaneck, NJ.

Parsons KM, Everett M, Park L (2018) Water, water everywhere: environmental DNA can unlock population structure in elusive marine species. *Royal Society Open Science*, **5**.

QGIS Development Team (2018) QGIS geographic information system.

R Core Team (2018) R: a language and environment for stastical computing.

von Rintelen K, Page TJ, Cai Y *et al.* (2012) Drawn to the dark side: A molecular phylogeny of freshwater shrimps (Crustacea: Decapoda: Caridea: Atyidae) reveals frequent cave invasions and challenges current taxonomic hypotheses. *Molecular Phylogenetics and Evolution*, **63**, 82–96.

Rognes T, Mahé F, Xflouris (2015) vsearch: VSEARCH version 1.0.16.

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, **4**, 406–425.

Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. *Molecular Ecology*, **21**, 1794–805.

Souza H V., Marchesin SRC, Itoyama MM (2016) Analysis of the mitochondrial COI gene and its informative potential for evolutionary inferences in the families Coreidae and Pentatomidae (Heteroptera). *Genetics and Molecular Research*, **15**, 1–14.

Taberlet P, Coissac E, Hajibabaei M, Rieseberg LH (2012) Environmental DNA. *Molecular ecology*, **21**, 1789–93.

Takahara T, Minamoto T, Doi H (2013) Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PloS one*, **8**, e56584.

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. In: *Proceedings of the National Academy of Sciences* , pp. 11030–11035.

Thermo Fisher Scientific (2017) ExoSAP-IT ™ PCR Product Cleanup. , 1–2.

Thomsen PF, Kielgast J, Iversen LL *et al.* (2011) Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular ecology*, **21**, 2555–2828.

Thomsen PF, Møller PR, Knudsen SW *et al.* (2016) Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, **1**, 0004.

U.S. Fish and Wildlife Service (1988) *Kentucky Cave Shrimp RecoveryPlan*. Atlanta, Georgia.

US Department of the Interior (2012) National Park Service: Mammoth Cave homepage.

Vörös J, Márton O, Schmidt BR *et al.* (2017) Surveying Europe's Only Cave-Dwelling Chordate Species (Proteus anguinus) Using Environmental DNA. *Plos One*, **12**, e0170945.

White WB, Culver DC, Herman JS, Kane TC, Mylroie JE (1995) Karst Lands. *American Scientist*, **83**, 450–459.

White WB, Watson RA, Pohl ER, Brucker R (1970) The central Kentucky karst. *Geographical Review*, **60**, 88–115.

Wright ES (2015) DECIPHER: Harnessing local sequence context to improve protein multiple sequence alignment. *BMC Bioinformatics*, **16**, 1–14.

Wright ES (2016) Using DECIPHER v2.0 to Analyze Big Biological Sequence Data in R. *The R Journal*, **8**, 352–359.

Yoccoz NG (2012) The future of environmental DNA in ecology. *Molecular ecology*, **21**, 2031–8.

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: A fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, **30**, 614–620.

APPENDICES

Appendix A: Tables

Table 1. A list of the known groundwater basins that have historically housed P. ganteri and their earliest known documented sighting within the basin.

Recognized *P. ganteri* Basins	First Known Historical Record	Reference
Echo River Basin	Hay, 1902	Lisowski 1983
Ganter Bluehole Basin	Leitheuser & Holsinger, 1983	U.S. Fish and Wildlife Service 1988
McCoy Bluehole Basin	Leitheuser & Holsinger, 1983	U.S. Fish and Wildlife Service 1988
Mile 205.7 Spring Basin	Leitheuser & Holsinger, 1983	U.S. Fish and Wildlife Service 1988
Pike Spring Basin	Barr, 1967	Lisowski 1983
River Styx Basin	Bolivar & Jeannel, 1928	Lisowski 1983
Running Branch Basin	Leitheuser & Holsinger, 1983	U.S. Fish and Wildlife Service 1988
Sand Cave Basin	Leitheuser, 1988	U.S. Fish and Wildlife Service 1988
Suds Cave Basin	Leitheuser, 1988	U.S. Fish and Wildlife Service 1988
Turnhole Spring Basin	Leitheuser, 1988	U.S. Fish and Wildlife Service 1988

Table 2. A table indicating the changes in percent duplication and sequencing ranges in data before and after the use of Trimmomatic.

	Pre-Trimmomatic Quality Filtering				Post Trimmomatic Quality Filtering			
Sample	**% Duplicate**	**% GC**	**Total Seq**	**Seq Ranges**	**% Duplicate**	**% GC**	**Total Seq**	**Seq Ranges**
McCoyBluehole_R1	84.4	43.0	1,083,992	35-300	81.0	45.0	615,882	125-300
McCoyBluehole_R2	83.3	43.0	1,083,992	35-300	81.0	45.0	615,882	125-300
Sud'sCave_R1	91.0	43.0	1,416,761	35-300	89.2	42.0	1,011,879	125-300
Sud'sCave_R2	90.3	43.0	1,416,761	35-300	89.1	42.0	1,011,879	125-300
RoaringRiver_R1	79.6	45.0	750,862	35-300	90.9	44.0	876,208	125-300
RoaringRiver_R2	79.2	45.0	750,862	35-300	91.0	44.0	876,208	125-300
RiverStyx_R1	87.4	42.0	1,261,387	35-300	88.3	47.0	689,551	125-300
RiverStyx_R2	87.3	42.0	1,261,387	35-300	88.4	47.0	689,551	125-300
RunningBranch_R1	88.9	44.0	1,056,637	35-300	85.5	43.0	848,700	125-300
RunningBranch_R2	88.5	44.0	1,056,637	35-300	85.5	43.0	848,700	125-300
GanterBlueHole_R1	87.0	47.0	865,232	35-300	94.1	43.0	1,007,792	125-300
GanterBlueHole_R2	85.7	47.0	865,232	35-300	94.2	43.0	1,007,792	125-300
OwlCave_R1	82.2	42.0	984,228	35-300	83.4	42.0	836,752	125-300
OwlCave_R2	81.6	42.0	984,228	35-300	83.4	42.0	836,752	125-300
EchoRiver_R1	85.1	46.0	992,397	35-300	86.0	46.0	841,584	125-300
EchoRiver_R2	84.5	46.0	992,397	35-300	86.0	46.0	841,584	125-300
PikeSpring_R1	90.6	43.0	1,069,786	35-300	91.8	43.0	907,042	125-300
PikeSpring_R2	89.9	43.0	1,069,786	35-300	91.8	43.0	907,042	125-300

Table 3. A summary of the number of sequences lost or combined in each sample during quality control steps within the bioinformatics pipeline. The final column shows the number of unique OTUs remaining for analysis after known contaminant removal.

	Sequencing Reads	Quality Trimming	Pair-End Assembly	Chimera Filter	OTU Clustering	Contaminant Filter
Echo River Spring	992,397	841,584	708,909	708,909	19,622	19,617
Roaring River	750,863	615,882	530,631	530,631	29,771	29,769
Ganter Blue Hole	865,232	689,551	561,407	561,407	12,121	12,118
McCoy Bluehole	1,083,992	848,700	728,494	728,494	20,609	20,580
Mile 205.7						
Pike Spring	1,069,786	907,042	759,268	759,268	5,291	5,283
River Styx Spring	1,261,387	1,011,879	895,203	895,203	25,346	24,764
Running Branch	1,056,637	876,208	766,510	766,510	7,772	7,755
Sud's Cave Spring	1,416,761	1,007,792	871,442	871,442	10,723	10,636
Owl Cave	984,228	836,752	738,632	738,632	21,824	21,813
Turnhole Spring						

Table 4. Sampling locations showing the dates of sampling and occurrence of positive detections for P. ganteri eDNA within each groundwater basin.

Groundwater Basin	Site Name	Sample #	Date	Latitude	Longitude	eDNA Detection	% OTU Match
Echo River	Echo River Spring	Site 2	9/29/12	37.179303	-86.108583	Negative	N/A
Echo River	Roaring River Shrimp Pool	Site 4	6/4/13	37.176335	-86.093219	Negative	N/A
Ganter Bluehole	Ganter Blue Hole	Site 9	9/14/13	37.187402	-86.147444	Positive	100.0%
McCoy Bluehole	McCoy Bluehole	Site 10	9/15/13	37.249912	-86.009918	Positive	99.4%
Mile 205.7	Mile 205.7 Spring	Site 6	7/21/13	37.225676	-86.039866		
Pike Spring	Pike Spring	Site 3	9/29/12	37.214321	-86.055469	Positive	99.5%
River Styx	River Styx Spring	Site 5	6/14/13	37.186938	-86.109034	Positive	100.0%
Running Branch	Running Branch Spring	Site 7	9/14/13	37.189443	-86.126111	Positive	100.0%
Sud's Cave	Sud's Cave Spring	Site 11	9/15/13	37.239168	-86.015335	Positive	100.0%
Turnhole Spring	Owl Cave	Site 1	9/29/12	37.150196	-86.154041	Positive	99.8%
Turnhole Spring	Turnhole Spring	Site 8	7/21/13	37.164722	-86.158333		

Table 5. A summary of the base substitutions found at each position from sequences identified as >98% similar to known P. ganteri reference haplotypes. All additional haplotypes were inferred from Illumina sequencing data. Red letters identify where base substitutions have caused changes in the resulting amino acid translation.

	69	130	199	204	222	233	251	252	286	343	351	354	441	484	506	512	534
HaplotypeA	A	G	T	G	T	A	T	C	A	G	C	T	C	C	T	C	C
HaplotypeB	G	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
HaplotypeC	—	A	.	.	.	.	.	.	.	.	T	.	—	—	—	—	—
HaplotypeD	.	.	C	.	.	.	C	.	.	.	.	.	—	—	—	—	—
HaplotypeE	—	—	.	.	.	.	.	.	.	.	.	.	.	.	C	T	.
HaplotypeF	—	—	.	.	.	.	.	.	G	.	.	.	.	.	.	.	.
HaplotypeG	—	—	.	A	C	.	.	T	.	A	T	C	T	G	.	.	T
HaplotypeH	.	.	.	.	.	G	.	.	.	.	.	.	—	—	—	—	—

Table 6. Sites where eDNA evidence confirmed the presence of a particular COI shrimp allele. P. ganteri DNA matches were also found at Roaring River and Ganter Blue Hole, but the amplified regions did not clearly specify a particular haplotype.

	Hap A	Hap B	Hap C	Hap D	Hap E	Hap F	Hap G	Hap H
PG1 Illumina Sample	X							
PG5 Illumina Sample		X						
Echo River Spring								
Roaring River								
Ganter Blue Hole								
McCoy Bluehole	X			X			X	
Pike Spring	X				X			
River Styx Spring	X						X	
Running Branch								
Sud's Cave Spring			X					
Owl Cave	X		X			X		X

Table 7. **A distance matrix generated with MEGA using the number of base**

Table 7. A distance matrix generated with MEGA using the number of base substitutions per site between sequences. Analyses were conducted using the Maxium Composite Likelihood model and data gaps were handled with pair-wise deletion.

	HaplotypeA	HaplotypeB	HaplotypeC	HaplotypeD	HaplotypeE	HaplotypeF	HaplotypeG	HaplotypeH
HaplotypeA		0.003	0.006	0.005	0.005	0.003	0.024	0.002
HaplotypeB	0.003		0.003	0.010	0.005	0.003	0.024	0.007
HaplotypeC	0.006	0.003		0.012	0.004	0.008	0.019	0.009
HaplotypeD	0.005	0.010	0.012		0.009	0.014	0.033	0.008
HaplotypeE	0.005	0.005	0.004	0.009		0.008	0.031	0.004
HaplotypeF	0.003	0.003	0.008	0.014	0.008		0.028	0.008
HaplotypeG	0.024	0.024	0.019	0.033	0.031	0.028		0.027
HaplotypeH	0.002	0.007	0.009	0.008	0.004	0.008	0.027	

Appendix B: Figures

Appendix B: Figures

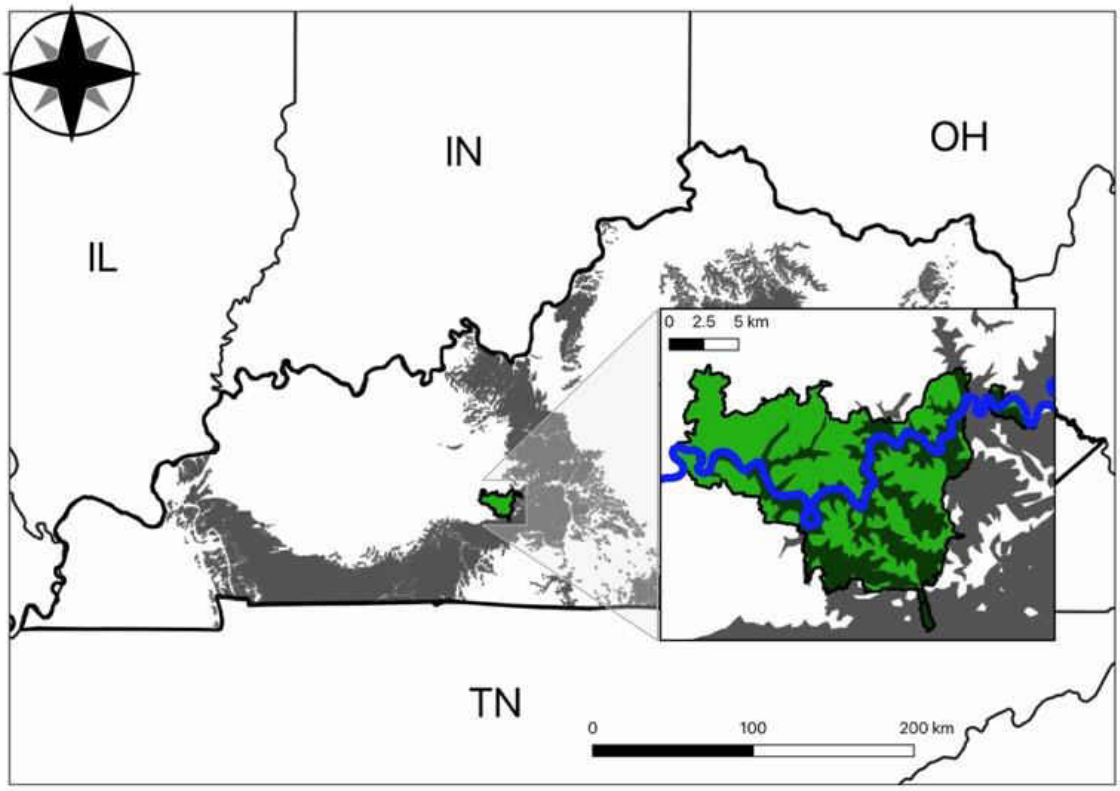

Figure 1. Kentucky Karst Water Regions
A map of the limestone belt that extends from Kentucky's southern border with Tennessee to its northern boundary with Indiana. The location of Mammoth Cave National Park and the WKU Green River Nature preserve are highlighted in Green.

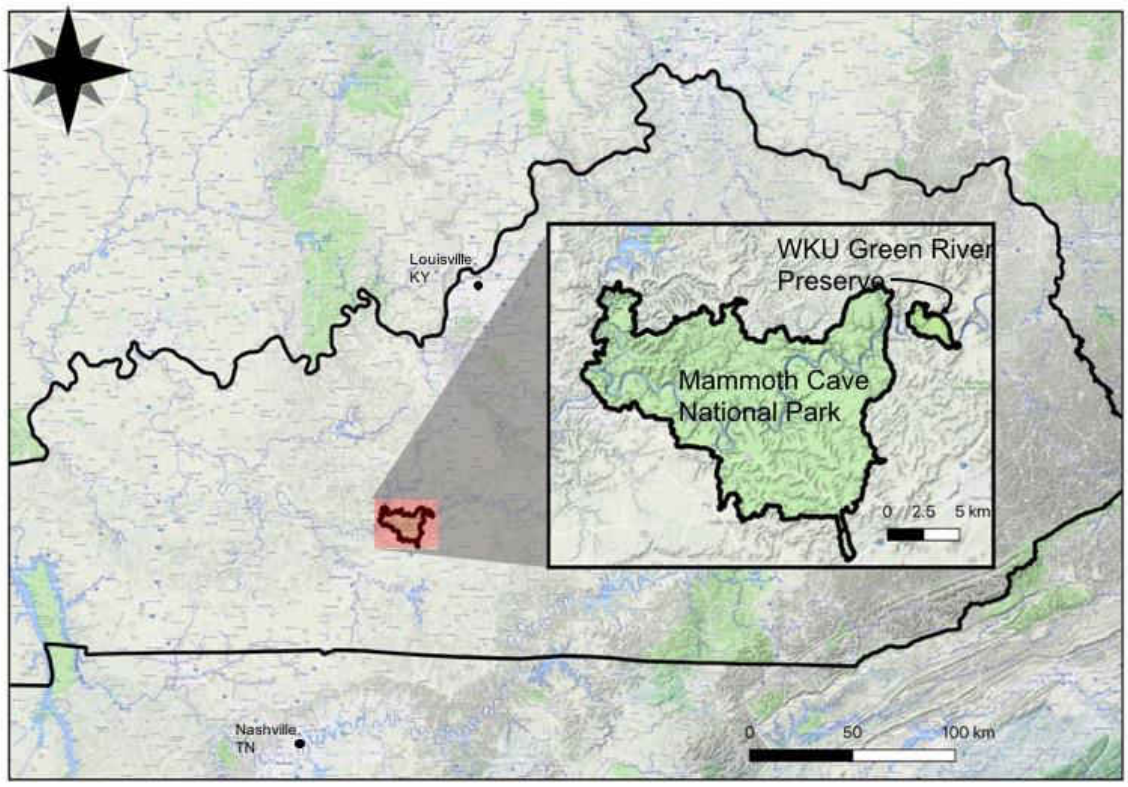

Figure 2. Mammoth Cave National Park and WKU Green River Preserve
The geographic location of Mammoth Cave National Park and the WKU Green River Preserve. Both are located along the borders of the Interior Plateau and Interior River Valleys and Hills ecoregions in the State of Kentucky. Mammoth Cave itself is approximately 160km north-northeast of Nashville, TN and 160km south-southwest of Louisville, KY.

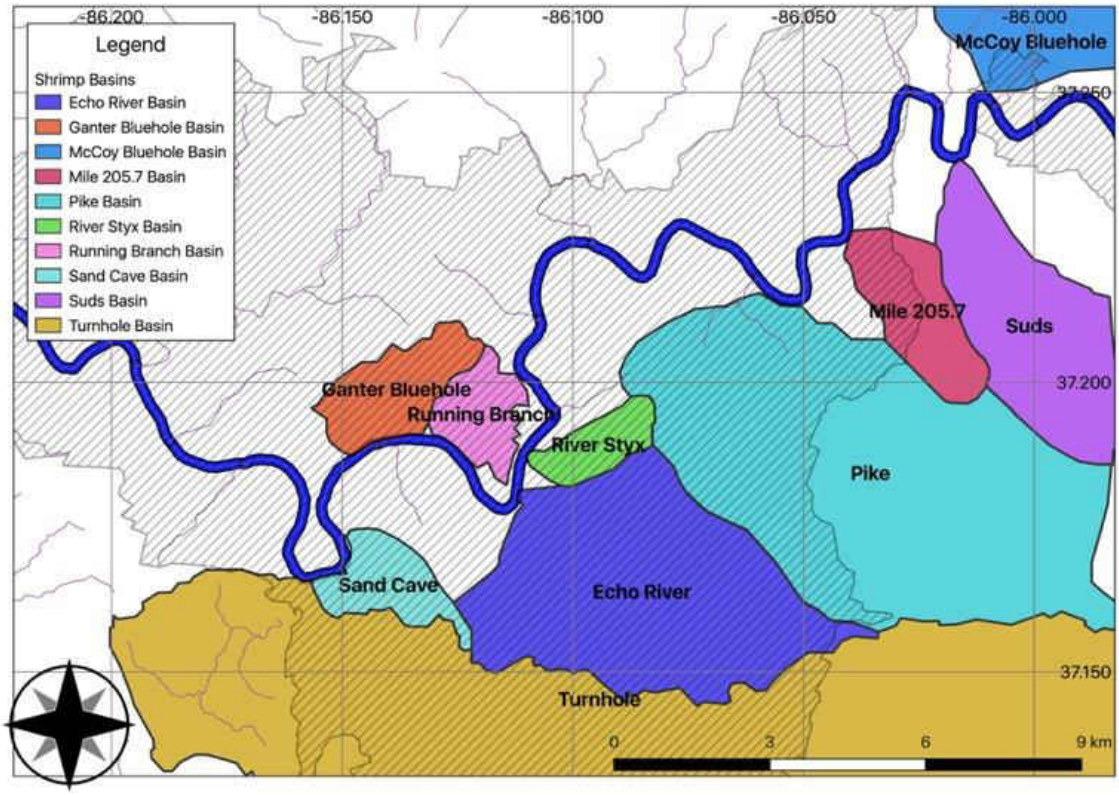

Figure 3. Kentucky Cave Shrimp Subterranean Basins
Ten karst, groundwater basins adjacent to the Green River known to house the federally endangered Kentucky Cave Shrimp, *P. ganteri* (Hay, 1901).

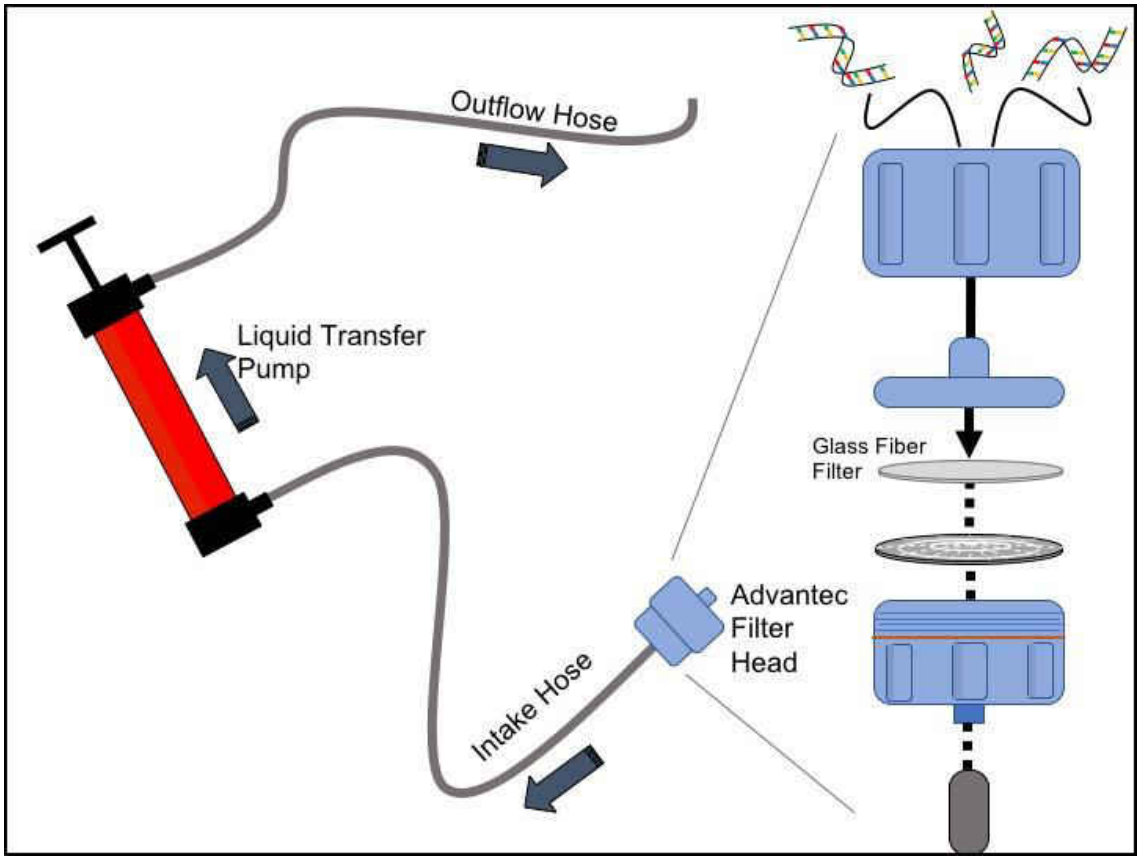

Figure 4. eDNA Field Filtering Equipment
A diagram illustrating the equipment used during this project to filter water samples on site. The Advantec filtering head was fitted with a glass-fiber filter and connected to a liquid transfer pump via standard polyvinyl tubing. All filtered water was collected and dumped away from the site during collection. After two liters of water were filtered through, the filter was simply removed, preserved, and a new filter was placed in the filtering head.

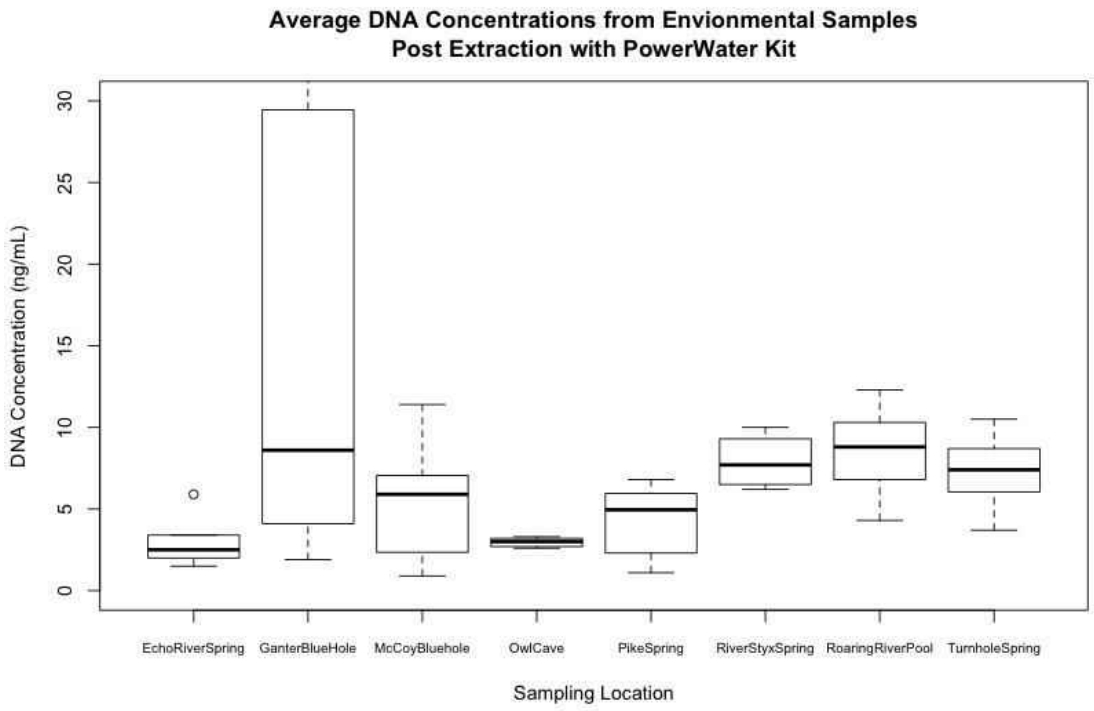

Figure 5. Boxplot of DNA Concentrations per Filter Extraction
A boxplot showing the range of DNA concentrations per filter extraction from water samples taken from Mammoth Cave. DNA concentration was measured by a NanoDrop spectrophotometer and used to select samples with higher DNA concentrations for downstream analysis.

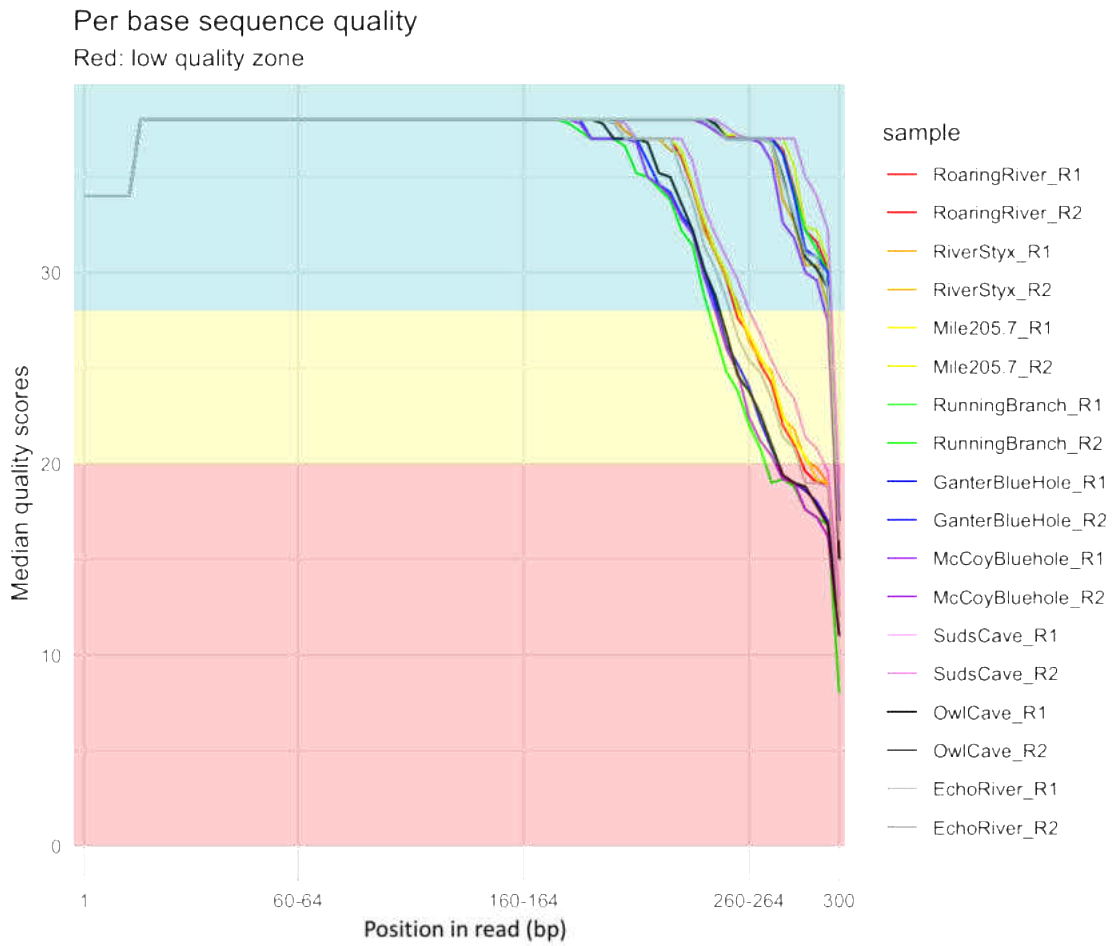

Figure 6. Per-Base Sequence Quality: Pre-Quality Filtering
The median per-base sequence quality (PHRED score) before quality filtering and trimming Illumina data. R1 and R2 denote the forward and reverse reads from each sampling site. In general, reverse reads had lower median quality scores as sequencing length increased past 165bp.

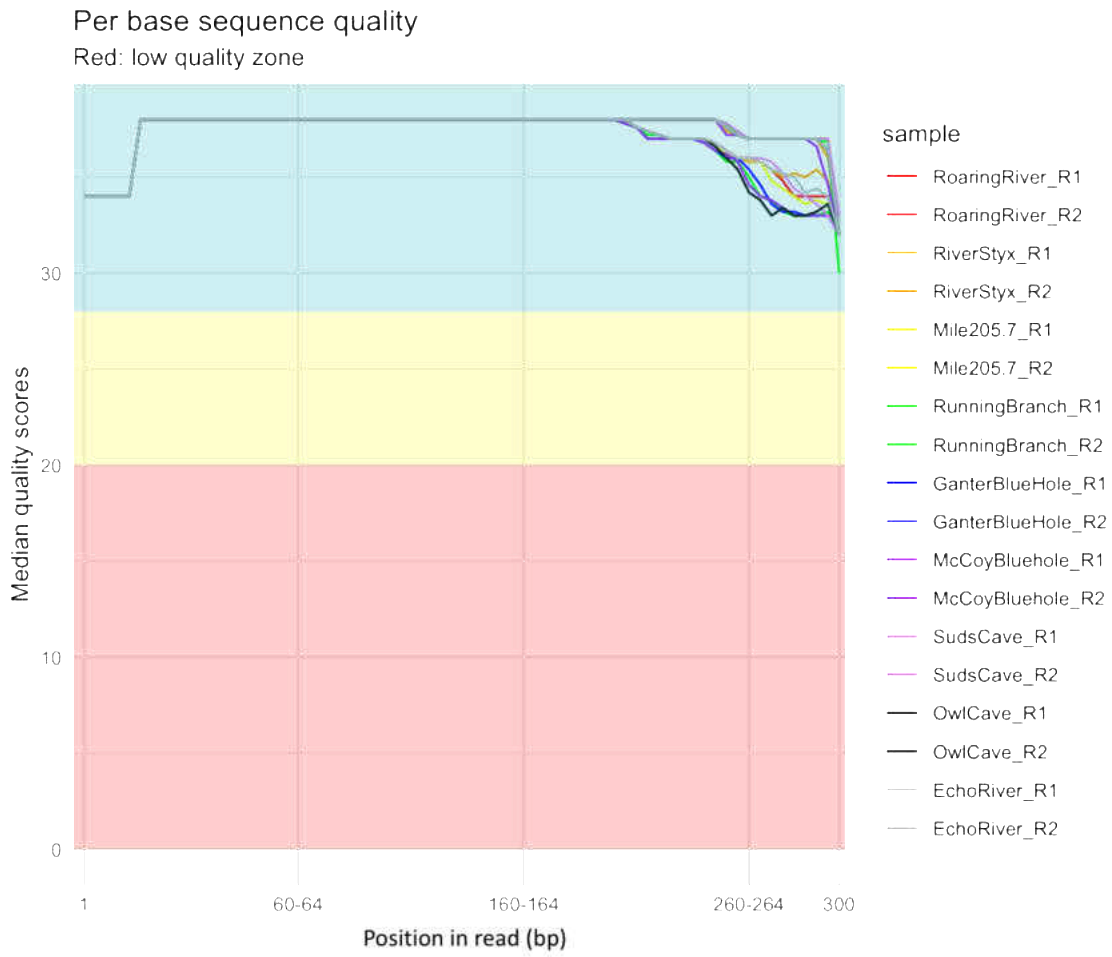

Figure 7. Per-Base Sequence Quality: Post-Quality Filtering
The median per-base sequence quality (PHRED score) after quality filtering and trimming Illumina data. R1 and R2 denote the forward and reverse reads from each sampling site. In general, reverse reads had lower median quality scores as sequencing length increased past 165bp.

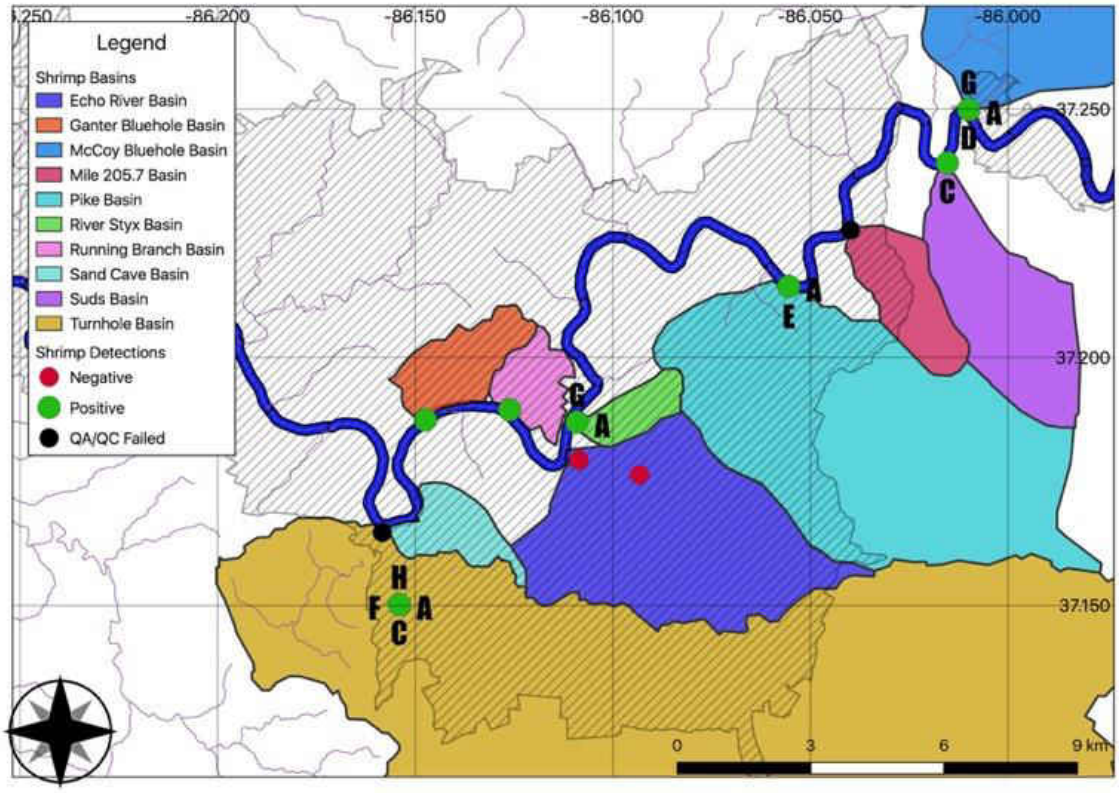

Figure 8. Positive Shrimp Detections by Groundwater Basin
A map of the karst groundwater basins known to house shrimp with the resulting *P. ganteri* detections from eDNA sampling at each of 11 locations. Green points signify positive shrimp detections while red points denote no detection. Two locations failed quality control standards and were excluded from analysis: they are shown as black points. Letters represent the identification of one of the eight haplotypes, inferred from Illumina sequencing data as a result of this study.

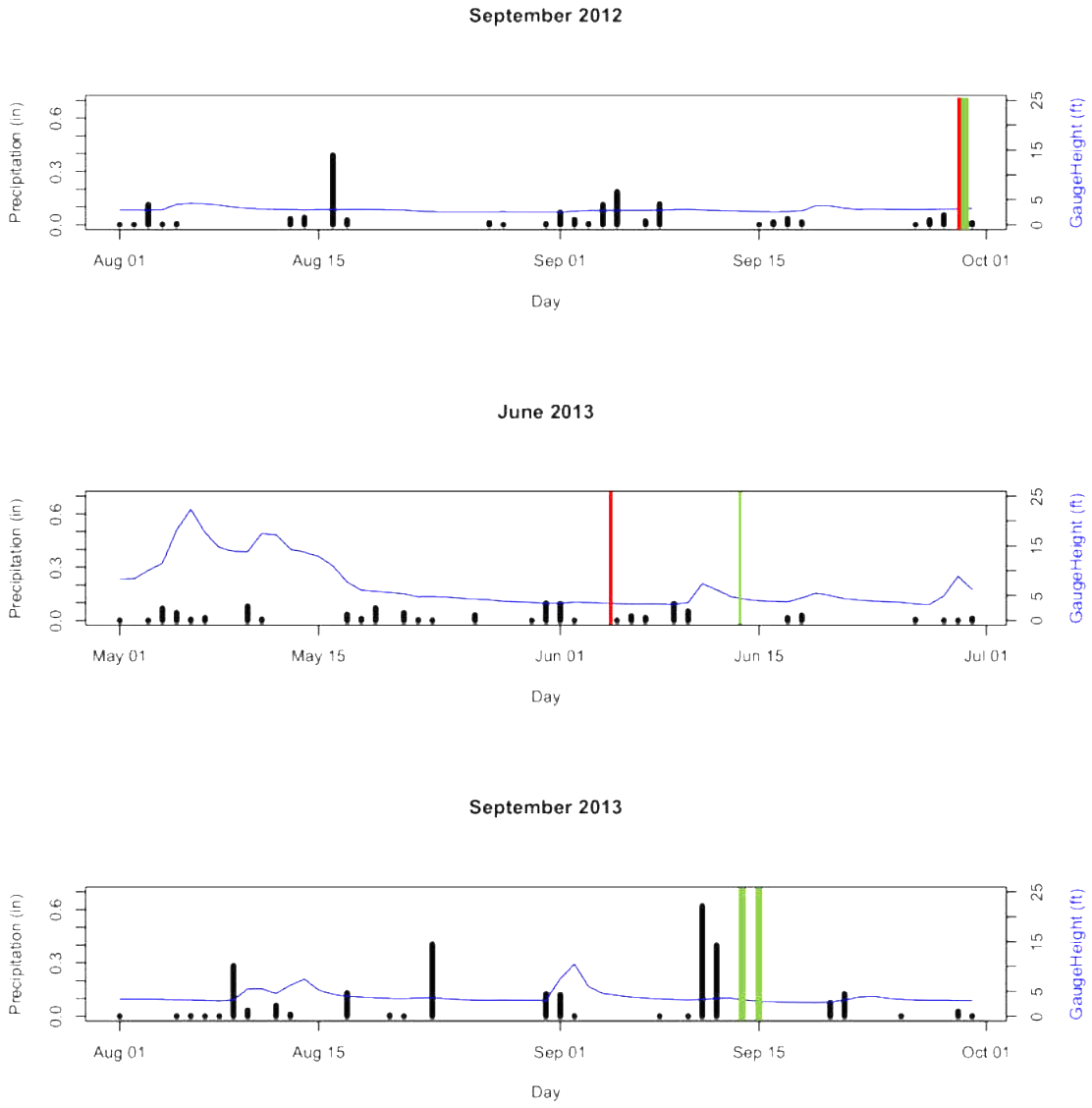

Figure 9. Precipitation and River Conditions During Sampling
Precipitation (in) and gauge height (ft) for the Brownsville gauge on the Green River during sampling months. Green vertical lines indicate positive shrimp detections while red vertical lines indicate a failure to detect shrimp DNA at a site.

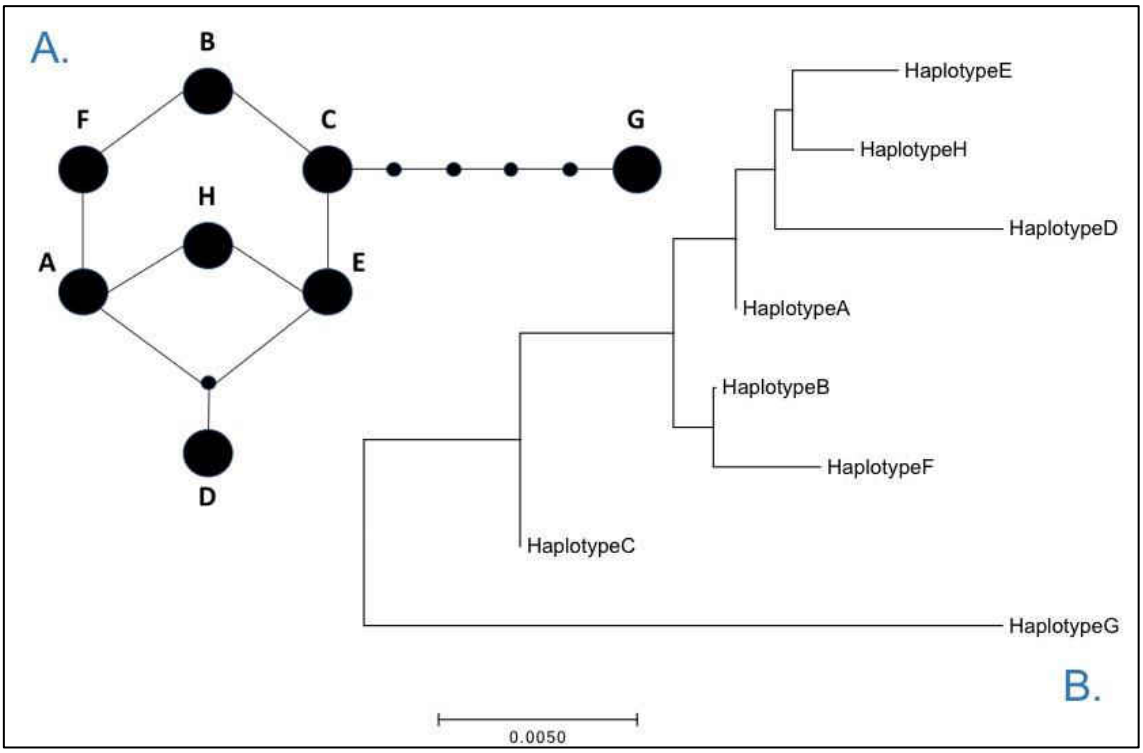

Figure 10. TCS Network and Optimal Neighbor-Joining Tree for Inferred Variants
A. (Left) A TCS network showing the population-level relationships between haplotypes identified at each of the seven karst water basins; **B. (Right)** An un-rooted optimal tree created in MEGA 7 using the Neighbor-Joining method and the evolutionary distances computed from a distance matrix (Table 7, Appendix A) using Maximum Composite Likelihood.

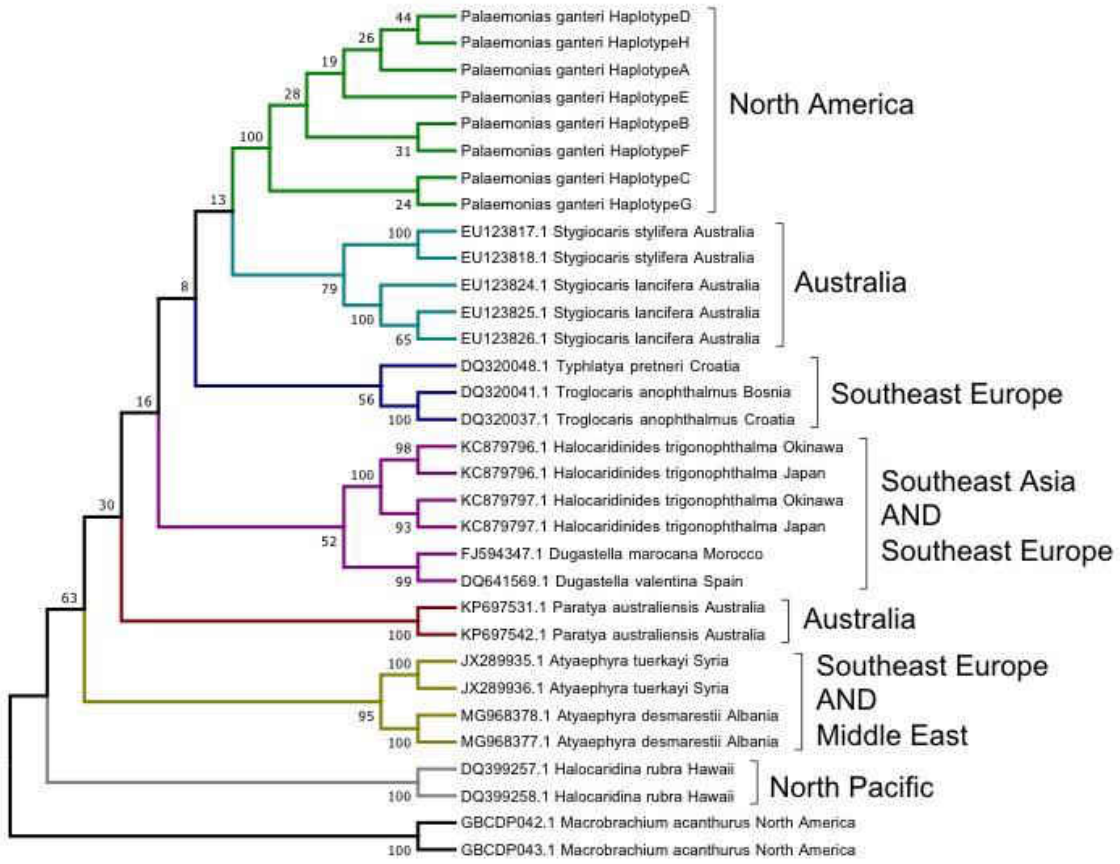

Figure 11. Atyid Consensus Tree and Geographic Region

A consensus tree inferring the evolutionary relationships using the COI gene region in closely related Atyids with the Neighbor-Joining algorithm. The tree was rooted using *Macrobranchium acanthurus* and bootstrap values indicate the percentage of replicate trees where associated taxa clustered together from 1000 replicates. This infers the evolutionary relationships between *P. ganteri* and other Atyid shrimp along with a summary of the geographic region for species used in analysis.

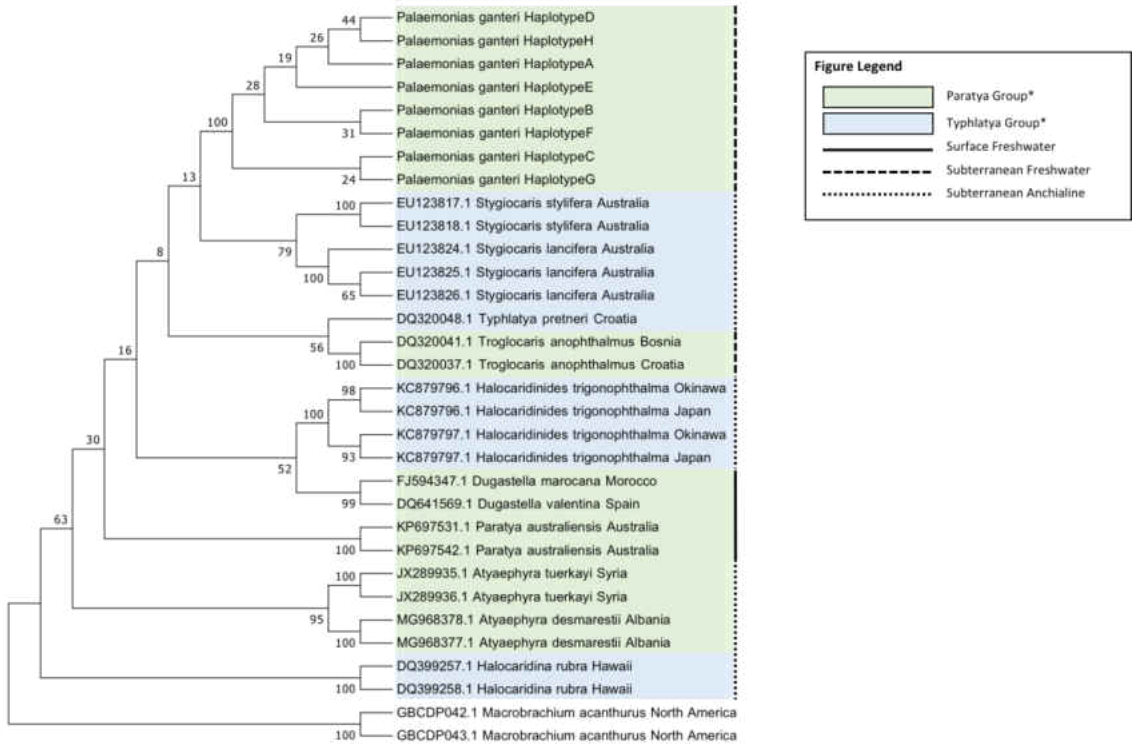

Figure 12. Atyid Consensus Tree Considering Paratya and Typhlatya Groups
A consensus tree inferring the evolutionary relationships using the COI gene region in closely related Atyids with the Neighbor Joining algorithm. The tree was rooted using *Macrobranchium acanthurus* and bootstrap values indicate the percentage of replicate trees where associated taxa clustered together from 1000 replicates. This infers the evolutionary relationships between *P. ganteri* and other Atyid shrimp in the Paratya (Green) and closely related Typhlatya (Blue) groups proposed by von Rintelen et al. 2012. In addition, some cladistic information is included.