2018

IMPROVING THE CONSERVATION OF A CRYPTIC ENDANGERED FRESHWATER MUSSEL (PARVASPINA COLLINA) THROUGH THE USE OF ENVIRONMENTAL DNA AND SPECIES DISTRIBUTION MODELING

Bonnie A. Roderique
Virginia Commonwealth University

IMPROVING THE CONSERVATION OF A CRYPTIC ENDANGERED FRESHWATER MUSSEL (*PARVASPINA COLLINA*) THROUGH THE USE OF ENVIRONMENTAL DNA AND SPECIES DISTRIBUTION MODELING

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Environmental Studies at Virginia Commonwealth University

by

Bonnie Roderique
Master of Science Degree in Environmental Studies

Director: Rodney J. Dyer, PhD
Director, Center for Environmental Studies

Virginia Commonwealth University
Richmond, Virginia
July 2018

Acknowledgement

Thank you to all the people who have helped make this project possible. To my advisor Dr. Rodney Dyer, for his support, guidance, and unwavering confidence over the last two years. To the members of my graduate committee for their crucial assistance during the planning of this project. Specifically, to Dr. Daniel McGarvey whose expertise was essential for the creation and interpretation of the Species Distribution Modeling. To Dr. Brian Verrelli, for providing valuable feedback on my project proposal, analyses, and helping to troubleshoot lab-work. To Brian Watson (Virginia Department of Game and Inland Fisheries), whose dedication to the protection of endangered mussels inspired me to pursue this avenue of research. His expertise on all things “mussel” was instrumental to the planning and execution of this project.

To my undergraduate advisor, Dr. John Styrsky who first exposed me to the research process and gave me the confidence to pursue a graduate degree. To the many members of the Dyer, Verrelli, and Eckert labs here at VCU for their friendship and motivation. To Alissa Nicholson, Alia Hamdan, Colleen O’Brien, and Rolang Piocuda for their invaluable help with fieldwork. To the Virginia Transportation Research Council (VRTC), and the VCU Rice Rivers Center who provided the funding for this project.

Finally, to Adam, without whom none of this would have been possible. From the bottom of my heart, Thank You.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

eDNA – Environmental DNA; a molecular technique that allows us to capture and identify DNA from the environment rather than from an individual

nDNA - Nuclear DNA; contained within the nucleus of eukaryotic organisms

MaxEnt - a program to create Species Distribution Models based on the maximum-entropy approach

mtDNA- Mitochondrial DNA; DNA contained within the mitochondria of eukaryotic organisms

PCR - polymerase chain reaction; a molecular technique used to amplify small numbers of DNA copies

qPCR- quantitative polymerase chain reaction; Also known as Real Time PCR, this molecular technique monitors and quantifies the amplification of DNA during PCR using fluorescent dyes or probes

SDM - Species Distribution Model

VDGIF- Virginia Department of Game and Inland Fisheries

Abstract

IMPROVING THE CONSERVATION OF A CRYPTIC ENDANGERED FRESHWATER MUSSEL (*PARVASPINA COLLINA*) THROUGH THE USE OF ENVIRONMENTAL DNA AND SPECIES DISTRIBUTION MODELING

Bonnie Roderique, Master of Science Degree in Environmental Studies

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at Virginia Commonwealth University.

Virginia Commonwealth University, 2018

Major Director: Dr. Rodney J. Dyer, Director, Center for Environmental Studies

Conservation efforts that involve habitat protection, population augmentation, and species reintroductions require knowledge of the habitat requirements, distribution, and abundance of a species—information that can be challenging to acquire, especially for rare organisms with patchy distributions. In this thesis, I develop a protocol for the use of environmental DNA (eDNA) and create a Species Distribution Model for the endangered James spinymussel, *Parvaspina collina* (Unionidae). The results of this work show that eDNA is a robust tool for identifying species presence but not for estimating the relative abundance of populations. This study found that *P. collina's* distribution is influenced by abiotic habitat characteristics related to sedimentation and runoff rather than by the distribution of its host fishes. The predicted habitat suitability

was used to identify locations of priority conservation concern and these results can be used to direct future sampling efforts, identify potential dispersal routes, and inform conservation decisions.

Chapter 1: Introduction

The proper management of endangered species often involves creating plans for habitat protection and restoration, population augmentation, and reintroduction of the species into its historic range. Accomplishing this necessitates a thorough understanding of the species' distribution and population sizes. To put it simply, this involves answering two important questions: *Where is it?* and *How many are there?* The answers to these questions provide key insights into biotic and abiotic factors necessary for species persistence facilitating the development of management and conservation plans that prioritize areas most in need of protection (Wilson et al. 2011). Once the presence of a species is documented, one can predict its potential distribution which can guide field sampling to discover existing populations and the identification of locales for future reintroduction (Seddon, 2010; Stoeckle et al. 2015). To determine which populations would benefit most from augmentation efforts, we must establish the size of the populations in question which can be used to give a more comprehensive picture of the species' viability. Specifically, populations that are small and fragmented can lose genetic diversity which can lead to local extirpation. Conservation managers can develop more effective mitigation plans with information describing the current distribution and abundance of the species in question.

While having adequate distribution and abundance information about populations is essential, obtaining this information can be challenging. For example, it is difficult to locate populations for species with large geographic ranges, and detection of cryptic

and low-density species is often not reliable. However, leveraging molecular genetic approaches such as environmental DNA (eDNA) has the potential to augment traditional surveying methods and reduce some of those challenges. In general, eDNA refers to any source of species DNA in the environment (Bohmann et al. 2014), thus eDNA sampling techniques involve collecting various types of samples (soil, water, sediment, etc.) and extracting the DNA to determine if the species of interest is present. This technique has been shown to effectively detect a variety of species (Thomsen & Willerslev, 2015), even those with large geographic distributions (Laramie *et al*. 2015; McKelvey *et al.* 2016) and small populations (Sigsgaard et al. 2015). Additionally, this technique has proven reliable in providing an estimate of species abundance (Takahara et al. 2012; Thomsen et al. 2011; Pilliod et al. 2013; Kelly et al. 2014; Klymus et al. 2015). Therefore, eDNA sampling allows us to gather important population characteristics in an efficient, non-invasive manner. Furthermore, the combination of eDNA techniques and traditional surveying has the potential to better inform the management of cryptic endangered species than either method alone.

Species Recovery Plans developed by the U.S. Fish and Wildlife Service often list population augmentation and species reintroduction into its historic range as a key Recovery Task (USFWS, 1990). These efforts may include releasing propagated individuals to boost population size, introducing new allelic diversity into inbred populations to increase genetic diversity, and translocating individuals into new habitat to expand or re-establish their historic distribution. Endangered species management efforts can include ongoing population monitoring surveys and captive breeding

programs that release hatchery-born individuals into streams with known presence. To enhance these conservation efforts, we must first develop a comprehensive understanding of the habitat requirements, distribution, and population abundance of a species which will allow us to make more informed decisions regarding recovery efforts.

Chapter 2: Development and Testing of Environmental DNA (eDNA) Protocols for the Endangered James Spinymussel (*Parvaspina collina*)

Abstract

Environmental DNA (eDNA) techniques alleviate challenges associated with locating rare, cryptic, or patchily distributed organisms and have been shown to accurately estimate population abundance. This study evaluated the effectiveness of using eDNA to detect the presence and estimate the abundance of an endangered James spinymussel, *Parvaspina collina,* using quantitative PCR methods. Detection probabilities at the level of the sampling locale were high (66.7 – 100 %) but varied between sampling seasons. Two sources of potential false negatives were identified and traced to qPCR inhibition and local census sizes below analytical limits of detection. DNA concentrations in the samples matched predicted levels of mussel activity but did not correlate to relative abundance. The extent to which at-site stream characteristics were predictive of DNA concentration was inconsistent across the sampling seasons. Overall, this is a robust technique for identifying species presence but the transport distance of DNA should be determined and inhibitors should be identified and removed before full application of this technique.

Introduction

Choosing the appropriate methodology for identifying the presence and abundance of species is difficult if the taxon is rare or cryptic in appearance. For freshwater mussels, survey approaches are commonly challenged by small, isolated populations within restricted geographic ranges (e.g., Smith et al. 2001; Strayer et al. 1996). Despite these

logistical challenges, appropriate sampling protocols are necessary to identify the presence of cryptic species, especially in locations with proposed or ongoing impacts to critical habitat. The development of technological advancements that augment physical survey approaches should contain two fundamental characteristics. First, any new approach must be able to increase either the detection probability at a particular locale or allow a broader number of locales to be examined. If novel approaches cannot increase precision or accuracy over physical sampling, they will not be utilized. Second, any novel approach should be developed with the ability to provide a probabilistic estimation of sampling error rates (e.g., not detecting the taxon even though it is present) to better serve the development of conservation and management plans.

This work outlines the development of environmental DNA techniques (hereafter eDNA) for identifying the presence and abundance of the James spinymussel (*Parvaspina collina*; Unionidae), an endangered freshwater mussel endemic to the James and Dan river basins in Virginia, West Virginia, and North Carolina. Due to the combined effects of habitat degradation, river impoundments, predation, and resource competition from invasive Asian clams (*Corbicula fluminea*), many *P. collina* populations experienced local extirpation within the last two decades (Clarke & Neves, 1984; Hove & Neves, 1994; USFWS, 1990). The species is now patchily distributed throughout the James River and Dan River basins in Virginia and North Carolina, a distribution which only encompasses approximately 10% of its historic range (USFWS, 1990). This sharp contraction of the range led to its listing on the Endangered Species List in 1988. At present, the Virginia Department of Game and Inland Fisheries (VDGIF) has

implemented an extensive conservation program designed to locate populations and conserve critical habitat.

Environmental DNA is a non-invasive means of detecting the presence of rare, endangered, or invasive species by isolating discrete pieces of nuclear DNA (nDNA) and mitochondrial DNA (mtDNA) from the water column (Bohmann et al. 2014; Ardura et al. 2015). Minute particles of tissue, either excreted or shed from individuals *in situ*, are used as raw template for DNA extraction and subsequent amplification using species specific genetic markers. For freshwater mussels, the source of this DNA is likely cells sloughed during filter feeding, gametes released into the water during breeding, and even DNA released from the shell material (Ardura et al. 2015; Geist et al. 2008). Example applications of this approach include identifying the presence of invasive species such as the silver carp in the Mississippi drainage (Hickcox 2011) and the spread of the American Bullfrog across Spain (Ficetola et al. 2008). This approach has also been used to identify cryptic species such as the Rocky Mountain Tailed Frog and the Idaho Giant Salamander (Goldberg et al. 2011).

The addition of eDNA approaches to existing sampling protocols may have several implications for ongoing monitoring programs and management policies. First, given the cryptic nature of these organisms, the current detection probabilities for this species range from 12% to 20% (Esposito 2015; VDGIF 2015) for mark and recapture of individual mussels. At the site level, physical detection is likely for large populations, but the ability to detect an individual is highly variable for locales with only few individuals,

especially if those individuals are patchily distributed within the stream. Molecular techniques based upon water sampling may help to augment these rates thereby increasing the confidence in where populations exist. Second, the life history of these organisms makes sampling efficiency temporally variable as the organisms are more accessible during certain times of the year. Despite their variable position within the substrate throughout the year however, they are continually in contact with the water column with the potential for providing assayable DNA samples independent of substrate position (e.g. Stoeckle et al. 2015). Finally, molecular approaches are very amenable to high throughput analysis (e.g., the processing of large numbers of samples). Evaluation of many locations can be assayed first using an eDNA approach thereby potentially reducing the number of areas and regions requiring physical field surveys (McKelvey et al. 2016). In this manner, eDNA approaches serve to create additional efficiencies in existing sampling protocols by allowing field technicians to prioritize the locations they sample.

This study aimed to 1) develop *de novo* molecular genetic markers that differentiate this species from other organisms that coexist in native streams of Virginia, 2) estimate the probability of species detection using eDNA techniques, and 3) determine the ability of this technique to estimate the relative abundance of populations based on the amount of DNA template in the water column.

Methods

Study Area and Sample Collection

The majority of known *P. collina* populations are in the upper James River drainage of Virginia so study sites within this distribution were selected to include well monitored streams representing a range of population sizes. Field collections were taken from stream reaches whose local densities are known to VDGIF biologists from previous surveys and ongoing mark-recapture studies (Table 1.1, Figure 1.1). Historical census estimates were provided by VDGIF biologist Brian Watson for five of the locations, though predation, length of time since last physical census, and ongoing demographic changes at these locales necessitated the use of ranked population sizes as a more realistic estimate of abundance when eDNA samples were collected.

At each site, 4-6 water samples were collected by submerging a sterile 1L Nalgene bottle approximately 5-10 cm below the surface until filled. All sample bottles were sealed and stored on ice during transport. To maximize species detection, samples were collected immediately downstream (1-2 m) of known *P. collina* populations in equal intervals across the width of each stream (Laramie et al. 2015). Stream characteristics such as flow rate, dissolved oxygen, pH, temperature, and turbidity were also measured at each site during the time of sample collection to determine if site-specific features may either inhibit or reduce the efficiency of DNA amplification and estimation of DNA concentration (Jane et al. 2015).

Samples (including negative distilled water controls) were filtered within 24 hours of collection through 0.45-micron nitrocellulose filters and stored in 100% ethanol at -20°C for subsequent DNA extraction. DNA was extracted from the filters following the protocol of Goldberg (2015) with the following modification: sterile disposable forceps were used to handle each filter rather than a pair of metal forceps to reduce the potential for contamination. A plain sterile filter was processed during each extraction as a control to detect any potential lab contamination. Extracted samples were stored at -20 °C for up to 2 weeks prior to amplification via qPCR. All equipment was sterilized under UV light for 10 minutes prior to use.

A total of three sampling events were performed throughout the summer, in the months of June, August, and October, coinciding with the peak and end of *P. collina's* reproductive period (Hove & Neves, 1994). Samples were collected across these time periods to determine whether the DNA concentrations or detection probabilities would be influenced by the expected levels of mussel activity. It was expected that DNA concentrations and detection probabilities should be highest when *P. collina* are most active at the surface in June and should decline as the mussels become less active during August and begin to burrow in October.

Genetic Marker and qPCR Assay Development

Species-specific primers were designed for qPCR assay targeting a 111 bp sequence within the NADH dehydrogenase 1 (ND1) region of the mitochondrial genome. This region has demonstrated high levels of interspecies variability while exhibiting a relatively high level of intraspecies similarity (Campbell et al. 2008), making it an ideal

target for species identification. Published sequences available through GenBank were used to create a consensus sequence for this region. Primers were designed for this consensus sequence using PrimerQuest (Integrated DNA Technologies, Coralville, IA) and the species specificity was confirmed using Primer-BLAST (Ye et al. 2012) which confirmed a 100% match to *P. collina* and at least 2 base pair mismatches between other freshwater mussel species with overlapping distributions. Primer specificity was verified *in vitro* by testing the primer on DNA from preserved individuals provided by Brian Watson (VDGIF biologist). The amplified NADH nucleotide region was sequenced via Sanger fluorescent dye sequencing at Nevada Genomics and compared to published GenBank sequences of this species.

Relative Abundance Estimation

We assessed the ability of this method to estimate the relative abundance of individuals along a stream reach using qPCR. This allows the density of DNA fragments to be tracked during the polymerase amplification process. More initial DNA content results in a more rapid increase of amplified products than less initial DNA template. Given known relative abundance estimated from field surveys, qPCR can provide a standardized curve for estimation of local population abundance in non-surveyed areas (Takahara et al. 2012; Pilliod et al. 2013). All products were amplified using the primers developed in this study (discussed below) and sample runs contained a negative plate control to detect any potential contamination.

A standard curve was included on each qPCR plate which consisted of a 5-fold serial dilution of *P. collina* DNA template from 10 ng per reaction to $10e^{-5}$ ng per reaction. The

data from this curve were fit to a log-normal function and was used to estimate the initial DNA concentration of each qPCR replicate. The DNA concentration for each water replicate was then calculated as the average DNA concentration of any of the 3 positive qPCR replicates. For each sampling season, the relationship between the concentration of DNA in the sample to the relative abundance of the population was determined using a general linear model.

To determine whether significant differences existed between the mean eDNA concentration of each stream within and across sampling seasons, a nonparametric Kruskal-Wallis rank sum test was performed for each season and stream using the positive qPCR samples as replicates within each treatment group (sampled stream). A Conover-Iman *post hoc* test was then performed to analyze the sample pairs for stochastic dominance and determine the directionality of any significant relationships (i.e., did larger population sizes relate to higher eDNA concentrations).

Inhibition of qPCR

Both biotic and abiotic inhibitors present in environmental samples can affect the reaction efficiency of qPCR by binding to nucleic acids, changing their chemical properties, or reducing the specificity of the primers (Abbaszadegan et al. 1993; John, 1992; Opel et al. 2010). For water samples, the most likely inhibitors present are dissolved or solid organic compounds such as fulmic acids, humic acids, metal ions, and polyphenol (Abbaszadegan et al. 1993; Ijzerman et al. 1997). In many cases, the effect of these inhibitors can be reduced either by diluting the sample or identifying and removing the specific inhibitors.

Based on the results of a pilot study, most inhibition within the collected water samples was effectively removed by diluting the sample 1:10 with ddH2O. However, it was also determined that this level of dilution could potentially reduce the concentration of DNA in the sample below the threshold necessary for qPCR, reducing the detection probability. To reduce qPCR inhibition while maximizing detection probability, samples from each trip were processed twice; once with the samples diluted 1:10 and again with the samples run “as-is”. The calculated DNA concentrations for each diluted sample were then multiplied by 10 to produce an estimate of the original DNA concentration of the undiluted sample. On both plates, samples were run in triplicate. The comparison between the diluted and non-diluted samples from the same stream allowed us to determine whether a sample that failed to amplify was inhibited or negative for eDNA. If the diluted sample amplified while the non-diluted sample did not, inhibition was likely present. If both samples failed to amplify, the sample was likely negative for eDNA. Each qPCR plate contained the samples from one of the sampling trips (including the negative controls) run in triplicate along with a negative plate control.

Detection Probability

Each 1-L water sample was treated as an at-site replicate (after Laramie et al 2015). Detection probability was calculated per site as the number of 1-L replicates that tested positive for *P. collina* eDNA (N = 0-6) divided by the number of replicates collected at that site (N = 4-6). The overall detection probability across sites was calculated as the total number of positive 1-L replicates (N = 0- 6) divided by the total number of water samples collected during each of the three sampling events (N = 4-6). The percentage

of false positives and false negatives was also determined. False positives could arise from laboratory contamination and false negatives occur when we fail to detect eDNA in a sample from a location with known species presence. All analyses were performed using R-Statistical Software (Version 3.4.1, 2017-06-30, The R Foundation for Statistical Computing).

Results

Study Area and Sample Collection

Water samples were collected from each locale in June, August, and October, and each sample was subdivided into replicates for DNA extraction and subsequent qPCR amplification. Likely due to small census population sizes, sites CF and AF did not yield any detectable mussel DNA and were not returned to during subsequent sampling trips (discussed below). The additional sites RIC and RC were added to increase replication for relative abundance estimation across sites. Sampled locales represented a broad range in flow rate, temperature, pH, and dissolved oxygen. Stream characteristics varied between all sampled locations and across sampling events (Table 1.2).

Genetic Marker and qPCR Assay Development

Several candidate primers were designed to minimize the potential for false positive PCR results for *P. collina,* while at the same time produced the most uniform qPCR profile. A total of 14 different primer combinations were developed and assayed (Appendix 1, Table A1). The primers found to be most effective at producing quality qPCR products (111 bp in length for the target species) had the following sequences:

ND1_pcbr1 (forward) 5’-GCGTAGCATTCTTTACCCTTCT-3’

ND1_pcbr1 (reverse) 5’-GAGCGTCTGCTAATGGTTGT-3’

Amplification was conducted on a CFX96 Touch Real-Time PCR Detection System (BIO-RAD, Hercules, California) using the following settings: SYBR green only, Denaturation at 98°C for 3 minutes, 40 amplification cycles of 98°C for 15 seconds, 53°C for 30 seconds, plate read at 65°C, with melt curve analysis set to instrument default. Based on the results of the primer specificity analysis, a sample was deemed positive for eDNA if any of the qPCR replicates amplified a fragment of the proper size and melting temperature (79 - 80°C). Any samples with ambiguous melting temperatures were subsequently sequenced using Sanger sequencing (Nevada Genomics) to ensure species identification.

Relative Abundance Estimation

Estimated DNA concentrations were neither commensurate with rank population size nor were their relative ranking in DNA concentration consistent across sampling trips throughout the season (Table 1.3; Figure 1.2). Assuming the census population sizes are correct, it was predicted that census population size should be directly influencing the amount of DNA template sampled in the water column. Independent of census size, it was also thought that the relative rank of DNA concentrations between sampling locations would be consistent across sampling events, which was also not observed (e.g., compare MC and JC concentrations in Figure 1.2). That is not to say that there

were not significant differences in template concentration, they were just not aligned with census size nor consistent in ordering.

In June (Kruskal-Wallace Test; $H_3 = 12.15$, $P = 0.01$), MC was found to have significantly higher eDNA concentrations than DC, JC, and LOC (Conover-Iman Test; $p < 0.01$), but the mean of the other sites when compared with each other was not significantly different (Figure 1.2). In August ($H_5 = 49.34$, $P < 0.0001$), eDNA concentrations at RIC were much higher than at all other locations ($p < 0.001$) and significant differences were also found between MC and all other sites: DC ($p < 0.0001$), JC ($p < 0.0001$), LOC ($p = 0.013$), and RC ($p< 0.0001$). In each pairwise comparison of MC to another site, the concentration of eDNA found at MC was lower than the other locations. While we expect MC to have lower DNA concentrations than LOC and RC based on the relative abundance, we do not expect it to have lower concentrations than DC and JC (see again Table 1.1). In October ($H_5 = 49.34$, $P < 0.0001$), JC had the highest overall concentrations ($p = 0.004 - 0.02$) and only RIC was significantly lower than RC ($p = 0.003$).

DNA concentrations did tend to match predicted levels of mussel activity at a given site that coincide with the peak and end of *P. collina* reproductive period (Hove & Neves, 1994; Figure 1.3). In DC, DNA concentrations were highest in June and lowest in October ($H_2 = 12.65$, $P < 0.001$). In LOC, concentrations decreased between June and August ($H_1 = 6.08$, $P = 0.01$) but were unavailable for October since no positive samples were recovered for that sampling event. In MC, concentrations were highest in June and

decreased in August ($H_2 = 17.61$, $P < 0.001$). However, there was no significant difference in the DNA concentrations between August and October. Finally, in RIC, DNA concentrations were highest in August and decreased significantly in October ($H_1 = 15.21$, $P < 0.0001$). No data were available for this stream in June because it was added to the sampling sites after CF and AF were removed. The remaining streams did not fit the pattern seen in the other locations. In JC, DNA concentrations did decrease between June and August, but were highest in October which was contrary to the pattern seen in the other sites ($H_2 = 9.71$, $P < 0.01$). There was also no significant difference in DNA concentrations between June and October ($p > 0.05$). In RC, no difference was found between August and October ($H_1 = 2.55$, $P = 0.11$). This site was not sampled in June, so data are only available for the last two sampling events.

The extent to which at-site stream characteristics were predictive of DNA concentration was inconsistent across the sampling seasons (Table 1.4). During June, pH was significantly correlated with template concentrations ($df = 2$, $r^2 = 0.98$, $p = 0.01$) yet this relationship was not found in samples from the remaining trips. There was also a high correlation between dissolved oxygen and DNA concentrations in June ($df = 2$, $r^2 = 0.93$, $p = 0.06$**)** but this correlation was not significant and was not seen in subsequent trips. In October, the highest correlation existed between the flow of the stream and the resulting DNA concentrations where high flows yielded lower concentrations of DNA ($df = 3$, $r^2 = -0.83$, $p = 0.08$) but this relationship was not significant and was also not seen in the previous two trips.

Inhibition of qPCR

Sampling locations CF and AF (June) and LOC (October) did not yield any assayable PCR products. These samples were further examined to determine if abiotic factors may be inhibiting PCR reactions. To identify the presence of inhibition, the samples were spiked with a known concentration of DNA (1 ng/mL of reference *P. collina* DNA) and then subjected to reanalysis (following Gibson et al. 2012). Samples with spiked DNA that continued to yield no genetic markers were classified as inhibited (i.e., it could not be determined from these samples if there were target sequences as the entire reaction was inhibited by abiotic compounds in the water).

All spiked samples from AF and CF amplified, indicating that the lack of amplification of the non-spiked samples was likely due to the absence of eDNA and not the result of inhibition. The lack of eDNA was likely related to the low population densities at these locations. Every spiked sample collected from LOC in October failed to amplify, indicating strong inhibition of the reaction. A dilution test was employed to determine if reducing the concentration of the unidentified environmental inhibitors might recover the qPCR products, though no product was observed.

Detection Probability

At the site level, detection probability varied by sampling season. In June, *P. collina* DNA was identified in four of the six sampled sites (66.7% detection probability). The two sites that did not yield DNA, AF and CF, were the populations with the smallest census size reported ($N_{AF} = 5$, $N_{CF} = 4$ individuals) during the most recent physical census in 2005 (Table 1.1). Lack of product may reflect a lower limit to detection for

qPCR. After replacing these sites with RC and RIC, at-site detection probability was 100% for the August sampling trip and decreased to 83.3% in October with the inhibition of amplification for LOC (see above). Samples from LOC were positively identified as being inhibited meaning that this approach could not indicate if there was target DNA in the samples.

The detection probability based on the percentage of positive qPCR replicates sampled within a site varied between both sites and seasons. Average at-site replicate detection probabilities were highest among samples in June (66.7%) and decreased to 47.7% in August and 24.3% in October. Within each season, there was significant variation in the percentage of positive qPCR replicates between streams. In June, detection probabilities for each stream ranged from 0 (CF and AF) to 75%. In August, these values ranged from 20.8 (LOC) to 79.2 (RIC) and in October, they ranged from a low of 0% (LOC) to a high of only 41.7% (RIC). None of the negative control samples from water collection, filtering, extraction, or qPCR amplification produced a positive *P. collina* result so the rate of false positives was 0.

Discussion

The objectives of this study were to determine whether eDNA can both detect the presence of and rank the relative abundance of DNA in the water column. Our results suggest that this method can be used to identify the presence of this species and that this protocol has the potential to be used for ongoing monitoring of cryptic aquatic taxa. The lack of positive samples at sampling locations CF and AF is assumed to indicate a lower limit to template detection for qPCR-based approaches. While LOC failed to

produce positive hits in October, it was easy to demonstrate that abiotic conditions in the sample water were inhibiting polymerase reactions. In these cases, lack of detection is not considered a false negative as it is impossible to determine if there is any DNA in the sample. Overall, this method appears to be robust at identifying the presence of the target species. This approach is most effective when samples are collected during the breeding season when individuals are broadcasting gametes into the water column. In both June and August, detection probabilities for sites were 100% for sites with a census count estimated to be greater than five resident individuals. These rates are higher than the 12-20% detection probabilities we expect with physical surveying techniques. However, it is important to note that these estimates are for the detection of previously marked individuals and as such, represent an individual based detection probability rather than the site or species-based detection probabilities that result from eDNA sampling (Esposito, 2015; VDGIF, 2015). It is unclear which method would result in higher species detection probabilities at *de novo* sites.

When comparing across streams, eDNA concentrations were not a good predictor of species relative abundance. While a relationship between DNA concentration and species density has been demonstrated in laboratory conditions for similar species (De Ventura et al. 2017), this relationship was not observed from these data. There are several potential contributing factors to this relationship. First, the amount of template DNA per unit volume of water may be quite variable. As such, six samples per site may not be sufficient to gain a reasonable estimate of mean DNA concentrations. There may also be differences between stream morphology that influences the distribution of

the mussels within each stream (i.e., the level of “patchiness”). As stream morphology was not measured, its influence could not be partitioned in our analyses. Finally, depending upon the size distribution of the resident population, relative abundance of individuals (and even census population size) may not be an effective proxy for biomass. The population-level estimates used for census size do not take into account the size distribution of each population. More up-to-date census data that includes either biomass estimates, or demographic data would be necessary to estimate the extent to which sampled DNA concentrations can predict at-site census size beyond a relative ranking.

More data are necessary to provide power to the associations between the average DNA concentration and the measured in-stream characteristics. Although it was not found to be significant, the consistent negative relationship between stream flow and DNA concentration implies that higher flows may either be flushing the DNA from the system or diluting the DNA in the water.

Despite these differences, DNA concentration did change in predictable ways across the season, reflecting expected “activity” levels based on the timing of the breeding season for this species.

Management Implications

Before this protocol can be implemented as either augmentation or replacement of other methodologies, additional study is required. Findings of non-detection in some samples highlight the next set of factors needed to be addressed prior to full implementation of

this work: factors causing inhibition and the limits of detectability for very small populations. While beyond the scope of this study's objectives, overcoming qPCR inhibition attributable to local conditions has not been shown in other examples to be too onerous a task and may be approached in several ways. Under the approach outlined in this study, samples from inhibited locales may be assayed for the presence of dissolved or solid organic compounds such as fulmic acids, humic acids, humic material, metal ions, and polyphenol (Abbaszadegan et al. 1993; Ijzerman et al. 1997). Once identified, the proper protocol for removing these compounds from samples may be integrated into the DNA extraction protocol (Shrader et al. 2010). Another approach would be to examine the effect that using alternative DNA extraction protocols have on removing inhibiting compounds. The extraction protocol used (Qiagen) is based on a silica purification protocol. While this is the most commonly used protocol, other approaches relying on detergents and other means may be more effective at removing compounds that may interfere with the polymerase reaction. Another approach may be to use non-polymerase DNA replication approaches to increase target DNA concentrations. Here, all DNA in the samples would be replicated to increase initial template density followed by qPCR of specific marker regions. When combined with dilution tests, this last approach may be the most efficient as it would not require knowledge of which specific compounds were inhibiting all PCR reactions.

The next issue to address should be the limits of detectability. Two of the populations in this project (AF and CF) did not produce positive qPCR products for any of the replicate samples and inhibition was not a factor in this particular instance of non-detection.

These populations were estimated during the 2010-2013 collection season to have as few as four individuals. Unpublished census work by VDGIF biologists have more recently verified that there are individuals at this location, though the last two physical sampling trips did not yield any positive identifications. The interpretation of the census data in this study and the molecular approaches take a conservative stance with the assumption that individuals are still onsite, which is why they were counted as failed identifications in the detection probabilities. Absence of evidence is not evidence of absence. However, the most important next step will be to test the limits of detection for eDNA protocols. The limits of detection will vary based on the study organism and the primer specificity and sensitivity. Previous work has demonstrated a lower limit of detection at 0.4 ng/µL and 100 pg using PCR and laboratory dilutions of template DNA for freshwater mussels (Ardura et al. 2015, Stoeckle et al. 2015). However, neither of these studies used the qPCR method so the concentration of DNA that was captured from the eDNA samples is unknown. One study using the qPCR method found the limit of detection for a freshwater mussel species to be 1 copy / mL (Sansom and Sassoubre 2017). It is important to note that these limits to detection were demonstrated in laboratory conditions with DNA template diluted in pure water, so it is not certain that these limits would translate to environmental samples that contain a mixture of species DNA along with inhibiting compounds.

In this work, 4-6L of water was sampled for each locale, and this may be an insufficient volume to yield enough DNA template if there are populations whose sizes are in the single digits. For endangered species, the presence of even one individual is of utmost

concern and subsequent work should be focused on methodological approaches that increase the specificity of the approach toward these lower limits. One potential avenue to pursue if this research is to continue may include increasing the volume of water sampled at a location which may yield sufficient template DNA from small populations. Another approach, more common in forensic DNA protocols, may be to use non-PCR-based DNA template enrichment protocols prior to qPCR. A detection probability is a site-wide feature and only one sample yielding a positive result triggers subsequent actions.

In addition to inhibition and limits of detectability, future efforts may be best served in determining the spatial extent by which template DNA may be detected. Regulatory constraints dictate a physical distance within which endangered species presence cause concern for road and bridge activities. In order to best employ this method, we need a better understanding of the transport distance of eDNA downstream from a known population. In this study, samples were collected directly downstream of known populations. Previous work has demonstrated the ability of eDNA to travel long distances in flowing water (e.g. Deiner and Altermatt 2014; Jane et al. 2015) while other studies have shown a loss of signal beyond 500 m (Stoeckle et al. 2015). Depending on the distance eDNA travels in this system, we may or may not have detected the population if we had sampled several meters or miles downstream.

Conclusion

The use of eDNA techniques appears to be a robust method for detecting the presence of this rare and endangered species of freshwater mussel. With better detection

probabilities than traditional surveys, it can certainly augment traditional surveying efforts. In areas of unknown occurrence, it can serve as the first pass, where a positive result could trigger additional visual surveys. This method has been shown to be both time cost effective which could allow for much larger areas to be surveyed than what is currently possible. However, before this methodology is fully adopted into a conservation plan, we must establish the limits of detection and identify and remove any compounds inhibiting the qPCR reaction. Additionally, this method should only be employed during the height of the breeding season in June to maximize the detection probability and minimize the potential for false negatives.

Tables

Table 1.1: Population Size Estimates for Sample Locales; Historical census counts were provided for each locale along with an estimate of current rank population size (1 = smallest, 6 = largest). Population census data were provided by the Virginia Department of Game and Inland Fisheries and reflect collections taken from 2010 through 2013. Population ranks relate to best estimates for current (2017) local population sizes based on recent surveys and population augmentation activities.

Site	Census Size	Rank
AF	5	0
CF	4	0
DC	430	1
JC	428	3
LOC	1125	5
MC	173[a]	4
RC	NA	6
RIC	NA	2

[a] This site was augmented with hatchery grown individuals between the last physical survey and the sampling for this study. At present, it is unknown what the real census size may have been.

Table 1.2: The following in-stream characteristics were measured at each site during each sampling event: water temperature (in °C), pH, turbidity (in nephelometric turbidity units, NTU), dissolved oxygen (percent), and stream flow (m^3/s). The mean DNA concentration (ng/µL) for each locale was calculated using the DNA concentration of each positive qPCR sample. No positive samples were obtained from AF or CF in June or from LOC in October.

Season	Site	Temperature	pH	Turbidity	Dissolved Oxygen	Flow	Mean DNA Concentration
June	AF	17.46	6.98	0.03	94	7.73	NA
	CF	17.88	6.87	0	94.8	11.35	NA
	DC	18.98	5.35	6.3	89.5	0.93	6.70E-04
	JC	16.09	5.58	3.6	94.1	1.77	7.97E-04
	LOC	18.1	5.8	7.2	93.6	0.63	1.42E-03
	MC	18.6	7.05	1.8	101.8	0.54	2.66E-03
August	DC	21.3	5.69	2	89.5	0.1	3.43E-04
	JC	20.97	4.8	NA	93.6	0.65	3.74E-04
	LOC	21.75	5.67	0.8	89.6	0.07	2.42E-04
	MC	22.72	6.12	7.7	94.8	0.36	7.64E-05
	RC	22.6	5.24	NA	97.8	0.02	4.38E-04
	RIC	21.68	6.135	2	93	0.07	1.60E-03
October	DC	18.36	8.4	2.875	87.5	0.07	1.04E-04
	JC	17.6	7.4	2.375	93.5	0.22	2.57E-08
	LOC	18.7	7.95	4.48	88.5	0.06	NA
	MC	19.15	7.9	0.925	88.7	0.11	7.34E-05
	RC	18	6.74	1.5	83.5	0.01	3.02E-04
	RIC	17.72	8.66	2.51	94.45	0.11	2.13E-04

Table 1.3: Conover-Iman pairwise comparisons of significance between estimated DNA concentrations from streams sampled in October. Bold values indicate p-value < 0.01.

	DC	**JC**	**MC**	**RC**
JC	**-2.15**	-	-	-
MC	0.67	**2.64**	-	-
RC	-1.34	0.84	-1.9	-
RIC	1.61	**3.72**	0.77	**2.95**

Table 1.4: The correlation between the average DNA concentration and the measured in stream characteristics at each location was calculated for each sampling season. The only significant correlation was found between the pH of the streams sampled in June and the average DNA concentration recovered from each location.

	DNA concentration	Temperature	pH	Turbidity	Dissolved Oxygen	Flow (m3/s)
June	0.34	**0.98***	-0.61	0.93	-0.65
August	-0.19	0.32	-0.36	0.03	-0.31
October	-0.18	-0.22	-0.15	-0.49	-0.83

Figures

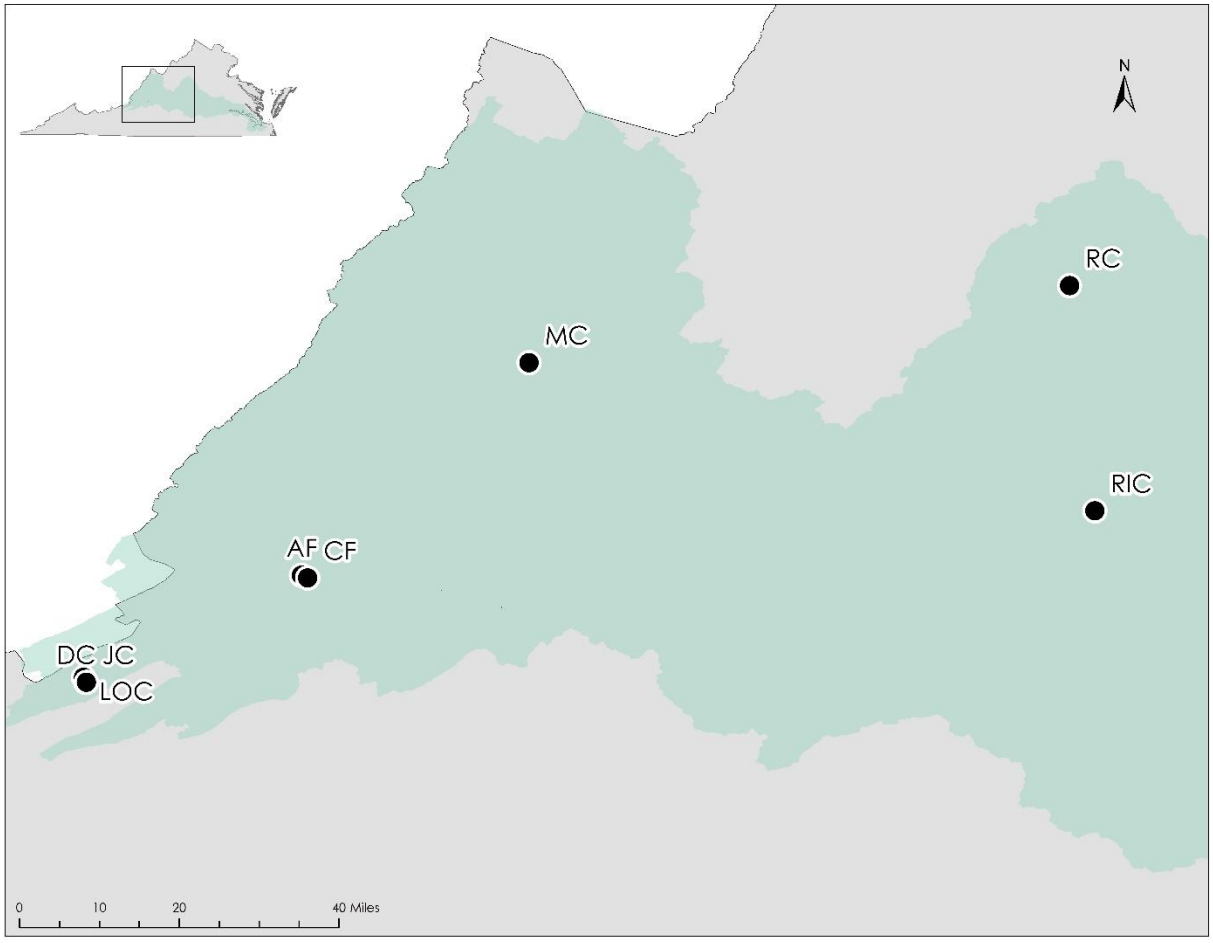

Figure 1.1: Study region and sampling locales. Points represent environmental DNA field collection sites across the James River Basin in Virginia. Exact locations and stream names are not reported given this species' status as endangered.

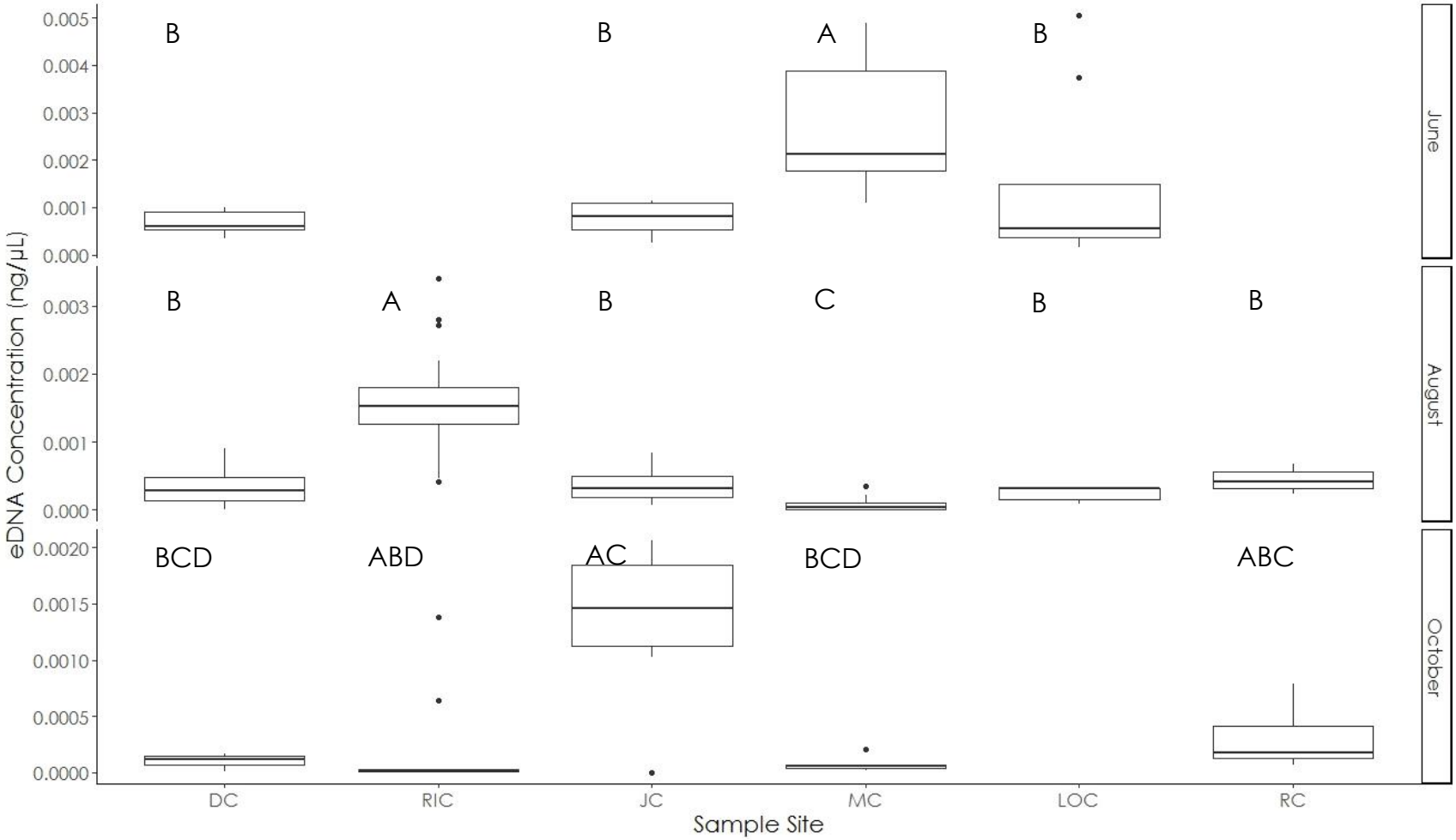

Figure 1.2: Estimation of *P. collina* DNA template concentration in water samples collected from the six sample sites in June, August, and October. Sample sites are ordered from smallest to largest (left to right) based on relative census population size. Groupings representing the mean concentrations not significantly different from each other are indicated using uppercase letters and are based on the results of a Kruskal-Wallace and Conover-Iman post hoc test for pairwise differences.

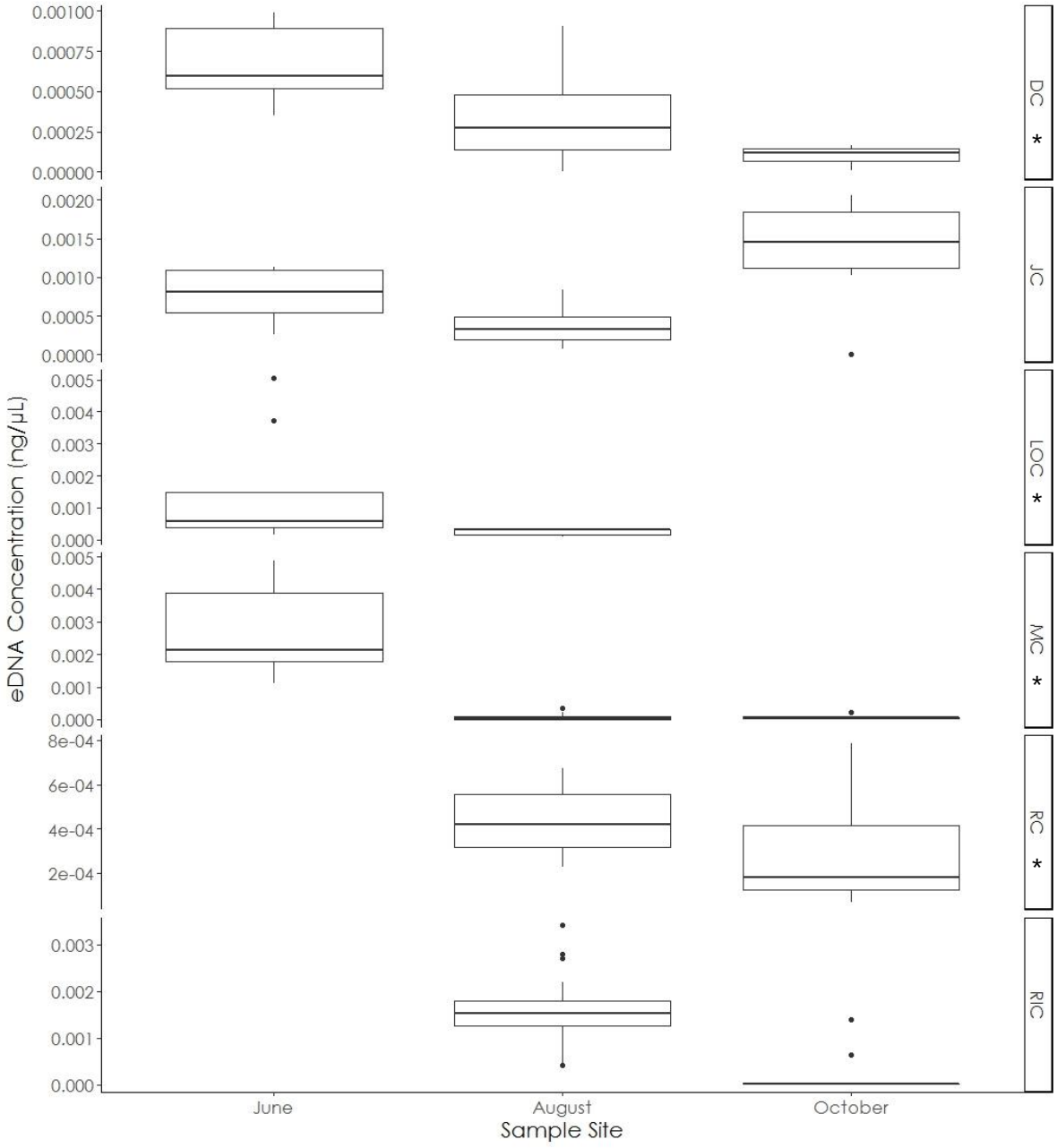

Figure 1.3: Change in DNA concentrations at sample sites across seasons. In four of the sampled streams (DC, LOC, MC, and RIC) DNA concentrations matched predicted levels of mussel activity. As predicted, DNA concentrations were significantly higher in June and August than in October (indicated by an *). No significant difference was found between seasons in RC. In JC, DNA concentrations were significantly higher in October than in August, contrary to the expected pattern.

References

Abbaszadegan, M., Huber, M.S., Gerba, C.P. and Pepper, I.L. (1993) Detection of enteroviruses in groundwater with the polymerase chain reaction. *Applied Environmental Microbiology 59*, 1318–1324.

Ardura, A., Zaiko, A., Martinez, J. L., Samulioviene, A., Semenova, A., & Garcia-Vazquez, E. (2015). eDNA and specific primers for early detection of invasive species - A case study on the bivalve Rangia cuneata, currently spreading in Europe. *Marine Environmental Research*, *112*, 48–55. DOI:10.1016/j.marenvres.2015.09.013

Bohmann, K., Evans, A., Gilbert, M. T. P., Carvalho, G. R., Creer, S., Knapp, M., … de Bruyn, M. (2014). Environmental DNA for wildlife biology and biodiversity monitoring. *Trends in Ecology and Evolution*, *29*(6), 358–367. DOI: 10.1016/j.tree.2014.04.003

Campbell, D. C., Johnson, P. D., Williams, J. D., Rindsberg, A. K., Serb, J. M., Small, K. K., & Lydeard, C. (2008). Identification of “extinct” freshwater mussel species using DNA barcoding. *Molecular Ecology Resources*, *8*(4), 711–724. DOI: 10.1111/j.1755-0998.2008.02108.x

Clarke, A.H., & R.J. Neves. (1984). Status survey of the James River spinymussel, *Canthyria collina*, in the James River, Virginia. Final Report, U.S. Fish and Wildlife Service,Newton Corner, MA. 73pp.

Deiner, K., and Altermatt, F. (2015). Transport Distance of Invertebrate Environmental DNA in a Natural River. *PLoS ONE, 9*(2): e88786. DOI: 10.1371/journal.pone.0088786

De Ventura, L., Kopp, K., Seppälä, K., & Jokela, J. (2017). Tracing the quagga mussel invasion along the Rhine river system using eDNA markers : early detection and surveillance of invasive zebra and quagga mussels. *Management of Biological Invasions*, *8*(1), 101–112.

Esposito, A.C. (2015). Using Capture-Mark-Recapture Techniques to Estimate DetectionProbabilities & Fidelity of Expression for the Critically Endangered James Spinymussel (*Pleurobema collina*). (2015) M.S. Thesis. James Madison University, Harrisonburg, VA.

Ficetola, G.F., Miaud, C., Pompanon, F., and Taberlet, P. (2008). Species Detection Using Environmental DNA From Water Samples. *Biology Letters*, *4*(4), 423-425.

Geist, J., Wunderlich, H., & Kuehn, R. (2008). Use of Mollusc Shells for DNA-Based Molecular Analyses. *Journal of Molluscan Studies*, *74*, 337–343. DOI: 10.1093/mollus/eyn025

Gibson, K. E., Schwab, K. J., Spencer, S. K., & Borchardt, M. A. (2012). Measuring and mitigating inhibition during quantitative real time PCR analysis of viral nucleic acid extracts from large-volume environmental water samples. *Water Research*, *46*, 4281–4291. DOI: 10.1016/j.watres.2012.04.030

Goldberg, C. S., Pilliod, D. S., Arkle, R. S., & Waits, L. P. (2011). Molecular detection of vertebrates in stream water: A demonstration using rocky mountain tailed frogs and Idaho giant salamanders. *PLoS ONE*, *6*(7), e22746. DOI: 10.1371/journal.pone.0022746

Goldberg, C. S., Strickler, K. M., & Pilliod, D. S. (2015). Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. *Biological Conservation*, *183*, 1–3. DOI: 10.1016/j.biocon.2014.11.040

Hickcox T., Hsu, F.C., Wong, R., and Brito-Robinson, T. (2011). eDNA Surveillance of Asian Carp on the St. Croix and Mississippi Rivers. *Environmental DNA Solutions*, Granger, IN

Hove, M. C., & Neves, R. J. (1994). Life history of the endangered James spinymussel *Pleurobema collina* (Conrad, 1837) (Mollusca: Unionidae). *American Malacological Bulletin*

Ijzerman, M.M., Dahling, D.R., and Fout, G.S. (1997). A Method to Remove Environmental Inhibitors Prior to the Detection of Waterborne Enteric Viruses by Reverse Transcription-Polymerase Chain Reaction. *Journal of Virological Methods*, *63*(1), 145-153.

Jane, S. F., Wilcox, T. M., Mckelvey, K. S., Young, M. K., Schwartz, M. K., Lowe, W. H., … Whiteley, A. R. (2015). Distance, flow and PCR inhibition: eDNA dynamics in two headwater streams. *Molecular Ecology Resources*, *15*(1), 216–227. DOI: 10.1111/1755-0998.12285

John, M.E. (1992). An Efficient Method for Isolation of RNA and DNA From Plants Containing Polyphenolics. *Nucleic Acids Research*, *20*(9), 2381.

Kelly, R.P., Port, J.A., Yamahara, K.M., Crowder, L.B. (2014). Using Environmental DNA to Census Marine Fishes in a Large Mesocosm. *PLoS ONE, 9*(1): e86175. DOI: 10.1371/journal.pone.0086175

Klymus, K. E., Richter, C. A., Chapman, D. C., & Paukert, C. (2015). Quantification of eDNA shedding rates from invasive bighead carp Hypophthalmichthys nobilis and silver carp Hypophthalmichthys molitrix. *Biological Conservation*, *183*, 77-84. DOI: 10.1016/j.biocon.2014.11.020

Laramie, M. B., Pilliod, D. S., & Goldberg, C. S. (2015). Characterizing the distribution of an endangered salmonid using environmental DNA analysis. *Biological Conservation*, *183*, 29–37. DOI: 10.1016/j.biocon.2014.11.025

Mckelvey, K. S., Young, M. K., Knotek, W. L., Carim, K. J., Wilcox, T. M., Padgett-Stewart, T. M., & Schwartz, M. K. (2016). Sampling large geographic areas for rare species using environmental DNA: A study of bull trout Salvelinus confluentus occupancy in western Montana. *Journal of Fish Biology*. DOI: 10.1111/jfb.12863

Opel, K. L., Chung, D., and McCord, B. R. (2010). A Study of PCR Inhibition Mechanisms Using Real Time PCR. *Journal of Forensic Science*, *55*(1), 25-33.

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., Waits, L. P., & Richardson, J. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(8), 1123–1130. DOI: 10.1139/cjfas-2013-0047

Sansom, B. J. and L. M. Sassoubre. (2017). Environmental DNA (eDNA) Shedding and Decay Rates to Model Freshwater Mussel eDNA Transport in a River. *Environmental Science & Technology, 51*, 14244-14253. DOI: 10.1021/acs.est.7b05199

Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012). PCR Inhibitors - Occurrence Properties and Removal. *Journal of Applied Microbiology*, *113*(5), 1014-1026.

Seddon, P. J. (2010). From Reintroduction to Assisted Colonization: Moving along the Conservation Translocation Spectrum. *Restoration Ecology*, *18*(6), 796–802. DOI: 10.1111/j.1526-100X.2010.00724.x

Sigsgaard, E. E., Carl, H., Møller, P. R., & Thomsen, P. F. (2015). Monitoring the near-extinct European weather loach in Denmark based on environmental DNA from water samples. *Biological Conservation*, *183*, 46–52. DOI: 10.1016/j.biocon.2014.11.023

Smith, D.R., Villella, R.F., and Lemarié, D.P. (2001). Survey Protocol for Assessment of Endangered Freshwater Mussels in the Allegheny River. *Journal of the American Benthological Society*, *20*(1), 118-132.

Stoeckle, B. C., Kuehn, R., & Geist, J. (2015). Environmental DNA as a monitoring tool for the endangered freshwater pearl mussel (Margaritifera margaritifera L.): A substitute for classical monitoring approaches? *Aquatic Conservation: Marine and Freshwater Ecosystems*, (October), 1–10. DOI: 10.1002/aqc.2611

Strayer, D.L., Sprague, S., and Claypool, S. (1996). A Range-Wide Assessment of Populations of *Alasmidonta Heterodon*, an Endangered Freshwater Mussel (Bivalvia: Unionidae). *Journal of the North American Benthological Society*, *15*(3), 308-317.

Takahara, T., Minamoto, T., Yamanaka, H., Doi H, Kawabata, Z. (2012). Estimation of Fish Biomass Using Environmental DNA. *PLoS ONE 7*(4): e35868. DOI:10.1371/journal.pone.0035868

Thomsen, P. F., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., … Willerslev, E. (2011). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 1–9. DOI:10.1111/j.1365-294X.2011.05418.x

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA - An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. DOI:10.1016/j.biocon.2014.11.019

USFWS. (1990). James Spinymussel *(Pleurobema collina)* Recovery Plan. <https://efotg.sc.egov.usda.gov/references/public/WV/James_Spinymussel_Recov_Plan.pdf>

VDGIF. (2015). Using Mark-Recapture Techniques to Estimate Population Parameters and Detection Probability for the Federally Endangered James Spinymussel. General Technical Report. Virginia Contract No. 2014-14665

Wilson, C. D., Roberts, D., & Reid, N. (2011). Applying species distribution modelling to identify areas of high conservation value for endangered species: A case study using Margaritifera margaritifera (L.). *Biological Conservation*, *144*, 821–829. DOI:10.1016/j.biocon.2010.11.014

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T. (2012). Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 13:134. DOI:10.1186/1471-2105-13-134

Chapter 3: Characterization of Habitat Requirements and Identification of Priority Conservation Areas for the Endangered James Spinymussel (*Parvaspina collina*)

Abstract

Ecologists and conservation managers increasingly rely on species distribution models (SDM) to predict habitat suitability for a species across a landscape but these models often don't consider potentially important interspecies biotic interactions. Here we develop a biologically relevant SDM for *Parvaspina collina* that identified important habitat characteristics influencing their distribution as well as priority conservation areas. Habitat suitability and the potential distribution of *P. collina* were best predicted by variables associated with land cover and anthropogenic effects related to siltation. Contrary to our predictions, the distribution of the host fish did not predict habitat suitability for *P. collina.* Only 19% of the streams in the study area are considered suitable for *P. collina.* Twenty six noncontiguous waterways were identified as priority conservation areas based on the total habitat suitability for *P. collina* and their fish hosts. These results can be used to direct future sampling efforts and identify dispersal corridors.

Introduction

Species distribution modeling is based on ecological niche theory (Kearney and Porter 2009; Phillips 2004) and has been used for a multitude of applications including

quantifying a species' environmental niche (Wharton and Kriticos 2004), assessing the risk of species invasions (Gama et al. 2016; Gormley et al. 2011), determining the impact of environmental disturbances (Devictor et al. 2008), predicting the effects of climate change (Erasmus et al. 2002; Chen et al. 2011), identifying areas of high conservation priority (Early et al. 2008; Wilson et al. 2011; Morato et al. 2014; Ferrer-Sanchez & Rodriguez-Estrella 2016), prioritizing locations to search for new populations (Jarvis et al. 2003), and estimating the spatial and genetic connectivity of populations (Wang et al. 2008; Koen et al. 2012; Poor et al. 2012; Tournant et al. 2013; Razgour et al. 2014).

Effective conservation requires knowledge of the habitat requirements and potential distribution of a species (i.e., its fundamental niche). A thorough understanding of these parameters allows managers to make more informed decisions regarding its conservation such as the boundaries of protected areas, locations of dispersal corridors, and proposed areas for species reintroductions (Guisan et al. 2013). While this knowledge is important, documenting the entire occupied range or niche of a species is challenging and often cost prohibitive, especially if the species is cryptic and has a broad or patchy distribution. To overcome these challenges, ecologists and conservation managers increasingly rely on species distribution models (also referred to as habitat distribution models or environmental niche models) to predict areas of occurrence and habitat suitability across a landscape for a given species.

In general, these models relate species location data to abiotic environmental predictor variables (Guisan and Thuiller 2005) to 1) identify important factors influencing where a species occurs and 2) predict its potential distribution across the landscape (Kearney and Porter 2009). The species location information used in these models is often in the form of presence only, presence-absence, or abundance data and can come from opportunistic sightings, structured field surveys, or museum records (Guisan and Thuiller 2005; Graham et al. 2004a). Presence-only data refers to a set of point locations based only upon positive identification of where individuals are found. These types of data are readily available from natural history records and free online data warehouses but often lacks accompanying information such as the sampling date, level of effort, or the number of individuals at the location. Also, ongoing debate exists over the types of inferences that can be made about species occurrence probability using this type of data (Hastie and Fithian 2013). Presence-absence data gives us valuable information about the areas where individuals are found as well as where they are *not* found. This provides additional insights over the assumption of "pseudo-absence" locations used in presence-only approaches. However, obtaining presence-absence data requires systematic and structured sampling which results in this data type being less prevalent. It is also important to note that an "absence" point may be misleading because a lack of detection does not necessarily mean the species does not or could not occupy that habitat patch (Mackenzie 2005). An absence point could result from our inability to detect a species at a location, species range shifts, temporal variation in occupancy, or barriers preventing movement (Elith and Leathwick 2009; Howard et al. 2014). Abundance data, i.e., population size estimates at given locations, allows us to

model the relative suitability of habitats more accurately than presence only or presence-absence data because the species abundance at a location should be indicative of the habitat quality (Howard et al. 2014). However, like presence-absence data, reliable information about species' abundance requires structured and repeated sampling so it is hard to come by and the time or budget constraints of a study may prohibit this type of data collection. The environmental predictor variables incorporated into the models can range from broad scale climatic variables to fine scale habitat characteristics. While there are a multitude of different modeling approaches, the framework that is ultimately chosen will depend on the specific goals and questions of the study, the study organism, as well as the format and spatial scale of available data (Guillera-Arroita et al. 2015).

Species distribution models have allowed us to use measurable abiotic parameters (such as those related to the climate in a region) to make inferences about habitat suitability for a species across a variety of spatial scales. This allows us to predict where it might occur across the landscape based on locations that fall within those putatively important parameters and make informed conservation decisions. However, until recently, these models have not included any information about potentially important biotic interactions between species. Ignoring these relationships, especially for organisms that exhibit obligate parasitism (Vaughn and Taylor 2000), is likely making our inferences about their habitat requirements and potential distributions less reliable, especially if the aim is to extrapolate to future climate conditions where novel species interactions may occur (Gilman et al. 2010; Van der Putten et al. 2010; Elith and

Leathwick 2009). It has been frequently suggested that models should consider potentially important biotic interactions (Pollock et al. 2014; Elith and Leathwick 2009) and in studies performed thus far, these relationships have indeed had an influence on the distribution of a species (Guisan et al. 2006; Heikkinen et al. 2007; Wharton and Kriticos 2004; leRoux et al. 2014). One approach to incorporating biotic interactions is to include the predictions of a species distribution model for one species as covariates when fitting a model for another species (e.g., Guillera-Arroita et al. 2015; Meier et al. 2010; Leathwick and Austin 2001).

The purpose of this study is to create a biologically relevant habitat distribution model for a federally endangered species of freshwater mussel, the James spinymussel (*Parvaspina collina*; Unionidae). Of particular interest here is that *P. collina* utilizes a suite of fishes to serve as hosts to its parasitic larvae. Given this biotic interaction, any efficient development of distribution models for the mussel must investigate the extent to which constraints in distributions of host fishes influence the presence of adult *P. collina*. Previous work has shown a propensity for freshwater mussels to have highly specific niche characteristics (e.g., Wilson et al 2011) and the distributions of similar freshwater mussels are influenced by watershed metrics like altitude, topographic relief, and soil characteristics (Wilson et al. 2011; Arbuckle and Downing 2002). Because of these factors, it is likely that the distribution of *P. collina* is influenced by a combination of watershed level environmental metrics and the distribution of their host fishes. The specific goals are threefold: 1) Identify environmental parameters that are influencing where *P. collina* occurs, 2) predict its full potential distribution based on those

parameters across the landscape, and 3) identify priority conservation areas based on the overall habitat suitability for *P. collina* and its fish hosts.

Methods

Study Area

The James River Basin in Virginia covers approximately 10,000 square miles across the center of the state (Figure 2.1). It is the largest river basin in Virginia and flows southeast from the Allegheny Mountains to the Chesapeake Bay. The James River Basin spans four physiographic provinces from East to West: Valley and Ridge, Blue Ridge, Piedmont, and the Coastal Plain. The Coastal Plain is separated from the Piedmont by a three mile stretch of the river called the Fall Zone (or Fall Line) where the river descends 84 feet from the hard bedrock of the Piedmont to the softer sediment of the Coastal Plain. The majority of the James River Basin is classified as forested (>65%) and approximately 12% is classified as urban, with the remaining area designated as cropland or pasture. Annual precipitation across the region averages 42.5 inches (DEQ 2015).

Study Species

Parvaspina collina is a federally endangered species of freshwater mussel whose historic range once extended throughout the entirety of the James River Basin above the Fall Line (Clark & Neves, 1984). However, due to the combined effects of habitat degradation, river impoundments, predation, and resource competition from invasive

Asian clams (*Corbicula fluminea*), many *P. collina* populations experienced local extirpation within the last two decades (Clarke & Neves, 1984; Hove & Neves, 1994; USFWS, 1990). The species is now patchily distributed throughout the James River and Dan River basins in Virginia and North Carolina, a distribution which only encompasses approximately 10% of its historic range (USFWS, 1990). Presently, there are 25 streams in Virginia and 1 in the Roanoke River drainage in North Carolina with extant populations of James spinymussel (Figure 2.2). However, the complete distribution of extant *P. collina* populations has not been established.

Life History of the James spinymussel

Like most other species of freshwater mussels, the juvenile stages of the James spinymussel is an obligate parasite of specific fishes (Hove 1990). During reproduction, sperm released by the males into the water column is taken in by females to fertilize eggs. Larval mussels (glochidia) are then released and encyst on the gills or fins of fish hosts where they will remain for a period of growth before dropping off to settle into new habitat. James spinymussel release their glochidia as conglutinates, a package of glochidia bound by a matrix of mucus (Watters 1999). Based on the breeding ecology of mussels and their reliance on host fish for reproduction and dispersal, mussel distribution and abundance patterns are likely constrained by a combination of historical effects, landscape-level abiotic factors, fish host availability, and in-stream environmental conditions (Vaughn and Taylor 2000).

For the James spinymussel, eight fish species in the family *Cyprinidae* (minnows) have been identified as suitable hosts (Hove 1990; Petty et al. 2005; Table 2.1). These fishes

are non-migratory and have relatively broad distributions across the James River Basin (Figure 2.3; Jenkins and Burkhead 1994). While *P. collina* has been found to successfully attach to and metamorphosize on all eight species, the primary fish host has been identified as *Nocomis leptocephalus* (bluehead chub). This species is an omnivore but feeds primarily by scraping algae from rocks in the stream substrate (Jenkins and Burkhead 1994), where it is likely to encounter the conglutinates full of larval *P. collina* during breeding season. The bluehead chub is broadly distributed across the Valley and Ridge, Blue Ridge, and Piedmont provinces but its eastern limit generally corresponds to the Fall Line (Figure 2.3 (E); Jenkins and Burkhead 1994).

Species Presence Points

Species presence records for *P. collina* were obtained from Brian Watson at the Virginia Department of Game and Inland Fisheries (VDGIF; N = 315). Each record included the latitude and longitude of a known location within the James River Basin inhabited by a population of James spinymussel. These records were accumulated from environmental site surveys, museum records, and on-going population monitoring surveys conducted by VDGIF. Only species presence records collected after 1980 (N = 263) were included in the models due in large part to the rapid range contraction seen over the last 20 years. All points were “snapped” to the digitized National Hydrography Dataset (NHDPlus ver. 2; McKay et al. 2015) and then spatially thinned so that only one presence record existed for each unique stream segment (COMID) resulting in 79 presence records.

Fish occurrence data for each of the 8 species were collected from the IchthyMaps database (Frimpong et al. 2016), which is a compilation of historical occurrence records of fish within the contiguous United States collected between the years 1950 and 1990. Occurrence records are listed by the unique identifier (COMID) of their corresponding digitized stream segment in the National Hydrography Dataset (NHDPlus ver. 2; McKay et al. 2012). Occurrence records were trimmed to the extent of the James River Basin study area and were further reduced to ensure there was only one presence record per species in each stream segment. Once thinned, the number of occurrences for each species ranged from 78 (rosefin shiner) to 671 (bluehead chub; Table 2.1).

Environmental Covariates

Environmental covariates were downloaded for the state of Virginia from the StreamCat database (Hill et al. 2016), which contains a total of 513 natural and anthropogenic landscape metrics for streams and their associated catchments. These data represent statistical summaries of GIS layers and were developed using the National Hydrography Dataset (NHDPlus ver. 2; McKay et al. 2012). Additional environmental variables were downloaded from the Northeast Stream Classification providing categorical information on the size, gradient, temperature, and geologic buffering capacity of waterways in Virginia (Olivero & Anderson, 2008). Only variables with heterogeneity in the study area were retained. A pairwise Spearman correlation was used to identify pairs of predictors with high levels of collinearity (e.g., $R^2 > 0.70$; Dormann et al. 2013). When two variables were found to be highly correlated, the more biologically relevant parameter was retained for the model. A “background file”, containing the set of all environmental

variables measured across the entire range was created for each of the host species as well as for *P. collina*.

Species distribution modelling

A Species Distribution Model (SDM) was estimated for host and mussel species using Maximum Entropy (MaxEnt software, ver. 3.3.3; Phillips et al. 2006, Phillips and Dudik 2008). This approach compares the environmental conditions at locations where the species was found to the frequency of those conditions sampled from across the same landscape (Guillera-Arroita 2017). If the distribution of environmental values where samples occur deviates from the background distribution of those same variables across the landscape (i.e., it is not a random sample), this is suggestive that this variable is likely to be either directly or indirectly influencing the ability of the organism to persist at this location. The MaxEnt method was chosen because it relies on presence only information and has been shown to perform well even with small sample sizes of fewer than 25 species presence points (Hernandez et al. 2006; Pearson et al. 2007). For all species models, 75% of the background points were randomly selected for model building and the remaining 25% were used for model validation. The models were fit using simple quadratic and hinge features. The MaxEnt raw scores (i.e., the predictions for habitat suitability) for each stream reach were then projected across the landscape for each species.

Host Fish Habitat Suitability Models

To identify the most relevant environmental parameters for each host species, the model was run using a combination of 60 variables at a time without attributing any *a*

priori knowledge of what might be considered “good” predictor variables. The variables used in the final model were selected through an iterative process where for each successive model, only the variables with the highest percent contribution and importance based on the Jackknife plots were kept. Final model selection was based on the Regularized Training Gain (RTG: a measure of the model’s ability to discriminate between locations with known occurrences from random background points) and the Area Under the Curve (AUC: an indication that the MaxEnt raw scores are higher at locations with known species occurrences than at random background points) (McGarvey et al. 2018).

Parvaspina collina Model Selection

The model for *P. collina* was created using the same iterative process. However, the predicted habitat suitability for the fish hosts (i.e., the MaxEnt raw predicted scores for each stream segment for each potential host species) were also included as environmental predictor variables. If the predicted habitat for any of the fish host contributes to the spatial distribution of the mussel, then the fish suitability scores should be included in the MaxEnt model for the mussel and we would expect those variables to have the highest percent contribution to the model when compared to that of the other environmental parameters. Model selection for *P. collina* proceeded as described for host fishes.

Habitat Suitability & Priority Conservation Sites

A lower threshold of values considered “suitable” *P. collina* habitat was established by identifying the lowest predicted habitat score associated with a stream segment with

documented *P. collina* presence. This lower threshold was then used to calculate the percentage of streams deemed suitable habitat across the watershed. Any stream segment with a value below this threshold was removed from consideration when identifying areas of highest conservation priority. Only stream segments whose habitat suitability scores for *P. collina* fell above this threshold value were used. Priority locations to search for new populations were identified based on the stream segments with the highest overall scores based on the *P. collina* model that did not have documented *P. collina* presence. A One-Way ANOVA was performed to compare the overall habitat suitability for the host fishes to that of *P. collina*. It was expected that *P. collina* should have lower average habitat scores since they have a much narrower distribution and therefore may have stricter habitat requirements leading to limited habitat suitability across the watershed.

Areas within the James River Basin that should be prioritized for conservation efforts were identified by summing the average MaxEnt raw scores produced by the models for each species of host fish and *P. collina*. Stream segment scores now represented the summed total of all habitat scores meaning stream segments whose scores fell within the top 99th percentile of all scores now indicated the most suitable sites overall. Once identified, the spatial distribution and overall connectivity of these proposed conservation sites was determined using ArcMap v. 10.4.1.

Results

Host Fish Habitat Suitability Models

The predictive performance of the models predicting host species habitat suitability were shown to be relatively high (AUC 0.715 - 0.863; Table 2.1), indicating a good fit of the models to the data. The ability to discriminate locations with known presence points from random background points based on the environmental parameters given varied between species (RTG 0.27 - 0.816; Table 2.1). Predictions of habitat suitability across the watershed for each species are shown in Appendix 2 (Figures B1-8).

The most important environmental parameters for each model varied for each fish species (Table 2.2). A description of the two most influential variables for each species are included here. *Campostoma anomalum* was most influenced by the mean catchment elevation (57.5%) and mean wetness index (19.6%) and occurs more frequently in areas of higher elevation with lower average wetness index. The distribution of *Clinostomus funduloides* was most influenced by the number of road crossings occurring on steep slopes (32.5%) and the mean soil thickness (31.6%). This species preferred areas with lower numbers of road crossings and mid to high values of soil thickness. The combination of these variables could lead to less runoff and sedimentation. *Cyprinella analostana* selected areas with lower levels of silicic residual material and colluvial sediment (25 and 22.5 percent contribution respectively). *Lythrurus ardens* was most influenced by the percent of the watershed classified as barren land (21.6%) as well as the amount of silicic residual material (17.3%), preferring areas with lower values of the former and higher values of the latter. *Nocomis*

leptocephalus occurred more often in areas with fewer road crossings on steep slopes (43.3%) but with higher values of the mean compressive strength in the surface geology (16.2%). *Notropis procne* preferred habitat with lower values of forest loss (27%) across the watershed but also occurs more frequently in areas with mid to high levels of high intensity land use (18.4%). *Chrosomus oreas* preferred areas with a lower density of road crossings (35.4%) and appeared to avoid places with warm water (30.6%). Finally, *Rhinichthys atratulus* preferred habitat characterized by a lower density of road stream crossings on steep slopes (38%) and mid to higher catchment elevation (32.7%). While the variables and percent contribution were different for each species, there was some overlap in the important variables. Eight of the variables related to land-cover, temperature, soils, elevation, and impervious surfaces were important to two of the host fish but had varying contributions to the models. One of the most important variables overall appears to be the density of stream crossings that occurred on steep slopes. It was an important predictor for four of the eight fish hosts, a possible indication that these fish species are highly influenced by levels of runoff in their environment.

There was variation in the overall distribution of stream scores for each species (Figure 2.4). Results of a one-way ANOVA on the log10 transformed values show that there were significant differences between the average habitat scores of the eight fish hosts (ANOVA; $F_{7,126910} = 358.1$, $p < 0.0001$; Figure 2.5).

Parvaspina collina Model

The performance of the habitat suitability model for *P. collina* was relatively high (AUC 0.957), indicating a good fit of the models to the data (Table 2.1). In fact, compared to

the models describing the fish hosts, *P. collina* model had the best fit overall. Predictions of habitat suitability across the watershed are shown in Appendix 2 (Figure B2).

Habitat suitability and potential distribution was best predicted by variables associated with land cover and anthropogenic effects (Table 2.2). *P. collina* preferred locations with a low percentage of the catchment area classified as agriculture and occurred more often in areas with low to moderate gradients and low levels of forest loss within the watershed. They appear to avoid areas with high levels of colluvial sediment and open water (i.e., less tree cover over waterways).

The minimum predicted habitat score for a stream segment with known *P. collina* presence was $1.97e^{-4}$. Based on this lower threshold value, a total of 9,705 stream segments in the study area (~81% of all segments) were removed indicating that only about 19% of the streams across the James River Basin are considered suitable habitat for *P. collina* (Figure 2.6).

As expected, the model predicted significantly higher habitat suitability in locations with documented *P. collina* presence ($W = 31,527$, $N_{P.\ collina} = 78$, $N_{Background} = 11,924$, $p < 0.0001$; Figure 2.7). Predicted habitat scores for *P. collina* were significantly smaller than those predicted for all of the host fish species ($F_{8,102,332} = 6,402$, $P < 0.0001$; Figure 2.8).

Priority Conservation Sites

A total of 122 stream segments (defined as locales in the 99th percentile) were identified as priority locations to search for new *P. collina* populations. These locations (Figure 2.9) had a high overall habitat suitability score yet have no documentation of *P. collina* presence. While these stream segments represented a total of 28 distinct waterways, they were not contiguous within a given river or stream. Sixteen of these stream segments were less than 1 river mile from a documented *P. collina* population and 31 of them were within 2 river miles of a *P. collina* population meaning the majority of the identified sites (75 total) are over 2 river miles beyond a documented population. The distribution of the summed scores across all stream segments was skewed to the left indicating a large frequency of relatively low scores (Figure 2.10).

A total of 121 stream segments were identified as the top 99th percentile of summed habitat suitability scores, all of which had scores greater than 5.7e^{-3}. These were the sites identified as priority conservation areas (Figure 2.11, Table B.1). These 121 stream segments represented a total of 26 distinct waterways but were not necessarily contiguous within a given river or stream. In several cases, the priority conservation sites were located upstream or downstream of a known *P. collina* population. A total of 53 stream segments fell within 2 miles of a known *P. collina* population, 41 of which fell within 1 mile of a known *P. collina* population. There was also significant overlap between the sites identified as priority conservation locations based on the fish + *P. collina* scores and those based on *P. collina* scores only (N = 102).

Discussion

Fish Habitat Suitability Models

Even though the host fish species are all in the same family Cyprinidae, the results of the models indicate they have different habitat requirements. These differences can potentially influence their individual responses to future habitat modifications and climate change.

There was significant variation in the model's accuracy among species, which is similar to what other studies have found (Tsoar et al. 2007). The worst performing model was for *Nocomis leptocephalus* (bluehead chub) while the best performing model was for *Cyprinella analostana* (satinfin shiner; Table 2.1). In general, there was an inverse relationship between the number of presence points and overall model performance. The best performing models were for species with the fewest presence points and the worst performing models were for species with the most presence points. This was expected given the underlying mechanism used by MaxEnt to identify good habitat. To do this effectively, there must be a detectable difference between the habitat at presence locations and the habitat available across the watershed. If the species occupies a wide range with many presence records, it is likely to inhabit a larger range of these habitat variables, making it harder for the algorithm to identify differences. Fish hosts that are specialists or have a smaller range may have a more negative response to landscape fragmentation and disturbance (Devictor et al. 2008). This could affect their distribution and abundance over time, leading to decreased opportunity for *P. collina* to successfully reproduce in the affected stream reaches.

Parvaspina collina Model

The combination of the important predictor variables for *P. collina* model likely relate to levels of runoff and sedimentation of the waterway. Higher gradient streams could lead to increased flows and areas with high forest loss and agriculture on steep slopes could lead to increased sedimentation, a condition known to be unsuitable for freshwater mussels by interfering with their ability to filter feed and respire (Box and Mossa 1999). Increased sedimentation has been shown to have cascading effects on food webs with direct and indirect effects at every trophic level (Henley et al. 2000). For freshwater mussels, this could result in decreased food availability (plankton) leading to decreases in growth or reproduction. Therefore, areas with habitat factors that may lead to increased sedimentation are likely to be considered bad habitat for *P. collina*. These results are supported by a previous study by Arbuckle and Downing (2002) who found that agricultural watersheds with high slopes impact freshwater mussel abundance and richness through siltation and destabilization of stream substrate.

These findings indicate that future habitat modifications that lead to increased runoff and sedimentation are likely to have a negative impact on the potential distribution of *P. collina*. Based on the model predictions, approximately 19% of the stream reaches in the James River Basin are currently considered suitable habitat for *P. collina*. If those areas experience increased agriculture or forest loss, conditions in the streams could worsen, ultimately driving the habitat suitability down below the threshold. However, if areas that fall just below the threshold value were targeted for restoration purposes, it

might be possible to increase the habitat suitability over time through replanting of forests and reducing the percentage of agricultural land in each watershed.

It was predicted that the most important variables for determining habitat suitability for *P. collina* would be a combination of the fish host habitat suitability and environmental parameters related to anthropogenic influences (Watters 1999). However, contrary to these predictions, the habitat suitability of the fish hosts was not the best predictor of *P. collina* habitat suitability.

Given that the distribution of fish hosts has been shown to be a significant influence on the distribution of freshwater mussels (Watters 1992), it is unlikely that the distribution of the host fish has *no* influence on the distribution of *P. collina*. However, this relationship might be scale dependent, having greater impact at the stream level than at the watershed level (Guisan et al. 2006; Heikkinen et al. 2007). It's possible that the interactions with fish host may have a strong influence on the distribution of *P. collina* within a stream reach but this relationship is less apparent when considering their respective distributions across the entire watershed. The presence points within a stream are given by the location of the stream reach which vary in length. We don't have information on where *P. collina* are in the stream, rather just that they exist somewhere along that reach. Because of this, we are unable to pick up on any "stream - scale" influences the fish distributions might be having. To do this, we would need information on the habitat characteristics within a stream, along with the locations of mussels and fish within the stream on a much finer scale.

The fact that the model selected for *P. collina* had the best fit when compared to the fish host models indicates *P. collina* has a narrower range of habitat tolerances and fewer areas of suitable habitat within the James River Basin than the host fish (Tsoar et al. 2007). Most of the fish hosts had relatively large distributions with large numbers of suitable stream reaches. Since *P. collina* is a generalist parasite (i.e., it can successfully reproduce using any of the eight identified hosts) and its populations exist in a much narrower range of habitat conditions than the fish hosts, it stands to reason that the regional distribution of *P. collina* is not necessarily limited by that of the fish hosts.

We must also consider the possibility that fish host abundance, rather than habitat suitability, is driving mussel abundance at the regional scale (Vaughn and Taylor 2000). Higher fish host abundance could lead to a higher probability of successful reproduction, in turn leading to higher *P. collina* abundances at locations with suitable habitat. This relationship could have a significant influence on the overall distribution of *P. collina* across the watershed.

Also, the temporal variation in the collection of presence samples could have led to a lack of congruence. The current distribution of *P. collina* represents populations of mussels detected since 1980 and the fish presence points represent locations with documented sightings ending in 1990 so there is only a 10 year overlap in the documented distributions. There may be a lag between a change in distribution of the

fish and a resulting change in the distribution of mussels. *P. collina* are long lived, up to about 20 years (Hove 1990). They rely on the fish hosts for reproduction, but they are able to survive without them. This means the population as a whole can withstand shifts in fish host distributions and abundance. It's possible that historical distributions of *P. collina* may have had a tighter relationship to the overall distribution or abundances of the host fish but it would not be possible to pick up on this potential historical relationship using the data available in this study.

A more likely explanation for the lack of relationship seen in this study is the fact that *P. collina* has experienced significant range contraction over the last several decades due to anthropogenic habitat modification. The presence points used in this study represent the known locations of *P. collina* populations observed since 1980, a distribution that only encompasses approximately 10% of their historic range (USFWS, 1990). Only the populations that historically existed in the most upstream tributaries remain, meaning the current distribution is driven almost entirely by the impacts of anthropogenic effects rather than the historical and biological relationships with the host fist. While it has been argued that species distribution models developed when a species is no longer found throughout its historic range may be biased, they should still indicate important environmental factors influencing the species' niche (Gibson et al. 2004).

It is important to note that presence only data are susceptible to estimation bias introduced by sampling bias because they do not contain any information about sampling effort (Guillera-Arroita 2017). The level of sampling effort can have a big

impact on the likelihood of finding the organism (detection probability), especially if that organism is cryptic or patchily distributed in the stream. Using MaxEnt modeling, all "presence" locations are treated as equal regardless of how long the survey was conducted at that site or how many individuals were located in the area. Because of this, presence only data cannot tell us whether sparse species records in an area are due to actual species rarity or a lack of survey effort (Guillera-Arroita 2017). If the former is true, then the habitat suitability at these particular locations may not be ideal yet the environmental conditions at these locations are still used with equal weight to construct the model, potentially influencing the relative contribution of each environmental variable and the resulting inferences about species habitat requirements and suitability for both *P. collina* and its fish hosts (Guillera-Arroita et al. 2015).

Priority Conservation Sites

The priority conservation sites were located in the central and eastern portion of the James River Basin. Some of the locations represent contiguous stretches of streams but most are broken up by smaller sections of less suitable habitat. This level of connectivity implies a relative lack of dispersal opportunities for *P. collina* on a broad scale. While the fact that *P. collina* is a host generalist gives it a greater opportunity for dispersal into new habitats (Douda et al. 2011), its ability to reach those locations might be limited by the distance and relative habitat quality between patches. The host fish and *P. collina* may not traverse the less suitable areas therefore limiting long distance dispersal. Most of the priority conservation locations are not well connected so the ability of *P. collina* to disperse unassisted to new habitat is likely limited.

This model can be used to direct future habitat protection efforts. Twenty-two waterways were identified as the highest priority conservation sites based both on *P. collina* + fish scores and *P. collina* scores alone, implying those locations are of utmost priority. These locations and their surrounding watersheds should be targeted for conservation and restoration efforts to maintain and improve habitat conditions for both the mussels and their host fish. Restoration efforts can include: reducing runoff, removal of dams and impoundments, and planting of riparian areas (Roni et al. 2001).

Future Directions

The results of these models can be utilized for additional studies. The habitat suitability values estimated from these models can be used to identify the least cost path between patches of suitable habitat which can help identify potential dispersal routes for both the fish and *P. collina* (McRae et al. 2008; Beier et al. 2011). As organisms move through the landscape (or "riverscape" in this case), there is a certain "cost" associated with traveling through areas of lower habitat suitability. Organisms should choose to disperse along the path that accrues the least cost. If we can identify the pathways between habitat patches that would accrue the least cost, we can prioritize those paths for habitat protection and restoration efforts. Once identified, the least cost path can also be combined with population genetic data collected from the extant mussel populations. Together, this information can be used to determine whether the habitat connectivity between locations is affecting the level of gene flow across the watershed (McCallum et al. 2014).

These models can also be projected into the future under different emission scenarios to predict future climate change effects on both the fish hosts and *P. collina* (Fischer et al. 2011; McCallum et al. 2014). Different environmental variables were identified as the best predictors of habitat suitability for all species modeled. Because of this, it is likely that future habitat alterations due to climate change and anthropogenic effects could affect each species differently. If the fish hosts are driven to new habitat to escape these effects, their distributions may no longer have the same levels of overlap with that of *P. collina*. This could result in decreased connectivity between the mussel populations, thereby increasing the genetic divergence between and levels of inbreeding within the populations, furthering their decline (Li *et al*. 2016). Understanding how future changes may affect both the fish hosts and *P. collina* will be an important factor in ensuring their long-term survival.

Finally, these models can also be used to direct future sampling efforts. The areas of highest habitat suitability for *P. collina* identified by these models should be the target of traditional surveys and non-invasive environmental DNA sampling techniques in an effort to discover undocumented populations. Given the patchy distribution of the *P. collina* across the James River Basin, this approach will allow conservation managers to prioritize resources to save time and money. This method will also allow for ground-truthing the model predictions. Modeling should be a dynamic process where as more populations are found, the model can be re-run to include those new locations thereby refining our understanding of their habitat requirements and potential distribution.

Overall, these models provide us with crucial insights into the habitat requirements and potential distributions of *P. collina* and its fish hosts, allowing us to prioritize resources and make more informed conservation decisions to ensure the long-term survival of these important species.

Tables

Table 2.1: Results of the MaxEnt models created for *Parvaspina collina* and its 8 fish hosts. Common names for each species are listed along with the total number of presence points used to create each species distribution model. Model results include the Regularized Training Gain (RTG: a measure of the model's ability to discriminate between locations with known occurrences from random background points) and the Area Under the Curve (AUC: an indication that the MaxEnt raw scores are higher at locations with known species occurrences than at random background points). The mean habitat score (MaxEnt raw score) for each species is also reported.

			Model Results		
Species	**Common name**	**$N_{Presence}$**	**RTG**	**AUC**	**Mean Habitat Score**
Campostoma anomalum	Central stoneroller	255	0.692	0.82	4.5e-4
Chrosomus oreas	Mountain redbelly dace	460	0.495	0.784	2.9e-4
Clinostomus funduloides	Rosyside dace	423	0.369	0.751	2.8e-4
Cyprinella analostana	Satinfin shiner	122	0.816	0.863	2.7e-4
Lythrurus ardens	Rosefin shiner	78	0.689	0.842	3.1e-4
Nocomis leptocephalus	Bluehead chub	671	0.27	0.715	2.7e-4
Notropis procne	Swallowtail shiner	177	0.707	0.831	2.6e-4
Rhinichthys atratulus	Eastern blacknose dace	260	0.382	0.753	2.6e-4
Parvaspina collina	James spinymussel	79	1.911	0.957	2.7e-4

Table 2.2: Percent contribution of the environmental variables to the MaxEnt models for *Parvaspina collina* and its 8 fish hosts (common names are listed below the scientific names). The variable name as it appeared in the original dataset, along with a description of the variable, is provided for each species.

Species	Variable Name	Percent Contribution	Description
Campostoma anomalum	ELEVCAT	57.7	Mean catchment elevation (m)
Central stoneroller	AVGWETINDXCAT	19.6	Mean Composite Topographic Index (CTI)[Wetness Index] within catchment
	PCTWDWET2006CAT	9.3	% of catchment area classified as woody wetland land cover (NLCD 2006 class 90)
	NWS	8.8	Mean % of lithological nitrogen (N) content in surface or near surface geology within watershed
	prG_BMMI	4.7	Predicted probability that a stream segment is in good biological condition based on a random forest model of the NRSA benthic invertebrate multimetric index (BMMI)
Chrosomus oreas	RDCRSCAT	35.4	Density of roads-stream intersections within catchment (crossings/square km)
Mountain redbelly dace	D_NETEMPCL	30.6	Temperature classification
	RCKDEPCAT	15	Mean depth (cm) to bedrock of soils (STATSGO) within catchment
	PCTIMP2006CATRP100	10.8	Mean imperviousness of anthropogenic surfaces (NLCD 2006) within catchment and within a 100-m buffer of NHD stream lines
	PCTCOLLUVSEDWS	8.3	% of watershed area classified as lithology type: colluvial sediment
Clinostomus funduloides	RDCRSSLPWTDCAT	32.5	Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines) multiplied by NHDPlusV21 slope within catchment
Rosyside dace	RCKDEPCAT	31.6	Mean depth (cm) to bedrock of soils (STATSGO) within catchment
	PCTMXFST2011CAT	11.3	% of catchment area classified as mixed deciduous/evergreen forest land cover (NLCD 2011 class 43)
	PCTGRS2011CAT	9.5	% of catchment area classified as grassland/herbaceous land cover (NLCD 2011 class 71)
	PCTCROP2006CAT	8.3	% of catchment area classified as crop land use (NLCD 2006 class 82)
	PCTDECID2011CATRP100	6.9	% of catchment area classified as deciduous forest land cover (NLCD 2011 class 41) within a 100-m buffer of NHD streams

Table 2.2 (continued)

Species	Variable Name	Percent Contribution	Description
Cyprinella analostana Satinfin shiner	PCTSILICICWS	25	% of watershed area classified as lithology type: silicic residual material
	PCTCOLLUVSEDCAT	22.5	% of catchment area classified as lithology type: colluvial sediment
	RDCRSSLPWTDCAT	20.3	Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines) multiplied by NHDPlusV21 slope within catchment (crossings*slope/square km)
	OMCAT	14.6	Mean organic matter content (% by weight) of soils (STATSGO) within catchment
	PCTIMP2006CATRP100	9.4	Mean imperviousness of anthropogenic surfaces (NLCD 2006) within catchment and within a 100-m buffer of NHD stream lines
	BFIWS	8.2	Base Flow Index within the watershed
Lythrurus ardens Rosefin shiner	PCTBL2011WS	21.6	% of watershed area classified as barren land cover (NLCD 2011 class 31)
	PCTSILICICCAT	17.3	% of catchment area classified as lithology type: silicic residual material
	KFFACTWS	17	Mean soil erodibility (Kf) factor (unitless) of soils within watershed. The Kf factor is used in the Universal Soil Loss Equation (USLE) and represents a relative index of susceptibility of bare, cultivated soil to particle detachment and transport by rainfall.
	PCTGRS2011CAT	16.6	% of catchment area classified as grassland/herbaceous land cover (NLCD 2011 class 71)
	D_NETEMPCL	16.3	Temperature classification
Nocomis leptocephalus Bluehead chub	TMIN8110WS	11.1	PRISM climate data - 30-year normal minimum temperature (C°): Annual period: 1981-2010 within the watershed
	COMPSTRGTHCAT	16.2	Mean lithological uniaxial compressive strength (megaPascals) content in surface or near surface geology within catchment
	PCTCONIF2011CATRP100	15.7	% of catchment area classified as evergreen forest land cover (NLCD 2011 class 42) within a 100-m buffer of NHD streams
	RCKDEPCAT	14.5	Mean depth (cm) to bedrock of soils (STATSGO) within catchment
	MAST_2014	10.3	Predicted mean annual stream temperature (Jan-Dec) for year 2014

Table 2.2 (continued):

Species	Variable Name	Percent Contribution	Description
Notropis procne	PCTFRSTLOSS2013WSRP100	27	% Forest cover loss (Tree canopy cover change) for 2013 within watershed and within 100-m buffer of NHD stream lines
Swallowtail shiner	PCTURBHI2011WS	18.4	% of watershed area classified as developed, high-intensity land use (NLCD 2011 class 24)
	PCTURBOP2011CAT	15.9	% of catchment area classified as developed, open space land use (NLCD 2011 class 21)
	ELEVWS	15.5	Mean watershed elevation (m)
	PCTAG2006SLP10CAT	14.4	% of catchment area classified as agricultural land cover (NLCD 2006 classes 81-82) occurring on slopes ≥ 10%
	PCTCROP2006CAT	8.8	% of catchment area classified as crop land use (NLCD 2006 class 82)
Rhinichthys atratulus	RDCRSSLPWTDCAT	38	Density of roads-stream intersections (2010 Census Tiger Lines-NHD stream lines) multiplied by NHDPlusV21 slope within catchment (crossings*slope/square km)
Eastern blacknose dace	ELEVCAT	32.7	Mean catchment elevation (m)
	KFFACTCAT	12.4	Mean soil erodibility (Kf) factor (unitless) of soils within catchment. The Kf factor is used in the Universal Soil Loss Equation (USLE) and represents a relative index of susceptibility of bare, cultivated soil to particle detachment and transport by rainfall.
	PCTAG2006SLP10CAT	9.7	% of catchment area classified as agricultural land cover (NLCD 2006 classes 81-82) occurring on slopes ≥ 10%
	PCTCONIF2011CAT	7.2	% of catchment area classified as evergreen forest land cover (NLCD 2011 class 42)
Parvaspina collina	PCTAG2006SLP20WS	27.6	% of catchment area classified as agricultural land cover (NLCD 2006 classes 81-82) occurring on slopes ≥ 20%
James spinymussel	D_NESLPCL	21.5	Gradient Classification from NAHCS
	PCTCOLLUVSEDWS	19.2	% of watershed area classified as lithology type: colluvial sediment
	PCTFRSTLOSS2013WS	17.4	% Forest cover loss (Tree canopy cover change) for 2013 within watershed
	PCTOW2011WS	14.4	% of watershed area classified as open water land cover (NLCD 2011 class 11) (all areas of open water, generally with less than 25% cover of vegetation/land cover.)

Figures

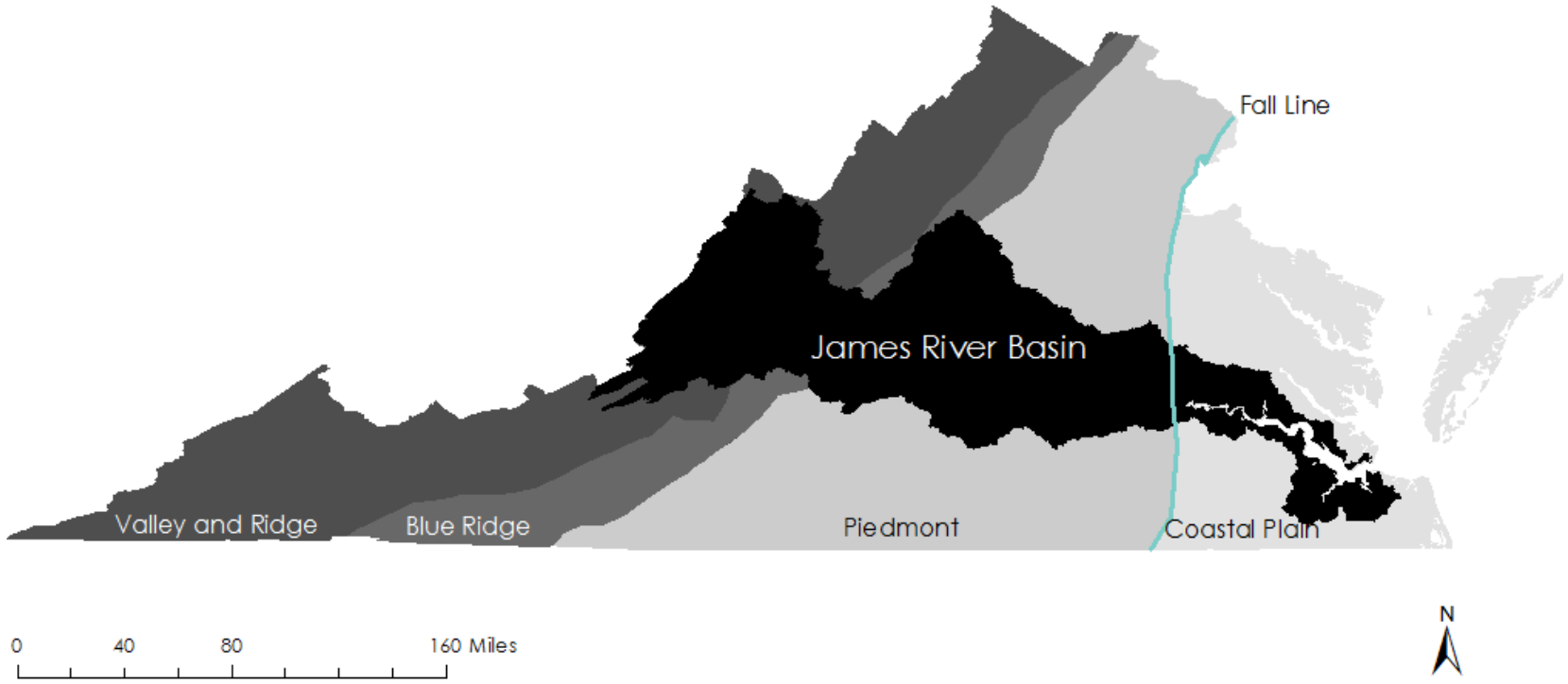

Figure 2.1: Overview of the Study Area. The James River Basin crosses four physiographic provinces across the central portion of the state.

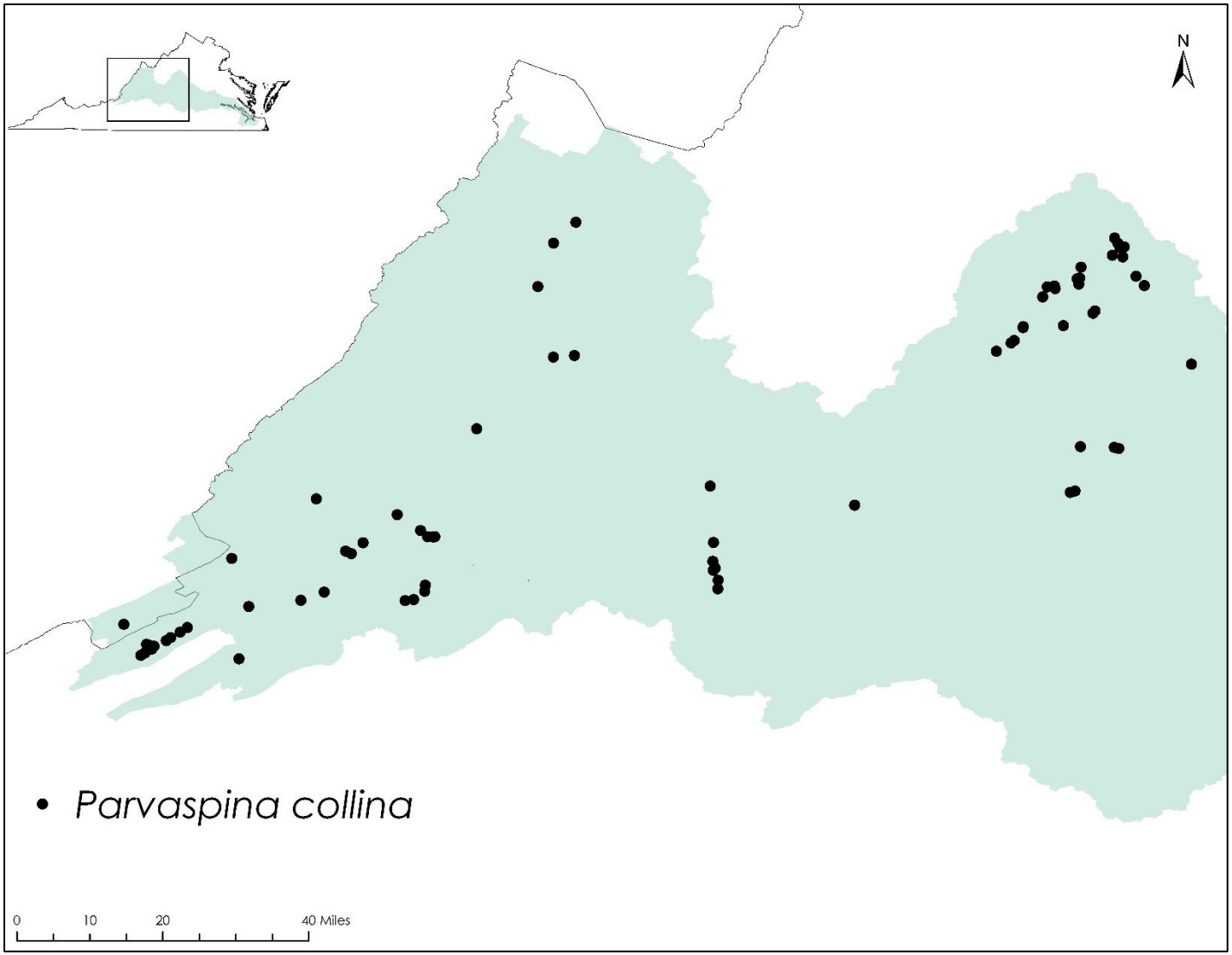

Figure 2.2: Distribution of *Parvaspina collina* across the James River Basin (shaded area) in Virginia. Points on the map represent species presence records collected since 1980 that were used to create the species distribution model (N = 78).

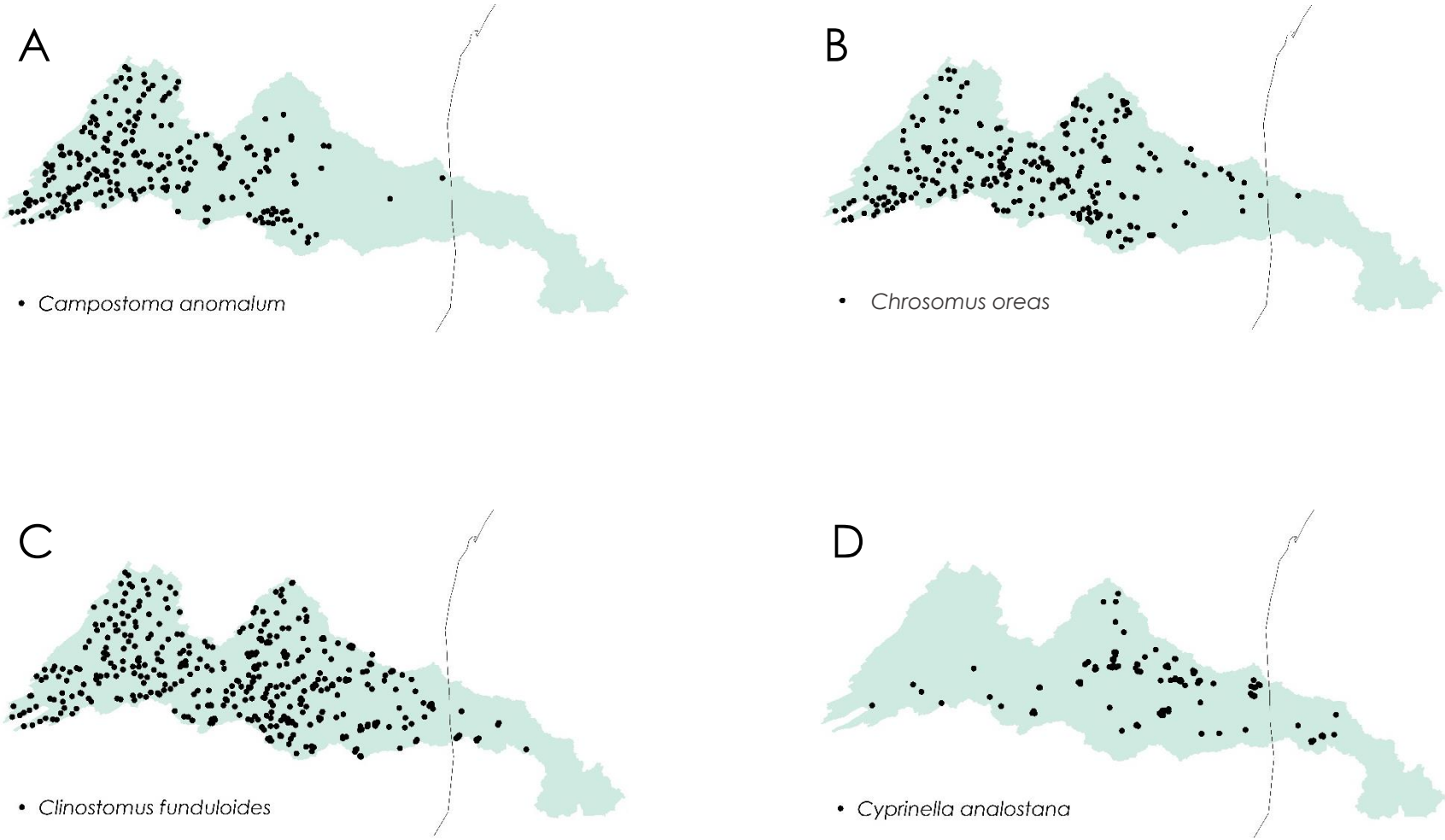

Figure 2.3 (A-D): Distributions of *P. collina* Fish Hosts in the James River Basin (shaded area) of Virginia. Points on map represent species presence records that were used to create the species distribution models. The Eastern distribution of several species appears limited by the Fall Line (dashed line).

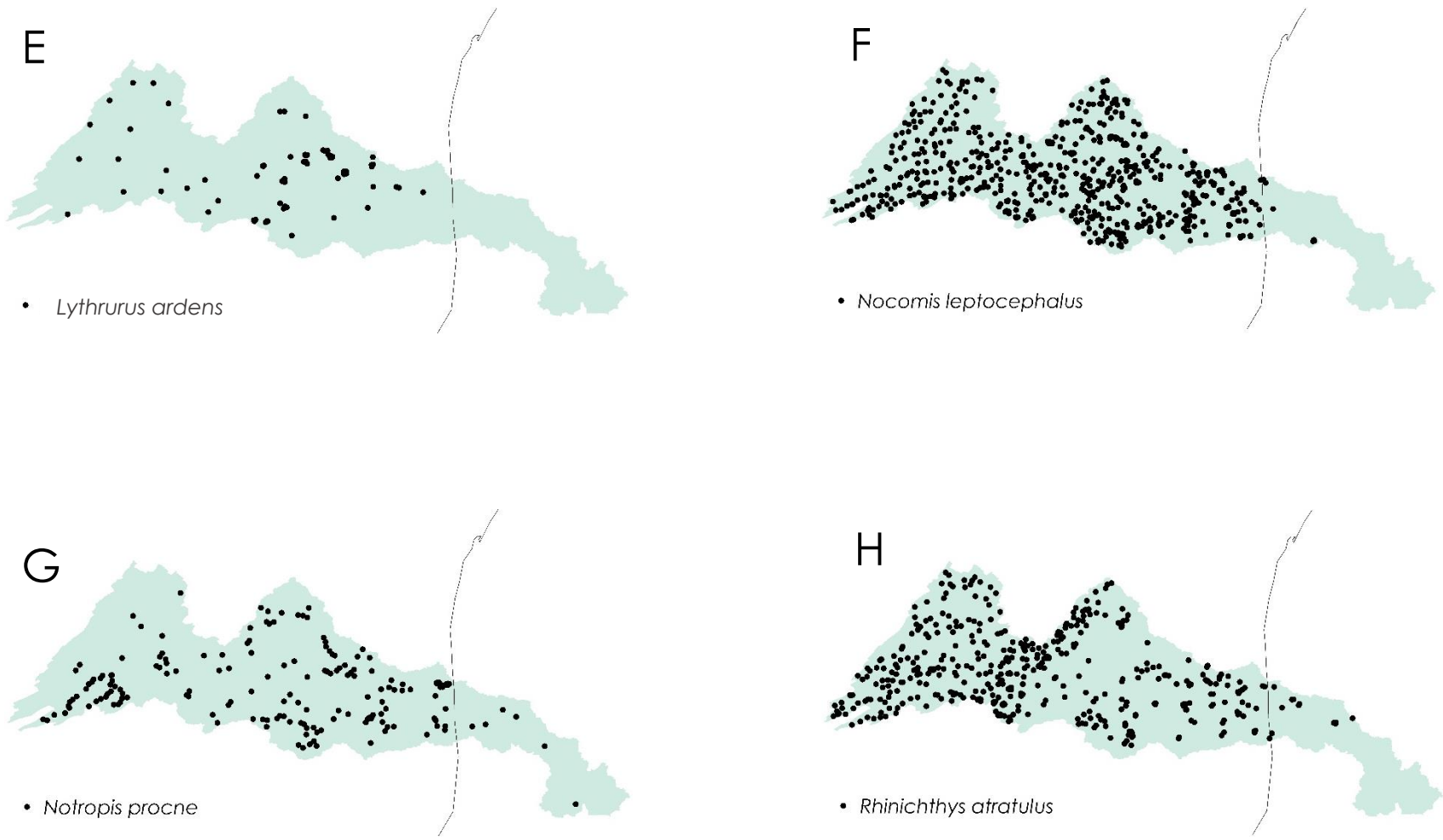

Figure 2.3 continued (E - H): Distributions of *P. collina* Fish Hosts in the James River Basin (shaded area) of Virginia. Points on map represent species presence records that were used to create the species distribution models. The Eastern distribution of several species appears limited by the Fall Line (dashed line).

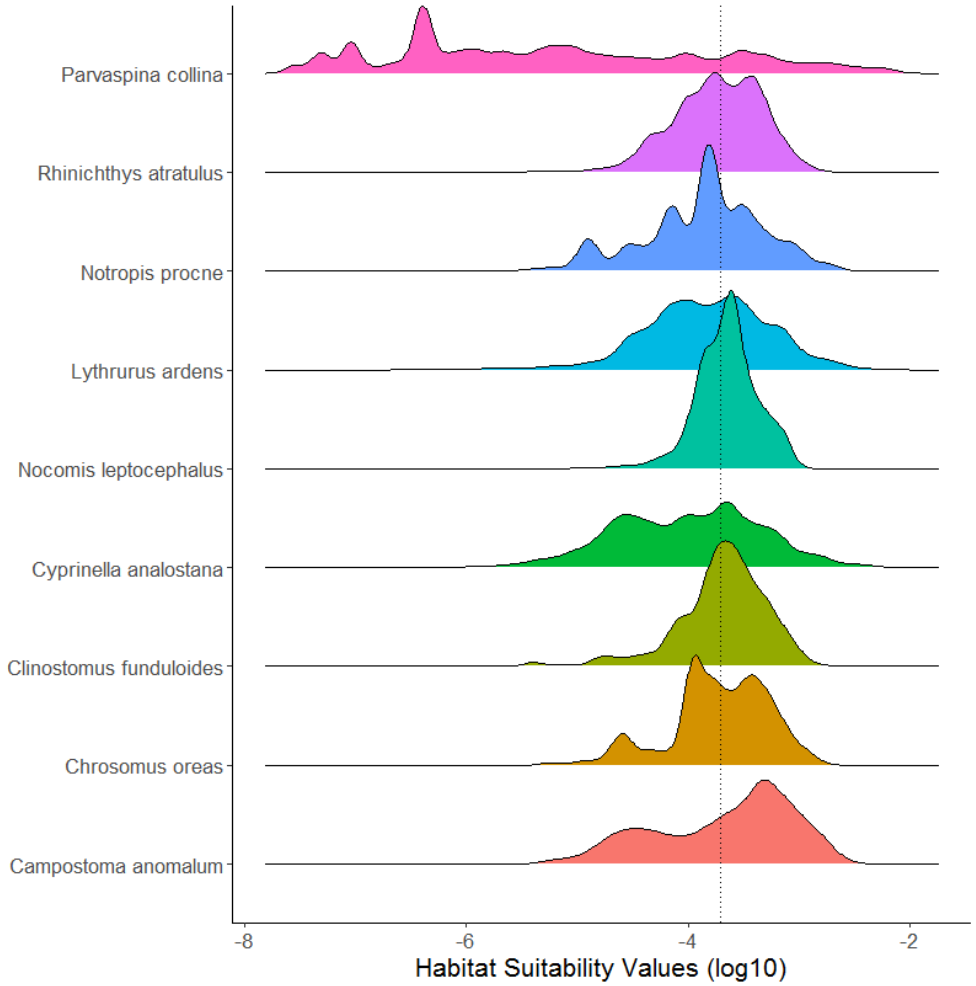

Figure 2.4: Comparison of the distribution of predicted habitat suitability values (log10 of MaxEnt raw scores) for each fish host species to that of *Parvaspina collina.* Dashed line shows the minimum predicted habitat score for a stream segment with known *P. collina* presence ($1.97e^{-4}$). Predicted habitat suitability across the James River Basin in Virginia was much lower overall for *P. collina* than any of the host fish species.

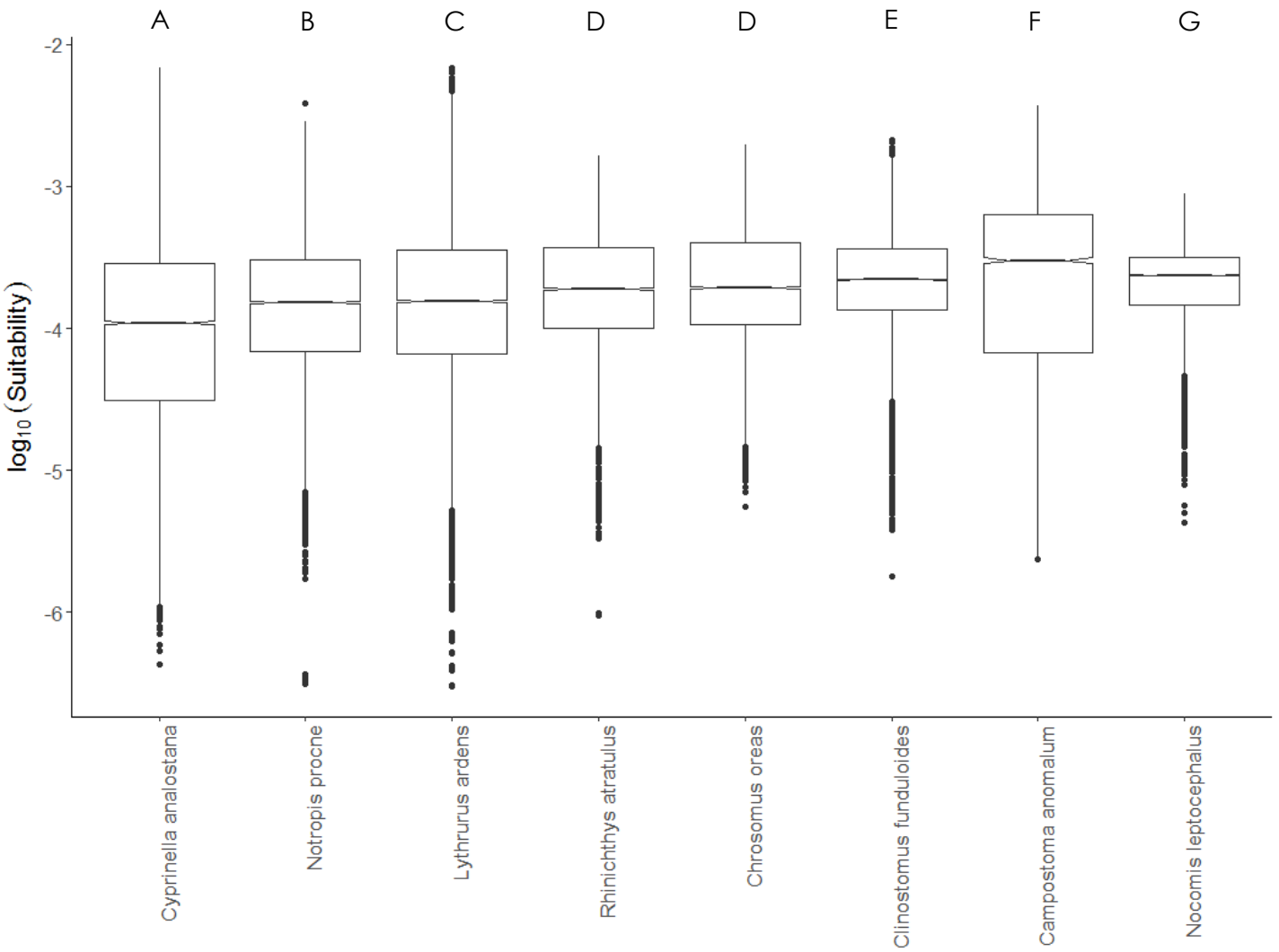

Figure 2.5: Suitability scores (log_{10} MaxEnt score) for each species of *P. collina* host fish ordered from smallest to largest mean values. Significant differences existed between pairwise comparisons of all species except *R. atratulus* and *C. oreas* (Group D).

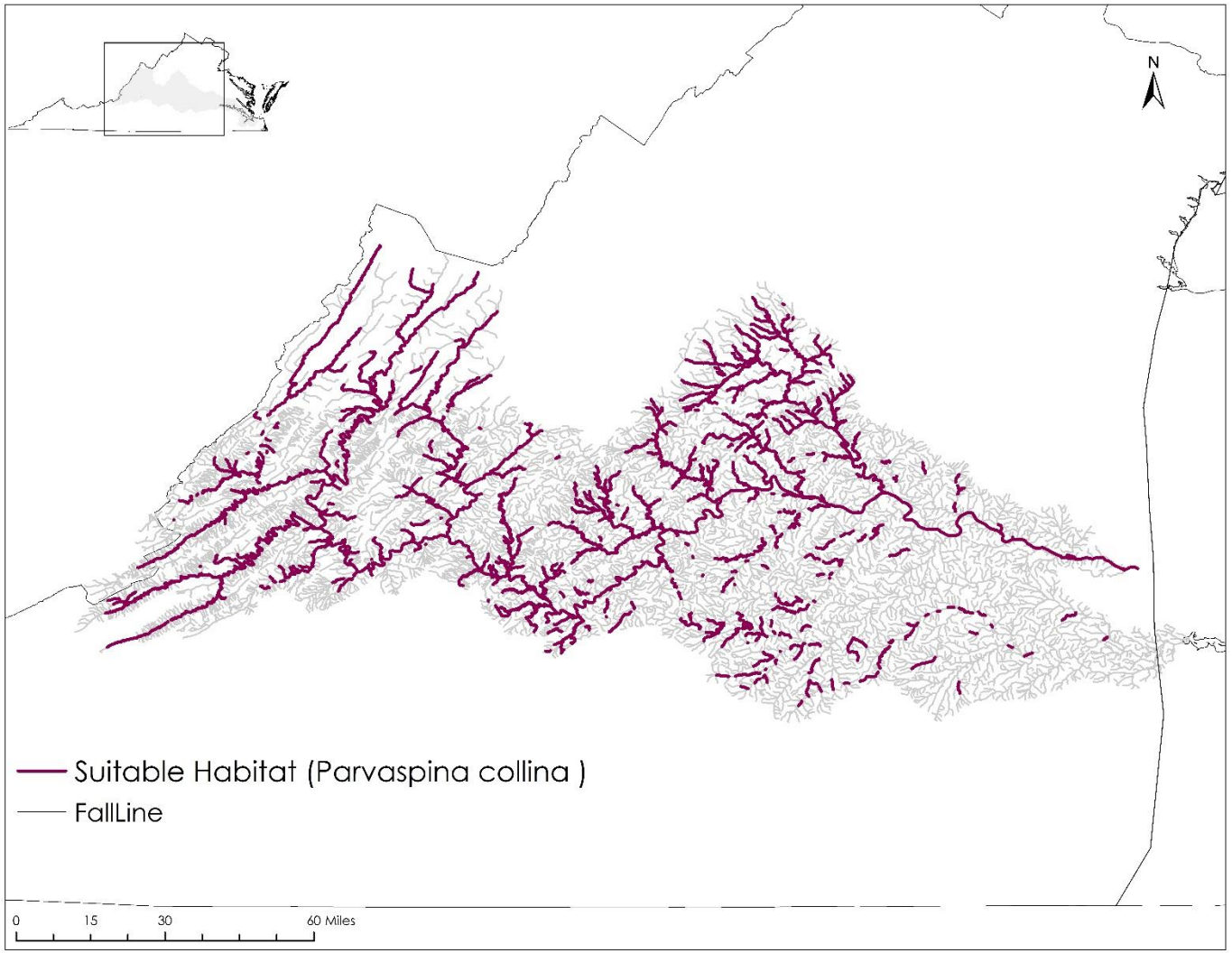

Figure 2.6: Stream segments within the James River Basin in Virginia identified as suitable habitat sites for *P. collina* based on the habitat suitability scores and minimum threshold value from the MaxEnt model. Approximately 19% of the total stream segments within the watershed were considered suitable habitat.

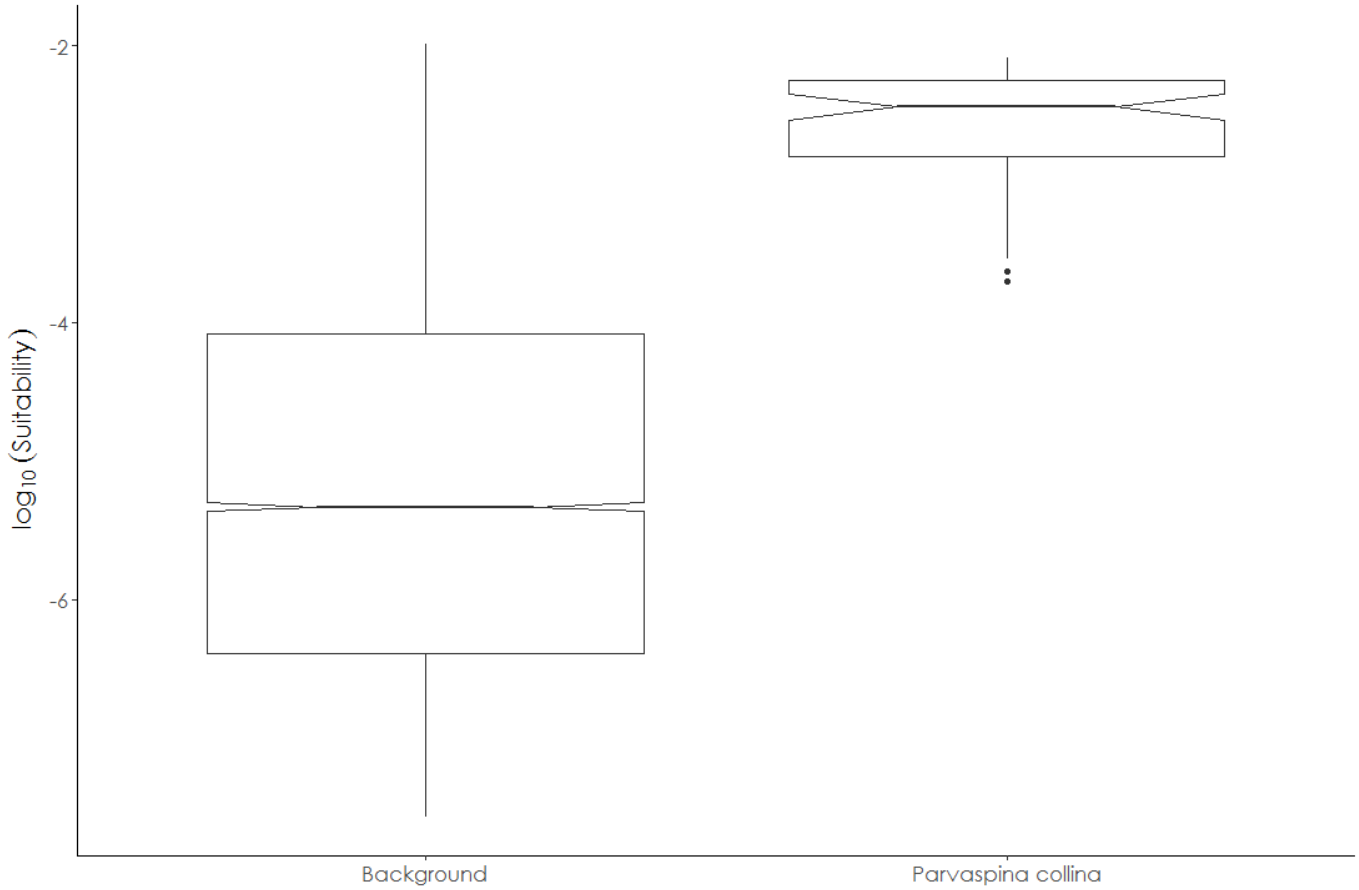

Figure 2.7: Comparison of mean habitat scores of streams with *Parvaspina collina* presence to those of the study area background. The mean habitat score among streams with known *Parvaspina collina* presence was significantly higher than the average predicted score for all background streams in the James River Basin.

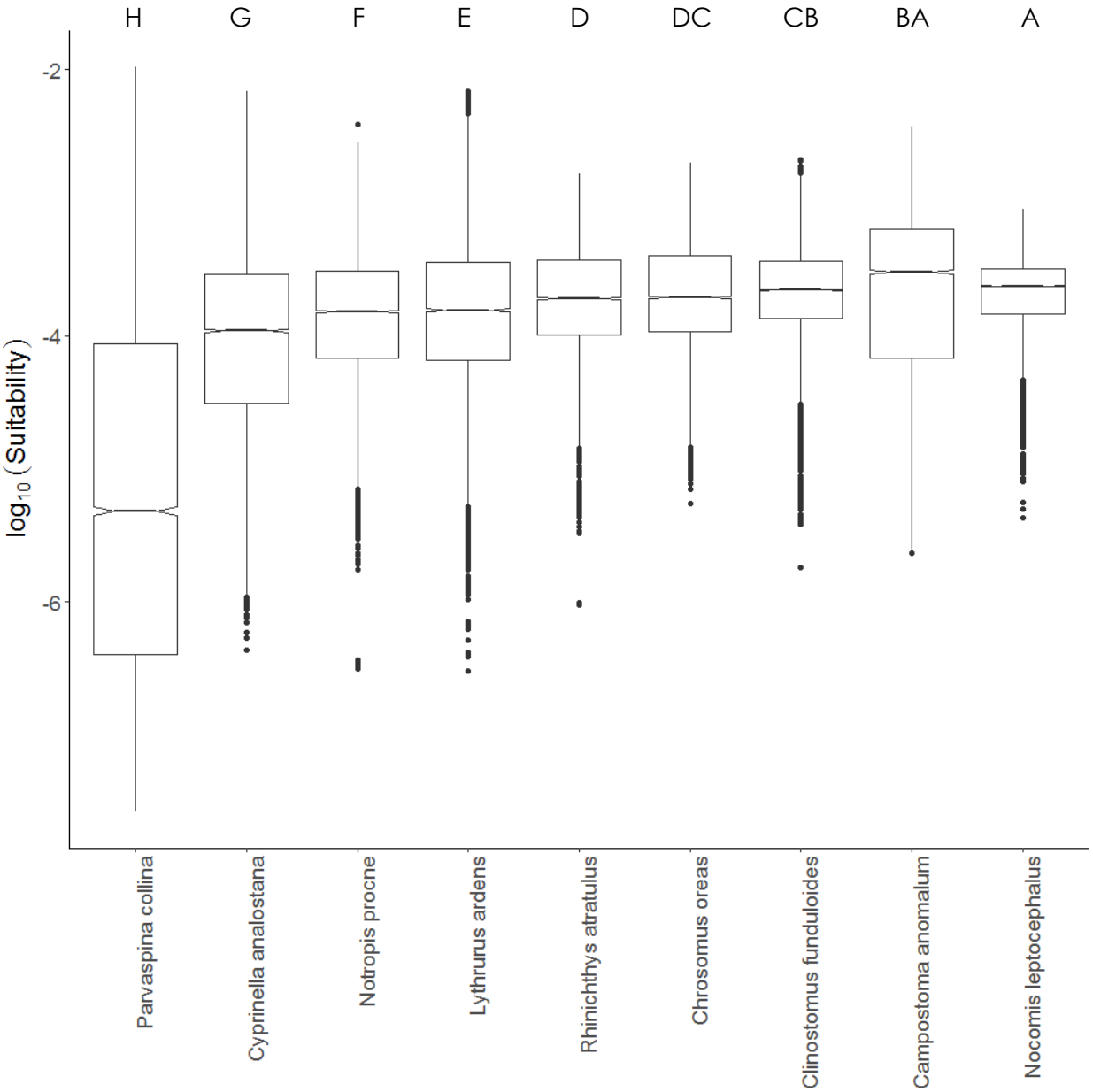

Figure 2.8: The predicted habitat scores (log_{10} MaxEnt score) for *P. collina* were significantly smaller than those predicted for the 8 species of host fish species ($F_{8, 102332} = 6402$, $P < 0.0001$). Groups were determined based on the results of a Tukey-HSD post hoc test for pairwise significance. Species are listed in ascending order from left to right based on the data means.

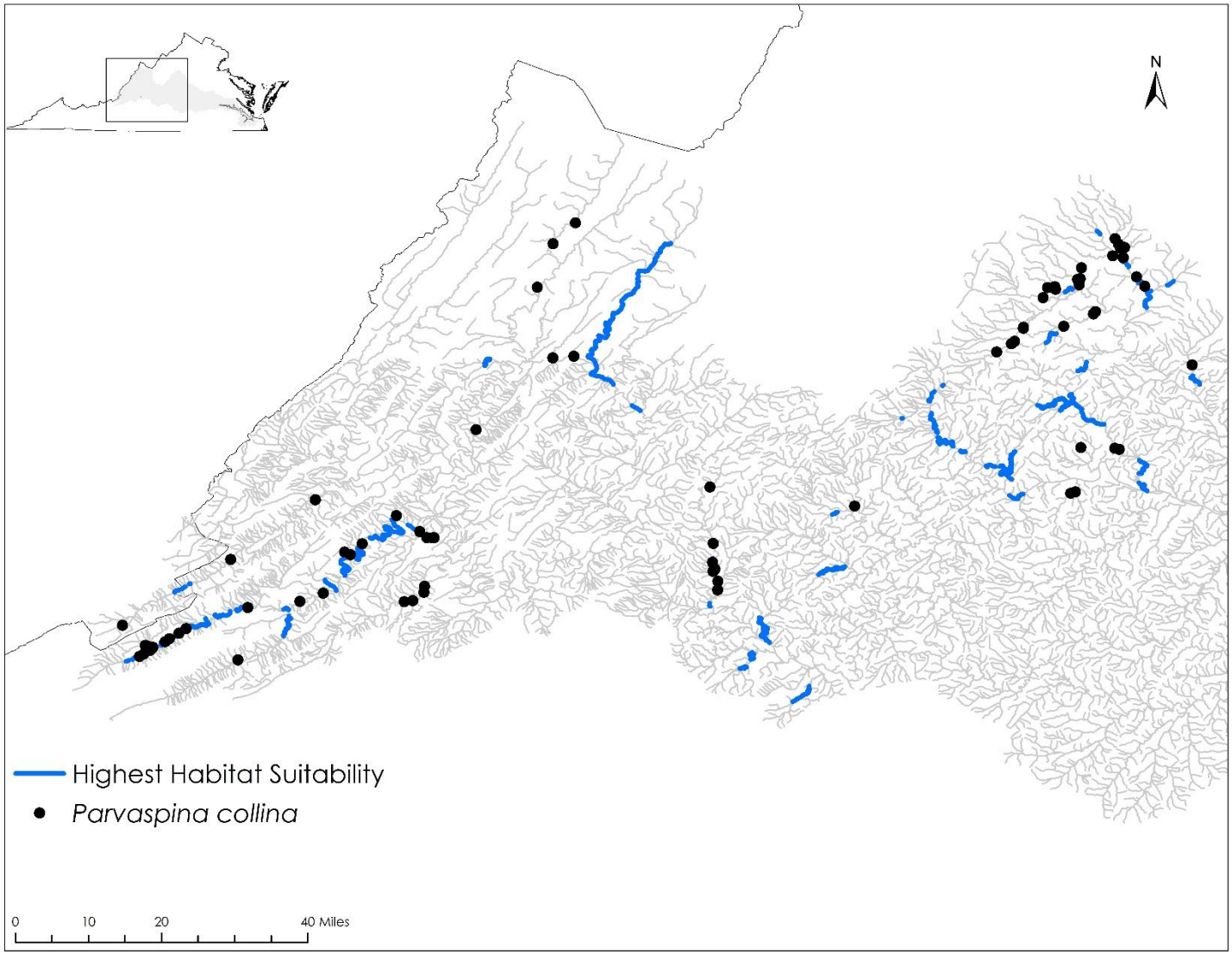

Figure 2.9: Map of locations with highest overall scores based on *P. collina* MaxEnt model that did not have documented species presence. These areas should be targeted for surveys to locate new populations.

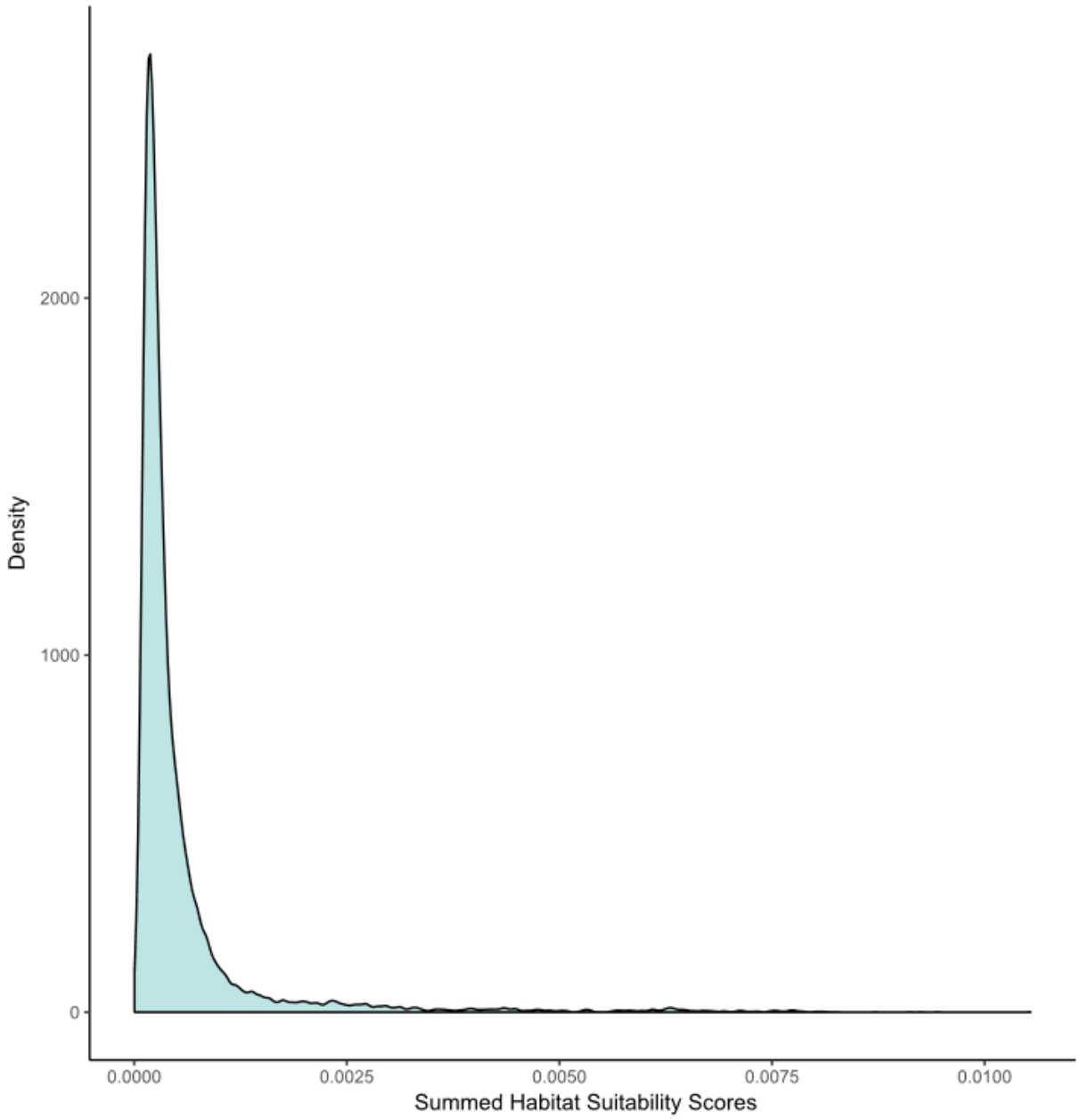

Figure 2.10: The distribution of the Fish Host + *P. collina* summed habitat suitability scores (MaxEnt raw values) across all stream segments in the James River Basin. These summed values were used to identify locations of highest conservation priority.

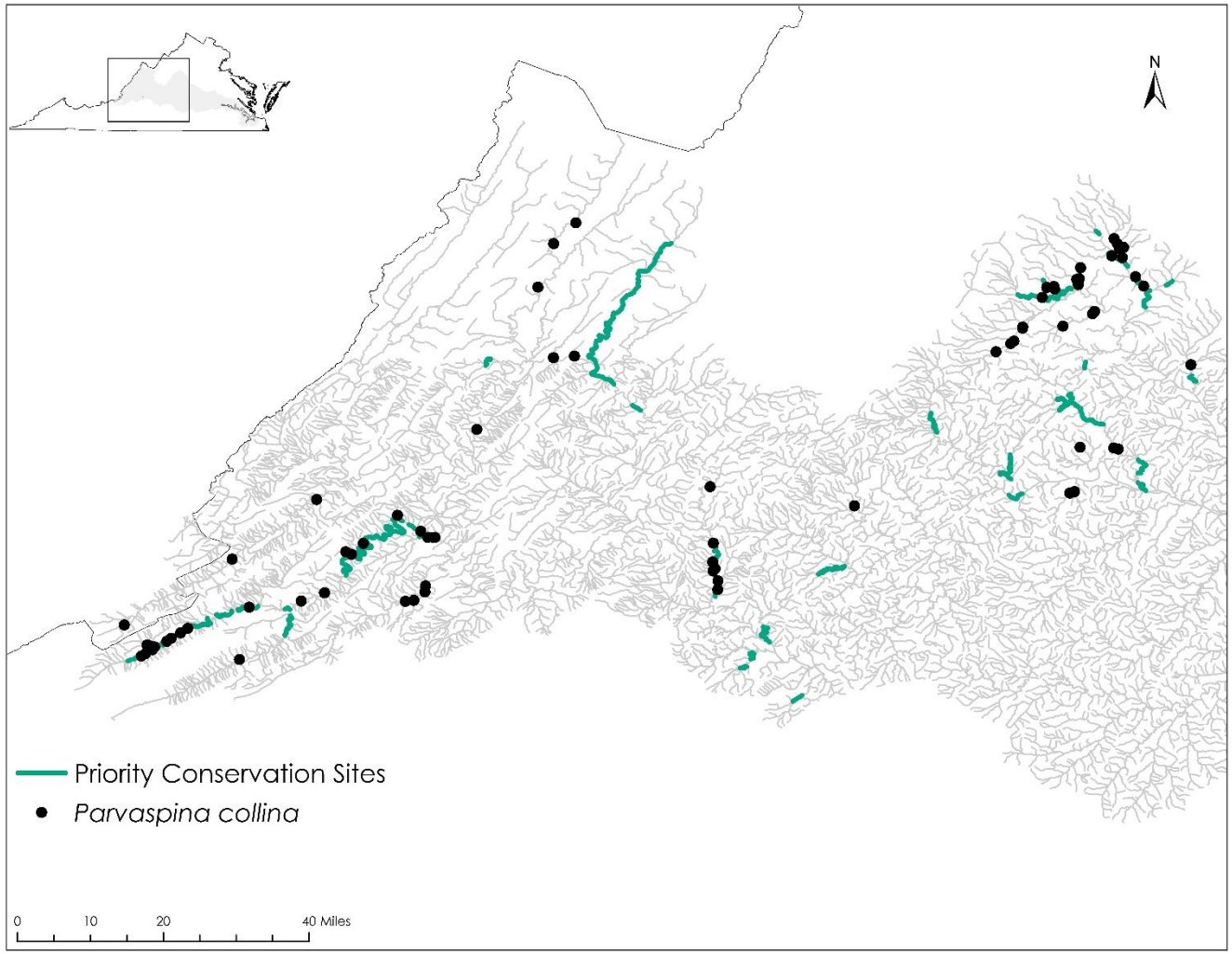

Figure 2.11: Priority Conservation Areas. Stream segments displayed in green had the highest overall habitat suitability for *P. collina* and its host fishes. These were identified as priority conservation areas that should be protected and targeted for restoration efforts. They represent 26 unique waterways.

References

Arbuckle, K. E., & Downing, J. A. (2002). Freshwater mussel abundance and species richness: GIS relationships with watershed land use and geology. *Canadian Journal of Fisheries and Aquatic Sciences*, *59*(2), 310–316. DOI: 10.1139/f02-006

Beier, P., Spencer, W., Baldwin, R. F., & McRae, B. H. (2011). Toward Best Practices for Developing Regional Connectivity Maps. *Conservation Biology*, *25*(5), 879–892. DOI: 10.1111/j.1523-1739.2011.01716.x

Box, J. B., & Mossa, J. (1999). Sediment , Land Use , and Freshwater Mussels : Prospects and Problems. *Journal of the North American Benthological Society*, *18*(1), 99–117.

Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid Range Shifts of Species Associated with High Levels of Climate Warming. *Science*, *333*, 1024–1027.

Clarke, A.H., & R.J. Neves. (1984). Status survey of the James River spinymussel, *Canthyria collina*, in the James River, Virginia. Final Report, U.S. Fish and Wildlife Service, Newton Corner, MA. 73pp.

Department of Environmental Quality (DEQ) 2015 Virginia State Water Resources Plan. Appendix B; 148-168

Devictor, V., Julliard, R., & Frederic, J. (2008). Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. *Oikos*, *117*, 507–514. DOI: 10.1111/j.2008.0030-1299.16215.x

Dormann, C. F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Lautenbach, S. (2013). Collinearity: A review of methods to deal with it and a simulation study evaluating their performance. *Ecography*, *36*(1), 27–46. DOI: 10.1111/j.1600-0587.2012.07348.x

Douda, K., VrtileK, M., Slavik, O., & Reichard, M. (2012). The role of host specificity in explaining the invasion success of the freshwater mussel Anodonta woodiana in Europe. *Biological Invasions*, *14*(1), 127–137. DOI: 10.1007/s10530-011-9989-7

Early, R., Anderson, B., & Thomas, C. D. (2008). Using habitat distribution models to evaluate large-scale landscape priorities for spatially dynamic species. *Journal of Applied Ecology*, *45*, 228–238. DOI: 10.1111/j.1365-2664.2007.01424.x

Elith, J., & Leathwick, J. R. (2009). Species Distribution Models : Ecological Explanation and Prediction Across Space and Time. *Annual Reviews of Ecology, Evolution, and Systematics*, *40*, 677–697. DOI: 10.1146/annurev.ecolsys.110308.120159

Erasmus, B. F. N., VanJaarsveld, A. S., Chown, S. L., Kshatriya, M., & Wessels, K. J. (2002). Vulnerability of South African animal taxa to climate change. *Global Change Biology*, *8*, 679–693.

Ferrer-sánchez, Y., & Rodríguez-estrella, R. (2016). How rare species conservation management can be strengthened with the use of ecological niche modelling : The case for endangered endemic Gundlach ' s Hawk and Cuban Black-Hawk. *Global Ecology and Conservation*, *5*, 88–99. DOI: 10.1016/j.gecco.2015.11.008

Fischer, D., Moeller, P., Thomas, S. M., Naucke, T. J., & Beierkuhnlein, C. (2011). Combining climatic projections and dispersal ability: A method for estimating the responses of sandfly vector species to climate change. *PLoS Neglected Tropical Diseases*, *5*(11), e1407. DOI: 10.1371/journal.pntd.0001407

Frimpong, E. A., Huang, J., and Liang, Y. (2016). IchthyMaps: a database of historical distributions of freshwater fishes of the United States. *Fisheries* 41, 590–599.

Gama, M., Crespo, D., Dolbeth, M., & Anastácio, P. (2016). Predicting global habitat suitability for Corbicula fluminea using species distribution models: The importance of different environmental datasets. *Ecological Modelling*, *319*, 163–169. DOI: 10.1016/j.ecolmodel.2015.06.001

Gibson, L.A., Wilson, B.A., Cahill, D.A., Hill, J. (2004). Spatial prediction of rufous bristlebird habitat in a coastal heathland: a GIS-based approach. *Journal of Applied Ecology* 41, 213–223.

Gilman, S. E., Urban, M. C., Tewksbury, J., Gilchrist, G. W., & Holt, R. D. (2010). A framework for community interactions under climate change. *Trends in Ecology & Evolution*, *25*(6), 325–331. DOI: 10.1016/j.tree.2010.03.002

Gormley, A. M., Forsyth, D. M., Griffioen, P., Lindeman, M., Ramsey, D. S. L., Scroggie, M. P., & Woodford, L. (2011). Using presence-only and presence – absence data to estimate the current and potential distributions of established invasive species. *Journal of Applied Ecology*, *48*, 25–34. DOI: 10.1111/j.1365-2664.2010.01911.x

Graham, C.H., Ferrier, S., Huettman, F., Moritz, C. & Peterson, A.T. (2004a). New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology and Evolution*, 19, 497–503

Guillera-Arroita, G. (2017). Modelling of species distributions, range dynamics and communities under imperfect detection: advances, challenges and opportunities. *Ecography*, *40*, 281–295. DOI: 10.1111/ecog.02445

Guillera-Arroita, G., Lahoz-Monfort, J. J., Elith, J., Gordon, A., Kujala, H., Lentini, P. E., … Wintle, B. A. (2015). Is my species distribution model fit for purpose ? Matching data and models to applications. *Global Ecology and Biogeography*, *24*, 276–292. DOI: 10.1111/geb.12268

Guisan, A., Lehmann, A., Ferrier, S., Austin, M., Overton, J. M. C. C., Aspinall, R., & Hastie, T. (2006). Making better biogeographical predictions of species ' distributions. *Journal of Applied Ecology*, *43*, 386–392. DOI: 10.1111/j.1365-2664.2006.01164.x

Guisan, A., & Thuiller, W. (2005). Predicting species distribution : offering more than simple habitat models. *Ecology Letters*, *8*, 993–1009. DOI: 10.1111/j.1461-0248.2005.00792.x

Guisan, A., Tingley, R., Baumgartner, J. B., Naujokaitis-Lewis, I., Sutcliffe, P. R., Tulloch, A. I. T., … Buckley, Y. M. (2013). Predicting species distributions for conservation decisions. *Ecology Letters*, *16*, 1424–1435. DOI: 10.1111/ele.12189

Hastie, T., & Fithian, W. (2013). Inference from presence-only data; the ongoing controversy. *Ecography*, *36*, 864–867. DOI: 10.1111/j.1600-0587.2013.00321.x

Heikkinen, R. K., Luoto, M., Virkkala, R., Pearson, R. G., & Körber, J.-H. (2007). Biotic interactions improve prediction of boreal bird distributions at macro-scales. *Global Ecology and Biogeography*, *16*, 754–763. DOI: 10.1111/j.1466-8238.2007.00345.x

Henley, W. F., Patterson, M. A., Neves, R. J., & Lemly A.D. (2000). Effects of Sedimentation and Turbidity on Lotic Food Webs: A Concise Review for Natural Resource Managers, *Reviews in Fisheries Science,* 8(2), 125-139, DOI: 10.1080/10641260091129198

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography*, *29*, 773–785.

Hill, R. A., Weber, M. H., Leibowitz, S. G., Olsen, A. R., & Thornbrugh, D. J. (2016). The Stream-Catchment (StreamCat) Dataset: A Database of Watershed Metrics for the Conterminous United States. *Journal of the American Water Resources Association*, *52*(1), 120–128. DOI: 10.1111/1752-1688.12372

Hove, M. (1990). Distribution and life history of the endangered James spinymussel, *Pleurobema collina* (Bivalvia: Unionidae). M.S. Thesis. Virginia Polytechnic Institute and State University, Blacksburg, Virginia. 113 pp.

Hove, M. C., & Neves, R. J. (1994). Life history of the endangered James spinymussel Pleurobema collina (Conrad, 1837) (Mollusca: Unionidae). *American Malacological Bulletin*.

Howard, C., Stephens, P. A., Pearce-Higgins, J. W., Gregory, R. D., & Willis, S. G. (2014). Improving species distribution models: The value of data on abundance. *Methods in Ecology and Evolution*, *5*, 506–513. DOI: 10.1111/2041-210X.12184

Jarvis, A., Williams, K., Williams, D., Guarino, L., Caballero, P. J., & Mottram, G. (2005). Use of GIS for optimizing a collecting mission for a rare wild pepper (*Capsicum flexuosum Sendtn*.) in Paraguay. *Genetic Resources and Crop Evolution*, *52*, 671–682. DOI: 10.1007/s10722-003-6020-x

Jenkins R. E. and N. M. Burkhead. (1994). *Freshwater Fishes of Virginia*. American Fisheries Society.

Kearney, M., & Porter, W. (2009). Mechanistic niche modelling : combining physiological and spatial data to predict species' ranges. *Ecology Letters*, *12*, 334–350. DOI: 10.1111/j.1461-0248.2008.01277.x

Koen, E.L., Bowman, J., Garroway, C.J., Mills, S.C., and Wilson, P.J. (2012). Landscape resistance and American marten gene flow. *Landscape Ecology*, 27, 29–43.

le Roux, P. C., Pellissier, L., Wisz, M. S., & Luoto, M. (2014). Incorporating dominant species as proxies for biotic interactions strengthens plant community models. *Journal of Ecology*, *102*, 767–775. DOI: 10.1111/1365-2745.12239

Leathwick, J. R., & Austin, M. P. (2001). Competitive Interactions between Tree Species in New Zealand 's Old-Growth Indigenous Forests. *Ecology*, *82*(9), 2560–2573.

Li, Y., Cooper, S. J. B., Lancaster, M. L., Packer, J. G., & Carthew, S. M. (2016). Comparative population genetic structure of the endangered southern brown bandicoot, Isoodon obesulus, in fragmented landscapes of Southern Australia. *PLoS ONE*, *11*(4), 1–21. DOI: 10.1371/journal.pone.0152850

Mackenzie, D. I. (2005). What Are the Issues with Presence-Absence Data for Wildlife Managers. *Journal of Wildlife Management*, *69*(3), 849–860.

McCallum, K. P. M., Guerin, G. R., Breed, M. F., & Lowe, A. J. (2014). Combining population genetics , species distribution modelling and field assessments to understand a species vulnerability to climate change. *Austral Ecology*, *39*, 17–28. DOI: 10.1111/aec.12041

McGarvey, D. J., Menon, M., Woods, T., Tassone, S., Reese, J., Vergamini, M., & Kellogg, E. (2018). On the use of climate covariates in aquatic species distribution models: are we at risk of throwing out the baby with the bath water? *Ecography*, *41*, 695–712. DOI: 10.1111/ecog.03134

McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., (2012). "NHDPlus Version 2: User Guide"

McRae, B. H., Dickson, B. G., Keitt, T. H., & Shah, V. B. (2008). Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology*, *89*(10), 2712–2724.

Meier, E. S., Kienast, F., Pearman, P. B., Svenning, J., Thuiller, W., Arau, M. B., … Zimmermann, N. E. (2010). Biotic and abiotic variables show little redundancy in explaining tree species distributions. *Ecography*, *33*, 1038–1048. DOI: 10.1111/j.1600-0587.2010.06229.x

Morato, R. G., Ferraz, K. M. P. M. dB., de Paula, R.C., Campos, C. (2014). Identification of Priority Conservation Areas and Potential Corridors for Jaguars in the Caatinga Biome, Brazil. *PLoS ONE,* 9(4), e92950. DOI:10.1371/journal.pone.0092950

Olivero, A. P., M. G. Anderson. (2008). Northeast Aquatic Habitat Classification System. The Nature Conservancy Eastern Regional Office. Boston, MA. 88pp.

Pearson, R. G., Raxworthy, C. J., Nakamura, M., & Peterson, A. T. (2007). Predicting species distributions from small numbers of occurrence records: A test case using cryptic geckos in Madagascar. *Journal of Biogeography*, *34*, 102–117. DOI: 10.1111/j.1365-2699.2006.01594.x

Petty, M. A., Neves, R. J., Hallerman, E. M., & Angermeier, P. (2005). Distribution, Genetic Characterization, and Life History of the James spinymussel, *Pleurobema collina* (Bivalvia: Unionidae), in Virginia and North Carolina. Virginia Polytechnic Institute. http://hdl.handle.net/10919/31682

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*, 231–259. DOI: 10.1016/j.ecolmodel.2005.03.026

Phillips, S. J. and Dudík, M. (2008). Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography*, 31, 161–175.

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A Maximum Entropy Approach to Species Distribution Modeling. In *Proceedings of the Twenty-First International Conference on Machine Learning* (pp. 655–662).

Pollock, L. J., Tingley, R., Morris, W. K., Golding, N., O'Hara, R. B., Parris, K. M., … Mccarthy, M. A. (2014). Understanding co-occurrence by modelling species simultaneously with a Joint Species Distribution Model (JSDM). *Methods in Ecology and Evolution*, *5*(5), 397–406. DOI: 10.1111/2041-210X.12180

Poor, E. E., Loucks, C., Jakes, A., & Urban, D. L. (2012). Comparing Habitat Suitability and Connectivity Modeling Methods for Conserving Pronghorn Migrations. *PLoS ONE*, *7*(11), e49390. http://doi.org/10.1371/journal.pone.0049390

Razgour, O., Rebelo, H., Puechmaille, S. J., Juste, J., Ibanez, C., Kiefer, A., … Jones, G. (2014). Scale-dependent effects of landscape variables on gene flow and population structure in bats. *Diversity and Distributions*, *20*, 1173–1185. DOI: 10.1111/ddi.12200

Roni, P., Beechie, T.J., Bilby, R.E., Leonetti, F.E., Pollock M.M., and Pess, G.R. (2001). A Review of Stream Resotration Techniques and a Hierarchial Strategy for Prioritizing Resortation in Pacific Northwest Watersheds. *North American Journal of Fisheries Management*, 22(1), 1-20.

Tournant, P., Afonso, E., Roué, S., Giraudoux, P., & Foltête, J. (2013). Evaluating the effect of habitat connectivity on the distribution of lesser horseshoe bat maternity roosts using landscape graphs. *Biological Conservation*, *164*, 39–49. DOI: 10.1016/j.biocon.2013.04.013

Tsoar, A., Allouche, O., Steinitz, O., Rotem, D., & Kadmon, R. (2007). A comparative evaluation of presence- only methods for modelling species distribution. *Diversity and Distributions*, *13*, 397–405. DOI: 10.1111/j.1472-4642.2007.00346.x

USFWS. (1990). James spinymussel *(Pleurobema collina)* Recovery Plan. <https://efotg.sc.egov.usda.gov/references/public/WV/James_spinymussel_Recov_Plan.pdf>

Van Der Putten, W., Der, W. H. Van, Macel, M., & Visser, M. E. (2010). Predicting species distribution and abundance responses to climate change : why it is essential to include biotic interactions across trophic levels. *Philisophical Transactions of the Royal Society B*, *35*(1549), 2025–2034. DOI: 10.1098/rstb.2010.0037

Vaughn, C. C., & Taylor, C. M. (2000). Macroecology of a host-parasite relationship. *Ecography*, *23*, 11–20.

Wang, Y., Yang, K., Bridgman, C. L., & Lin, L.-K. (2008). Habitat suitability modelling to correlate gene flow with landscape connectivity. *Landscape Ecology*, *23*, 989–1000. DOI: 10.1007/s10980-008-9262-3

Watters, G. T. (1992). Unionids , Fishes , and the Species-Area Curve. *Journal of Biogeography*, *19*(5), 481–490.

Watters, G. T. (1999). Freshwater mussels and water quality : A review of the effects of hydrologic and instream habitat alterations. In *Proceedings of the First Freshwater Mollusk Conservation Society Symposium* (pp. 261–274).

Wharton, T. N., & Kriticos, D. J. (2004). The fundamental and realized niche of the Monterey Pine aphid , Essigella californica (Essig) (Hemiptera : Aphididae): implications for managing softwood plantations in Australia. *Diversity and Distributions*, *10*, 253–262.

Wilson, C. D., Roberts, D., & Reid, N. (2011). Applying species distribution modelling to identify areas of high conservation value for endangered species: A case study using Margaritifera margaritifera (L.). *Biological Conservation*, *144*, 821–829. DOI: 10.1016/j.biocon.2010.11.014

Appendix 1: Environmental DNA Resources

Table A -1: Tested ND1 (*NADH dehydrogenase-1*) sequences for both forward and reverse primers selected for full evaluation. The primers pair ND1_pcbr1 (denoted in bold) were found to be the most effective at identifying *Parvaspina collina.*

Primer ID	Direction	Primer Sequence (5' - 3')
ND1Fq_1	Forward	CAATTCGATCAATTAATGCC
ND1Rq_1	Reverse	TTTCTGCTAAAATAACAAC
ND1Fq_2	Forward	GTTGTTATTTTAGCAGAAA
ND1Rq_2	Reverse	CAGAGACTAATTCTGACT
ND1Fq_3	Forward	TTAGCAGACGCTCTAAAGC
ND1Rq_3	Reverse	AATAAACGGTAAATAGTTCG
ND1Fq_4	Forward	CGAACTATTTACCGTTTATT
ND1Rq_4	Reverse	TGTATAGACGGTTAAAGAAG
ND1Fq_d1	Forward	AGCCATAGCCCARACCATCT
ND1Rq_d1	Reverse	AATGRCTAATGGTGCGCMGA
ND1Fq_d2	Forward	CGAGCCATAGCCCWRACCA
ND1Rq_d2	Reverse	ATGRCTAATGGTGCGCMGA
ND1Fq_d3	Forward	GAGCCATAGCCCWRACCATCT
ND1Rq_d3	Reverse	TGRCTAATGGTGCGCMGAG
ND1Fq_d4	Forward	CTCKGCGCACCATTAGYCA
ND1Rq_d4	Reverse	TTCGATGTTGAACMCAGAGAC
ND1Fq_d5	Forward	CKGCGCACCATTAGYCATTA
ND1Rq_d5	Reverse	TTCGATGTTGAACMCAGAGA
ND1Fq_d6	Forward	CKGCGCACCATTAGYCATTA
ND1Rq_d6	Reverse	TTCGATGTTGAAYCMCAGAGA
ND1_Gen2_F	Forward	ACCCTTCTAGAACGCAAAGC
ND1_Gen2_R	Reverse	TCTGCTAATGGTTGTGGGATTC
ND1_Gen3_F	Forward	TGCGCACCATTAGCCATTA
ND1_Gen3_R	Reverse	TTCGATGTTGAACCCAGAGAC
ND1_Gen4_F	Forward	CCATTTGACTTTGCTGAAGGAG
ND1_Gen4_R	Reverse	GCCATGAATAGGAAGGCAAAG
ND1_pcbr1	**Forward**	**GCGTAGCATTCTTTACCCTTCT**
ND1_pcbr2	**Reverse**	**GAGCGTCTGCTAATGGTTGT**

Appendix 2: Species Distribution Modeling Resources

Table B.1: Names and lengths of the streams designated as Priority conservation Sites based on overall habitat suitability for *Parvaspina collina* and its host fishes.

Stream Name	Total Length (km)
Beaver Creek	2.206
Biscuit Run	1.484
Blackwater Creek	5.359
Calfpasture River	47.039
Craig Creek	48.994
Dicks Creek	0.969
Eppes Creek	2.846
Green Creek	6.559
Hardware River	17.57
Harris Creek	7.242
Johns Creek	34.133
Maury River	3.434
Mechunk Creek	2.35
Mill Creek	2.692
Moormans River	14.139
North Fork Hardware River	2.648
North Fork Rivanna River	11.135
Parker Branch	0.961
Patterson Creek	2
Pedlar River	10.71
Preddy Creek	1.467
Rockfish River	11.826
Rocky Creek	3.591
Rutledge Creek	8.6
South Fork Hardware River	4.605
Wards Creek	2.827

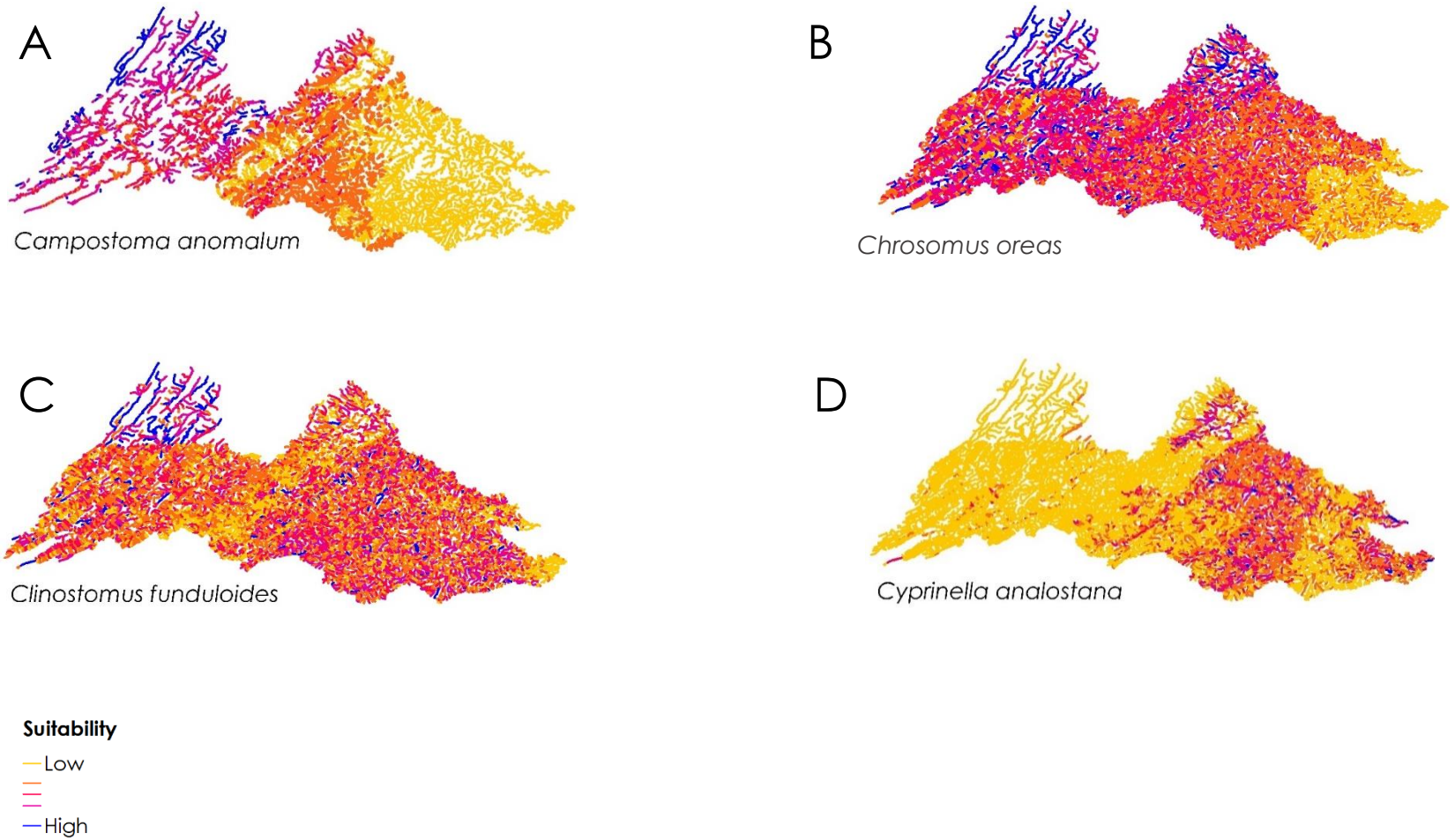

Figure B1 (A:D): MaxEnt model predictions for habitat suitability for *P. collina* fish hosts in the James River Basin of Virginia.

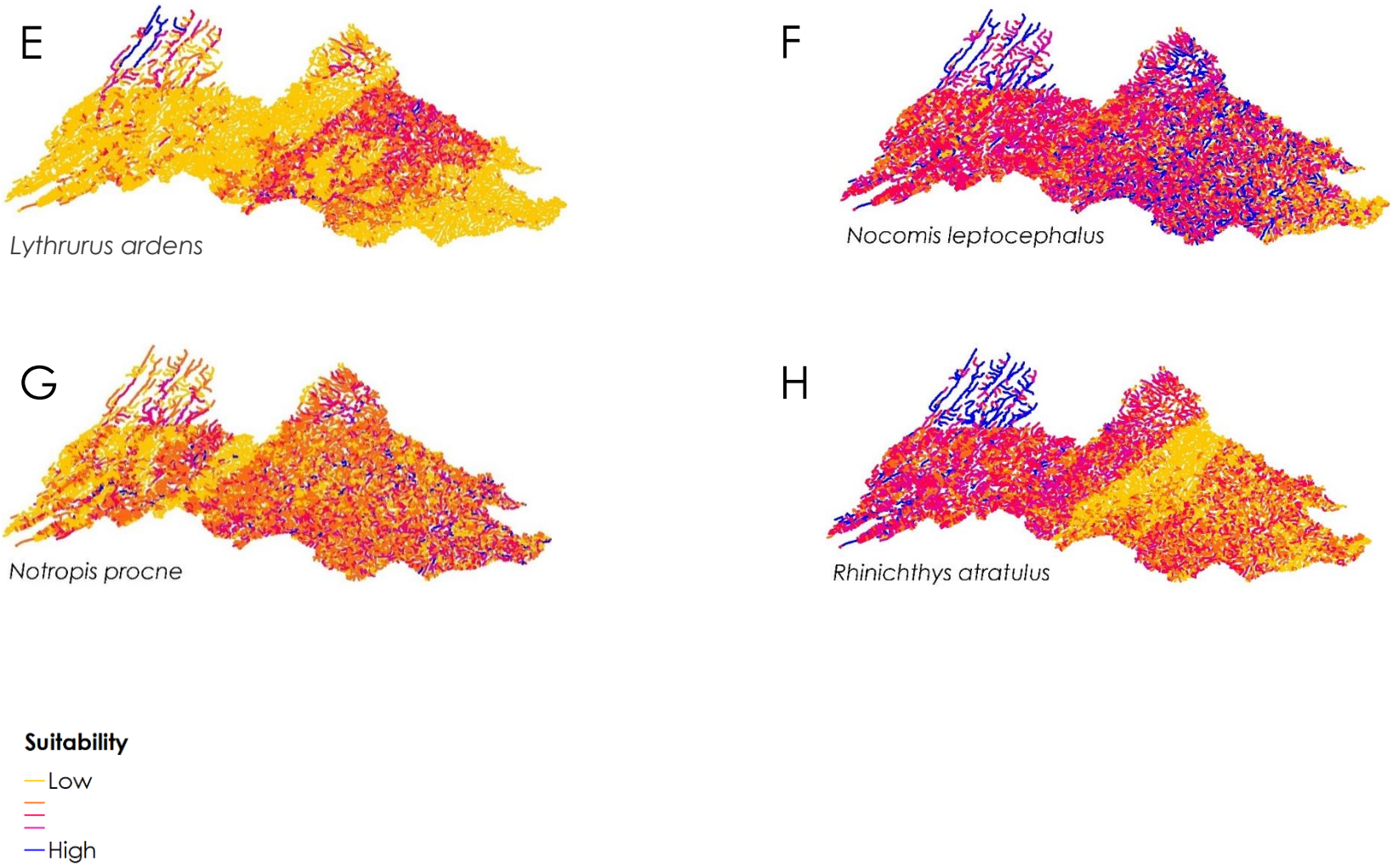

Figure B1 (E:H; continued): MaxEnt model predictions for habitat suitability for *P. collina* fish hosts in the James River Basin of Virginia.

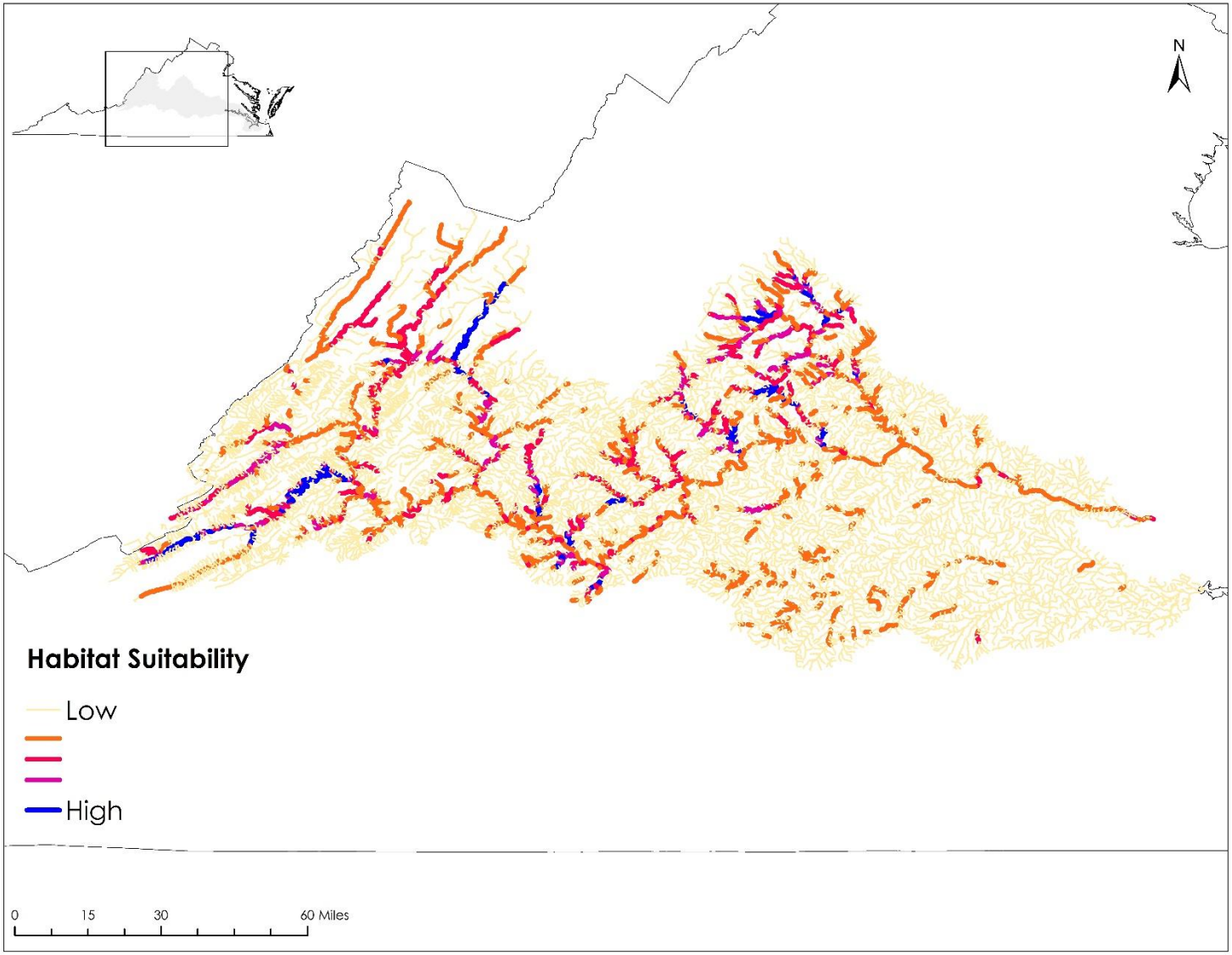

Figure B2: Predicted habitat suitability across the James River Basin in Virginia for *Parvaspina collina*.

Vita

Bonnie April Roderique was born December 8th 1990 in Lynchburg, Virginia and is an American citizen. She attended Lynchburg College where she received the Mahan - Osborne Award for Outstanding Senior in Biology and graduated Magna Cum Laude with a Bachelor's of Science degree in Biology. She worked as an Aquatics Technician for the Virginia Department of Game and Inland Fisheries for one year assisting with population monitoring surveys and hatchery propagation of *Parvaspina collina*. She received a Master's of Science in Environmental Studies and a Certificate in Geographic Information Systems from Virginia Commonwealth University in 2018.